

SPECTRAL ANALYSIS OF THE DEGREE-CORRECTED LAPLACIAN

by

John Jeongsoo Park

Copyright © John Jeongsoo Park 2025

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MATHEMATICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2025

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: John Park titled:

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Ning Hao

Ning Hao

Date: Nov 26, 2025

Yue Niu

Yue Niu

Date: Nov 26, 2025

Xueying Tang

Xueying Tang

Date: Nov 26, 2025

Giorgio Cipolloni

Giorgio Cipolloni

Date: Nov 27, 2025

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Ning Hao

Ning Hao
Mathematics

Date: Nov 26, 2025



ACKNOWLEDGMENTS

First and foremost, thank you to my advisor, Ning Hao for his guidance, support, and encouragement. He has modeled every quality I wish to have as a researcher, and I consider myself incredibly lucky to have worked with him during my graduate career. Special thanks to the rest of my committee members, Selena Niu, Xueying Tang, Sunder Sethuraman, and Giorgio Cipolloni, for their input and feedback throughout this process. I would also like to thank Kevin Lin and Shay Gilpin, both of whom displayed endless patience for my many questions about graduate and post-graduate life, and inspire me to the type of coworker whose door is always open.

I am grateful to many other members of the mathematics department. Sean Zhu and Christopher Mount, for their consistent friendship and countless hours of conversation. Jacob Quintero, Jackson Zariski, and Mina Dutra, for their geography, football, and pop-culture knowledge, and for the many laughs during trivia. There are too many others to list, but I would like to thank Kristine Van, Alexis Vizzerra, Jeffrey Mei, and Brandi Lu, whose friendship helped make the Tucson heat feel more bearable.

Finally, thanks to my family. To my partner Jason, who moved across the country without question to support me. I would not have made it here without him. To Patti, for the companionship curled up at my feet during the many late nights, and for reminding me to take frequent breaks for belly rubs. To my sister, Rachel, who has been a perfect role model, mentor, and friend. Most importantly, thanks to my parents, Jong and Hyesun, who love me unconditionally and taught me to work hard, aim high, and to always drop off your luggage at the hotel before dinner.

LAND ACKNOWLEDGMENT

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O'odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

TABLE OF CONTENTS

LIST OF FIGURES	8
ABSTRACT	10
CHAPTER 1 Introduction	11
CHAPTER 2 Background and Setting	16
2.1 The Random Dot Product Graph Model	17
<i>2.1.1 Erdős-Rényi Model</i>	18
<i>2.1.2 The Stochastic Block Model</i>	18
<i>2.1.3 The Degree-Corrected Stochastic Block Model</i>	19
2.2 Identifiability of Block Models	20
2.3 Community Detection and Spectral Clustering	26
<i>2.3.1 Spectral Clustering Under the Stochastic Block Model</i>	26
<i>2.3.2 Spectral Clustering Under the Degree-Corrected Stochastic Block Model</i>	28
2.4 Adjacency Spectral Embedding	30
2.5 Laplacian Spectral Embedding	34
CHAPTER 3 Spectral Clustering with the Generalized Laplacian	38
3.1 Population Analysis of the Degree-A Laplacian	39
3.2 Consistency of Spectral Clustering	42
<i>3.2.1 Sparsity Condition</i>	43
<i>3.2.2 Main Results on Consistency</i>	44
3.3 Degree-A Laplacian Spectral Embedding	46
3.4 Proof Sketch	49

CHAPTER 4 Spectral Clustering with the Generalized Laplacian	52
4.1 Regularization	52
<i>4.1.1 Consistency of Spectral Clustering with the Regularized Laplacian</i>	53
4.2 Additional Modifications to Spectral Clustering	54
<i>4.2.1 Selecting Additional Eigenvectors</i>	55
<i>4.2.2 Scaling Eigenvectors</i>	56
CHAPTER 5 Comparison of Methods	58
5.1 Theoretical Comparisons	58
<i>5.1.1 Two-Community Balanced SBM</i>	61
<i>5.1.2 Two-Community Imbalanced SBM</i>	64
5.2 Numerical Comparisons	64
5.3 Real Data	71
<i>5.3.1 University of Arizona KMAP Dataset</i>	71
CHAPTER 6 Discussion and Conclusion	75
APPENDIX A Appendix to Chapter 2	77
A.1 Proof of Theorem 2	77
APPENDIX B Appendix to Chapter 3	82
B.1 Proofs of Theorem 8 and 9	82
B.2 Proofs of Lemmas 15 and 16	87
APPENDIX C Appendix to Chapter 4	90
C.1 Proof of Lemma 10	90
C.2 Proof of Lemma 11	96
C.3 Proof of Theorem 10	100
C.4 Investigation of Modifications to Spectral Clustering	105
<i>C.4.1 Additional Eigenvectors</i>	105
<i>C.4.2 Eigenvalue Scaling</i>	105
C.5 Lemma 24 and Proof	108

APPENDIX D Appendix to Chapter 5	109
D.1 Derivation of Equation (5.3)	109
D.2 Description of Real-World Networks	113
D.3 List of Departments and Grouping	114
REFERENCES	123

LIST OF FIGURES

2.1	A representation of the latent positions of specialized RDPGs	20
2.2	A network of 4 departments from the University of Arizona. The nodes are colored by department and edges represent a collaboration between academics. The right shows the graph of the values of the eigenvectors by index for the top 4 eigenvectors.	28
2.3	The rows of $\hat{U}\hat{\Lambda}^{1/2}$ for a two-community SBM using the Adjacency Spectral Embedding. The colors correspond to the community of the node. The solid lines represent the theoretical 95% level curves predicted by Theorem 4 and the dashed lines represent the empirical 95% level curves computed using the EM algorithm.	34
3.1	The rows of $\hat{U}\hat{\Lambda}^{1/2}$ for a three-community DCSBM using the Degree-Corrected Laplacian Spectral Embedding in the setting $\rho_n = 1$. The colors correspond to the community of the node. For each community, 95% level curves corresponding to degree-correction parameter 0.4, 0.7, and 1 are drawn. See Section 3.3 for more information.	49
3.2	A brief summary of the orders of the Davis-Kahan theorem for various Laplacian matrices. The order of the bound is independent of α	51
5.1	The optimal α for the two-community SBM model described in Section 5.1.1. The dark blue region corresponds to $\alpha = 0$, adjacency spectral clustering being preferred. The lighter colors represent higher α being preferred, with degree-corrected spectral clustering being preferred in very sparse regions. . .	63

5.2	The ratio $\rho_{0,\alpha}^*$ for the imbalanced model described in Section 5.1.2 The values for α are $\alpha = 0.25$ (first column), $\alpha = 0.50$ (second column) and $\alpha = 0.75$ (third column). In the first row, $\pi_1 = 1/3$ and in the second row, $\pi_1 = 1/4$. Values larger than 1 represent parameter values where adjacency spectral clustering ($\alpha = 0$) is preferred. The black line corresponds to the level curve $\rho_{0,\alpha}^* = 1$	65
5.3	Numerical results of Experiment 1.	66
5.4	Numerical results of Experiment 2.	67
5.5	Numerical results of Experiment 3.	69
5.6	Numerical results of Experiment 4	70
5.7	Scores for the estimated community labels on the KMAP dataset.	74
5.8	A visualization of the full KMAP network: a A diagram of each node colored by the community labeling that results from DCSC with $K = 4$. b Examples of representative departments for each grouping.	74
C.1	The effect of including additional eigenvectors for various spectral clustering algorithms. Each column corresponds to a different algorithm, and each row corresponds to a different model, with increasing signal from the top row to the bottom row. Each graph shows boxplots of 50 replicates, using K , $K + 1$, or $K + 2$ eigenvectors for analysis.	106
C.2	The effect of scaling eigenvectors for various spectral clustering algorithms. Each column corresponds to a different algorithm, and each row corresponds to a different model, where the top row has a higher signal than the second. For each algorithm, the eigenvectors are scaled by $\hat{\Lambda}_{\alpha,\tau}^\beta$ for $\beta \in 0, 0.5, 1$	107

ABSTRACT

In network analysis, community detection aims to partition the nodes into meaningful groups based on their connections. To study this problem, random graph models such as the Stochastic Block Model (SBM) and the Degree-Corrected Stochastic Block Model (DCSBM) are widely used. A common approach for estimating communities is spectral clustering, a dimension reduction technique that maps the nodes to a lower-dimensional space while preserving the relevant information. Specifically, the eigenvectors of a particular matrix are used as a representation of each node. These methods are popular due to their simplicity and relative effectiveness.

In this thesis, we investigate the identifiability of the SBM and DCSBM, addressing a gap in the existing literature. We then generalize the matrices commonly used in spectral methods, and introduce the degree-corrected Laplacian, which accounts for the degree heterogeneity introduced by the DCSBM. We prove that the spectral clustering method induced by this matrix is consistent, and show that it enjoys the same asymptotic properties as its classical counterparts. Finally, we propose a spectral clustering algorithm based on the degree-corrected Laplacian and evaluate its performance through simulated and real-world network data.

Chapter 1

Introduction

Network analysis describes and models relationships between agents in populations of interest, with applications spanning a wide range of domains. In social networks, we may analyze connections between people to understand how information and misinformation spread (DeGroot, 1974). In biochemical networks, understanding complex interactions between genes and proteins can facilitate more efficient drug development (D’haeseleer, 2005). In ecology, organisms exhibit hierarchical organization, in which populations can be split into communities which further split into subcommunities (Krause et al., 2003).

At a high level, a network can be represented by a binary matrix indicating the presence or absence of edges between n nodes. Many statistical models exist for network data, but the Stochastic Block Model (SBM) is among the most popular due to its simple structure, which facilitates theoretical analysis (Holland et al., 1983). In the SBM, the n nodes are partitioned into $K < n$ communities, and edges are drawn between the nodes independently, with probabilities that depend only on the group memberships. This simple model can simulate group dynamics well, but it assumes that all nodes in the same group behave in exactly the same way. To better account for real-world dynamics, the Degree-Corrected Stochastic Block Model (DCSBM) was introduced to account for individual differences within a group (Karrer and Newman, 2011). In the DCSBM, each node is assigned a degree-correction parameter that represents its propensity for forming connections. Several other extensions of the SBM have been proposed, including models for directed networks (Wang and Wong, 1987) and networks for nodes with mixed memberships (Airoldi et al., 2008). These models are well-studied, particularly from the perspective of community detection, which aims to recover the unobserved communities from a single realization of a network. The DCSBM remains

a popular framework due to its balance between model flexibility and analytic tractability, and many modern community detection algorithms have been proposed based on it (Qin and Rohe, 2013; Jin et al., 2022; Qing and Wang, 2024; Chen et al., 2018; Pal and Coates, 2019). In these models, self-loops are typically disallowed (Jin, 2015; Lei and Rinaldo, 2015; Qin and Rohe, 2013), which leads to a subtle identifiability issue. We clarify the source of the issue and provide a simple sufficient condition for identifiability (Park et al., 2025).

Many community detection techniques focus on efficiently searching possible partitions for one that maximizes a modularity score (Clauset et al., 2004; Blondel et al., 2008). These techniques become impractical as the number of nodes grows, as this results in extremely high-dimensional representations. This high dimensionality can obscure simple relationships among nodes. Embedding the nodes into a lower-dimensional latent space can facilitate analysis and lead to more interpretable conclusions. A natural dimension reduction technique is to assume that the nodes are organized into distinct communities that behave similarly, as in the SBM and DCSBM. In these models, the matrices associated with the graphs are low-rank, and they have convenient spectral properties that aid in community detection (Von Luxburg, 2007). These techniques exploit the assumed community structure to embed the nodes in a lower-dimensional space. These representations are constructed from the spectral properties of matrices associated with the network, and are thus generally referred to as spectral methods.

In general, spectral methods use the eigenvectors and eigenvalues of some graph matrix to infer community membership. Assuming K is known, the eigenvectors corresponding to the K largest eigenvalues are computed and stacked column-wise into a matrix. The rows of this matrix are then clustered using k -means or similar rotation-invariant clustering method (Ng et al., 2001). Spectral methods are celebrated for their simple but effective approach, particularly in the SBM. While many standard approaches can detect communities consistently in the dense case, when the number of edges is relatively large, real-world networks often display sparsity, in which the number of nodes is much higher relative to the number of edges. Spectral clustering is well-known to achieve consistent community detection in the sparse case, when the expected degree of the nodes grows at least as fast as $\log n$ (Rohe et al., 2011; Jing et al., 2022). In addition to providing consistent results, its computational

efficiency makes it an ideal candidate for initializing more expensive, iterative algorithms, which can achieve consistency in even sparser cases (Gao et al., 2017).

Despite these advantages, spectral methods face several drawbacks. The techniques extend naturally to the DCSBM, but the flexibility of the model complicates the simple structure of the eigenvectors. The low-rank structure of the expected adjacency matrix is preserved, but the resulting eigenvectors are scaled by the degree-correction parameters, which necessitates a normalization of the rows before clustering (Ng et al., 2001). The proof techniques used to prove consistency for these modifications impose a penalty on the overall convergence rate (Lei and Rinaldo, 2015).

Additionally, most literature on spectral methods focuses on embeddings derived from two specific matrices — the adjacency matrix and the symmetric Laplacian (Abbe, 2018; Tang and Priebe, 2018). These matrices are natural choices for analysis because they are well-understood representations of graphs, and the resulting community detection algorithms provide intuitive mathematical interpretations. In particular, the symmetric Laplacian is similar to a row-stochastic matrix, providing a natural probabilistic interpretation for the induced spectral clustering (Chung, 1997). However, recent work has shown that modified forms of these matrices can yield better results under certain conditions. Modell and Rubin-Delanchy (2021) consider the row-stochastic matrix itself, and Qing and Wang (2024) consider a spectral method called Dual-Regularized spectral clustering (DRSC), which utilizes a Laplacian with the similarity matrix applied twice.

This thesis is motivated by the observation that the diagonal degree matrix admits fractional powers. This leads to the definition of the *degree- α Laplacian*, a one-parameter family of graph operators that naturally generalizes the adjacency and symmetric Laplacian matrices. We show that for all $\alpha \in [0, 1]$, the degree- α Laplacian preserves the low-rank structure of the underlying block model, and enjoys the same spectral properties as its classical counterparts. In particular, the case $\alpha = 1$ is of special interest: the resulting *degree-corrected Laplacian* automatically removes the effects of degree heterogeneity in the DCSBM.

Building on these theoretical insights, we introduce *Degree-Corrected Spectral Clustering (DCSC)*, a practical algorithm that uses the degree-corrected Laplacian for community detection. The algorithm is simple to implement, computationally efficient, and robust across a

broad range of network regimes. Our theoretical analysis establishes conditions under which DCSC achieves consistent community recovery. Numerical studies and applications to empirical datasets show that DCSC offers measurable improvements over traditional spectral clustering, particularly when communities are highly imbalanced or networks are sparse.

In addition to modifying the matrix used in spectral clustering, recent work has shown there are several post-processing techniques that can improve performance of the algorithms on real-world datasets. [Jin et al. \(2022\)](#) suggest using additional eigenvectors for analysis, which can mitigate the impact of eigenvector reordering in noisy regimes. [Qing and Wang \(2024\)](#) consider weighting the eigenvectors prior to clustering, which can also improve performance. Finally, [Joseph and Yu \(2016\)](#) consider the impact of regularization on spectral clustering, showing it can improve theoretical guarantees. We investigate the impact of these modifications when applied to the degree- α Laplacian.

The remainder of this dissertation is organized as follows. Chapter 2 introduces the random graph models used throughout the thesis, including the SBM, DCSBM, and their degree-corrected variants. We also discuss a subtle identifiability issue that arises when self-loops are absent and present a simple sufficient condition that resolves this gap in the literature ([Park et al., 2025](#)). The chapter concludes with a review of spectral embedding methods and existing results on their asymptotic behavior.

In Chapter 3, we introduce a degree- α Laplacian and develop analogous consistency results for the spectral clustering that arise from this matrix. We show that the induced embedding shares a simple distributional structure with the adjacency and symmetric Laplacian embeddings. In Chapter 4, we discuss practical guidelines for spectral clustering, introducing several modifications that can improve the performance of the algorithm in real-world networks. In particular, we show that a regularized version of the degree- α Laplacian also induces a consistent spectral clustering algorithm. In Chapter 5, we compare the various spectral clustering algorithms introduced through theoretical arguments, numerical simulations, and real-world datasets. Finally, Chapter 6 summarizes the main contributions of the dissertation and discusses several directions for future research.

We collect here the notation used throughout the dissertation. For a matrix M , $\|M\|$ will represent the spectral norm of the matrix. If any other norm is used, the appropriate

subscript will be used to denote the norm. We use $\|M\|_F$ for the Frobenius norm and $\|M\|_{2,\infty} = \max_i \|M_{i*}\|_2$ to denote the $2 \rightarrow \infty$ norm (Cape et al., 2019b). $\|v\|_2$ denotes the ℓ_2 norm of a vector.

Graphs are represented by their adjacency matrices $A = (a_{ij})$, with corresponding degree matrix $D = \text{diag}(d_1, \dots, d_n)$, where $d_i = \sum_j a_{ij}$. Random quantities estimated from data are indicated by hats, such as \hat{D} .

We say an event \mathcal{A}_n occurs “with high probability” if, for any constant c_1 , there exists a constant c_2 (which may depend on c_1) such that $P(\mathcal{A}_n) \geq 1 - c_2 n^{-c_1}$. We say $f(n) = O(g(n))$ if there exists a constant c_3 independent of n and an n_0 such that for all $n > n_0$, $f(n) \leq c_3 g(n)$. If this statement is followed by “with high probability,” the constant c_3 may depend on c_1 . We say that $f(n) = \Omega(g(n))$ if there exists a constant c_4 independent of n and an n_1 such that for all $n > n_1$, $f(n) > c_4 g(n)$. Similarly, we say $f(n) = o(g(n))$ if for every constant c there exists an n_0 such that for all $n > n_0$, $f(n) \leq c g(n)$ and $f(n) = \omega(g(n))$ if for every constant c there exists an n_0 such that for all $n > n_0$, $f(n) \geq c g(n)$. Finally, we say that $f(n) \asymp g(n)$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Additional notation specific to individual models or results is introduced as needed in later chapters.

Chapter 2

Background and Setting

In this chapter, we review the background for community detection in networks. We begin with an overview of the primary statistical framework for this field, the random graph model, providing the general definition and several specific, illustrative examples. In this review, we also address a subtle gap that exists in the literature regarding the identifiability of these models. Next, we introduce spectral clustering, a widely used technique that utilizes the spectral properties of graphs for community detection. Finally, we review some existing results that characterize the distribution of the eigenvectors of random graphs. In particular, these results show that the eigenvector deviations are asymptotically normal, with means and covariances that can be explicitly described in terms of the model parameters.

In their most general form, random graph models begin with n nodes and draw edges between them at random. Many models of varying complexity and flexibility have been proposed ([Athreya et al., 2018](#)). However, it is particularly convenient to assume the edges are drawn independently. We will focus on undirected graphs without self loops, so the resulting adjacency matrices will be symmetric with diagonal entries of the adjacency matrix will be set to 0. To that end, the graph models we will study will have the following form.

Definition 1 (Random Graph). *Let $P = (p_{ij})$ be an $n \times n$ matrix with entries in $[0, 1]$. We may define a random graph by setting the entries of its adjacency matrix $A = (a_{ij})$. For $i < j$, the entries satisfy*

$$a_{ij} \sim \text{Bern}(p_{ij}).$$

The remaining entries satisfy

$$a_{ij} = \begin{cases} 0 & \text{if } i = j \\ a_{ji} & \text{if } i > j. \end{cases}$$

It is also convenient to assume the expected adjacency matrix P is low rank, as this admits a simple decomposition of P . To that end, we do not set the diagonal of P to be 0, instead filling the diagonal with the entries that will allow P to be low rank. We begin with a particularly flexible model that satisfies these requirements, the random dot product graph model (RDPG).

2.1 The Random Dot Product Graph Model

In the random dot product graph model, each node is assigned a latent position \mathbb{R}^d , and the probability of a connection between two nodes is given by the dot product between the latent positions (Young and Scheinerman, 2007; Hoff et al., 2002; Handcock et al., 2007). These models are flexible, encompassing a wide range of other graph models while maintaining interpretability. In the latent space, nodes at a similar angle behave similarly. Additionally, distance to the origin represents a node’s propensity for connection. We provide a formal definition below.

Definition 2 (Random Dot Product Graph Model). *A graph follows a random dot product graph model with latent positions $X_1, \dots, X_n \in \mathbb{R}^d$ if*

$$p_{ij} = \langle X_i, X_j \rangle \in [0, 1].$$

Equivalently, if we define $X = [X_1 | X_2 | \dots | X_n]^\top$, $P = XX^\top$.

There are two sources of non-identifiability in the latent positions of an RDPG. First, the latent positions may lie in S , a d' dimensional subspace of \mathbb{R}^d with $d' < d$. If this is the case, we may express each latent position in terms of an orthonormal basis of S without changing the inner products. We thus assume that the span of the latent positions is \mathbb{R}^d . Equivalently, we assume that the rank of P is exactly d . The other source of non-identifiability is that any

orthogonal transformation of the latent positions will produce the same model. The results of this thesis will thus generally be up to an orthogonal transformation Q . These assumptions are standard when studying RDPGs (Tang and Priebe, 2018; Modell and Rubin-Delanchy, 2021).

We now introduce some random graph models which are special cases of the random dot product graph model and analyze the geometry of their latent positions. Figure 2.1 depicts the latent positions of each model geometrically.

2.1.1 Erdős-Rényi Model

The prototypical example of a random graph is the Erdős-Rényi graph, in which the presence of any edge is an independent Bernoulli random variable with success probability p (Gilbert, 1959). This deceptively simple model gives rise to quite complex behavior, and its properties have been studied extensively.

A highly relevant property of the graph is its connectedness, and it was shown that an Erdős-Rényi graph is almost surely connected if $np = \omega(\log n)$ and almost surely disconnected otherwise (Erdős and Rényi, 1960). This phase transition is relevant in community modeling, which we describe in the next section. As a random dot product graph, the latent positions are one-dimensional, and each node shares the same position, $X_i = \sqrt{p}$. A geometric depiction of the latent positions is given in Figure 2.1a.

2.1.2 The Stochastic Block Model

We now begin to model community structure, in which the n nodes are partitioned into K communities, and the behavior of the node depends only on its community. To that end, we define a community membership matrix Z that partitions the nodes, and a community connection matrix B that describes how communities relate.

Definition 3 (Community Membership Matrix). *Z is an $n \times K$ binary matrix where $Z_{ik} = 1$ if node i belongs to community k , and 0 otherwise. We use z_i to denote the community label of node i . That is, $z_i = k$ if and only if $Z_{ik} = 1$. Denote by $\mathbb{M}_{n,K}$ the collection of all possible community membership matrices for n nodes and K communities.*

Definition 4 (Community Connection Matrix). B is a symmetric, positive-definite $K \times K$ matrix where $0 \leq B_{kl} \leq 1$. B_{kl} represents the intensity of connection between nodes in community k and community l .

With these definitions, we may define the stochastic block model (Holland et al., 1983).

Definition 5 (Stochastic Block Model (SBM)). Given a community membership matrix Z and a community connection matrix B , a graph follows a stochastic block model with parameters (Z, B) if $P = ZBZ^\top$.

When detecting communities in this model, one of the primary assumptions is that the graph is connected. If not, we may consider each connected component to be a cluster of communities, and detect the communities on each component separately. The connectedness requirement in the previous section suggests that as long as the expected node degree is $\omega(\log n)$, then the graph is almost surely connected. This result holds more generally when the expected node degree in an Erdős-Rényi graph is replaced with the average expected node degree in an SBM (Abbe, 2018).

As a random dot product graph, the latent positions are K -dimensional, and each node lies on one of K distinct positions, corresponding to its community. A geometric depiction of the latent positions of an SBM is given in Figure 2.1b.

2.1.3 The Degree-Corrected Stochastic Block Model

In an SBM, it is assumed that all nodes within a community behave in exactly the same way. However, in real-world networks, there is heterogeneity present within the communities. The DCSBM was introduced as a generalization of the SBM account for this heterogeneity (Karrer and Newman, 2011). We may define a degree-correction matrix that describes the propensity for a given node to form any connection.

Definition 6 (Degree Correction Matrix). Θ is a strictly positive $n \times n$ diagonal matrix, where $\Theta_{ii} = \theta_i > 0$. θ_i represents the degree parameter for node i .

We now define the Degree-Corrected Stochastic Block Model (DCSBM).

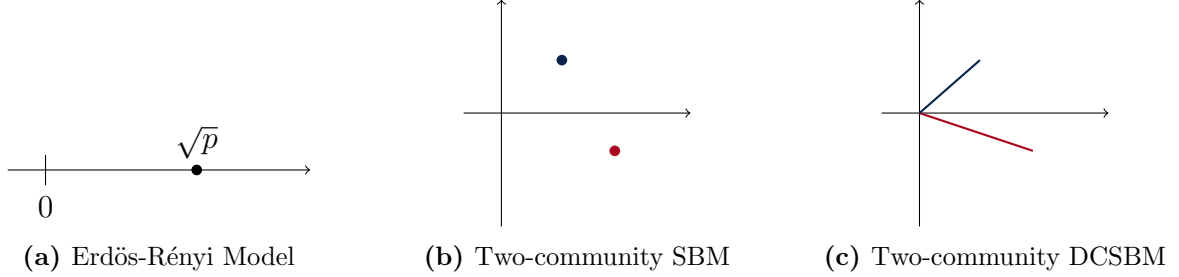


Figure 2.1: A representation of the latent positions of specialized RDPGs

Definition 7 (Degree-Corrected Stochastic Block Model (DCSBM)). *Given a community membership matrix Z , a community connection matrix B and a degree correction matrix Θ , a graph follows a degree-corrected stochastic block model with parameters (Z, Θ, B) if $P = \Theta Z B Z^\top \Theta$.*

Note that setting $\Theta = I_n$ in the model recovers the SBM. As a random dot product graph, the latent positions are K -dimensional, and the nodes lie on one of K distinct rays, with each direction corresponding to one community. A geometric depiction of the latent positions of a DCSBM is given in Figure 2.1c.

2.2 Identifiability of Block Models

With the models defined, we turn to the question of identifiability in SBMs and DCSBMs. In most literature on community detection, it is generally understood that a model is identifiable only up to a permutation of the labels. That is, two SBMs (Z, B) and (\tilde{Z}, \tilde{B}) define the same model if there exists a $K \times K$ permutation matrix Π such that $\tilde{Z} = Z\Pi$ and $\tilde{B} = \Pi^\top B\Pi$. It is also generally understood that the diagonal entries of an adjacency matrix can be ignored. Typically, this is done by asserting the graph either has no self loops, in which case the diagonal is set to zero (Tang and Priebe, 2018), or by setting the diagonal to another constant (Lei and Rinaldo, 2015). Most literature focuses on the asymptotic properties of norms, in which the diagonal part of the matrix becomes asymptotically insignificant as $n \rightarrow \infty$. However, even in simple cases, ignoring the diagonal can introduce identifiability issues beyond a permutation of the labels.

Formally, it is usually assumed that two SBMs parameterized by (Z, B) and (\tilde{Z}, \tilde{B}) are equivalent if and only if the expected adjacency matrices are the same, i.e., if $ZBZ^\top = \tilde{Z}\tilde{B}\tilde{Z}^\top$. However, setting the diagonals to a constant introduces a weaker condition. Letting \mathbb{P} denote the natural projection from $n \times n$ matrices to $n \times n$ matrices with zero diagonals, two SBMs parameterized by (Z, B) and (\tilde{Z}, \tilde{B}) are equivalent if and only if

$$\mathbb{P}(ZBZ^\top) = \mathbb{P}(\tilde{Z}\tilde{B}\tilde{Z}^\top). \quad (2.1)$$

It is generally understood, but rarely explicitly stated, that this minor modification introduces an additional identifiability issue that arises if the smallest community has size 1. This may be seen by the example below.

Example 1. *Consider the SBMs parameterized by*

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \frac{1}{10} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

$$\tilde{Z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \tilde{B} = \frac{1}{10} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then two matrices

$$ZBZ^\top = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 \end{bmatrix} \quad \text{and} \quad \tilde{Z}\tilde{B}\tilde{Z}^\top = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.1 \end{bmatrix}$$

share the same off-diagonal components, but $B \neq \tilde{B}$.

In this simple example, even if the population adjacency matrix is known, it is not possible to determine the community structure of the nodes. However, we introduce a mild condition for SBMs on the size of the smallest community that resolves this issue. Let $\mathbf{1}_m$ be the vector in \mathbb{R}^m consisting of all ones.

Condition 1. For an SBM parameterized by (Z, B) , assume that each community has at least two members. That is, $Z^\top \mathbf{1}_n \geq 2 \cdot \mathbf{1}_K$, where the inequality holds entry-wise.

Theorem 1. Assume that two SBMs parameterized by (Z, B) and (\tilde{Z}, \tilde{B}) satisfy Condition 1. These parameters define the same model, i.e., (2.1) holds, if and only if $\tilde{Z} = Z\Pi$ and $\tilde{B} = \Pi^\top B\Pi$, where Π is a $K \times K$ permutation matrix.

Proof. The “only if” part is trivial, because

$$\tilde{Z}\tilde{B}\tilde{Z}^\top = Z\Pi\Pi^\top B\Pi\Pi^\top Z^\top = ZBZ^\top.$$

For the “if” part, choose i such that $z_i = 1$. For each $k \in \{1, \dots, K\}$, it is possible to choose a j such that $z_j = k$. Then

$$B_{z_i z_j} = e_i^\top ZBZ^\top e_j = e_i^\top \tilde{Z}\tilde{B}\tilde{Z}^\top e_j = \tilde{B}_{\tilde{z}_i \tilde{z}_j}$$

and the first row of B is equal to the \tilde{z}_i -th row of \tilde{B} . This can be repeated for each row of B , which implies that up to a permutation of the rows and columns, $B = \tilde{B}$. In other words, $\tilde{Z} = Z\Pi$ and $\tilde{B} = \Pi^\top B\Pi$. \square

In the DCSBM, the degree-correction matrix Θ introduces additional sources of non-identifiability. First, Θ and B are only identifiable up to a scalar. Let D be a positive $K \times K$ diagonal matrix. Then (Z, Θ, B) and $(\Theta \text{diag}(ZD^{-1}\mathbf{1}_K), Z, D^{-1}B)$ produce the same expected adjacency matrix. Intuitively, we may multiply the degree-correction parameters in each community by c , then divide the corresponding row in B by c to produce the same model. To address this point, it can be assumed that the degree-correction parameters are normalized to the largest entry within each community (Lei and Rinaldo, 2015). Let ψ_k be the $n \times 1$ vector

$$\psi_k = \begin{cases} \theta_i & \text{if } z_i = k \\ 0 & \text{otherwise} \end{cases}.$$

In other words, ψ_k is the k -th column of ΘZ . We will assume without loss of generality that Θ has the form

$$\text{diag}\left(\frac{\theta_1}{\max \psi_{z_1}}, \dots, \frac{\theta_n}{\max \psi_{z_n}}\right). \quad (2.2)$$

It is typically assumed that two DCSBMs parameterized by (Z, Θ, B) and $(\tilde{Z}, \tilde{\Theta}, \tilde{B})$ are equivalent if and only if there exists a $K \times K$ permutation matrix and a $K \times K$ positive diagonal matrix D such that $\tilde{Z} = Z\Pi$, $\tilde{B} = \Pi^\top D B D \Pi$, and $\tilde{\Theta} = \Theta \text{diag}(Z D^{-1} \mathbf{1}_K)$. We introduce an analogue of Equation (2.1) for DCSBMs. Two DCSBMs parameterized by (Z, Θ, B) and $(\tilde{Z}, \tilde{\Theta}, \tilde{B})$ are equivalent if and only if

$$\mathbb{P}(\Theta Z B Z^\top \Theta) = \mathbb{P}(\tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta}). \quad (2.3)$$

Again, the weaker condition introduced by Equation (2.3) introduces additional sources of non-identifiability.

Example 2. Consider the DCSBMs parameterized by

$$\begin{aligned} \Theta &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, & Z &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, & B &= \frac{1}{40} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}; \\ \tilde{\Theta} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}, & \tilde{Z} &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, & \tilde{B} &= \frac{1}{40} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \end{aligned}$$

Then two matrices

$$\Theta Z B Z^\top \Theta = \begin{bmatrix} 0.2 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.2 \end{bmatrix} \quad \text{and} \quad \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} = \begin{bmatrix} 0.05 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.8 \end{bmatrix}$$

share the same off-diagonal components.

Example 3. Consider the DCSBMs parameterized by

$$\Theta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \frac{1}{10} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix};$$

$$\tilde{\Theta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \quad \tilde{Z} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \tilde{B} = \frac{1}{10} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Then the two matrices

$$\Theta Z B Z^\top \Theta = \begin{bmatrix} 0.1 & 0.1 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & 0.4 & 0.4 \end{bmatrix} \quad \text{and} \quad \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} = \begin{bmatrix} 0.1 & 0.1 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0 \\ 0 & 0 & 0.2 & 0.4 \\ 0 & 0 & 0.4 & 0.8 \end{bmatrix}$$

share the same off-diagonal components.

These examples demonstrate that Condition 1 is no longer sufficient to guarantee identifiability of the model. Introducing a slightly stronger version of Condition 1 resolves this issue.

Condition 2. For an SBM parameterized by (Z, B) , assume that each community has at least three members. That is, $Z^\top \mathbf{1}_n \geq 3 \cdot \mathbf{1}_K$, where the inequality holds entry-wise.

Theorem 2. Assume that two DCSBMs parameterized by (Z, Θ, B) and $(\tilde{Z}, \tilde{\Theta}, \tilde{B})$ satisfy Condition 2. These parameters define the same model, i.e., Equation (2.3) holds, if and only if $\tilde{Z} = Z\Pi$, $\tilde{B} = \Pi^\top D B D \Pi$, and $\tilde{\Theta} = \Theta \text{diag}(Z D^{-1} \mathbf{1}_K)$ where Π is a $K \times K$ permutation matrix and D is a $K \times K$ positive diagonal matrix.

A proof is given in Appendix A.1, and uses the assumption that the community connection matrix B is positive-definite. This requirement cannot be relaxed, as shown by the example below.

Example 4. Consider the DCSBMs parameterized by

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \Theta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \frac{1}{10} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix};$$

$$\tilde{Z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \tilde{\Theta} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad \tilde{B} = \frac{1}{10} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Then two matrices

$$\Theta Z B Z^\top \Theta = \begin{bmatrix} 0.4 & 0.4 & 0.2 & 0.2 \\ 0.4 & 0.4 & 0.2 & 0.4 \\ 0.2 & 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.4 & 0.4 \end{bmatrix} \quad \text{and} \quad \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} = \begin{bmatrix} 0.4 & 0.4 & 0.2 & 0.2 \\ 0.4 & 0.4 & 0.2 & 0.4 \\ 0.2 & 0.2 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.4 & 0.4 \end{bmatrix}$$

share the same off-diagonal components, but the community structures are not the same.

The previous example can easily be extended to two communities with an arbitrary number of members in each community, so a condition on the minimal size of the community will not remedy this situation. We summarize the results of this section in the below condition.

Condition 3. For a DCSBM parameterized by (Z, Θ, B) assume that:

- Each community has at least three members. That is, $Z^\top \mathbf{1}_n \geq 3 \cdot \mathbf{1}_K$, where the inequality holds entry-wise.
- The degree-correction parameters have the form described in Equation (2.2).
- The matrix B is positive-definite.

This condition guarantees that the model (Z, Θ, B) is identifiable up to a scalar and permutation of the labels.

2.3 Community Detection and Spectral Clustering

In this section, we review the community detection problem and established spectral clustering methods for block models. Given an observed adjacency matrix A assumed to follow an SBM or DCSBM, the goal of community detection is to recover Z , up to a permutation of the labels. Since A is a noisy observation of P , which has a predictable block structure, the intuitive algorithm is to cluster the rows of A . However, the number of nodes n , and thus the dimension of A , is often very large, rendering this idea intractable in practice. Spectral clustering aims to reduce the dimensionality of the problem by using the eigenvalues of A or a related matrix rather than working with the matrix directly. This method begins with the observation that the eigenvectors of the expected adjacency matrix P capture the underlying community structure.

2.3.1 Spectral Clustering Under the Stochastic Block Model

Consider as a simple example an SBM parameterized by (Z, B) . Let $P = U\Lambda U^\top$ be the eigendecomposition of P , where U is an $n \times K$ matrix whose columns are the unit eigenvectors of P , and Λ is a $K \times K$ diagonal matrix whose diagonal entries are the eigenvalues. Since P has K distinct rows, the equality of these two representations implies that U must also have K distinct rows. The rows of U thus reflect the community structure, which is formally stated in the lemma below (Lei and Rinaldo, 2015).

Lemma 1. *Let (Z, B) be an SBM with K communities, and let $U\Lambda U^\top$ be the eigendecomposition of the expected adjacency matrix $P = ZBZ^\top$. Then $U = ZX$ for some $K \times K$ matrix X .*

Proof. For each community k , let $n_k = \|Ze_k\|_1$, where e_k is the k -th standard basis vector of \mathbb{R}^K . Thus, n_k is the number of nodes in community k . Define $M = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_K})$. Then

$$P = ZBZ^\top = ZM^{-1}MBMM^{-1}Z^\top.$$

Additionally,

$$M^{-1}Z^\top ZM^{-1} = I_K.$$

Let $V\tilde{\Lambda}V^\top = MBM$ be the eigendecomposition of MBM . Define $U = ZM^{-1}V$ and $\Lambda = \tilde{\Lambda}$. Then

$$P = ZM^{-1}V\tilde{\Lambda}V^\top M^{-1}Z = U\Lambda U^\top.$$

Since

$$U^\top U = V^\top M^{-1}Z^\top ZM^{-1}V = I_K,$$

the spectral theorem implies that $U\Lambda U^\top$ is the eigendecomposition of P . Letting $X = M^{-1}V$ completes the proof. \square

The previous lemma suggests a simple algorithm to recover the communities. If we define \hat{U} to be the $n \times K$ matrix whose columns consist of the eigenvectors of A corresponding to the K top eigenvalues, then \hat{U} should roughly be the same as U , whose rows can be perfectly clustered. A reasonable labeling would then be obtained by clustering the rows of \hat{U} . In practice, k -means clustering is a popular choice, though other algorithms have been suggested (Lei and Rinaldo, 2015; Rubin-Delanchy et al., 2022).

Algorithm 1 Adjacency Spectral Clustering

Require: Observed adjacency matrix A , number of communities K

Ensure: Estimated community membership Matrix \hat{Z}

- 1: Compute $\hat{U} \in \mathbb{R}^{n \times K}$, whose columns consist of the K top eigenvectors of A .
 - 2: Apply k -means to the rows of \hat{U} .
 - 3: **return** \hat{Z} .
-

The technique of spectral clustering can be traced back to Donath and Hoffman (1973), though the general idea of using the spectral properties of a graph to partition nodes has been discovered independently several times. More recently, Shi and Malik (2000) and Ng et al. (2001) popularized the technique in the machine learning community. Since then, spectral clustering has been widely studied and generalized extensively. The technique is easy to describe, readily studied from multiple perspectives, and computationally feasible. Furthermore, it performs well across a wide range of regimes. We refer the reader to Von Luxburg (2007) and Abbe (2018) for a comprehensive overview the technique.

We pause here to present a practical example of a network that provides some visual intuition for spectral clustering. Figure 2.2a shows a graph of 4 departments at the University

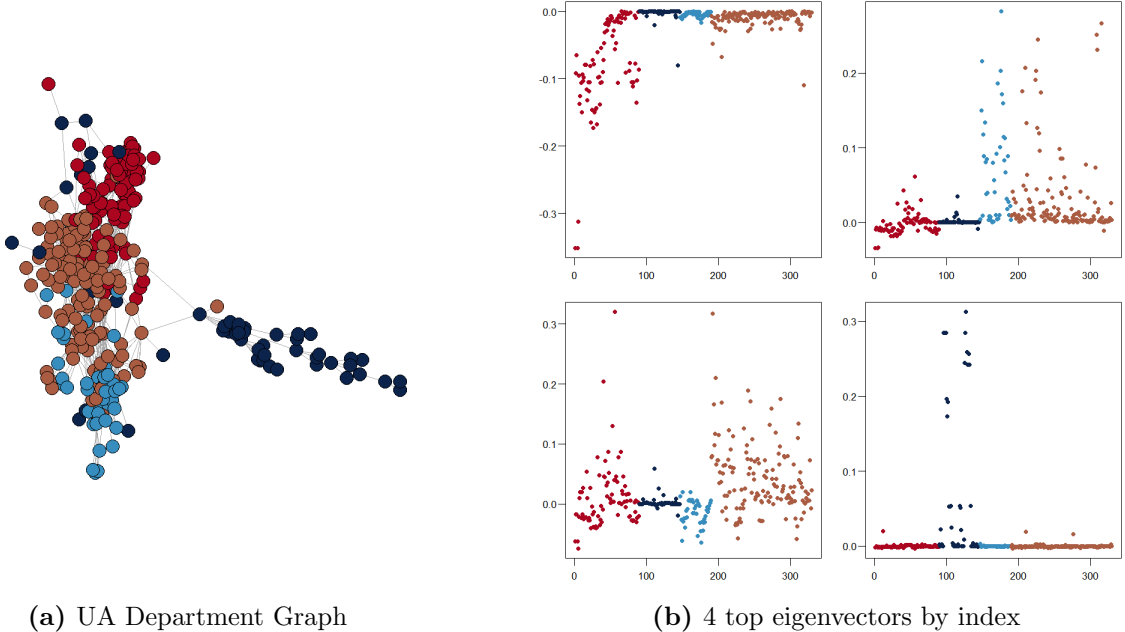


Figure 2.2: A network of 4 departments from the University of Arizona. The nodes are colored by department and edges represent a collaboration between academics. The right shows the graph of the values of the eigenvectors by index for the top 4 eigenvectors.

of Arizona, in which nodes represent people and edges represent collaborations between academics. (See Appendix D.2 for more information about this network.) Figure 2.2b plots the 4 top eigenvectors of A by index, where the indices of the eigenvectors are ordered by department. Visually, each of the eigenvectors distinguishes at least one department from the others, providing some information on how to cluster the nodes. In spectral clustering, these four eigenvectors are concatenated to form an $n \times K$ matrix, whose rows are viewed as points in \mathbb{R}^4 .

2.3.2 Spectral Clustering Under the Degree-Corrected Stochastic Block Model

We now consider a DCSBM parameterized by (Z, Θ, B) satisfying Condition 3. The idea of spectral clustering may be extended to this setting, but the rows of the eigendecomposition now cluster around rays instead of points (Jin, 2015; Lei and Rinaldo, 2015). We state an analogue of Lemma 1 for DCSBMs.

We first define several useful quantities. Recall that ψ_k is the k -th column of ΘZ . Define $\psi = \sum_{k=1}^K \psi_k$ and a $K \times K$ matrix $\Psi = \text{diag}(\|\psi_1\|, \dots, \|\psi_K\|)$.

Lemma 2. *Let (Z, Θ, B) be a DCSBM with K communities, and let USU^\top be the eigendecomposition of the expected adjacency matrix $P = \Theta Z B Z^\top \Theta$. Then $U = \Theta Z X$ for some $K \times K$ matrix X .*

Proof. We first write

$$P = \Theta Z \Psi^{-1} \Psi B \Psi \Psi^{-1} Z^\top \Theta.$$

Through direct computation, we may see that

$$\Psi^{-1} Z^\top \Theta \Theta Z \Psi^{-1} = I_K.$$

Let $\tilde{V} \tilde{\Lambda} \tilde{V}^\top$ be the eigendecomposition of $\Psi B \Psi$. Define $U = \Theta Z \Psi^{-1} \tilde{V}$ and $\Lambda = \tilde{\Lambda}$. Then

$$P = \Theta Z \Psi^{-1} \Psi B \Psi \Psi^{-1} Z^\top \Theta = \Theta Z \Psi^{-1} \tilde{V} \tilde{\Lambda} \tilde{V}^\top \Psi^{-1} Z^\top \Theta = U \Lambda U^\top.$$

Since

$$U^\top U = \tilde{V}^\top \Psi^{-1} Z^\top \Theta \Theta Z \Psi^{-1} \tilde{V} = I_K,$$

the spectral theorem implies that $U \Lambda U^\top$ is the eigendecomposition of P . Letting $X = \Psi^{-1} \tilde{V}$ completes the proof. \square

The previous lemma implies that when performing spectral clustering under the DCSBM, some form of normalization of the rows of \hat{U} is required. Several normalizations are possible, but unit normalization is an intuitive and natural choice (Qin and Rohe, 2013; Lei and Rinaldo, 2015). Under this model, any nodes with small degree correction parameter will have the corresponding latent position lie close to the origin. After normalization, any noise present in the observation will be amplified, making clustering more difficult. We will refer to any method that requires normalization as a normalized spectral clustering method, as in Algorithm 2.

Normalized spectral clustering methods have also been well studied, and their theoretical guarantees are competitive with those of other spectral clustering methods. However, it has been noted that in certain cases, the proof techniques required to deal with the extra normalization can affect the obtained asymptotic convergence rate (Lei and Rinaldo, 2015).

Algorithm 2 Normalized Adjacency Spectral Clustering

Require: Observed adjacency matrix A , number of communities K

Ensure: Estimated community membership Matrix \hat{Z}

- 1: Compute $\hat{U} \in \mathbb{R}^{n \times K}$, whose columns consist of the K top eigenvectors of A .
 - 2: Compute \hat{U}^* , the row-normalized version of \hat{U} .
 - 3: Apply k -means to the rows of \hat{U}^* .
 - 4: **return** \hat{Z}
-

2.4 Adjacency Spectral Embedding

The spectral clustering methods in the previous section provide a straightforward framework for community detection. These approaches use the eigenvectors of a graph’s adjacency matrix to represent each node as a low-dimensional vector, effectively reducing the dimensionality from \mathbb{R}^n to \mathbb{R}^K . This is formalized by the notation of the RDPG (Definition 2), in which each node is assigned a latent position. Given a symmetric, positive-definite expected adjacency matrix P , we may write $P = XX^\top$, and the rows of X may be thought of as the latent positions. It is natural to ask if we may decompose A in the same way, $A = \hat{X}\hat{X}^\top$, and examine the statistical properties of \hat{X} .

Early research on spectral methods focused on bounding the number of mis-labeled nodes from the algorithms. However, these early results are intrinsically tied to the choice of the clustering algorithm. Motivated by this shortcoming, [Tang and Priebe \(2018\)](#) studied the distributions of the rows of \hat{X} , finding that the asymptotic distribution of the rows could be explicitly computed. In particular, they find that the rows of \hat{X} are uniformly consistent estimators of the latent positions, and that the distributions of the rows are asymptotically normal, with means and covariances that can be explicitly calculated in terms of the model parameters. We first define the spectral embedding.

Definition 8 (Adjacency Spectral Embedding). *Let A be an $n \times n$ adjacency matrix and let $\hat{U}\hat{\Lambda}\hat{U}^\top$ be the eigendecomposition of A . That is, if $|\hat{\lambda}_1| \geq |\hat{\lambda}_2| \geq \dots \geq |\hat{\lambda}_K|$ are the largest K eigenvalues of A by magnitude, and $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_K$ are the corresponding orthonormal eigenvectors, then \hat{U} is the $n \times K$ matrix whose columns are $\hat{u}_1, \dots, \hat{u}_K$ and $\hat{\Lambda} = \text{diag}\left(|\hat{\lambda}_1|, \dots, |\hat{\lambda}_K|\right)$. The adjacency spectral embedding of A is $\hat{X} = \hat{U}\hat{\Lambda}^{1/2}$.*

We now formalize the notion that the adjacency spectral embedding is a consistent estimate of the latent positions. In particular, the deviations of the embedding are asymptotically normal. To see this, consider a sequence of graphs $\{A_N\}_{N \in \mathbb{N}}$. The sequence of graphs will be generated by the following process, which will be referenced frequently throughout this thesis.

Model 1. *In this model, we consider a sequence of graphs $\{A_N\}_{N \in \mathbb{N}}$ constructed according to the following process. Consider a distribution F on \mathbb{R}^d such that for all $x, y \in \text{Supp}(F)$, $x^\top y \in [0, 1]$. For each $N \in \mathbb{N}$, draw n vectors $\{\xi_{1,N}, \dots, \xi_{n,N}\}$ independently from F . Define $\{\rho_N\}_{N \in \mathbb{N}}$ to be a sequence of sparsity factors, and let $X_{i,N} = \rho_N^{1/2} \xi_{i,N}$. Finally, let A_N follow a random dot product graph corresponding to the latent positions $X_{i,N}$.*

We remark that the quantity $n\rho_N$ roughly scales with the expected degree, and it is typically assumed that this scales at least logarithmically in n . In general, the asymptotic results of this thesis assume that the number of nodes n diverges as $N \rightarrow \infty$. By taking an appropriate subsequence of graphs and re-indexing if necessary, we may thus assume that $N = n$. For ease of exposition, we will use the subscript n to denote the entry in the sequence of graphs as well as the number of nodes.

We are now ready to formally state the results, which show the adjacency spectral embedding is a consistent estimator of the latent positions, and provides an explicit formula for the asymptotic distribution of the deviations ([Tang and Priebe, 2018](#)).

Theorem 3. *Let A_n be generated according to Model 1, and suppose that the sparsity factor satisfies $n\rho_n = \omega(\log^{4c} n)$ for a universal constant $c > 0$. There exists a sequence of orthogonal transformations Q_n such that the adjacency spectral embedding satisfies*

$$\left\| Q_n \hat{X}_n - X_n \right\|_{2,\infty} = O\left(\frac{\log^c n}{n^{1/2}}\right)$$

with high probability.

Theorem 4. *Assume the setting of Theorem 3 and define the quantities*

$$\Upsilon = \mathbb{E}[\xi\xi^\top],$$

$$\Gamma_\rho(x) = \begin{cases} \mathbb{E} [\langle x, \xi \rangle (1 - \langle x, \xi \rangle) \xi \xi^\top] & \text{if } \rho_n = 1 \\ \mathbb{E} [\langle x, \xi \rangle \xi \xi^\top] & \text{if } \rho_n \rightarrow 0 \end{cases},$$

$$\Sigma(x) = \Upsilon^{-1} \Gamma_\rho \Upsilon^{-1}.$$

For each fixed index i , conditional on $\xi_i = x_i$, the random vector

$$\sqrt{n} \left(Q_n \hat{U}_{n,i^*} - U_{n,i^*} \right)$$

converges in distribution to a mean-zero normal random vector with covariance matrix given by $\Sigma(x_i)$.

We note that Theorem 4 is quite general and classify the behavior of RDPGs even when the latent positions have little structure. When we apply them to SBMs, the results simplify greatly due to the observations in Section 2.1.

Consider as an example an SBM parameterized by (Z, B) , where

$$B = \begin{bmatrix} 0.5 & 0.45 \\ 0.45 & 0.5 \end{bmatrix}, \quad \pi = (0.6, 0.4).$$

Here, the vector $\pi = (\pi_1, \pi_2)$ denotes the proportions of nodes in community k .

Given B and π , we may use Lemma 1 to compute the latent positions ξ , assuming that $\rho_n = 1$. Although no simple formula exists due to the imbalanced communities, they are numerically calculated to be roughly

$$U \approx \frac{1}{\sqrt{n}} \begin{bmatrix} 1.01 & -0.801 \\ 0.987 & 1.24 \end{bmatrix}, \quad \Lambda \approx n \begin{bmatrix} 0.476 & 0 \\ 0 & 0.0234 \end{bmatrix}, \quad X \approx \begin{bmatrix} 0.697 & -0.123 \\ 0.681 & 0.190 \end{bmatrix}.$$

The latent positions are then the rows of X , $X_1 \approx (0.697, -0.123)$ and $X_2 \approx (0.681, 0.190)$, and the distribution F is discrete probability distribution supported on those two points, with associated probabilities $p_1 = 0.6$ and $p_2 = 0.4$. The quantities from Theorem 4 may be

explicitly calculated. In particular,

$$\Upsilon = 0.6 (X_1 X_1^\top) + 0.4 (X_2 X_2^\top) \approx \begin{bmatrix} 0.476 & 0 \\ 0 & .024 \end{bmatrix}.$$

At $x = X_1$,

$$\begin{aligned} \Gamma_\rho(X_1) &= \mathbb{E} [\langle x, \xi \rangle (1 - \langle x, \xi \rangle) \xi \xi^\top] \\ &= 0.6 [(X_1^\top X_1) (1 - X_1^\top X_1) (X_1 X_1^\top)] + 0.4 [(X_1^\top X_2) (1 - X_1^\top X_2) (X_2 X_2^\top)] \\ &= 0.6 [(0.5)(0.5) (X_2 X_2^\top)] + 0.4 [(0.45) (0.55) (X_2 X_2^\top)] \\ &\approx \begin{bmatrix} 0.119 & 0.0001 \\ 0.0001 & 0.006 \end{bmatrix}, \end{aligned}$$

and

$$\Sigma(X_1) \approx \begin{bmatrix} 0.523 & 0.011 \\ 0.011 & 10.38 \end{bmatrix}.$$

Similarly, we may compute

$$\Sigma(X_2) \approx \begin{bmatrix} 0.523 & -0.011 \\ -0.011 & 10.38 \end{bmatrix}.$$

For a fixed n , we therefore have that rows of \hat{X} are distributed approximately normally with parameters that depend on their community. The two distributions are

$$N_1 = \mathcal{N}(\sqrt{n}X_1, \Sigma(X_1)), \quad N_2 = \mathcal{N}(\sqrt{n}X_2, \Sigma(X_2)).$$

In particular, as n increases, the means of the two distributions separate, while the covariance matrices remain constant, demonstrating the asymptotic separation of the two clusters.

Figure 2.3 shows a plot of the rows of \hat{X} for $n \in \{1000, 2000, 3000\}$. The solid lines represent the 95% level curves for the asymptotic distribution predicted by Theorem 4, and the dashed lines represent the 95% level curves for the empirical distributions using the EM algorithm via the R package ‘mclust’.

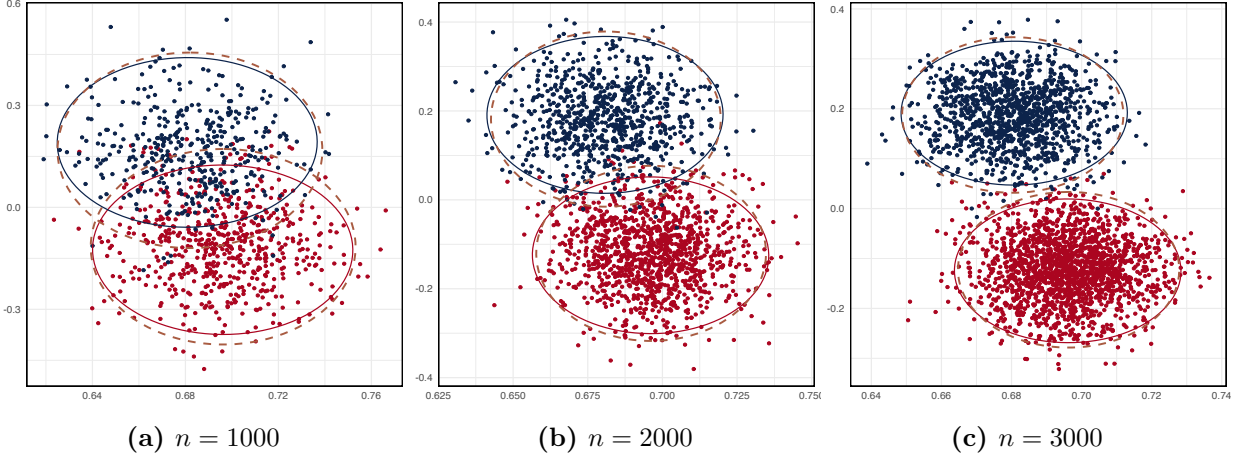


Figure 2.3: The rows of $\hat{U}\hat{\Lambda}^{1/2}$ for a two-community SBM using the Adjacency Spectral Embedding. The colors correspond to the community of the node. The solid lines represent the theoretical 95% level curves predicted by Theorem 4 and the dashed lines represent the empirical 95% level curves computed using the EM algorithm.

2.5 Laplacian Spectral Embedding

In this section, we define the symmetric Laplacian and describe analogous consistency results for the resulting embeddings. We begin with a definition. Let

$$t_i = \sum_{j=1}^n p_{ij}$$

denote the expected degree of node i and define an $n \times n$ diagonal matrix $T = \text{diag}(t_1, \dots, t_n)$. The population symmetric Laplacian is defined to be

$$L_{\text{sym}} = T^{-1/2}PT^{-1/2}.$$

Note that L_{sym} is well-defined in general, as we assume the graph is connected. We may define an analogous observed Laplacian matrix as $\hat{L}_{\text{sym}} = D^{-1/2}AD^{-1/2}$. We note that this definition of the symmetric Laplacian slightly differs from the definition sometimes found in graph theory literature, where it is defined to be $I - L_{\text{sym}}$ (Von Luxburg, 2007). In the context of community detection, we are generally concerned about the spectral properties of the matrices, for which these two definitions are equivalent. We may also decompose the

symmetric Laplacian as

$$L_{\text{sym}} = T^{-1/2} X X^\top T^{-1/2} := \tilde{X} \tilde{X}^\top,$$

and the rows of \tilde{X} may be viewed as the original latent positions of the nodes after scaling by $T^{-1/2}$.

The following lemma states that if P has the structure of a DCSBM, then L_{sym} shares a similar structure. We prove this lemma in more generality in Section 3.1.

Lemma 3. *Under a DCSBM parameterized by (Z, Θ, B) , define a $K \times K$ diagonal matrix $\tilde{T} = \text{diag}(\tilde{t}_1, \dots, \tilde{t}_k)$ and a $K \times K$ matrix \tilde{B} by*

$$\tilde{t}_l = \sum_{k \in K} \|\psi_k\|_1 b_{lk}, \quad \tilde{B} = \tilde{T}^{-1/2} B \tilde{T}^{-1/2}.$$

Then

$$L_{\text{sym}} = \Theta^{1/2} Z \tilde{B} Z^\top \Theta^{1/2}.$$

Lemma 2 then implies that the eigenvectors of L_{sym} encode the community structure in the same way the adjacency matrix does, i.e., the rows of \hat{U} point in K directions corresponding to the K communities. Motivated by this observation, we define an analogous spectral embedding for L_{sym} .

Definition 9 (Symmetric Laplacian Spectral Embedding). *Let $L_{\text{sym}}^{\hat{}}$ be the symmetric Laplacian matrix and let $\hat{U} \hat{S} \hat{U}^\top$ be the eigendecomposition of \hat{L} . That is, if $|\hat{\lambda}_1| \geq |\hat{\lambda}_2| \geq \dots \geq |\hat{\lambda}_K|$ are the largest K eigenvalues of L_{sym} by magnitude, and $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_k$ are the corresponding orthonormal eigenvectors, then \hat{U} is the $n \times K$ matrix whose columns are $\hat{u}_1, \dots, \hat{u}_k$ and $\hat{S} = \text{diag}(|\hat{\lambda}_1|, \dots, |\hat{\lambda}_k|)$. The symmetric Laplacian spectral embedding of L_{sym} is $\hat{X} = \hat{U} \hat{S}^{1/2}$.*

The next theorems are analogues of Theorems 3 and 4 for the symmetric Laplacian (Tang and Priebe, 2018). The proofs of the theorems require that the expected degrees of the graph grow like $n\rho_n$, which requires an additional condition.

Condition 4. *Let A_n be generated according to Model 1, and define $\mu = E(\xi)$. For all $x \in \text{Supp}(F)$, assume that $\langle x, \mu \rangle$ is bounded away from 0.*

The following lemma shows that this condition is sufficient for this to occur (Modell and Rubin-Delanchy, 2021).

Lemma 4. *Let A_n be generated according to Model 1, and assume that $\langle x, \mu \rangle$ is bounded away from 0 for all $x \in \text{Supp}(F)$. Then for each $i \in \{1, \dots, n\}$ and sufficiently large n ,*

$$t_{i,n} \asymp n\rho_n$$

with high probability.

Theorem 5. *Let A_n be generated according to Model 1 satisfying Condition 4. Additionally, suppose that the sparsity factor satisfies $n\rho_n = \omega(\log^{4c} n)$ for a universal constant $c > 0$. There exists a sequence of orthogonal transformations Q_n such that the symmetric Laplacian spectral embedding satisfies*

$$\left\| Q_n \hat{X}_n - (T_n^{-1/2} X_n) \right\|_{2,\infty} = O\left(\frac{\log^c n}{n\rho_n^{1/2}}\right)$$

with high probability.

Theorem 6. *Assume the setting of Theorem 5 and define the quantities*

$$\mu = \mathbb{E}[\xi] \quad \tilde{\Upsilon} = \mathbb{E}\left[\frac{\xi\xi^\top}{\langle \xi, \mu \rangle}\right]$$

$$\Gamma_\rho(x) = \begin{cases} \mathbb{E}\left[\langle x, \xi \rangle (1 - \langle x, \xi \rangle) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\tilde{\Upsilon}x}{2\langle x, \mu \rangle}\right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\tilde{\Upsilon}x}{2\langle x, \mu \rangle}\right)^\top\right] & \text{if } \rho_n = 1 \\ \mathbb{E}\left[\langle x, \xi \rangle \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\tilde{\Upsilon}x}{2\langle x, \mu \rangle}\right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\tilde{\Upsilon}x}{2\langle x, \mu \rangle}\right)^\top\right] & \text{if } \rho_n \rightarrow 0 \end{cases}$$

$$\tilde{\Sigma}(x) = \frac{\tilde{\Upsilon}^{-1} \Gamma_\rho \tilde{\Upsilon}^{-1}}{\langle x, \mu \rangle}$$

For each fixed index i , conditional on $\xi_i = x_i$, the random vector

$$n\rho_n^{1/2} \left(Q_n \hat{X}_{n,i^*} - T_n^{-1/2} X_{n,i^*} \right)$$

converges in distribution to a mean-zero normal random vector with covariance matrix given by $\tilde{\Sigma}(x_i)$.

As a final remark, we note that the results in this section were framed in terms of the latent positions defined by the random dot product graph, which are equal to $U\Lambda^{1/2}$. This contrasts with Algorithm 1, in which utilizes the matrix \hat{U} for clustering. We may consider an alternative spectral clustering that utilizes the rows of $\hat{U}\hat{\Lambda}^{1/2}$, which may be seen as weighting the j -th column of \hat{U} by the square root of the corresponding eigenvalue. We consider the effect of scaling the matrix in this way in Section 4.2.

Chapter 3

Spectral Clustering with the Generalized Laplacian

We have so far focused on the statistical properties of the adjacency and symmetric Laplacian matrices and their respective spectral clustering algorithms. The symmetric Laplacian in particular enjoys a connection to stochastic processes, as it may be viewed as a symmetrization of the row-stochastic matrix $D^{-1}A$ (Chung, 1997). Furthermore, the induced spectral clustering algorithm has an intuitive interpretation as an approximation to an optimization problem over all possible partitions of the nodes (Von Luxburg, 2007; Hagen and Kahng, 1992; Shi and Malik, 2000). Although this interpretation is appealing, it is natural to ask if alternative normalizations could lead to improvements in the resulting clustering algorithms. To this end, in this chapter, we introduce the idea of the degree- α Laplacian and a corresponding spectral clustering. Recall that under a DCSBM, the latent positions of the nodes lie on K rays instead of K points, which necessitates normalization in the induced spectral clustering algorithm. We focus on the case when $\alpha = 1$, which we name the degree-corrected Laplacian, that “collapses” the rays. The latent positions again fall onto K distinct points, bypassing the need for normalization.

First, we introduce the degree- α Laplacian and analyze the structure of the population matrix. We then show that the spectral clustering algorithm arising from this matrix is consistent and derive a bound on the error based on the model parameters. Finally, we prove an analogue to Theorem 3, showing that the spectral embedding that arises from this matrix is a uniformly consistent estimator of the latent positions; and Theorem 4, showing that the errors are asymptotically normal, giving explicit formulas for the means and covariances.

Throughout this chapter, we will work with the DCSBM, parameterized by (Z, Θ, B) (Section 2.1.3). Recall that $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$, a diagonal matrix consisting of the degree-

correction parameters of the nodes. Z is an $n \times K$ community membership matrix such that $z_{ik} = 1$ if and only if node i belongs to community k . B is a $K \times K$ positive-definite community connection matrix that describes how the communities relate. Then the expected adjacency matrix $P = \Theta Z B Z^\top \Theta$. Finally, the diagonal matrix $T = \text{diag}(t_1, \dots, t_n)$ is the expected degree matrix, where $t_i = \sum_{j=1}^n p_{ij}$.

3.1 Population Analysis of the Degree- α Laplacian

We begin with the definition of the degree- α Laplacian. Given an expected adjacency matrix P , we define the degree- α Laplacian to be

$$L_\alpha = T^{-\alpha} P T^{-\alpha}$$

for $\alpha \in [0, 1]$. Several particular values of α are worth highlighting. When $\alpha = 0$, L_α is the expected adjacency matrix; when $\alpha = 1/2$, L_α is the standard symmetric Laplacian; when $\alpha = 1$, L_α exhibits special behavior under the DCSBM. We will refer to this last matrix as the degree-corrected Laplacian, L_{DC} , which we will study more closely. The corresponding observed degree- α Laplacian is given by $\hat{L}_\alpha = D^{-\alpha} A D^{-\alpha}$.

We now study the theoretical properties of L_α , showing that it exhibits block structure similar to adjacency matrices of DCSBMs. This was stated in Lemma 3 for the symmetric Laplacian, and we now prove it for the degree- α Laplacian. We first recall that ψ_k is the $n \times 1$ vector whose entries are the k -th column of ΘZ . Then $\|\psi_k\|_1$ is the sum of the degree-correction parameters in group k .

Lemma 5. *Under a DCSBM parameterized by (Z, Θ, B) , define a $K \times K$ diagonal matrix $\tilde{T} = \text{diag}(\tilde{t}_1, \dots, \tilde{t}_k)$ and a $K \times K$ matrix \tilde{B} by*

$$\tilde{t}_l = \sum_{k \in K} \|\psi_k\|_1 b_{lk}, \quad \tilde{B}_\alpha = \tilde{T}^{-\alpha} B \tilde{T}^{-\alpha}.$$

Then

$$L_\alpha = \Theta^{1-\alpha} Z \tilde{B}_\alpha Z^\top \Theta^{1-\alpha}.$$

Proof. The proof of the lemma will follow from two steps. First, we show that $T^{-\alpha}\Theta = \Theta^{1-\alpha} \text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)$, where $\mathbf{1}_K$ is the $K \times 1$ vector consisting of all ones. Second, we show that $\text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)Z = Z\tilde{T}^{-\alpha}$. Then

$$\begin{aligned} L_\alpha &= T^{-\alpha}\Theta Z B Z^\top \Theta T^{-\alpha} \\ &= \Theta^{1-\alpha} \text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right) Z B Z^\top \text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right) \Theta^{1-\alpha} \\ &= \Theta^{1-\alpha} Z\tilde{T}^{-\alpha} B\tilde{T}^{-\alpha} Z^\top \Theta^{1-\alpha}, \end{aligned}$$

which proves the lemma.

For both claims, we show equality of the entries directly. We first rewrite the entries of $D^{-\alpha}$ as

$$d_i = \sum_{j=1}^n \theta_i \theta_j b_{z_i z_j} = \theta_i \sum_{k=1}^K \sum_{j \in k} \theta_j b_{z_i z_k} = \theta_i \sum_{k=1}^K b_{z_i z_k} \sum_{j \in k} \theta_j = \theta_i \sum_{k=1}^K b_{z_i z_k} \|\psi_k\|_1.$$

Then

$$\left(D^{-\alpha}\Theta\right)_i = \theta_i \left(\theta_i \sum_{k=1}^K b_{z_i z_k} \|\psi_k\|_1\right)^{-\alpha} = \theta_i^{1-\alpha} \left(\sum_{k=1}^K b_{z_i z_k} \|\psi_k\|_1\right)^{-\alpha} = \theta_i^{1-\alpha} \tilde{t}_i^{-\alpha}.$$

It is easy to check that $\left(\text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)\right)_i = \tilde{t}_i^{-\alpha}$, so

$$\left(\Theta^{1-\alpha} \text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)\right)_i = \theta_i^{1-\alpha} \tilde{t}_i^{-\alpha}$$

and $D^{-\alpha}\Theta = \Theta^{1-\alpha} \text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)$.

For the second claim,

$$\left(\text{diag}\left(Z\tilde{T}^{-\alpha}\mathbf{1}_K\right)Z\right)_{ik} = \tilde{t}_i^{-\alpha} Z_{ik} = Z\tilde{T}^{-\alpha}.$$

□

The previous lemma suggests that, given a DCSBM parameterized by (Z, Θ, B) , L_α is related to the DCSBM parameterized by $(Z, \Theta^{1-\alpha}, \tilde{B})$. This simple relationship has been

noted before in the case of the symmetric Laplacian (Qin and Rohe, 2013), but we emphasize that the degree- α Laplacian is explicitly related to this alternative DCSBM. Additionally, the generalization to the degree- α makes the significance of the degree-corrected Laplacian clear. By Lemma 2, the rows of the K top eigenvectors of L_α enjoy the same properties as eigenvectors a general DCSBM, i.e., the rows of U_α point in the same direction based on the node's community. Applying this to the degree *corrected* Laplacian, the $\Theta^{1-\alpha}$ term becomes the identity matrix, implying that it has a block structure identical to that of an SBM. Thus normalized spectral clustering can be applied to a general L_α , while standard spectral clustering is appropriate for the degree-corrected Laplacian.

We may also describe the eigenvectors of L_α explicitly. Using the notation from Section 2.3.2,

$$L_\alpha = U_\alpha \Psi \tilde{B} \Psi U_\alpha^\top,$$

where the columns of U_α are $n \times 1$ unit vectors. We may then describe U_α in terms of the model parameters and the eigenvectors of $\Psi \tilde{B} \Psi$.

Lemma 6. *Under a DCSBM parameterized by (Z, Θ, B) , suppose all eigenvalues of $\Psi \tilde{B} \Psi$ are simple. Let $\lambda_1 > \lambda_2 > \dots > \lambda_K$ be the largest K eigenvalues of $\Psi \tilde{B} \Psi$ and v_1, \dots, v_K be the corresponding unit eigenvectors. Then the K nonzero eigenvalues of L_α are $\lambda_1, \dots, \lambda_K$ and their associated unit eigenvectors are*

$$u_{k,i} = \left(\frac{\theta_i^{1-\alpha}}{\|\psi_{z_i}\|} \right) v_{k,i}.$$

Proof. Note that

$$L_\alpha U v_i = U \Psi \tilde{B} \Psi v_i = U \lambda_i v_i = \lambda_i U v_i$$

so $U v_i$ is an eigenvector of L_α corresponding to λ_i . Additionally,

$$\langle U v_i, U v_i \rangle = \langle v_i, U^\top U v_i \rangle = \langle v_i, v_i \rangle = 1.$$

□

3.2 Consistency of Spectral Clustering

In this section, we describe a spectral clustering algorithm for the degree- α Laplacian, and bound its overall error rate in terms of the parameters of the DCSBM. In doing so, we provide specific criteria for the overall error rate to go to 0 as n diverges. The algorithm is outlined in Algorithm 3.

Algorithm 3 Degree- α Laplacian Spectral Clustering

Require: Degree α Laplacian, \hat{L}_α , number of communities K

Ensure: Estimated community membership matrix \hat{Z}

- 1: Compute $\hat{U}_\alpha \in \mathbb{R}^{n \times K}$, whose columns consist of the K top eigenvectors of \hat{L}_α .
 - 2: Compute $\hat{\Lambda}_\alpha \in \mathbb{R}^{n \times n}$, a diagonal matrix whose diagonal entries consist of the K top eigenvalues of \hat{L}_α .
 - 3: Compute $\hat{X} = \hat{U}_\alpha \hat{\Lambda}_\alpha^{1/2}$
 - 4: **if** $\alpha \neq 1$ **then**
 - 5: Compute \hat{X}_α^* , the row-normalized version of \hat{U}_α .
 - 6: **else**
 - 7: Set $\hat{X}_\alpha^* = \hat{X}_\alpha$.
 - 8: **end if**
 - 9: Apply k -means to the rows of \hat{U}_α^* .
 - 10: **return** \hat{Z}
-

A basic requirement of any algorithm is that the number of errors remains reasonably bounded. In community detection, our aim is to recover the true community membership matrix Z given an adjacency matrix A and number of communities K . There are several ways to define “consistency” in the context of community detection. Here, we specify our definition, noting that it corresponds to the “almost exact recovery” described in [Abbe \(2018\)](#).

Recall that we use the notation $z_i = k$ to denote the community of a node i . Any $n \times K$ community matrix Z then naturally induces an $n \times 1$ vector of communities l by defining $l_i = z_i$. We may then consider the total number of errors to be the number of indices where $l_i \neq \hat{l}_i$. Since community labels are arbitrary up to a permutation, we evaluate the error over all possible permutations of the labels and take the minimum. Formally, we define the misclassification rate

$$L(\hat{Z}, Z) = \frac{1}{n} \min_{\Pi \in S_K} \sum_{i=1}^n \mathbb{1}\{l_i \neq \Pi(\hat{l}_i)\},$$

where S_K is the set of all $K \times K$ permutation matrices. We say an algorithm is consistent under certain conditions if $L(Z, \hat{Z}) \rightarrow 0$ in probability as $n \rightarrow \infty$. In other words, the proportion of errors should go to 0 as the number of nodes diverges.

We remark that in the Algorithm 3, we cluster the rows of the eigenvectors scaled by the square root of the corresponding eigenvalues, $\hat{U}_\alpha \hat{\Lambda}_\alpha^{1/2}$. This is to explicitly connect the spectral clustering algorithm to the consistency results of the spectral embedding, which sets the latent positions as $X = U\Lambda^{1/2}$. In practice, the consistency of the algorithm is insensitive to the choice of β when scaling by Λ^β (Section 4.2.2).

3.2.1 Sparsity Condition

We now define five quantities that describe the extremal values of the model, which will be useful in the upcoming sections. Defining e_k to be the k -th standard basis vector in \mathbb{R}^k ,

$$\delta = \min_i t_i; \quad \Delta = \max_i t_i; \quad \theta_{\min} = \min_i \theta_i; \quad n_{\max} = \max_k \|Ze_k\|_1; \quad n_{\min} = \min_k \|Ze_k\|_1$$

We may understand δ to be the smallest expected degree of a node and Δ to be the largest expected degree. Recall that θ_{\max} is set to 1 when Condition 3 is satisfied. Then θ_{\min} is the smallest degree-correction parameter. Finally, n_{\max} is the size of the largest community and n_{\min} is the size of the smallest community.

In this subsection, we describe a condition that controls the sparsity and degree heterogeneity of the model. Recall that several assumptions are required for the identifiability of the DCSBM, Condition 3, which we impose on the models in this section. In particular, we assume that the degree-correction parameters within each community are normalized, so $\theta_{\max} = 1$. For convenience, we define $B = \rho_n B_0$, where the maximum entry of B_0 is 1. With these preliminaries in place, we make an assumption on the network sparsity.

Condition 5. *For a DCSBM parameterized by (Z, Θ, B) , assume that $\Delta \geq \log n$.*

This assumption requires that the maximum expected degree of the graph to grow at least logarithmically. This may be compared to the assumptions for the results on spectral embeddings, where it was assumed that the expected degrees grow at least polylogarithmically

to get tighter bounds on the Laplacian concentration inequalities. We study Condition 5 in more detail below.

Observe that

$$t_i = \sum_{j=1}^n \theta_i \theta_j b_{z_i z_j} \leq \sum_{j=1}^n \theta_j \rho_n = \|\theta\|_1 \rho_n$$

and $\|\theta\|_1 \rho_n \geq \Delta$. Thus, the condition immediately implies $\|\theta\|_1 \rho_n \geq \log n$. On the other hand, it is easy to see that $n \geq \|\theta\|_1 > K$. Thus, the condition requires that the sum of the degree-correction parameters does not shrink too quickly relative to ρ_n . If we make additional assumptions on the regularity of the parameters, we may frame the bound in terms of δ . Define $S_k = \sum_{z_i=k} \theta_i$, the sum of the degree-correction parameters in community k , and let $S_{\max} = \max_k S_k$ and $S_{\min} = \min_k S_k$. Suppose that $\theta_{\min} > c_1$, $S_{\min} \geq c_2 S_{\max}$, and $\min_i \sum_j b_{0,ij} \geq c_3$ for absolute constants c_1, c_2 and c_3 . Then for all i ,

$$t_i = \sum_{j=1}^n \theta_i \theta_j b_{z_i z_j} \geq \theta_{\min} \rho_n \sum_{k=1}^K S_k b_{0,z_i k} \geq c_1 S_{\min} \rho_n \sum_{k=1}^K b_{0,z_i k} \geq c_1 c_2 c_3 S_{\max} \rho_n \geq \frac{c_1 c_2 c_3 \|\theta\|_1 \rho_n}{K}.$$

If K is assumed to be $O(1)$, Condition 5 then implies that δ grows at least logarithmically.

3.2.2 Main Results on Consistency

The proof of the consistency of Algorithm 3 follows from two lemmas. Lemma 7 shows that \hat{L}_α is close to L_α . Lemma 8 shows that \hat{X}_α^* is close to X_α^* . Finally, Theorem 7 shows that \hat{Z} is close to Z .

Lemma 7. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, the following bound holds with high probability.*

$$\|L_\alpha - \hat{L}_\alpha\| = O\left(\frac{\Delta \|\theta\|_1 \rho_n \log n}{\delta^{2+2\alpha}} + \frac{\Delta \sqrt{\|\theta\|_1 \rho_n \log n}}{\delta^{1+2\alpha}} + \frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{\delta^{2\alpha}}\right).$$

For convenience, we set the right-hand side of the previous lemma to η_α .

By Lemmas 2 and 5, the rows of U_α^* may be perfectly clustered. The matrix $X_\alpha^* = U_\alpha^* \Lambda_\alpha^*$ is a rescaling of this matrix, so its rows may also be perfectly clustered. The following lemma bounds the distance between X_α^* and \hat{X}_α^* , which shows that the rows of \hat{X}_α^* are close

to perfectly capturing the community information. We have so far used λ_i to denote the i -th eigenvalue of a matrix when the matrix is clear from context. In the following lemma and for the rest of this thesis, we use an additional subscript to denote to the matrix when the context is not clear, e.g., $\lambda_{i,P}$ will denote the i -th eigenvalue of the expected adjacency matrix.

For notational convenience, we define

$$\Delta_\alpha = \begin{cases} \Delta & \text{if } \alpha < 1/2 \\ \delta & \text{if } \alpha \geq 1/2 \end{cases}.$$

Lemma 8. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, the following bound holds with high probability.*

$$\left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F = O \left(\left(\frac{\sqrt{K}}{\sqrt{\eta_\alpha} \Delta_\alpha^{\alpha-1/2}} + 2\sqrt{2}K \right) \frac{\Delta^{3\alpha} \eta_\alpha \sqrt{n_{\max}}}{\Delta_\alpha^{\alpha-1/2} \lambda_{K,P}^{3/2} \theta_{\min}^{1-\alpha}} \right).$$

We remark that, in general $\theta_{\min} \leq 1$. Thus, increasing α reduces the effect of this term on the asymptotic convergence of the left-hand term. Again, for the degree-corrected Laplacian, when $\alpha = 1$, the term reduces to 1.

The following theorem bounds the error rate of Algorithm 3 under mild conditions, and shows the rate in terms of explicit constants related to the model.

Theorem 7. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, the following bounds hold with high probability.*

If $\alpha < 1/2$,

$$L(\hat{Z}, Z) = O \left(\frac{\lambda_{1,P} K^2 \Delta^{6\alpha+3/2} n_{\max} \sqrt{\|\theta\|_1 \rho_n \log n}}{\delta^{6\alpha-1/2} \lambda_{K,P}^4 \theta_{\min}^{2-2\alpha} n} \right)$$

If $\alpha \geq 1/2$,

$$L(\hat{Z}, Z) = O \left(\frac{\lambda_{1,P} K^2 \Delta^{8\alpha+1/2} n_{\max} \sqrt{\|\theta\|_1 \rho_n \log n}}{\delta^{8\alpha-3/2} \lambda_{K,P}^4 \theta_{\min}^{2-2\alpha} n} \right).$$

In the next section, we introduce more general statements of Lemmas 7 and 8 and Theorem 7. The corresponding proofs are given in Appendix C, and the proofs of the lemmas and theorems in this section may be viewed as special cases of these more general results.

To better understand the result of Theorem 7, consider a DCSBM parameterized by $(Z, \Theta, \rho_n B_0)$, in which B_0 and K are fixed. Suppose the community sizes and degree heterogeneity parameters are balanced in the sense that

$$\frac{n_{\max}}{n_{\min}} = O(1) \text{ and } \frac{\theta_{\max}}{\theta_{\min}} = \frac{1}{\theta_{\min}} = O(1).$$

This implies that $\Delta/\delta = O(1)$, $\delta \geq (n\rho_n)$, and Assumption 5 becomes $n\rho_n \geq \log(n)$. Additionally, since $\lambda_{K,P} \geq n_{\min}\rho_n\lambda_{K,B_0}$ and $\lambda_{1,P} \leq n_{\max}\rho_n\lambda_{1,B_0}$.

If $\alpha \geq 1/2$, the bound in Theorem 7 becomes

$$L(\hat{Z}, Z) = O\left(\frac{(n_{\max}\rho_n\lambda_{1,B_0})K^2(n\rho_n)^2n_{\max}\sqrt{n\rho_n\log n}}{(n_{\min}\rho_n\lambda_{K,B_0})^4n}\right) = O\left(\sqrt{\frac{\log n}{n\rho_n}}\right). \quad (3.1)$$

Thus, Algorithm 3 is consistent. A similar argument gives the same result when $\alpha < 1/2$. This is in line with analogous results for the adjacency spectral clustering (Lei and Rinaldo, 2015) and symmetric Laplacian spectral clustering (Qin and Rohe, 2013).

3.3 Degree- α Laplacian Spectral Embedding

The results of the previous section bound the error of Algorithm 3 in terms of the model parameters, using proof techniques that rely on specific properties of the k -means solution. In this section, we extend the results from Sections 2.4 and 2.5 to the degree- α Laplacian. This will allow us to draw conclusions about the spectral properties of the resulting embeddings independently of a specific clustering algorithm. In particular, we show that the degree- α Laplacian spectral embedding is a uniformly consistent estimator of the latent positions, and that the deviations of the eigenvectors is asymptotically normal.

Definition 10 (Degree- α Laplacian Spectral Embedding). *Let A be an $n \times n$ adjacency matrix and let $\hat{U}_\alpha \hat{\Lambda}_\alpha \hat{U}_\alpha^\top$ be the eigendecomposition of \hat{L}_α . That is, if $|\hat{\lambda}_{\alpha,1}| \geq |\hat{\lambda}_{\alpha,2}| \geq \dots \geq |\hat{\lambda}_{\alpha,K}|$ are the K largest eigenvalues of \hat{L}_α by magnitude, and $\hat{u}_{\alpha,1}, \hat{u}_{\alpha,2}, \dots, \hat{u}_{\alpha,K}$ are the corresponding orthonormal eigenvectors, then \hat{U}_α is the $n \times K$ matrix whose columns are $\hat{u}_{\alpha,1}, \dots, \hat{u}_{\alpha,K}$ and $\hat{\Lambda}_\alpha = \text{diag}\left(|\hat{\lambda}_{\alpha,1}|, \dots, |\hat{\lambda}_{\alpha,K}|\right)$. The spectral embedding of \hat{L}_α is $\hat{X}_\alpha = \hat{U}_\alpha \hat{\Lambda}_\alpha^{1/2}$.*

Theorem 8. *Let A_n be generated according to Model 1 satisfying Condition 4. Additionally, suppose that the sparsity factor satisfies $n\rho_n = \omega(\log^{4c} n)$ for a universal constant $c > 0$. There exists a sequence of orthogonal transformations Q_n such that the degree α Laplacian embedding satisfies*

$$\left\| Q_n \hat{X}_{n,i^*} - T_n^{-\alpha} X_{n,i^*} \right\|_{2,\infty} = O\left(\frac{\log^c n}{n^{\alpha+1/2} \rho_n^\alpha}\right)$$

with high probability.

The below theorem is a central limit theorem for the eigenvectors of the degree α Laplacian. We note that in the case that $\alpha = 1$, the degree-corrected Laplacian, the vectors lie in a $d - 1$ -dimensional subspace of \mathbb{R}^d with probability 1. In this case, the covariance matrices have rank $d - 1$.

Theorem 9. *Assume the setting of Theorem 8 and define the quantities*

$$\mu = \mathbb{E}[\xi] \quad \tilde{\Upsilon} = \mathbb{E}\left[\frac{\xi\xi^\top}{\langle \xi, \mu \rangle}\right]$$

$$\Gamma_{\rho,\alpha}(x) = \begin{cases} \mathbb{E}\left[\langle x, \xi \rangle (1 - \langle x, \xi \rangle) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle}\right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle}\right)^\top\right] & \text{if } \rho_n = 1 \\ \mathbb{E}\left[\langle x, \xi \rangle \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle}\right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle}\right)^\top\right] & \text{if } \rho_n \rightarrow 0 \end{cases}$$

$$\Sigma_\alpha(x) = \frac{\tilde{\Upsilon}^{-1} \Gamma_{\rho,\alpha} \tilde{\Upsilon}^{-1}}{\langle x, \mu \rangle^{2\alpha}}$$

For each fixed index i , conditional on $\xi_i = x_i$, the random vector

$$n^{\alpha+1/2} \rho_n^\alpha \left(Q_n \hat{X}_{n,i^*} - T_n^{-\alpha} X_{n,i^*} \right)$$

converges in distribution to a mean-zero normal random vector with covariance matrix given by $\Sigma_\alpha(x_i)$.

A high-level overview of the techniques used in the proofs of Theorems 8 and 9 is provided in Section 3.4, and formal proofs are given in Appendix B.

We may apply Theorem 9 to DCSBMs to get explicit formulas for the means and covariances of the asymptotic distributions. Consider the following sequence of random graphs,

each of which follows a DCSBM. Suppose the community membership for each node i is drawn at random with probability π_1, \dots, π_K . The degree-correction parameter θ_i is drawn at random from a probability distribution f , which has support on $(0, 1]$. Let $B_n = \rho_n B_0$ be the community connection matrix for a fixed symmetric, positive definite matrix $B_0 = XX^\top$. Define x_1, \dots, x_k to be the rows of X .

Corollary 1. *Let A_n be generated according to Model 1 satisfying Condition 4. Additionally, suppose that the sparsity factor satisfies $n\rho_n = \omega(\log^{4c} n)$ for a universal constant $c > 0$. Define the quantities*

$$d_l = \mathbb{E} [\theta] \sum_{m=1}^K \pi_m B_{lm} \quad \tilde{Y} = \mathbb{E} [\theta] \sum_{k=1}^K \frac{\pi_k x_k x_k^\top}{d_k}$$

$$\Gamma_{\rho, \alpha}(k, \lambda, l) = \begin{cases} \mathbb{E} \left[\theta B_{kl} (1 - \lambda \theta B_{kl}) \left(\frac{x_l}{d_l} - \frac{\alpha \tilde{Y} x_k}{d_k} \right) \left(\frac{x_l}{d_l} - \frac{\alpha \tilde{Y} x_k}{d_k} \right)^\top \right] & \text{if } \rho_n = 1 \\ \mathbb{E} \left[\theta B_{kl} \left(\frac{x_l}{d_l} - \frac{\alpha \tilde{Y} x_k}{d_k} \right) \left(\frac{x_l}{d_l} - \frac{\alpha \tilde{Y} x_k}{d_k} \right)^\top \right] & \text{if } \rho_n \rightarrow 0 \end{cases}$$

$$\Sigma_\alpha(k, \lambda) = \frac{\sum_{l=1}^K \pi_l \tilde{Y}^{-1} \Gamma_{\rho, \alpha}(k, \lambda, l) \tilde{Y}^{-1}}{\lambda^{2\alpha-1} d_k^{2\alpha}}$$

where all expectations are taken with respect to $\theta \sim f$.

For a fixed index i , conditional on $z_{i,n} = z_i$ and $\theta_{i,n} = \lambda_i$, there exists a sequence of orthogonal transformations Q_n such that the random vector

$$n^{\alpha+1/2} \rho_n^\alpha \left(Q_n \hat{X}_{\alpha, i^*} - \frac{\lambda_i x_i}{\rho_n^{1/2} \left(\lambda_i \sum_{j=1}^n \theta_{j,n} b_{z_{i,n}, z_{j,n}} \right)^\alpha} \right)$$

converges in distribution to a mean-zero normal random vector with covariance matrices given by $\Sigma_\alpha(z_i, \lambda_i)$.

Note that when $\alpha = 1$, corresponding to the degree-corrected Laplacian, the means of the distributions depend only on the communities and not on λ . Furthermore, if $\rho_n \rightarrow 0$, λ affects the scale of the covariance matrices, but not the shape.

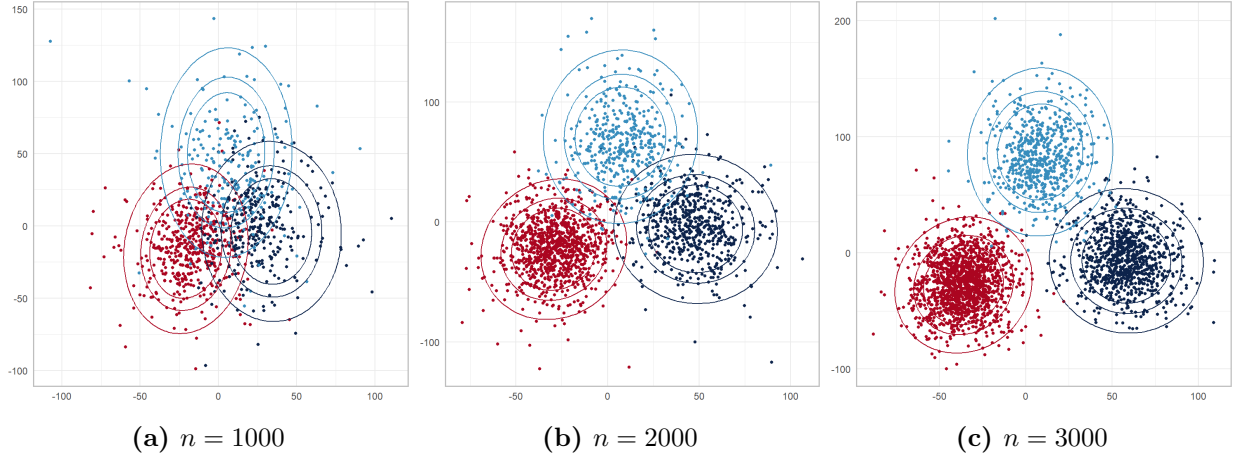


Figure 3.1: The rows of $\hat{U}\hat{\Lambda}^{1/2}$ for a three-community DCSBM using the Degree-Corrected Laplacian Spectral Embedding in the setting $\rho_n = 1$. The colors correspond to the community of the node. For each community, 95% level curves corresponding to degree-correction parameter 0.4, 0.7, and 1 are drawn. See Section 3.3 for more information.

Consider as an example a DCSBM parameterized by (Z, Θ, B) , where

$$B = \begin{bmatrix} 0.48 & 0.28 & 0.28 \\ 0.28 & 0.40 & 0.28 \\ 0.28 & 0.28 & 0.40 \end{bmatrix} \quad \theta \sim \text{Unif}(0.4, 1) \quad \pi = (0.5, 0.3, 0.2).$$

In Figure 3.1, we show the projection of the degree-corrected Laplacian spectral embedding ($\alpha = 1$) for a three-community DCSBM. In this case, Theorem 9 predicts that the means and covariances will lie in a 2-dimensional subspace of \mathbb{R}^3 , though this subspace is unique only up to an orthogonal matrix Q . Figure 3.1 shows an example of the distribution of the rows of $\hat{U}_\alpha \hat{\Lambda}_\alpha^{1/2}$ after projection to one choice of this subspace for $n \in \{1000, 2000, 3000\}$.

3.4 Proof Sketch

In this section, we provide a high-level overview of the primary techniques used in the proofs of Theorems 8 and 9. We again consider the simplified DCSBM introduced at the end of Section 3.2 for intuition. The assumptions are reproduced here for convenience. The

community sizes and degree heterogeneity parameters are balanced, in the sense that

$$\frac{n_{\max}}{n_{\min}} = O(1) \text{ and } \frac{\theta_{\max}}{\theta_{\min}} = \frac{1}{\theta_{\min}} = O(1).$$

In this case, the smallest and largest expected degrees are roughly balanced in the sense that $\Delta/\delta = O(1)$, and we may define a “typical” expected node degree d .

There are three main ideas used to prove the results from Sections 2.4 and 2.5. First, that \hat{L}_{sym} is strongly concentrated around L_{sym} . Second, the Davis-Kahan theorem implies that the eigendecompositions are roughly the same — that is, \hat{U}_{sym} is close to U_{sym} . Finally, a linearization of $D^{-1/2}$ is used to simplify the matrix $D^{-1/2}\hat{L}_{\text{sym}}D^{-1/2}$.

We may use a Taylor series approximation to linearize the matrix D^γ . By letting γ vary between $-1/2$ and $1/2$ and using the fact that $\hat{L}_\alpha = D^{1/2-\alpha}\hat{L}_{\text{sym}}D^{1/2-\alpha}$, this allows us to generalize the results of Tang and Priebe (2018) to the degree α Laplacian.

We begin with the Laplacian concentration. Assuming d scales polylogarithmically, a result of Lu and Peng (2013) states that the symmetric Laplacian satisfies

$$\left\| \hat{L}_{\text{sym}} - L_{\text{sym}} \right\| = O_P \left(\sqrt{\frac{1}{d}} \right).$$

Compare this to the results of Lemma 7. In the current mild setting, the result states that

$$\left\| \hat{L}_{\text{sym}} - L_{\text{sym}} \right\| = O_P \left(\sqrt{\frac{\log(n)}{d}} \right).$$

In the setting of Lemma 7, d is only assumed to grow logarithmically, but the concentration is weaker by a factor of $\sqrt{\log(n)}$. The stronger concentration is required to ensure the noise remains well-controlled.

The second result utilizes the Davis-Kahan $\sin \theta$ theorem. A convenient form is provided in Lei and Rinaldo (2015).

Lemma 9. *Let S be a rank K $n \times n$ symmetric matrix with smallest eigenvalue λ_K and let \hat{S} be any symmetric matrix. Define $U, \hat{U} \in \mathbb{R}^{n \times K}$ to be the matrices whose columns are the K top eigenvectors of S and \hat{S} , respectively. Then there exists a $K \times K$ orthogonal matrix*

Matrix	α	$\ \hat{L}_\alpha - L_\alpha\ $	λ_K	D-K Bound
Adjacency	0	$d^{1/2}$	d	$d^{-1/2}$
Laplacian	1/2	$d^{-1/2}$	1	$d^{-1/2}$
Degree- α Laplacian	α	$d^{1/2-2\alpha}$	$d^{-\alpha}$	$d^{-1/2}$
Degree-Corrected Laplacian	1	$d^{-3/2}$	d^{-1}	$d^{-1/2}$

Figure 3.2: A brief summary of the orders of the Davis-Kahan theorem for various Laplacian matrices. The order of the bound is independent of α .

Q such that

$$\left\| \hat{U} - UQ \right\|_F \leq \frac{2\sqrt{2K}}{\lambda_K} \left\| \hat{S} - S \right\|.$$

Consider the below quantity for the expected and observed adjacency matrices,

$$\frac{\|A - P\|}{\lambda_K}.$$

The order of the numerator is well known and is \sqrt{d} (Lei and Rinaldo, 2015). On the other hand, λ_K scales roughly like n and is thus on the order of d . Therefore, the quantity above is on the order of $d^{-1/2}$.

Intuitively, when we consider the symmetric Laplacian, we are “dividing” the numerator by d , and the Laplacian concentration is on the order of $d^{-1/2}$, as shown in the first row of Table 3.2. On the other hand, via an application of the Courant-Fisher min-max theorem, the smallest eigenvalue of L_{sym} is also scaled by d , and the order of the ratio remains roughly the same. This idea holds true for any α we choose, and the eigenspaces converge at the same rate regardless of the α chosen. This behavior is summarized in Table 3.2.

The final technique is to use a Taylor series approximation of $D^{-\alpha}$ to linearize the matrix L_α . This result follows from a straightforward application of the mean value theorem, which states that for any β , there exists a ξ_i between t_i and d_i such that

$$d_i^\beta = t_i^\beta + \beta t_i^{\beta-1}(d_i - t_i) + \frac{1}{2}\beta(\beta + 1)\xi_i^{\beta-2}(d_i - t_i)^2.$$

When d_i concentrates around t_i , we may ignore the quadratic term, and the above expression allows us to replace the matrix D with the matrix T plus a linear term.

Chapter 4

Spectral Clustering with the Generalized Laplacian

4.1 Regularization

The spectral clustering algorithms outlined in the previous sections work well in theory and show satisfactory performance across a wide range of conditions. However, there are several modifications that have been shown to empirically improve performance, particularly when the graph contains nodes of very low degree (Amini et al., 2013; Jin, 2015). Recall that for an observed adjacency matrix A , the degree- α Laplacian is given by $L_\alpha = D^{-\alpha}AD^{-\alpha}$, where D is the diagonal matrix consisting of the observed degrees. Intuitively, when degrees are small, we are “dividing” by a small number, which increases the variance of L_α . A standard ridge regularizer may be implemented to mitigate this effect (Chaudhuri et al., 2012). For $\tau \geq 0$, define the regularized degree matrix as $T_\tau = T + \tau I$. The regularized graph Laplacian is then given by

$$L_\tau = T_\tau^{-1/2}PT_\tau^{-1/2}.$$

This matrix naturally induces a spectral clustering, identical to Algorithm 3, using the regularized Laplacian instead of the standard Laplacian. Qin and Rohe (2013) show that under the DCSBM, this algorithm is consistent and bound the error rate in terms of the model parameters. The result of regularized spectral clustering is generally insensitive to the exact choice of τ ; in practice, the average node degree is used.

Similarly, we may define a regularized degree- α Laplacian.

$$L_{\alpha,\tau} = T_\tau^{-\alpha}PT_\tau^{-\alpha}.$$

Each of these definitions extend naturally to their observed counterparts.

4.1.1 Consistency of Spectral Clustering with the Regularized Laplacian

We now consider the spectral clustering algorithm induced by the regularized degree- α Laplacian, and show that it is consistent. To that end, we introduce analogues of each of the lemmas and theorems of Section 3.2, showing that the results change predictably with the introduction of the parameter τ . Note that these results may be viewed as corollaries to the results in this section by setting $\tau = 0$. We list each of the analogues, and refer the reader to Section 3.2 for more context. The proofs of these theorems and lemmas are found in Appendix C.

Lemma 10. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, if $\|\theta\|_1 \rho_n \geq \delta + \tau$, the following bound holds with high probability.*

$$\|L_{\alpha, \tau} - \hat{L}_{\alpha, \tau}\| = O\left(\frac{\Delta \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{2+2\alpha}} + \frac{\Delta \sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{1+2\alpha}} + \frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{2\alpha}}\right).$$

We may see that the introduction of τ can improve the asymptotic concentration of the degree- α Laplacian, which ultimately reduces the number of errors in the corresponding spectral clustering. The additional requirement in this lemma is that $\|\theta\|_1 \geq \delta + \tau$. When $\tau = 0$, this is vacuously true, because $\|\theta\|_1 \rho_n \geq \Delta$. The condition thus requires that τ not be too large relative to the sparsity factor.

Define the error term in the previous lemma to be $\eta_{\alpha, \tau}$ and

$$\Delta_{\alpha, \tau} = \begin{cases} \Delta + \tau & \text{if } \alpha < 1/2 \\ \delta + \tau & \text{if } \alpha \geq 1/2 \end{cases}.$$

We now state analogues of Lemma 8 and Theorem 7.

Lemma 11. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, if $\|\theta\|_1 \rho_n \geq \delta + \tau$, the following bound holds with high probability.*

$$\|\hat{X}_{\alpha, \tau}^* - X_{\alpha, \tau}^*\|_F \leq 2 \left(\frac{\sqrt{K}}{\sqrt{\eta_{\alpha, \tau}} \Delta_{\alpha, \tau}^{\alpha-1/2}} + 2\sqrt{2}K \right) \frac{(\Delta + \tau)^{3\alpha} \eta_{\alpha, \tau} \sqrt{n_{\max}}}{\Delta_{\alpha, \tau}^{\alpha-1/2} \lambda_{K, P}^{3/2} \theta_{\min}^{1-\alpha}}$$

Theorem 10. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, if $\|\theta\|_1 \rho_n \geq \delta + \tau$, the following bounds hold with high probability.*

If $\alpha < 1/2$,

$$L(\hat{Z}, Z) = O\left(\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{6\alpha+1} n_{\max} \sqrt{\Delta \|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{6\alpha-1/2} \lambda_{K,P}^4 \theta_{\min}^{2-2\alpha} n}\right)$$

If $\alpha \geq 1/2$,

$$L(\hat{Z}, Z) = O\left(\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{8\alpha} n_{\max} \sqrt{\Delta \|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{8\alpha-3/2} \lambda_{K,P}^4 \theta_{\min}^{2-2\alpha} n}\right).$$

We now study the previous result in more detail. For each case, we may compute the value of τ that minimizes the order of the error.

$$\tau_{\text{opt}} = \begin{cases} 0 & \text{if } \alpha \leq \frac{1}{12} \\ \min\left(0, \frac{2}{3} \left((6\alpha + 1)\Delta - (6\alpha - \frac{1}{2})\delta\right)\right) & \text{if } \frac{1}{12} < \alpha \leq \frac{1}{2} \\ \min\left(0, \frac{2}{3} \left((8\alpha - \frac{3}{2})\Delta - 8\alpha\delta\right)\right) & \text{if } \frac{1}{2} < \alpha \leq 1 \end{cases}.$$

Heuristically, τ should be large when α is large or when the degree-heterogeneity is extreme. In practice, this depends on the unknown quantities Δ and δ , and in practice, the average node degree is used. Regardless of the theoretical optimal τ , by exactly the same argument given in Equation (3.1), under the same mild conditions, the bound in Theorem 10 becomes

$$L(\hat{Z}, Z) = O\left(\sqrt{\frac{\log n}{n\rho_n}}\right). \quad (4.1)$$

4.2 Additional Modifications to Spectral Clustering

Recall that under the DCSBM, the latent positions of the nodes collapse when we use the degree-corrected Laplacian. Under the regularized degree-corrected Laplacian $L_{\alpha,\tau}$, this no longer occurs. Therefore, under the DCSBM, if a regularization parameter is used, normalization will be required regardless of the α value chosen. As mentioned previously, it

is standard to use the average node degree for τ , and we use this as its default value (Qing and Wang, 2024; Qin and Rohe, 2013).

Other than regularization, recent studies have shown that two additional modifications can improve performance in spectral clustering algorithms (Qing and Wang, 2024; Jin et al., 2022). First, we may consider using more than the K top eigenvectors for clustering. Second, we may scale the columns of \hat{U}_α by $\hat{\Lambda}_\alpha$, clustering the rows of $\hat{U}_\alpha \hat{\Lambda}_\alpha$. We study the effects of these two modifications using numerical studies in Appendix C.4.

4.2.1 Selecting Additional Eigenvectors

We first examine the effect of selecting additional eigenvectors for clustering. For the rank K population matrix L_α , only the top K eigenvectors contain information about the community structure. The remaining eigenvectors span the null space of the matrix, and thus do not provide any useful information for distinguishing between communities. This observation motivates the rationale behind selecting the top K eigenvectors of \hat{L}_α . We may view the top K eigenvectors as “signal” eigenvectors and the remaining $n - K$ to be “noise” eigenvectors.

As α increases, the factor of $D^{-\alpha}$ shrinks the eigenvalues of \hat{L}_α toward zero. This reduces the separation between two consecutive eigenvalues. If the noise level of the observed matrix is too large relative to this separation, it becomes more difficult to guarantee that the order of the eigenvalues remains the same between the population and observed matrices. In these situations, it becomes more likely that the K top eigenvectors contain noise eigenvectors. Selecting additional eigenvectors can mitigate this risk. This strategy is thus most effective when the smallest non-zero eigenvalue is small and when α is large. Conversely, selecting additional eigenvectors is less effective in the case when α is small, as the resulting gaps between the eigenvalues are relatively larger.

We adopt the convention given in Jin et al. (2022), who suggest investigating the ratio $\hat{\lambda}_{K+1}/\hat{\lambda}_K$. They analyze the weakness of the signal and include an additional eigenvector if and only if

$$1 - \hat{\lambda}_{K+1}/\hat{\lambda}_K \leq 0.1.$$

	α	Regularization	$\hat{\Lambda}$ Scaling	Total Eigenvectors	Normalization
ASC	0	No	0	K	Yes
RSC	1/2	Yes	0	K	Yes
DCSC	1	Yes	1	Adaptive (Sec. 4.2.1)	Yes

Table 4.1: A summary of the settings for selected algorithms.

4.2.2 Scaling Eigenvectors

We now consider the effect of scaling the columns of \hat{U}_α by $\hat{\Lambda}_\alpha$. We note that Algorithm 3 scales the i -th column of \hat{U}_α by $\hat{\lambda}_{\alpha,i}^{1/2}$. This is motivated by the fact that the spectral embedding central limit theorems of Chapter 2 are stated in terms of $\hat{U}_\alpha \hat{S}_\alpha^{1/2}$. However, there is some evidence that a more appropriate scaling is $\hat{U}_\alpha \hat{S}_\alpha$ (Jin et al., 2022; Qing and Wang, 2024). This scaling has two benefits. First, Abbe et al. (2020) states that the first order approximations of \hat{U}_{sym} for the symmetric Laplacian are given by

$$\hat{U}_{\text{sym},i*} \approx \frac{1}{\lambda_i} U_{\text{sym},i*}.$$

In other words, the eigenvectors associated with smaller eigenvalues are noisier approximations, and the noise level is on the order of λ_i^{-1} . Scaling the eigenvectors by the eigenvalues then aligns the noise levels. Second, when adding additional eigenvectors for clustering as considered previously, the effect of adding a “noise” vector is mitigated by down-weighting the corresponding column by a small eigenvalue.

In the theoretical setting, the eigenvalues are usually assumed to have the same order — recall that in the statement of Theorems 7 and 10, there is a factor of $\lambda_{1,P}/\lambda_{K,P}$ that affects the bound given in Algorithm 3. In this case, the exact choice of scaling factor generally does not affect the final clustering result. However, when the scales of the largest and smallest eigenvalue are different, choosing a scaling factor of $\hat{\Lambda}$ can empirically improve performance. We summarize the potential modifications to spectral clustering algorithms in Table 4.1 and provide our recommended choices for each parameter.

As a final remark, we note that the modifications described in this section do not affect the consistency of the algorithm, though it does affect the asymptotic order of the resulting bound. Intuitively, because $\hat{\Lambda}$ is a consistent estimate of Λ , $\hat{U}_{\alpha,\tau} \hat{\Lambda}_{\alpha,\tau}^\beta$ is a consistent estimate

of $U_{\alpha,\tau}\Lambda_{\alpha,\tau}^\beta$ for any β . Additionally, all eigenvalues beyond the K -th eigenvalue of $L_{\alpha,\tau}$ is identically 0, which “cancels” the effect of introducing additional eigenvectors in the analysis. We provide a short lemma and proof of this fact in Appendix C.

Algorithm 4 Degree Corrected Spectral Clustering (DCSC)

Require: Regularized Degree Corrected Laplacian, $\hat{L}_{\tau,\text{DC}}$, number of communities K

Ensure: Estimated community membership matrix \hat{Z}

- 1: Compute $1 - \hat{\lambda}_{K+1}/\hat{\lambda}_K$, and set $\kappa = K$ if it is greater than 0.1 and $\kappa = K + 1$ otherwise
 - 2: Compute $\hat{U}_\tau \in \mathbb{R}^{n \times \kappa}$, whose columns consist of the κ top eigenvectors of $\hat{L}_{\tau,\text{DC}}$.
 - 3: Compute $\hat{\Lambda}_\tau \in \mathbb{R}^{\kappa \times \kappa}$, a diagonal matrix whose diagonal entries consist of the κ top eigenvalues of $\hat{L}_{\tau,\text{DC}}$.
 - 4: Compute $\hat{X}_\tau = \hat{U}_\tau \hat{\Lambda}_\tau$
 - 5: Normalize the rows of \hat{X}_τ to obtain \hat{X}_τ^* .
 - 6: Apply k -means to the rows of \hat{X}_τ^* .
 - 7: **return** \hat{Z}
-

Chapter 5

Comparison of Methods

5.1 Theoretical Comparisons

In this section, we walk through a theoretical analysis of spectral clustering with the degree- α Laplacian. Consider an SBM with two communities. Given $\alpha \in [0, 1]$, assume the two asymptotic distributions from Theorem 9 are the true distributions of the rows of \hat{X}_α . In other words, they are multivariate normal distributions with known means and covariances. In this setting, any clustering algorithm may be thought of as deciding between

$$F_1 = N(\mu_1, \Sigma_1) \quad \text{and} \quad F_2 = N(\mu_2, \Sigma_2).$$

Additionally, the parameters of this distribution depend on α . It is natural to ask if certain choices of α yield a better ability to distinguish between the two distributions. Chernoff information provides a way to quantify the difficulty of distinguishing between two distributions.

Definition 11 (Chernoff Information). *Let F_1 and F_2 be two absolutely continuous multivariate distributions on \mathbb{R}^d and let f_1 and f_2 be their densities, respectively. Chernoff divergence is defined to be*

$$C_t(F_1, F_2) = -\log \left(\int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x) dx \right).$$

Chernoff information is defined to be

$$\begin{aligned} C(F_1, F_2) &= -\log \left(\inf_{t \in (0,1)} \int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x) dx \right) \\ &= \sup_{t \in (0,1)} \left(-\log \int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x) dx \right) \\ &= \sup_{t \in (0,1)} C_t(F_1, F_2) \end{aligned}$$

This quantity may be explicitly computed for multivariate normal distributions. Define $\Sigma_{12}(t) = t\Sigma_1 + (1-t)\Sigma_2$ and let $F_1 = N(\mu_1, \Sigma_1)$ and $F_2 = N(\mu_2, \Sigma_2)$. Then

$$C(F_1, F_2) = \frac{1}{2} \sup_{t \in (0,1)} \left((t(1-t)(\mu_1 - \mu_2)^\top \Sigma_{12}^{-1}(t)(\mu_1 - \mu_2)) + \log \frac{|\Sigma_{12}(t)|}{|\Sigma_1|^t |\Sigma_2|^{1-t}} \right). \quad (5.1)$$

Note that $(\mu_1 - \mu_2)^\top \Sigma_{12}^{-1}(t)(\mu_1 - \mu_2)$ is the square of the Mahalanobis norm with respect to $\Sigma_{12}(t)$. We thus rewrite Equation (5.1) as

$$C(F_1, F_2) = \frac{1}{2} \sup_{t \in (0,1)} \left(\left(t(1-t) \|\mu_1 - \mu_2\|_{\Sigma_{12}^{-1}(t)}^2 \right) + \log \frac{|\Sigma_{12}(t)|}{|\Sigma_1|^t |\Sigma_2|^{1-t}} \right).$$

As noted in [Cape et al. \(2019a\)](#), the Chernoff divergence falls into the family of f -divergences. Any f -divergence is invariant with respect to invertible linear transformations ([Liese and Vajda, 2006](#)), so the choice of Chernoff information is somewhat arbitrary. Other divergences, such as the Kullback-Liebler divergence, could be used for analysis, but the Chernoff divergence is chosen due to its relationship to the Bayes risk, which we make explicit below.

Let Y_1, \dots, Y_m be m independent and identically distributed variables drawn from a distribution F , which is equal to either F_1 or F_2 . We are interested in testing

$$H_0 : F = F_1 \quad \text{vs.} \quad H_1 : F = F_2.$$

The Neyman-Pearson Lemma ([Neyman and Pearson, 1933](#)) implies that given a threshold k_m , the most uniformly most powerful level α_m test is the likelihood ratio test, which rejects H_0 when

$$\frac{\prod_{i=1}^m f_1(y_i)}{\prod_{i=1}^m f_2(y_i)} \leq k_m.$$

Here α_m refers to the Type-I error, following standard notation. We note that it is unrelated to the exponent in the degree- α Laplacian. To avoid confusion, we will always use the subscript m when referring to Type-I error, and note that this notation only appears in this section. Given an α_m , this test minimizes the type-II error β_m . Assume that the prior probability of $H_0 = \pi$, i.e., the proportion of community 1 is π . For a given $\alpha_m^* \in (0, 1)$, let $\beta_m^* = \beta_m^*(\alpha_m^*)$ be the Type-II error associated with the likelihood ratio test such that the Type-I error is at most α_m^* . Then the Bayes risk in deciding between H_0 and H_1 is given by

$$\inf_{\alpha_m^* \in (0,1)} \pi \alpha_m^* + (1 - \pi) \beta_m^*.$$

Chernoff related this quantity to the Chernoff information through the following formula ([Chernoff, 1952](#)).

$$\lim_{m \rightarrow \infty} \frac{1}{m} \left[\inf_{\alpha_m^* \in (0,1)} \log (\pi \alpha_m^* + (1 - \pi) \beta_m^*) \right] = -C(F_1, F_2).$$

In other words, the Chernoff information between F_1 and F_2 is the exponential rate at which the Bayes risk decreases as $m \rightarrow \infty$. Higher Chernoff information thus reflects better ability to distinguish between the two distributions.

Returning to our example of a two community SBM, we may use the previous definitions to compute the Chernoff information for the distributions of the eigenvalues of the degree α Laplacian. Using the Theorems from [3.3](#), for the degree- α spectral embedding, we may compute the Chernoff information between $\Sigma_{\alpha,1}$ and $\Sigma_{\alpha,2}$. Here we remark that the Chernoff information is defined for fixed distributions F_1 and F_2 , and so the expression ρ_α also depends on n . For reasons that will become clear, we suppress this dependence for now.

$$\rho_\alpha = \sup_{t \in (0,1)} \left[t(1-t) \|\mu_1 - \mu_2\|_{\Sigma_{\alpha,12}^{-1}(t)}^2 + \frac{1}{2} \log \left(\frac{|\Sigma_{\alpha,12}(t)|}{|\Sigma_{\alpha,1}|^t |\Sigma_{\alpha,2}|^{1-t}} \right) \right].$$

From Theorem 9, the means of the normal distributions are given by the rows of

$$n^{\alpha+1/2} \rho_n^\alpha (T^{-\alpha} X_\alpha).$$

If $t_i \asymp n\rho_n$, then $(n^{\alpha+1/2} \rho_n^\alpha (T^{-\alpha} X_\alpha))_{i^*} \asymp n$. Thus, the quantity $\mu_1 - \mu_2$ is on the order of n . On the other hand, the covariance matrices are independent of n . Therefore, for large n , the second term in the previous equation is dominated by the first. We then have a large-sample approximation of the Chernoff information as

$$\rho_\alpha^* \approx \sup_{t \in (0,1)} \left[t(1-t) \|\mu_1 - \mu_2\|_{\Sigma_{\alpha,12}^{-1}(t)}^2 \right]. \quad (5.2)$$

Equation (5.2) implies that the Chernoff information increases as n increases. Consider L_α and $L_{\alpha'}$. The quantity $\rho_{\alpha,\alpha'}^*$ is independent of n , making it a natural choice for determining the relative efficiency of choosing the degree α Laplacian over the degree α' Laplacian. We now study the behavior of $\rho_{\alpha,\alpha'}^*$ under several two-community SBMs.

5.1.1 Two-Community Balanced SBM

Consider the two-community SBM model with

$$B = \begin{bmatrix} p & q \\ q & p \end{bmatrix}, \quad \pi = \left(\frac{1}{2}, \frac{1}{2} \right)$$

where $p > q$ and π is a $K \times 1$ vector such that π_i represents the proportion of nodes in community i . In this simple case, the probability of connection between two nodes depends only on whether they belong to the same community. Since P admits a simple eigendecomposition, we may compute a closed form for ρ_α . Via an application of Theorem 9, the rows of \hat{X}_α are asymptotically normal with means

$$\mu_1 = \begin{bmatrix} \sqrt{\frac{n(p+q)}{2(p+q)^\alpha}} \\ \sqrt{\frac{n(p-q)}{2(p+q)^\alpha}} \end{bmatrix} \quad \mu_2 = \begin{bmatrix} \sqrt{\frac{n(p+q)}{2(p+q)^\alpha}} \\ -\sqrt{\frac{n(p-q)}{2(p+q)^\alpha}} \end{bmatrix}$$

and covariances

$$\Sigma_{\alpha,1} = \begin{bmatrix} \frac{4^\alpha(\alpha-1)^2(p(1-p)+q(1-q))}{(p+q)^{2\alpha+1}} & -\frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} \\ -\frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} & \frac{4^\alpha(p(1-p)\gamma_2^2+q(1-q)\gamma_1^2)}{(p-q)(p+q)^{2\alpha+2}} \end{bmatrix}$$

and

$$\Sigma_{\alpha,2} = \begin{bmatrix} \frac{4^\alpha(\alpha-1)^2(p(1-p)+q(1-q))}{(p+q)^{2\alpha+1}} & \frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} \\ \frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} & \frac{4^\alpha(p(1-p)\gamma_2^2+q(1-q)\gamma_1^2)}{(p-q)(p+q)^{2\alpha+2}} \end{bmatrix},$$

where $\gamma_1 = a - \alpha(a - b) + b$ and $\gamma_2 = a - \alpha(a - b) + b$.

Using these quantities, $\rho_{\alpha,\alpha'}^*$ may be computed explicitly. Appendix D.1 contains further details on calculations.

$$\rho_{\alpha,\alpha'}^* = \frac{C_2\alpha^2 + C_1\alpha + C_0}{C_2\alpha'^2 + C_1\alpha' + C_0} \quad (5.3)$$

where

$$\begin{aligned} C_2 &= (p-q)^2(p^2 - p + q^2 - q) \\ C_1 &= -2(p-q)^2(p+q)(p+q-1) \\ C_0 &= (p+q)^2(p^2 - p + q^2 - q). \end{aligned}$$

Note that when $\alpha = \alpha'$, the ratio simplifies to 1, as expected. When $\alpha = 0$ (corresponding to the adjacency matrix) and $\alpha' = 1/2$ (corresponding to the symmetric Laplacian), the ratio simplifies to Equation 14 from [Cape et al. \(2019a\)](#). If we fix $\alpha' = 0$, we may determine when the ratio is largest, indicating the α at which the relative Chernoff information is maximized. For a fixed p and q , the ratio is clearly maximized at $-C_1/2C_2$. Thus, for each p and q the optimal α may be computed to be

$$\alpha_{\text{opt}} = \begin{cases} 0 & \text{if } -\frac{C_1}{2C_2} < 0 \\ 1 & \text{if } -\frac{C_1}{2C_2} > 1 \\ -\frac{C_1}{2C_2} & \text{otherwise} \end{cases} \quad (5.4)$$

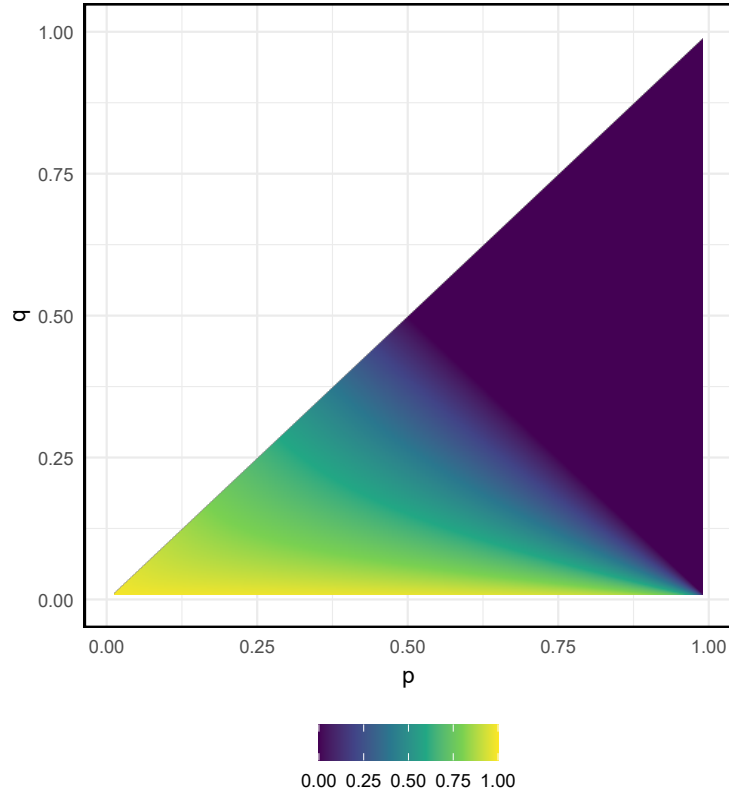


Figure 5.1: The optimal α for the two-community SBM model described in Section 5.1.1. The dark blue region corresponds to $\alpha = 0$, adjacency spectral clustering being preferred. The lighter colors represent higher α being preferred, with degree-corrected spectral clustering being preferred in very sparse regions.

Figure 5.1 shows a plot of this graph over the lower right half of the domain $[0, 1]^2$. At each point, the optimal α is computed using Equation (5.4). The most striking feature of the graph is the phase transition, corresponding to the case when $C_1 = 0$. This occurs when either $p = q$, which corresponds to the degenerate model, or when $p+q = 1$, which is visible as the diagonal line in the figure. Above this boundary, adjacency spectral clustering provides higher Chernoff information for clustering and below this boundary, some amount of degree correction is preferred. In general, as sparsity increases, increasing the degree-correction parameter α will provide the highest Chernoff information.

5.1.2 Two-Community Imbalanced SBM

We may also investigate this model when the classes are imbalanced, i.e. for a two-block SBM model with

$$B = \begin{bmatrix} p & q \\ q & p \end{bmatrix}, \quad \pi = (\pi_1, 1 - \pi_1)$$

In this case, a simple formula for ρ^* is not easily calculable. Instead, we may numerically solve the optimization problem in Equation (5.3) to get an estimate of ρ^* . For this setting, we compute the ratio of the Chernoff information for adjacency spectral clustering to the Chernoff information for degree- α spectral clustering for a fixed α when $\pi_1 = 1/3$ and $\pi_2 = 1/4$. Thus, adjacency spectral clustering is preferred when the value is larger than 1 and the degree α Laplacian is preferred when the value is less than 1.

Figure 5.2 shows 6 plots, corresponding α values of 0.25, 0.5 and 0.75 for $\pi = 1/3$ and $\pi_1 = 1/4$. We remark that the scale in this figure is different from that of Figure 5.1. In these plots, the color represents the ratio of the Chernoff information relative to the adjacency matrix, with values larger than one indicating that the adjacency matrix is preferred. The plots show again the degree α Laplacian is generally preferred when the graph is sparser. Furthermore, as α increases, the values near the bottom of the graph, corresponding to sparser regimes, decrease. This indicates again that higher values of α are preferred when the networks are sparse. Finally, as the class sizes become more imbalanced, the effect of α becomes less pronounced, and the relative Chernoff information becomes more balanced.

5.2 Numerical Comparisons

In this section, we review the numerical experiments conducted to compare the performance of several algorithms. In all experiments, for each parameterization, we run 50 replications and report the mean error rate. We compare our method, DCSC (Algorithm 4), with Regularized Spectral Clustering (RSC) (Qin and Rohe, 2013), Spectral Clustering on Ratios of Eigenvectors (SCORE) (Jin, 2015), Dual-Regularized Spectral Clustering (DRSC) (Qing and Wang, 2024), and Adjacency Spectral Clustering (ASC, Algorithm 2). Each of these algorithms was re-implemented in R for consistency.

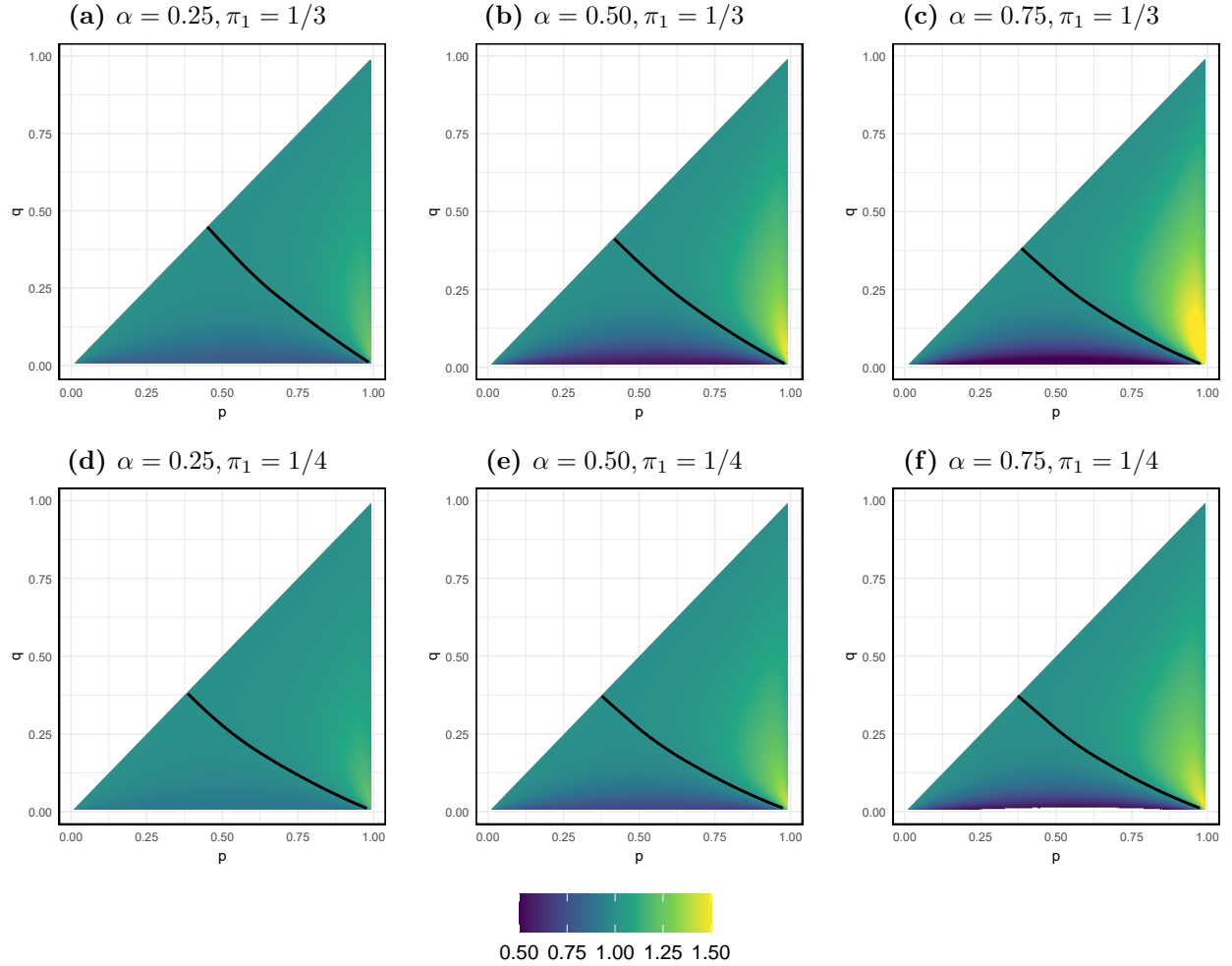


Figure 5.2: The ratio $\rho_{0,\alpha}^*$ for the imbalanced model described in Section 5.1.2. The values for α are $\alpha = 0.25$ (first column), $\alpha = 0.50$ (second column) and $\alpha = 0.75$ (third column). In the first row, $\pi_1 = 1/3$ and in the second row, $\pi_1 = 1/4$. Values larger than 1 represent parameter values where adjacency spectral clustering ($\alpha = 0$) is preferred. The black line corresponds to the level curve $\rho_{0,\alpha}^* = 1$.

In Experiment 1, we compare the algorithms when $K = 2$ and $K = 3$ with increasing n . The values used were $n \in \{100, 150, \dots, 500\}$. This reproduces the setup used in Experiment 1 of [Qing and Wang \(2024\)](#).

For Experiment 1A, we let $K = 2$ and assign each node to a community with equal probability. The community connection matrix is defined to be

$$B = \begin{bmatrix} 0.09 & 0.126 \\ 0.126 & 0.441 \end{bmatrix}.$$

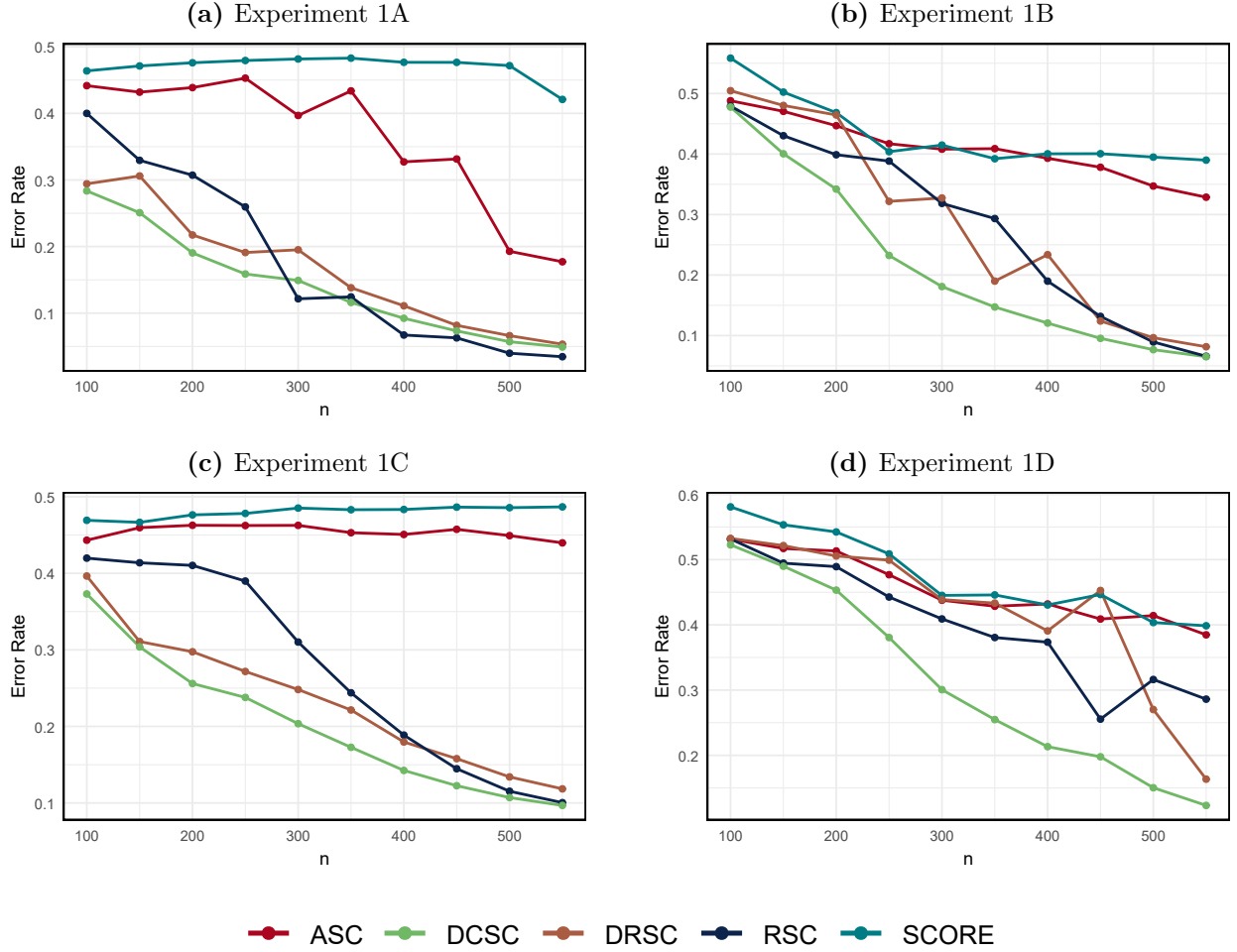


Figure 5.3: Numerical results of Experiment 1.

For Experiment 1B, we let $K = 3$ and assign each node to a community with equal probability. The community connection matrix is defined to be

$$B = \begin{bmatrix} 0.09 & 0.126 \\ 0.126 & 0.441 \end{bmatrix}.$$

Experiments 1C and 1D share the same setup as experiments 1A and 1B, respectively, with the addition of degree-correction parameters. The degree correction parameters are selected from a $\text{Unif}(0.7, 1)$ distribution. We remark that in these experiments, the model is not strictly assortative, as $b_{11} < b_{12}$. However, because the matrix B is positive-definite, the model may be viewed as predominantly assortative.

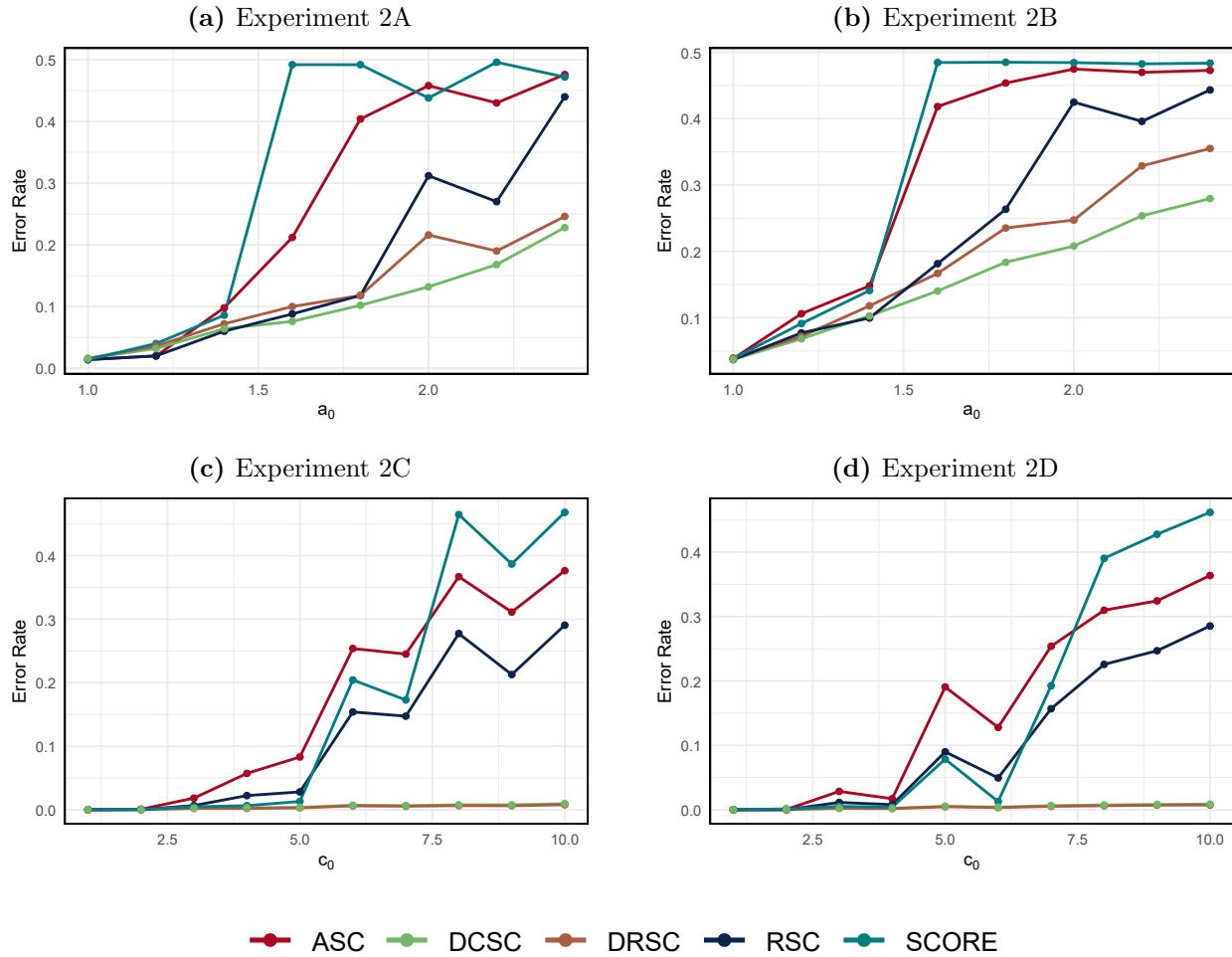


Figure 5.4: Numerical results of Experiment 2.

The results of Experiment 1 are shown in Figure 5.3. From Figure 5.3a, we see that when $K = 2$, under the SBM, the error rate of DRSC, ASC, DCSC decrease quickly, while the error rates for SCORE and ASC remain high. However, DCSC outperforms all other algorithms when n is small and remains competitive as n increases. In the DCSBM and $K = 3$ cases, the advantage of DCSC over all other algorithms is clear, and it achieves the lowest error rate across all n in each case. From Figures 5.3a and 5.3c, we note that DCSC still remains the best algorithm even in the SBM case, although its development was motivated by the degree-corrected model. This shows that DCSC can be effective across a wide range of models.

In Experiment 2, we compare the performance of the algorithms in block models when the ratios of the entries of B are changed and when the community sizes are imbalanced. We fix $K = 2$ and $n = 500$ for each experiment. For Experiments 2A and 2B, we assign each node to either community with equal probability and set

$$B = \begin{bmatrix} 0.25 & 0.15/a_0 \\ 0.15/a_0 & 0.225/a_0^2 \end{bmatrix},$$

for $a_0 \in \{1, 1.2, \dots, 2.4\}$. In Experiment 2A we assume no degree-correction parameter and in 2B we select the degree-correction parameters from a $\text{Unif}(0.7, 1)$ distribution. This reproduces the setup used in Experiment 2A of [Qing and Wang \(2024\)](#).

In Experiment 2C and 2D we set

$$B = \begin{bmatrix} 0.36 & 0.324 \\ 0.324 & 0.729 \end{bmatrix}.$$

We assign $\text{round}(n/(c_0 + 1))$ nodes to community 1 and the remaining nodes to community 2 for $c_0 \in \{1, 2, \dots, 10\}$. As c_0 increases, the community sizes become increasingly unbalanced, and community detection becomes more difficult. In Experiment 2C we assume no degree-correction parameter and in 2D we select the degree-correction parameters from a $\text{Unif}(0.7, 1)$ distribution. This reproduces the setup used in Experiment 2D of [Qing and Wang \(2024\)](#).

The results of Experiment 2 are shown in Figure 5.4. Figures 5.4a and 5.4b show that DCSC clearly outperforms all other algorithms in this setting. Additionally, Figures 5.4c and 5.4d demonstrate that DCSC and DRSC both are unaffected by community imbalance, while the performance of the other algorithms degrades as the imbalance becomes severe. In all cases, the presence of degree heterogeneity affects the performance of the algorithms negatively, although it affects RSC, ASC, and SCORE more severely.

In Experiment 3, we analyze the effect of degree heterogeneity on performance of the algorithms in homogeneous block models. We may parameterize homogeneous block models in the following way. Define I_K to be the $K \times K$ identity matrix and let $\mathbf{1}_K$ be the $K \times 1$ vector consisting of all ones. For two parameters $\gamma, \lambda \in (0, 1)$ we may define a community

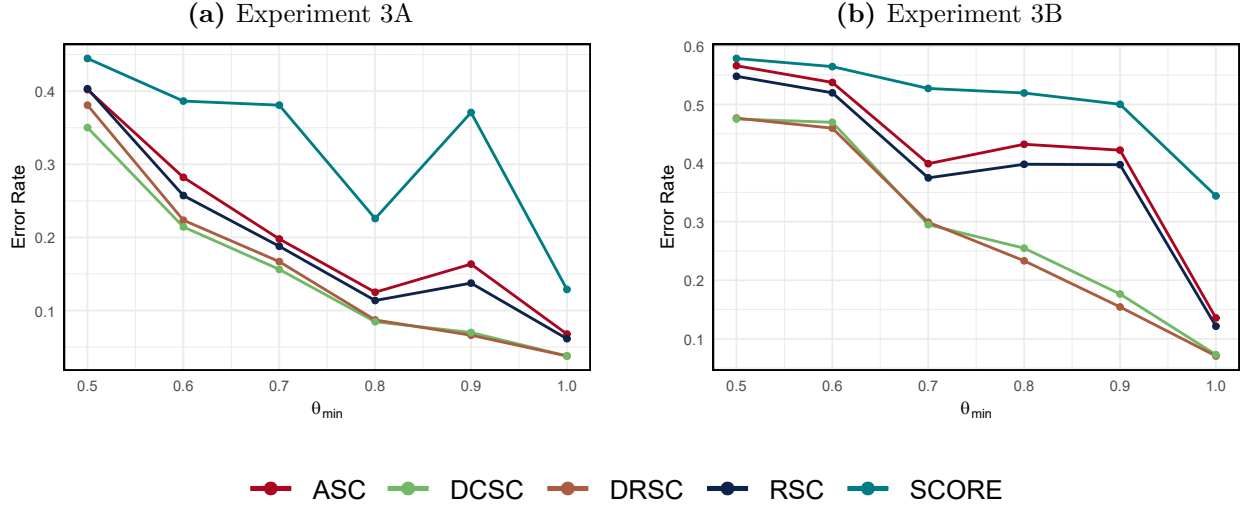


Figure 5.5: Numerical results of Experiment 3.

connection matrix B by

$$B = \gamma (\lambda I_K + (1 - \lambda) \mathbf{1}_K \mathbf{1}_K^\top).$$

Without degree-correction parameters, nodes within a community are connected with probability γ and nodes between communities are connected with probability $\gamma(1 - \lambda)$. In this model, γ controls the relative sparsity of the graph, with smaller γ corresponding to sparser graphs. λ controls the similarity of within-community and between-community interactions, with smaller λ corresponding to more similar behavior. In general, when γ or λ is close to 0, community detection becomes more difficult.

For Experiment 3A, $K = 3$ and $n = 300$, and for Experiment 3B, $K = 4$ and $n = 400$. The community connection matrix is a homogeneous matrix with $\gamma = 0.6$ and $\lambda = 0.3$. The degree correction parameters are chosen from a $\text{Unif}(\theta_{\min}, 1)$ distribution, where θ_{\min} is chosen from $\{0.5, 0.6, \dots, 1\}$. Thus, increasing θ_{\min} corresponds to less degree heterogeneity, and community detection is easier.

The results of Experiment 3 are shown in Figure 5.5. We see that for all algorithms, decreasing degree heterogeneity improves performance. Additionally, DRSC and DCSC show superior performance under this setup, and we can see that increasing the number of groups further amplifies the performance differences of these two methods when compared to the others.

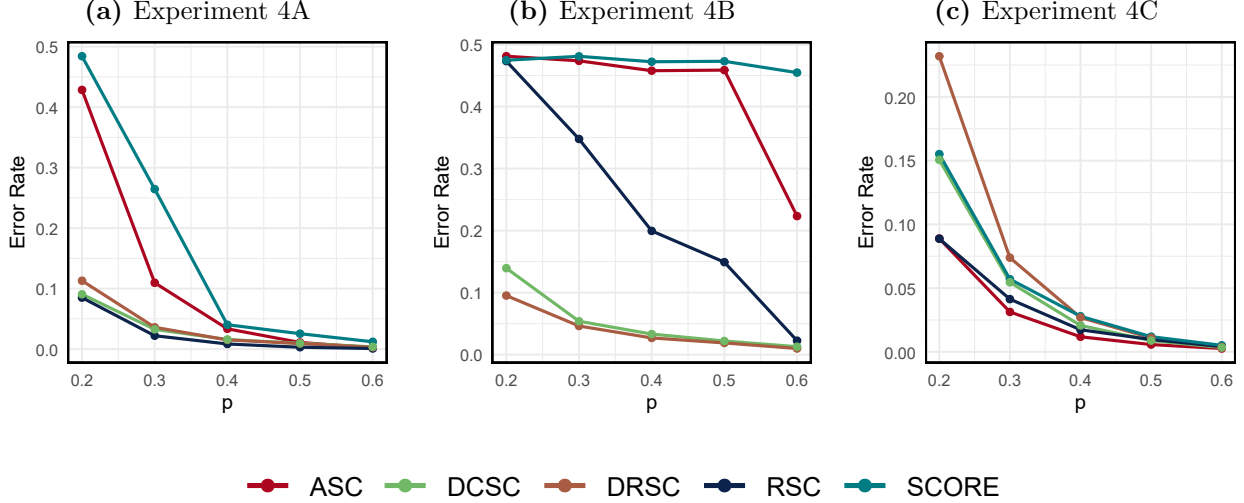


Figure 5.6: Numerical results of Experiment 4

In Experiment 4, we analyze a block model whose community connection matrix is of the form

$$B = \begin{bmatrix} p & q \\ q & q \end{bmatrix},$$

known as a core-periphery (Cape et al., 2019a) or planted clique (Lei and Rinaldo, 2015) model. In this model, there is a “core” network, denoted by community 1, that is connected to itself with high probability. The other community is weakly connected to itself and to community 1. We fix $n = 400$, set $q = p/4$, p varies from $\{0.2, 0.3, \dots, 0.6\}$. In Experiment 3A, we consider balanced community sizes. In Experiment 3B, we consider a large core, where the probability of being in community 1 is $3/4$. In Experiment 3C, we consider a small core, where the probability of being in community 1 is $1/4$.

We can see in this setup that DCSC and DRSC perform well when there are balanced communities, outperforming the other three algorithms. In a large core network, DCSC and DRSC outperform the other three algorithms, indicating that larger degree-correction is appropriate in this case. In the small core network, ASC outperforms the other algorithms, indicating that a smaller degree-correction is preferred.

Algorithm	Karate	Football	Polbooks	Polblogs	UKFaculty	Caltech	UA
n	34	110	92	1222	79	590	329
ASC	0	4	3	64	6	178	88
RSC	0	5	3	64	0	170	86
SCORE	0	14	1	58	34	181	128
DRSC	3	3	3	61	2	158	88
DCSC	0	3	4	59	2	96	70

Table 5.1: Algorithm comparison on real world data sets

5.3 Real Data

In this section, we analyze real network datasets and compare the algorithms. The datasets are used in standard benchmarks in community detection (Jin et al., 2022; Qing and Wang, 2024). For each of the datasets, the true communities are known and labeled by the researchers. See Appendix D.2 for a brief description of each dataset.

The results are summarized in Table 5.1, and the smallest number of errors for each graph is bolded. The table suggests that DCSC is competitive with the listed algorithms for each of the datasets, and can outperform the other methods in the Caltech and UA datasets. This again supports the idea that the advantage of DCSC is apparent when the network is low signal, as in Caltech dataset, or when the community sizes are unbalanced, as in the UA dataset.

5.3.1 University of Arizona KMAP Dataset

In this section, we return to the University of Arizona KMAP dataset and analyze the full network. The dataset consists of 7246 nodes and 2134 connected components. For computational efficiency, we first remove any nodes of degree less than five and consider the largest connected component of the resulting subgraph. We then remove any communities with fewer than two total nodes (see Section 2.2). The final graph consists of 2742 nodes in 166 communities. Although there is a ground-truth labeling available for this graph in the form of the departments, the data are too noisy for any of the previously described algorithms to achieve good performance. However, it is known that social networks often

exhibit hierarchical community structures, in which smaller communities combine to form larger ones. Using the ideas of completeness and Newman’s modularity, we may try to detect superstructures in the KMAP dataset that consists of groups of academic departments.

We first introduce the ideas of Newman’s modularity and completeness. Each of the algorithms outlined previously require that the number of communities K be known. In real-world networks, K is often unknown and must be estimated from the data. A strategy for estimating K based on Newman’s modularity is outlined below (Qing and Wang, 2024).

Let Z be a community membership matrix and recall that we use the notation $z_i = k$ to denote that node i is in community k . Let m be the total number of edges in the graph,

$$m = \frac{1}{2} \sum_{i=1}^n d_i.$$

Then for a given adjacency matrix A and labeling Z , Newman’s modularity is defined to be

$$Q = \frac{1}{2m} \sum_{i,j} \left(a_{ij} - \frac{d_i d_j}{2m} \right) \delta_{z_i z_j}.$$

Intuitively, the second term in the summand represents the expected number of edges between nodes i and j if the graph had no community structure. If this were true, the first and second terms would cancel and the Newman’s modularity would evaluate to 0. On the other hand, if we tend to observe more edges between nodes of the same community, then the first term would be larger than the first and the Newman’s modularity would be positive. In general, the quantity is bounded between -1 and 1 , and larger positive values indicate that the community structure is supported by the edge structure. We remark that defining modularity in this way inherently assumes the data are assortative.

We may now develop a simple procedure to estimate K for any algorithm that produces an estimated community labeling. For an algorithm \mathbb{A} and a hypothesized K , let \hat{Z} be the estimated community label and define a score

$$Q(\mathbb{A}, K) = \frac{1}{2m} \sum_{i,j} \left(a_{ij} - \frac{d_i d_j}{2m} \right) \delta_{\hat{z}_i \hat{z}_j}.$$

We may then define a set of candidates for the number of communities $\mathcal{K} \subset \mathbb{N}$ and let

$$\hat{K} = \arg \max_{k \in \mathcal{K}} Q(\mathbb{A}, k).$$

We now introduce the idea of completeness, a measure of how well a partition of nodes respects existing community labels. Let C be the set of true clusters and K be the set of predicted clusters. Let $n_{c,k}$ be the number of nodes in community c assigned to community k and define

$$H(C | K) = - \sum_{k \in K} \frac{n_k}{n} \sum_{c \in C} \frac{n_{c,k}}{n_k} \log \left(\frac{n_{c,k}}{n_k} \right)$$

and

$$H(C) = - \sum_{c \in C} \frac{n_c}{n} \log \left(\frac{n_c}{n} \right).$$

Then completeness is

$$1 - \frac{H(C | K)}{H(C)}.$$

Roughly speaking, completeness measures how well the true community labels are “kept together” in K , with 1 being a perfect score.

We first apply DCSC with the number of communities $K \in \{2, \dots, 25\}$. For each estimated community labeling, we compute both Newman’s modularity and the completeness score with respect to the departments. Figure 5.7a shows that Newman’s modularity increases sharply as K increases from 2 to 4. Beyond this point, the modularity continues to increase, but only marginally. Figure 5.7b shows that the completeness is maximized when $K = 4$. These observations suggest that the departments naturally organize into 4 broad categories. The full dataset, colored according to these categories, is shown in Figure 5.8a. The resulting communities are well-distinguished, providing empirical evidence of a higher-level organizational structure of the academic departments. Furthermore, a qualitative review of the groups show that the higher-level structures generally correspond to well-known academic divisions. The four categories are Arts/Administration, STEM, Medicine, and Astronomy/Steward Observatory. See Appendix D.3 for a complete listing of each department and the assigned group.

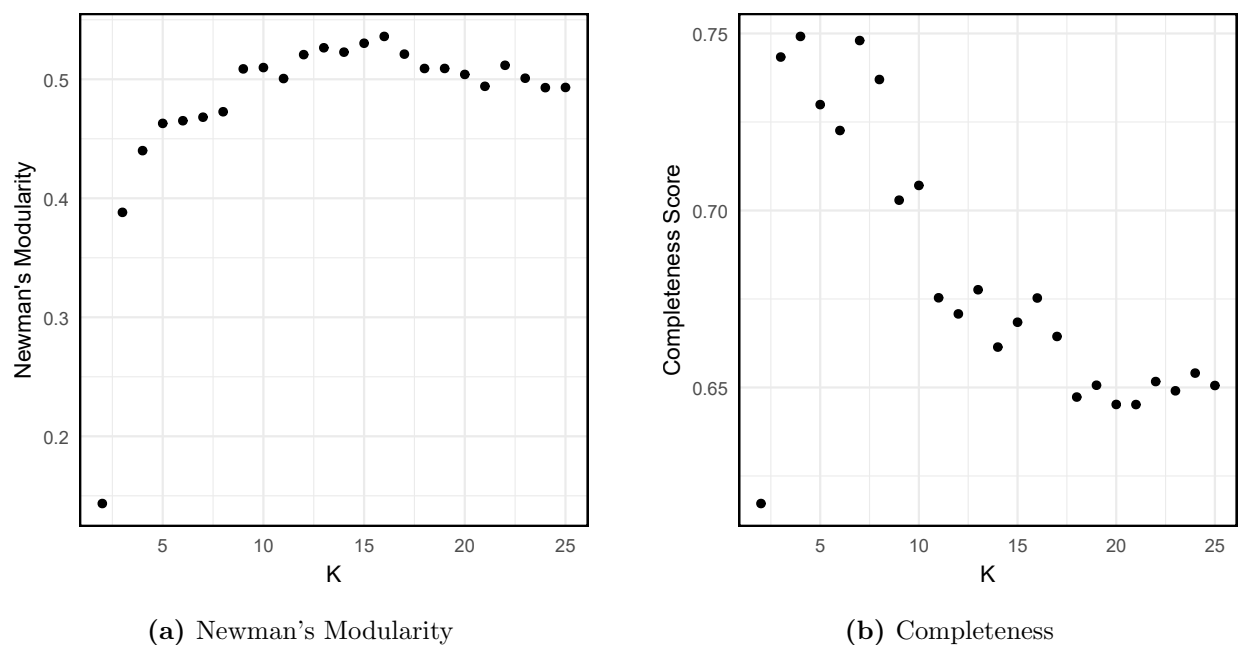


Figure 5.7: Scores for the estimated community labels on the KMAP dataset.

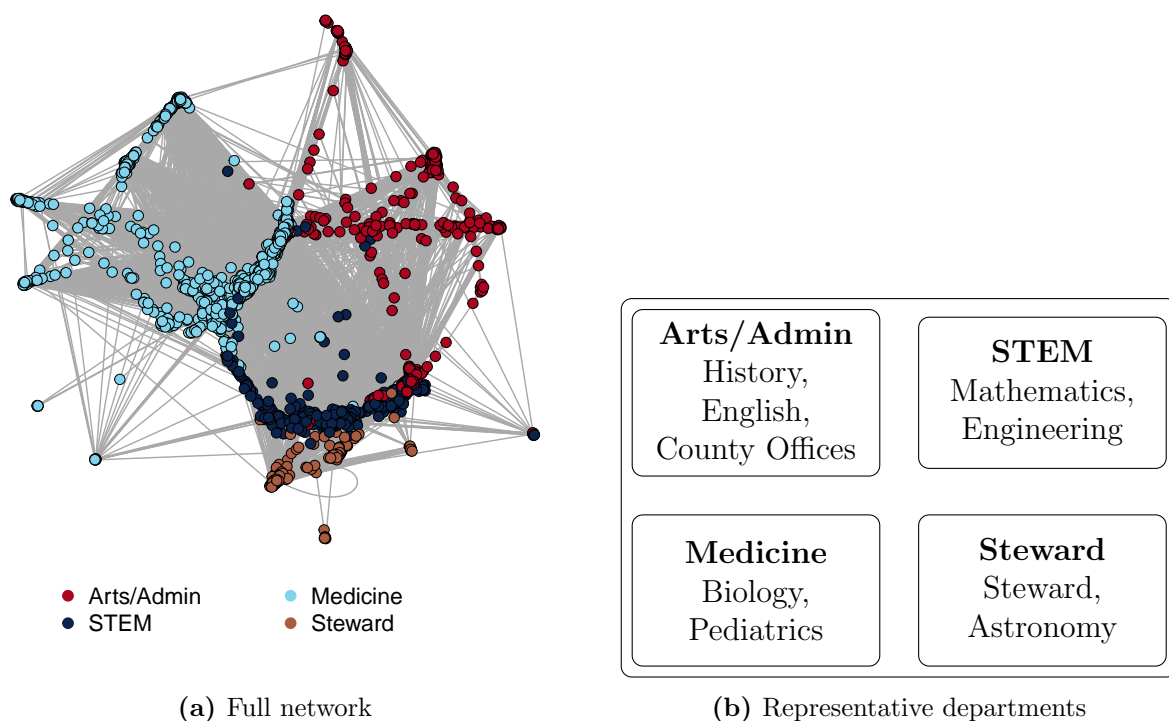


Figure 5.8: A visualization of the full KMAP network: **a** A diagram of each node colored by the community labeling that results from DCSC with $K = 4$. **b** Examples of representative departments for each grouping.

Chapter 6

Discussion and Conclusion

This dissertation presents a family of graph matrices, the degree- α Laplacian, that generalizes the adjacency and symmetric Laplacian matrices. We study the spectral properties of this family, studying how degree-normalization can influence the induced spectral embeddings and spectral clustering techniques. At the population level, these matrices preserve the low-rank structure of the SBM and the DCSBM. At the sample level, we establish that the eigenvectors are uniformly consistent estimators of the latent positions, generalizing the existing results on adjacency and Laplacian embeddings.

We focus on the case $\alpha = 1$, the degree-corrected Laplacian, which additionally has the property of correcting for the degree heterogeneity introduced by the DCSBM. Using this Laplacian we have introduced Degree-Corrected Spectral Clustering, which utilizes the eigenvectors of the degree-corrected Laplacian to perform community detection. The algorithm is robust across a wide range of conditions and can outperform other algorithms. Numerical studies show that it enjoys an advantage when communities are imbalanced or when networks are sparse. This conclusion is supported by case studies applied to real-world network data that display these features. Notably, DCSC also displays competitive performance under the standard SBM, highlighting the value of the broader degree- α framework even in simpler settings.

There are several natural extensions of this work. The introduction of the ridge regularizer changes the asymptotic distribution of the spectral embedding for the Laplacian matrix. Examining the limiting behavior of the regularized Laplacian would be useful for understanding how the distributions change with τ , and could provide additional insights into tuning the parameter.

Additionally, the results on spectral embedding arise from the study of the random dot product graph, which produces expected adjacency matrices that are necessarily positive-definite. This generally corresponds to the case where graphs are assortative, in which intra-community connections are more likely than inter-community connections. The generalized random dot product graph (GRDPG), introduced by [Rubin-Delanchy et al. \(2022\)](#), relaxes this restriction by allowing expected adjacency matrices with negative eigenvalues, accommodating disassortative networks. Many of the results in this thesis could likely be extended to this setting, broadening the applicability of the theory.

We have shown eigenvectors of the degree- α Laplacians have rows that are asymptotically normal. Based on this assumption, we may consider using discriminant analysis instead of k -means to separate the data points. In particular, quadratic discriminant analysis (QDA) is appropriate when the covariance matrices are not assumed to be identical. Given explicit formulas of the asymptotic covariance matrices, it would be interesting to seek examples in which the asymptotic distributions display high heteroscedasticity to determine if QDA shows better classification results than standard k -means.

Finally, the effect parameter α on clustering was studied through simulation and the lens of Chernoff information. These results lead naturally to two questions for future investigation. The first is whether the results of clustering is sensitive to the choice of α . The second is whether there is a data-driven method for tuning this parameter. A naive node-splitting technique for cross-validation technique is difficult, because splitting the data inherently severs the connections between the folds. Instead, a network cross-validation technique could be implemented to tune α ([Chen and Lei, 2018](#)).

Appendix A

Appendix to Chapter 2

A.1 Proof of Theorem 2

We begin by introducing some useful notation and technical lemmas. We use $[n]$ to denote the set of integers $\{1, \dots, n\}$. A partition of $[n]$ is a collection of nonempty, mutually disjoint subsets of $[n]$ whose union is $[n]$. Every equivalence relation on a set defines a partition of that set, and vice versa. For a matrix Q , we define its i -th row and j -th row to be equivalent if they are identical. Using this equivalence relation, any matrix Q with n rows defines a partition of $[n]$.

Lemma 12. *An $n \times K$ community membership matrix Z defines a partition of $[n]$ based on row equivalence. Two community membership matrices Z and \tilde{Z} define the same partition if and only if $\tilde{Z} = ZP$ for a permutation matrix P .*

Proof of Lemma 12: Define subset $C_k \subset [n]$ as $\{i \in [n] \mid Z_{ik} = 1\}$. It is straightforward to verify that $\{C_1, \dots, C_K\}$ is a partition of $[n]$. Similarly, we can define another partition $\{\tilde{C}_1, \dots, \tilde{C}_K\}$ based on \tilde{Z} . The partitions $\{C_1, \dots, C_K\}$ and $\{\tilde{C}_1, \dots, \tilde{C}_K\}$ are identical if and only if there exists a permutation $\pi : [K] \rightarrow [K]$ such that $\tilde{C}_{\pi(k)} = C_k$. This permutation π induces a permutation matrix P with $P_{k,\pi(k)} = 1$, and it follows that $\tilde{Z} = ZP$. \square

Lemma 13. *Let Z and \tilde{Z} be two $n \times K$ community membership matrices, and R and \tilde{R} be two full row-rank $K \times K'$ matrices where $K' \geq K$. The matrices ZR and $\tilde{Z}\tilde{R}$ define the same partition of $[n]$ based on row equivalence if and only if $\tilde{Z} = ZP$ for a permutation matrix P .*

Proof of Lemma 13: All rows of R are distinct, since R has full row rank. The rows of ZR are replications of rows of R , and the i -th row of ZR is identical to the k -th row of R

if and only if $Z_{ik} = 1$. As a result, ZR and Z define the same partition of $[n]$ based on row equivalence. The same conclusion holds for $\tilde{Z}\tilde{R}$ and \tilde{Z} .

Thus, ZR and $\tilde{Z}\tilde{R}$ define the same partition of $[n]$ if and only if Z and \tilde{Z} define the same partition, and by Lemma 12, if and only if $\tilde{Z} = ZP$ for a permutation matrix P . \square

For a matrix Q , we define its i -th row and j -th row to be proportionally equivalent if they are proportional to each other up to a positive scaling factor. Any matrix Q with n rows defines a partition of $[n]$ via this proportional equivalence.

Lemma 14. *The matrices Z , \tilde{Z} , R and \tilde{R} are defined as in Lemma 13. Θ and $\tilde{\Theta}$ are two $n \times n$ positive diagonal matrices. The matrices ΘZR and $\tilde{\Theta}\tilde{Z}\tilde{R}$ define the same partition of $[n]$ based on row proportional equivalence if and only if $\tilde{Z} = ZP$ for a permutation matrix P . Moreover, if $\Theta ZR = \tilde{\Theta}\tilde{Z}\tilde{R}$, and $\{C_1, \dots, C_K\}$ is the partition defined by ΘZR and $\tilde{\Theta}\tilde{Z}\tilde{R}$, then $\Theta_{ii}^{-1}\tilde{\Theta}_{ii} = \Theta_{jj}^{-1}\tilde{\Theta}_{jj}$ whenever $i, j \in C_k$.*

Proof of Lemma 14: No two rows of R are proportional to each other since R has full row rank. Therefore, the partition defined by ZR via row equivalence and the partition defined by ΘZR via row proportional equivalence are the same, and the first conclusion follows Lemma 13. To show the second statement, we write

$$ZR = \Theta^{-1}\tilde{\Theta}\tilde{Z}\tilde{R},$$

where $\Theta^{-1}\tilde{\Theta}$ is a positive diagonal matrix. It is straightforward to check when $i, j \in C_k$, the i -th and j -th rows of ZR are identical, and the i -th and j -th rows of $\tilde{Z}\tilde{R}$ are identical, which implies $\Theta_{ii}^{-1}\tilde{\Theta}_{ii} = \Theta_{jj}^{-1}\tilde{\Theta}_{jj}$. \square

We now prove Theorem 2. To show the “if” part, we calculate

$$\begin{aligned} \tilde{\Theta}\tilde{Z}\tilde{B}\tilde{Z}^\top\tilde{\Theta} &= \Theta \text{diag}(ZD^{-1}\mathbf{1}_K) ZPP^\top DBDPP^\top Z^\top \text{diag}(ZD^{-1}\mathbf{1}_K) \Theta \\ &= \Theta \text{diag}(ZD^{-1}\mathbf{1}_K) ZDBDZ^\top \text{diag}(ZD^{-1}\mathbf{1}_K) \Theta. \end{aligned}$$

To obtain Equation (2.3), it suffices to show

$$\text{diag}(ZD^{-1}\mathbf{1}_K) ZD = Z.$$

We verify this by checking each entry. As both $\text{diag}(ZD^{-1}\mathbf{1}_K)$ and D are diagonal matrices, the (i, k) entry on the left is

$$[\text{diag}(ZD^{-1}\mathbf{1}_K) ZD]_{ik} = [\text{diag}(ZD^{-1}\mathbf{1}_K)]_{ii} Z_{ik} D_{kk}.$$

When $z_i \neq k$, both $[\text{diag}(ZD^{-1}\mathbf{1}_K) ZD]_{ik}$ and Z_{ik} are zero. When $z_i = k$, $Z_{ik} = 1$, it is easy to check $[\text{diag}(ZD^{-1}\mathbf{1}_K)]_{ii} = D_{kk}^{-1}$, so

$$[\text{diag}(ZD^{-1}\mathbf{1}_K)]_{ii} Z_{ik} D_{kk} = D_{kk}^{-1} Z_{ik} D_{kk} = Z_{ik}.$$

For the “only if” portion, we first show that Equation (2.3) and Condition 2 imply that

$$\Theta Z B Z^\top \Theta = \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta}. \quad (\text{A.1})$$

Note that Equation (2.3) implies that there exists a diagonal matrix C such that

$$\Theta Z B Z^\top \Theta = \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} + C,$$

which leads to

$$\Theta^{-1} C \Theta^{-1} = Z B Z^\top - \Theta^{-1} \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} \Theta^{-1}. \quad (\text{A.2})$$

Let e_i denote the i -th coordinate vector in \mathbb{R}^n , defined as the vector with a 1 in the i -th position and 0 elsewhere. Let i and j be two nodes with $z_i = z_j$. Multiplying $e_i - e_j$ on both sides of (A.2), we have

$$\begin{aligned} \Theta^{-1} C \Theta^{-1} (e_i - e_j) &= \left(Z B Z^\top - \Theta^{-1} \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} \Theta^{-1} \right) (e_i - e_j) \\ &= -\Theta^{-1} \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} \Theta^{-1} (e_i - e_j), \end{aligned}$$

because $Z^\top (e_i - e_j) = 0$. The last equation further implies

$$\tilde{\Theta}^{-1} C \Theta^{-1} (e_i - e_j) = -\tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} \Theta^{-1} (e_i - e_j) = \tilde{Z} w, \quad (\text{A.3})$$

where $w = -\tilde{B}\tilde{Z}^\top\tilde{\Theta}\Theta^{-1}(e_i - e_j)$.

If w has a nonzero component, say $w_l \neq 0$, then $u = \tilde{Z}w$ has at least 3 nonzero components by Condition 1 since $u_s = w_l$ for all nodes s with $\tilde{z}_s = l$. In contrast, the left-hand side of equation (A.3) has at most 2 nonzero components. Thus, equation (A.3) indicates $w = 0$, and hence $C_{ii} = C_{jj} = 0$. We can repeat the argument with arbitrary i and j from the same community and conclude $C = 0$.

To complete the proof, we now show that $\Theta Z B Z^\top \Theta = \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta}$ implies $\tilde{Z} = Z \Pi$, $\tilde{B} = \Pi^\top D B D \Pi$, and $\tilde{\Theta} = \Theta \text{diag}(Z D^{-1} \mathbf{1}_K)$.

First, we calculate the rank of $\Theta Z B Z^\top \Theta$ as

$$\text{rank}(\Theta Z B Z^\top \Theta) = \text{rank}(Z B Z^\top) = \text{rank}(B) = K.$$

The first equal sign follows that Θ is a positive definite diagonal matrix. The second equal sign holds because the membership matrix Z contains $K \times K$ identity matrix I_K as a submatrix (up to a permutation).

Now consider the eigenvalue decomposition $\Theta Z B Z^\top \Theta = U \Lambda U^\top$, where Λ is a $K \times K$ diagonal matrix that contains all nonzero eigenvalues of $\Theta Z B Z^\top \Theta$, U is a $n \times K$ matrix whose columns are eigenvectors corresponding to nonzero eigenvalues. By the eigenvalue decomposition, we can write

$$U = \Theta Z B Z^\top \Theta U \Lambda^{-1} = \Theta Z B Z^\top \Theta U \Lambda^{-1} = \Theta Z R, \quad (\text{A.4})$$

where $R = B Z^\top \Theta U \Lambda^{-1}$ is a full rank $K \times K$ matrix. We conduct similar operations with $(\tilde{Z}, \tilde{\Theta}, \tilde{B})$, and express U via both parameter system

$$U = \Theta Z R = \tilde{\Theta} \tilde{Z} \tilde{R}, \quad (\text{A.5})$$

where $\tilde{R} = \tilde{B} \tilde{Z}^\top \tilde{\Theta} U \Lambda^{-1}$. (A.5) implies that $\Theta Z R$ and $\tilde{\Theta} \tilde{Z} \tilde{R}$ define the same partition via row proportional equivalence. By Lemma 14, we have $\tilde{Z} = Z P$ for a permutation matrix P . In addition, we denote the partition by $\{C_1, \dots, C_K\}$ where $C_k = \{i \in [n] \mid z_i = k\}$. Lemma 14 also implies that $\Theta_{ii}^{-1} \tilde{\Theta}_{ii} = \Theta_{jj}^{-1} \tilde{\Theta}_{jj}$, when $i, j \in C_k$.

Let D be a $K \times K$ diagonal matrix, where $D_{kk} = \tilde{\Theta}_{ii}^{-1} \Theta_{ii}$ for any i with $z_i = k$, which is well-defined by the previous paragraph. It can be verified directly that $\Theta^{-1} \tilde{\Theta} = \text{diag}(ZD^{-1} \mathbf{1}_K)$, and $\text{diag}(ZD^{-1} \mathbf{1}_K) Z = ZD^{-1}$.

Note that we have shown $\tilde{Z} = Z\Pi$. Equation (A.1) with $\tilde{Z} = Z\Pi$ leads to

$$\begin{aligned} ZBZ^\top &= \Theta^{-1} \tilde{\Theta} \tilde{Z} \tilde{B} \tilde{Z}^\top \tilde{\Theta} \Theta^{-1} \\ &= \Theta^{-1} \tilde{\Theta} Z \Pi \tilde{B} \Pi^\top Z^\top \tilde{\Theta} \Theta^{-1} \\ &= \text{diag}(ZD^{-1} \mathbf{1}_K) Z \Pi \tilde{B} \Pi^\top Z^\top \text{diag}(ZD^{-1} \mathbf{1}_K) \\ &= ZD^{-1} \Pi \tilde{B} \Pi^\top D^{-1} Z^\top, \end{aligned}$$

which indicates $B = D^{-1} \Pi \tilde{B} \Pi^\top D^{-1}$, or equivalently $\tilde{B} = \Pi^\top D B D \Pi$. We have thus shown that if two parameter sets produce the same expected adjacency matrices, then $\tilde{Z} = Z\Pi$, $\tilde{B} = \Pi^\top D B D \Pi$, and $\tilde{\Theta} = \Theta \text{diag}(ZD^{-1} \mathbf{1}_K)$. \square

Appendix B

Appendix to Chapter 3

B.1 Proofs of Theorem 8 and 9

We first restate some definitions for convenience. We define P to be an expected adjacency matrix and A to be a realization of a random graph with expected value P . We use t_i to denote the expected degree of node i , $t_i = \sum_j p_{ij}$ and d_i to denote the observed degree, $d_i = \sum_j a_{ij}$. Then T and D to denote the diagonal matrices whose ii -th entry is given by t_i and d_i , respectively.

The expected adjacency matrix is given by P . We let X be the $n \times K$ matrix consisting of the latent positions, i.e., $P = XX^\top$. The symmetric Laplacian is given by $L_{\text{sym}} = T^{-1/2}PT^{-1/2}$. We define $\tilde{\Lambda} \in \mathbb{R}^{K \times K}$ to be a diagonal matrix whose diagonal entries consist of the K largest eigenvalues of L_{sym} and $\tilde{U} \in \mathbb{R}^{n \times K}$ to be a matrix whose columns consist of the corresponding eigenvectors. We then define $\tilde{X} = \tilde{U}\tilde{\Lambda}^{1/2}$ so that $L_{\text{sym}} = \tilde{X}\tilde{X}^\top$, and \tilde{X} may be thought of as the latent positions of L_{sym} .

The degree α Laplacian is given by $L_\alpha = T^{-\alpha}PT^{-\alpha}$. We define $X_\alpha = T^{1/2-\alpha}\tilde{X}$, so that

$$X_\alpha(X_\alpha)^\top = T^{1/2-\alpha}\tilde{X}\tilde{X}^\top T^{1/2-\alpha} = T^{1/2-\alpha}T^{-1/2}PT^{-1/2}T^{1/2-\alpha} = L_\alpha$$

and X_α may be thought of as the latent positions of L_α . Each of these definitions extend naturally to their observed counterparts. We finally note that the graph generation process outlined in Section 2.4 implies that $\|X\|_{2,\infty} = \Theta(\rho_n^{1/2})$. Assumption 4 implies that $t_i = \Theta(n\rho_n)$, and $\tilde{X} = T^{-1/2}X$, so $\|\tilde{X}\|_{2,\infty} = \Theta(n^{-1/2})$.

We begin with two technical lemmas, which are proven in Section B.2. The first lemma is a bound on the deviations of the observed degrees from the true degrees.

Lemma 15. *Consider a sequence of graphs generated by the process described in Section 2.4 with the additional assumption described by Condition 4. Then*

$$\|T - D\|_\infty = O\left((n\rho_n)^{1/2} \log n\right)$$

with high probability.

The second lemma linearizes the quantity $D^\beta - T^\beta$ via Taylor expansion.

Lemma 16. *Consider a sequence of graphs generated by the process described in Section 2.4 with the additional assumption described by Condition 4. Then*

$$D^\beta - T^\beta = -\beta T^{\beta-1} (T - D) + R_1$$

for a diagonal matrix R_1 satisfying

$$\|R_1\|_\infty = O\left((n\rho_n)^{\beta-1} \log(n)\right)$$

with high probability.

We now begin the proof of Theorem 8. For notational convenience, we suppress the dependence of all matrices on n in this section. Recall Theorem 5, which states that there exists a sequence of orthogonal transformations Q such that the symmetric Laplacian spectral embedding satisfies

$$\max_{i \in \{1, \dots, n\}} \left\| \left(\hat{X} Q^\top - \tilde{X} \right)_{i*} \right\|_2 = \left\| \left(\hat{X} Q^\top - \tilde{X} \right) \right\|_{2, \infty} = O\left(\frac{\log^{c'} n}{n\rho_n^{1/2}} \right)$$

with high probability.

We are interested in bounding the quantity

$$\left\| \hat{X}_\alpha Q^\top - X_\alpha \right\|_{2, \infty}.$$

To that end, using Lemma 16, we may write

$$\begin{aligned}\hat{X}_\alpha Q^\top - X_\alpha &= D^{1/2-\alpha} \hat{X} Q^\top - T^{1/2-\alpha} \tilde{X} \\ &= \left(T^{1/2-\alpha} + \left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) + R_1 \right) \hat{X} Q^\top - T^{1/2-\alpha} \tilde{X}\end{aligned}$$

where

$$\|R_1\| = O\left((n\rho_n)^{-1/2-\alpha} \log n\right)$$

with high probability.

We may simplify this expression to get

$$\begin{aligned}&= T^{1/2-\alpha} (\hat{X} Q^\top - \tilde{X}) + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) + R_1 \right) \hat{X} Q^\top. \\ &= T^{1/2-\alpha} (\hat{X} Q^\top - \tilde{X}) + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) \right) \tilde{X} \\ &\quad + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) + R_1 \right) (\hat{X} Q^\top - \tilde{X}) + R_1 \tilde{X} \\ &= T^{1/2-\alpha} (\hat{X} Q^\top - \tilde{X}) + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) \right) \tilde{X} + R_2\end{aligned}\tag{B.1}$$

where

$$R_2 = \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2} (T - D) + R_1 \right) (\hat{X} Q^\top - \tilde{X}) + R_1 \tilde{X}.$$

Using Lemmas 15 and 16, as well as Theorem 5, we have that

$$\begin{aligned}\|R_2\|_{2,\infty} &\leq \left(\alpha - \frac{1}{2} \right) (\|T^{-\alpha-1/2}\| \|T - D\| + \|R_1\|) \|\hat{X} W^\top - \tilde{X}\|_{2,\infty} + \|R_1\| \|\tilde{X}\|_{2,\infty} \\ &= O\left(\frac{\log^{c'+1} n}{n^{\alpha+1} \rho_n^{\alpha+1/2}}\right)\end{aligned}$$

with high probability.

Thus,

$$\begin{aligned}
& \left\| \hat{X}^\alpha Q^\top - X^\alpha \right\|_{2,\infty} \\
& \leq \|T^{1/2-\alpha}\| \left\| \hat{X} W^\top - \tilde{X} \right\|_{2,\infty} + \left| \alpha - \frac{1}{2} \right| \|T^{-\alpha-1/2}\| \|T - D\| \left\| \tilde{X} \right\|_\infty + \|R_2\|_{2,\infty} \\
& = O\left(\frac{\log^{c'+1} n}{n^{(\alpha+1/2)} \rho_n^\alpha} \right)
\end{aligned}$$

and the proof of Theorem 8 is complete.

We now move to Theorem 9. We begin with a result derived in Appendix B.1 of [Tang and Priebe \(2018\)](#) that linearizes $\hat{X}Q^\top - \tilde{X}$. We have

$$\hat{X}Q^\top - \tilde{X} = T^{-1/2}(A - P)T^{-1/2}\tilde{X} \left(\tilde{X}^\top \tilde{X} \right)^{-1} + \frac{1}{2}T^{-1}(T - D)\tilde{X} + R_3,$$

where R_3 is a matrix such that $\left\| n\rho_n^{1/2}(R_3)_{i*} \right\|_2 \rightarrow 0$ almost surely for all i .

Plugging this into Equation (B.1) and recalling that $\tilde{X} = T^{-1/2}X$ gives

$$\begin{aligned}
\hat{X}_\alpha Q^\top - X_\alpha &= T^{1/2-\alpha}(\hat{X}Q^\top - \tilde{X}) + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2}(T - D) \right) \tilde{X} + R_2 \\
&= T^{1/2-\alpha} \left(T^{-1/2}(A - P)T^{-1/2}\tilde{X} \left(\tilde{X}^\top \tilde{X} \right)^{-1} + \frac{1}{2}T^{-1}(T - D)\tilde{X} + R_3 \right) \\
&\quad + \left(\left(\alpha - \frac{1}{2} \right) T^{-\alpha-1/2}(T - D) \right) \tilde{X} + R_2 \\
&= T^{-\alpha}(A - P)T^{-1}X \left(\tilde{X}^\top \tilde{X} \right)^{-1} + \alpha T^{-\alpha-1}(T - D)X + R
\end{aligned}$$

where $R = T^{-1/2-\alpha}R_3 + R_2$. Note that R satisfies $\left\| n^{\alpha+1/2}\rho_n^\alpha(R)_{i*} \right\|_2 \rightarrow 0$ almost surely for all i .

Denote by ζ_i the i -th row of $n^{\alpha+1/2}\rho_n^\alpha \left(\hat{X}_\alpha Q^\top - X_\alpha \right)$. Let r denote an arbitrary random vector with the property $\|r\|_2 \rightarrow 0$ almost surely. Then

$$\begin{aligned} \zeta_i &= \left(\tilde{X}^\top \tilde{X} \right)^{-1} \frac{n^{\alpha+1/2}\rho_n^\alpha}{t_i^\alpha} \left(\sum_j \frac{a_{ij} - p_{ij}}{t_j} X_j \right) + \frac{\alpha n^{\alpha+1/2}\rho_n^\alpha (t_i - d_i)}{t_i^{\alpha+1}} X_i + r \\ &= \left(\tilde{X}^\top \tilde{X} \right)^{-1} \frac{n^{\alpha+1/2}\rho_n^\alpha}{t_i^\alpha} \left(\sum_j \frac{\rho_n^{1/2} (a_{ij} - p_{ij})}{t_j} \xi_j \right) + \frac{\alpha n^{\alpha+1/2}\rho_n^{\alpha+1/2} (t_i - d_i)}{t_i^{\alpha+1}} \xi_i + r \\ &= \left(\tilde{X}^\top \tilde{X} \right)^{-1} \left(\frac{n\rho_n}{t_i} \right)^\alpha \left(\sum_j \frac{(n\rho_n)^{1/2} (a_{ij} - p_{ij}) \xi_j}{t_j} \right) + \alpha \left(\frac{n\rho_n}{t_i} \right)^{\alpha+1} \sum_j \frac{(a_{ij} - p_{ij}) \xi_i}{(n\rho_n)^{1/2}} + r \end{aligned}$$

We now use two results from [Tang and Priebe \(2018\)](#) that simplify each of the sums. Recalling that $\mu = \mathbb{E}[\xi]$ and $\tilde{\Upsilon} = \mathbb{E} \left[\frac{\xi \xi^\top}{\langle \xi, \mu \rangle} \right]$,

$$\sum_j \frac{(n\rho_n)^{1/2} (a_{ij} - p_{ij})}{t_j} \xi_j = \sum_j \frac{(a_{ij} - p_{ij}) \xi_j}{(n\rho_n)^{1/2} \langle \xi_j, \mu \rangle} + r$$

and

$$\left(\frac{n\rho_n}{t_i} \right) \sum_j \frac{(a_{ij} - p_{ij}) \xi_i}{(n\rho_n)^{1/2}} = \left(\tilde{X}^\top \tilde{X} \right)^{-1} \frac{\tilde{\Upsilon} \xi_i}{\langle \xi_i, \mu \rangle} \sum_j \frac{(a_{ij} - p_{ij})}{(n\rho_n)^{-1/2}} + r.$$

Thus,

$$\begin{aligned} \zeta_i &= \left(\tilde{X}^\top \tilde{X} \right) \left(\frac{n\rho_n}{t_i} \right)^\alpha \left(\sum_j \frac{(n\rho_n)^{1/2} (a_{ij} - p_{ij}) \xi_j}{t_j} \right) + \alpha \left(\frac{n\rho_n}{t_i} \right)^{\alpha+1} \sum_j \frac{(a_{ij} - p_{ij}) \xi_i}{(n\rho_n)^{1/2}} + r \\ &= \left(\tilde{X}^\top \tilde{X} \right) \left(\frac{n\rho_n}{t_i} \right)^\alpha \sum_j \frac{(a_{ij} - p_{ij})}{(n\rho_n)^{1/2}} \left(\frac{\xi_j}{\langle \xi_j, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} \xi_i}{\langle \xi_i, \mu \rangle} \right) + r \\ &= \left(\tilde{X}^\top \tilde{X} \right) \left(\frac{n\rho_n}{t_i} \right)^\alpha \sum_j \frac{(a_{ij} - \rho_n \xi_i^\top \xi_j)}{(n\rho_n)^{1/2}} \left(\frac{\xi_j}{\langle \xi_j, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} \xi_i}{\langle \xi_i, \mu \rangle} \right) + r \end{aligned}$$

For each index i , conditional on $\xi_i = x_i$, the quantity

$$\sum_j \frac{(a_{ij} - \rho_n \xi_i^\top \xi_j)}{(n\rho_n)^{1/2}} \left(\frac{\xi_j}{\langle \xi_j, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} \xi_i}{\langle \xi_i, \mu \rangle} \right) \quad (\text{B.2})$$

is a sum of independent, identically distributed, mean 0 random variables. By the multivariate central limit theorem, conditional on $\xi_i = x_i$, this converges in distribution to a multivariate normal distribution with mean 0 and covariance

$$\mathbb{E} \left[\langle x, \xi \rangle (1 - \rho_n \langle x, \xi \rangle) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle \xi, \mu \rangle} \right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle \xi, \mu \rangle} \right)^\top \right].$$

Finally, we use another result from [Tang and Priebe \(2018\)](#), which states that $(\tilde{X}^\top \tilde{X})^{-1}$ converges almost surely to $\tilde{\Upsilon}^{-1}$. Additionally,

$$\frac{t_i}{n\rho_n} = \frac{\sum_j p_{ij}}{n\rho_n} = \frac{\sum_j \xi_i^\top \xi_j}{n} = \xi_i^\top \frac{\sum_j \xi_j}{n},$$

which converges to $\langle \xi_i, \mu \rangle$ almost surely. By the continuous mapping theorem, $\left(\frac{n\rho_n}{t_i}\right)^\alpha$ converges almost surely to $\langle \xi_i, \mu \rangle^{-\alpha}$. An application of Slutsky's theorem then gives that the vector ζ_i converges in distribution to a multivariate normal random variable with mean 0 and covariance

$$\tilde{\Upsilon}^{-1} \mathbb{E} \left[\frac{\langle x, \xi \rangle (1 - \rho_n \langle x, \xi \rangle)}{\langle x, \mu \rangle^{2\alpha}} \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle \xi, \mu \rangle} \right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle \xi, \mu \rangle} \right)^\top \right] \tilde{\Upsilon}^{-1},$$

which completes the proof.

B.2 Proofs of Lemmas 15 and 16

Consider the i -th diagonal elements of D and T , denoted d_i and t_i . We will prove Lemma 16 and establish Lemma 15 in an intermediate step. We begin with a Taylor series approximation, which states that for any β , there exists a ξ_i between t_i and d_i such that

$$d_i^\beta = t_i^\beta + \beta t_i^{\beta-1} (d_i - t_i) + \frac{1}{2} \beta(\beta+1) \xi_i^{\beta-2} (d_i - t_i)^2$$

We will show that the last term is bounded by $(n\rho_n)^{\beta-1} \sqrt{\log n}$ for each index i with high probability.

We first show that with high probability, $\xi_i = \Theta(t_i)$. A standard multiplicative Chernoff bound (for example, Chapter 2 of [Vershynin \(2018\)](#)) gives

$$\mathbb{P}(|d_i - t_i| \geq bt_i) \leq 2 \exp\left(-\frac{b^2 t_i}{2+b}\right). \quad (\text{B.3})$$

By Assumption 4, $t_i = \Theta(n\rho_n)$. The assumption that $n\rho_n = \omega(\log^{4c} n)$ then implies that $t_i = \omega(\log n)$. Choosing $c_1 = (3 + \sqrt{35})/2$ gives

$$\mathbb{P}\left(\left|\frac{d_i}{t_i} - 1\right| \geq c_1\right) \leq 2 \exp\left(-\frac{c_1^2 t_i}{2+c_1}\right) \leq 2 \exp\left(-\frac{c_1^2 \log n}{2+c_1}\right) = 2n^{-c_1^2/(2+c_1)} = 2n^{-3}. \quad (\text{B.4})$$

We now show that with high probability, $|t_i - d_i| = O_P(\sqrt{n\rho_n \log n})$. Letting $b = s/t_i$ in Equation (B.3) gives

$$\mathbb{P}(|d_i - t_i| \geq s) \leq 2 \exp\left(-\frac{s^2}{2t_i + s}\right). \quad (\text{B.5})$$

Let $c_2 = 3 + 2\sqrt{3}$. Choosing $s = c_2\sqrt{t_i \log n}$ gives

$$\begin{aligned} \mathbb{P}\left(|d_i - t_i| \geq c_2\sqrt{t_i \log n}\right) &\leq 2 \exp\left(-\frac{c_2^2 t_i \log n}{t_i + 2c_2\sqrt{t_i \log n}/3}\right) \\ &\leq 2 \exp\left(-\frac{c_2^2 \log n}{1 + 2c_2\sqrt{\log n/t_i}}\right). \end{aligned}$$

Again recalling that $t_i = \omega(\log n)$,

$$\sqrt{\frac{\log n}{t_i}} < 1.$$

Thus,

$$\begin{aligned} \mathbb{P}\left(|d_i - t_i| \geq c_2\sqrt{t_i \log n}\right) &\leq 2 \exp\left(-\frac{c_2^2 \log n}{1 + 2c_2\sqrt{\log n/t_i}}\right) \\ &\leq 2 \exp\left(-\frac{c_2^2 \log n}{1 + 2c_2}\right) \\ &= 2n^{-c_2^2/(1+2c_2)} \\ &= 2n^{-3} \end{aligned} \quad (\text{B.6})$$

We may apply a union bound to the previous equation to get that

$$\begin{aligned} \mathbb{P}\left(\max_i |d_i - t_i| \geq s\right) &\leq \sum_i \mathbb{P}(|d_i - t_i| \geq s) \\ &\leq 2n^{-2}. \end{aligned}$$

This establishes Lemma 15, as

$$\|D - T\|_\infty = \max_i |d_i - t_i|.$$

Letting $c_3 = c_1 c_2$, Equations (B.4) and (B.6) imply

$$\mathbb{P}\left(\xi_i^{-\beta-2} |d_i - t_i|^2 \leq c_3 t_i^{-\beta-1} \log n\right) \leq 2n^{-3} + 2n^{-3}. \quad (\text{B.7})$$

We may apply a union bound to the previous equation to get that

$$\mathbb{P}\left(\max_i \left(\xi_i^{-\beta-2} |d_i - t_i|^2\right) \leq c_3 t_i^{-\beta-1} \log n\right) \leq 4n^{-2}$$

Since $t_i = \Theta(n\rho_n)$,

$$\xi_i^{\beta-2} |d_i - t_i|^2 = O\left((n\rho_n)^{\beta-1} \log n\right)$$

for each index i with high probability, and the proof is complete.

Appendix C

Appendix to Chapter 4

C.1 Proof of Lemma 10

We first recall that $T_\tau = T + \tau I$ and $D_\tau = D + \tau I$. Then

$$\begin{aligned} \left\| \hat{L}_{\alpha, \tau} - L_{\alpha, \tau} \right\| &= \left\| D_\tau^{-\alpha} A D_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha} \right\| \\ &= \left\| D_\tau^{-\alpha} A D_\tau^{-\alpha} - T_\tau^{-\alpha} A T_\tau^{-\alpha} + T_\tau^{-\alpha} A T_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha} \right\| \\ &\leq \left\| D_\tau^{-\alpha} A D_\tau^{-\alpha} - T_\tau^{-\alpha} A T_\tau^{-\alpha} \right\| + \left\| T_\tau^{-\alpha} A T_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha} \right\|. \end{aligned}$$

For the first term, we begin with a lemma that bounds the operator norm of $\|A - P\|$ (Theorem 5.2 in (Lei and Rinaldo, 2015)).

Lemma 17. *Let A be the adjacency matrix of a random graph on n nodes in which edges occur independently, and let $P = \mathbb{E}[A]$. Assume that $n \max p_{ij} \leq d$ for $d \geq c_0 \log n$ and $c_0 > 0$. Then $\|A - P\| = O(\sqrt{d})$ with probability at least $1 - n^{-2}$.*

An application of the triangle inequality gives that $\|A\| \leq \|P\| + \|A - P\|$. We assume that $\Delta \geq \log n$, so we may apply the previous lemma to get that $\|A - P\| = O(\sqrt{\Delta})$ with probability at least $1 - n^{-1}$. Combining this with the fact that $\|P\| \leq \Delta$ gives that $\|A\| = O(\Delta)$ with probability $1 - n^{-2}$.

Using the previous lemma, the following inequalities hold with probability at least $1 - n^{-2}$.

$$\begin{aligned}
& \|D_\tau^{-\alpha} A D_\tau^{-\alpha} - T_\tau^{-\alpha} A T_\tau^{-\alpha}\| \\
&= \|T_\tau^{-\alpha} (T_\tau^\alpha D_\tau^{-\alpha} A D_\tau^{-\alpha} T_\tau^\alpha - A) T_\tau^{-\alpha}\| \\
&\leq \|T_\tau^{-2\alpha}\| \|T_\tau^\alpha D_\tau^{-\alpha} A D_\tau^{-\alpha} T_\tau^\alpha - A\| \\
&= (\delta + \tau)^{-2\alpha} \|(T_\tau^\alpha D_\tau^{-\alpha} - I) A D_\tau^{-\alpha} T_\tau^{-\alpha} + A (D_\tau^\alpha T_\tau^{-\alpha} - I)\| \\
&= (\delta + \tau)^{-2\alpha} (\|(T_\tau^\alpha D_\tau^{-\alpha} - I)\| \|A\| \|D_\tau^{-\alpha} T_\tau^{-\alpha}\| + \|A\| \|(D_\tau^\alpha T_\tau^{-\alpha} - I)\|) \\
&\leq \Delta (\delta + \tau)^{-2\alpha} (\|(T_\tau^\alpha D_\tau^{-\alpha} - I)\| \|D_\tau^{-\alpha} T_\tau^{-\alpha}\| + \|(D_\tau^\alpha T_\tau^{-\alpha} - I)\|) \\
&\leq \Delta (\delta + \tau)^{-2\alpha} (\|(T_\tau^\alpha D_\tau^{-\alpha} - I)\| \|D_\tau^{-\alpha} T_\tau^{-\alpha} - I + I\| + \|(D_\tau^\alpha T_\tau^{-\alpha} - I)\|) \\
&\leq \Delta (\delta + \tau)^{-2\alpha} (\|(T_\tau^\alpha D_\tau^{-\alpha} - I)\| (\|D_\tau^{-\alpha} T_\tau^{-\alpha} - I\| + 1) + \|(D_\tau^\alpha T_\tau^{-\alpha} - I)\|) \\
&\leq \Delta (\delta + \tau)^{-2\alpha} \left(\|(I - T_\tau^{-\alpha} D_\tau^\alpha)\|^2 + 2 \|(I - D_\tau^\alpha T_\tau^{-\alpha})\| \right) \tag{C.1}
\end{aligned}$$

To get a bound on the two summands, we may apply a concentration inequality (for example, (Lu and Chung, 2006)). For $u > 0$,

$$\begin{aligned}
\mathbb{P}(|d_i - t_i| \geq u) &\leq \exp\left(-\frac{u^2}{2t_i}\right) + \exp\left(-\frac{u^2}{2t_i + \frac{2}{3}u}\right) \\
&\leq \exp\left(-\frac{u^2}{2t_i + \frac{2}{3}u}\right)
\end{aligned}$$

Letting $u = s(t_i + \tau)$ for $s > 0$ gives

$$\begin{aligned}
\mathbb{P}(|d_i - t_i| \geq s(t_i + \tau)) &\leq 2 \exp\left(-\frac{(s(t_i + \tau))^2}{2t_i + \frac{2}{3}s(t_i + \tau)}\right) \\
&\leq 2 \exp\left(-\frac{(s(t_i + \tau))^2}{2(t_i + \tau) + \frac{2}{3}s(t_i + \tau)}\right) \\
&\leq 2 \exp\left(-\frac{(s(t_i + \tau))^2}{(2 + 2s/3)(t_i + \tau)}\right) \\
&= 2 \exp\left(-\frac{s^2(\delta + \tau)}{2 + 2s/3}\right)
\end{aligned}$$

Choosing

$$s = c \frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{\delta + \tau},$$

$$\begin{aligned}
\mathbb{P}(|d_i - t_i| \geq s(t_i + \tau)) &\leq 2 \exp\left(-\frac{s^2(\delta + \tau)}{2 + 2s/3}\right) \\
&\leq 2 \exp\left(-\frac{c^2(\delta + \tau)^{-1} \|\theta\|_1 \rho_n \log n}{2 + c(\delta + \tau)^{-1} \sqrt{\|\theta\|_1 \rho_n \log n}}\right) \\
&\leq 2 \exp\left(-\frac{c^2(\delta + \tau)^{-1} \|\theta\|_1 \rho_n \log n}{2 + c(\delta + \tau)^{-1} \sqrt{\|\theta\|_1 \rho_n \log n}}\right) \\
&\leq 2 \exp\left(-\frac{c^2 \log n}{2 \frac{(\delta + \tau)}{\|\theta\|_1 \rho_n} + c \sqrt{\frac{\log n}{\|\theta\|_1 \rho_n}}}\right) \\
&\leq 2 \exp\left(-\frac{c^2 \log n}{2 + c}\right).
\end{aligned}$$

In the last step, we use the assumptions that $\|\theta\|_1 \rho_n > \delta + \tau$ and $\|\theta\|_1 \rho_n > \Delta > \log n$. Choosing $c = \frac{3 + \sqrt{35}}{2}$ and a union bound over all indices i shows

$$\mathbb{P}\left(\max_i |d_i - t_i| \geq s(t_i + \tau)\right) \leq 2n^{-2}.$$

Since

$$\|I - T_\tau^{-\alpha} D_\tau^\alpha\| = \max_i \left|1 - \frac{(d_i + \tau)^\alpha}{(t_i + \tau)^\alpha}\right| \leq \max_i \left|1 - \frac{d_i + \tau}{t_i + \tau}\right|,$$

we have

$$\|(I - T_\tau^{-\alpha} D_\tau^\alpha)\| = O\left(\frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{\delta + \tau}\right)$$

with probability at least $1 - 2n^{-2}$. Plugging this into Equation (C.1) gives

$$\|D_\tau^{-\alpha} A D_\tau^{-\alpha} - T_\tau^{-\alpha} A T_\tau^{-\alpha}\| = O\left(\frac{\Delta \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{2+2\alpha}} + \frac{\Delta \sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{1+2\alpha}}\right) \quad (\text{C.2})$$

with probability at least $1 - 2n^{-2}$.

For the second term, we use Theorem 5 of [Chung and Radcliffe \(2011\)](#).

Lemma 18. *Let X_1, X_2, \dots, X_m be independent random $n \times n$ Hermitian matrices. Moreover, assume that $\|X_i - E(X_i)\| \leq M$ for all i , and let $\sigma^2 = \|\sum \text{Var}(X_i)\|$. Let $X = \sum X_i$.*

Then for any $a > 0$,

$$\mathbb{P}(\|X - \mathbb{E}[X]\| \geq a) \leq 2n \exp\left(-\frac{a^2}{2(\sigma^2 + Ma/3)}\right).$$

Define E_{ij} to be the $n \times n$ matrix with a 1 in the ij and ji position and 0s everywhere else, and let

$$\begin{aligned} X_{ij} &= T_\tau^{-\alpha} ((a_{ij} - p_{ij}) E_{ij}) T_\tau^{-\alpha} \\ &= ((t_{ii} + \tau)(t_{jj} + \tau))^{-\alpha} (a_{ij} - p_{ij}) E_{ij}. \end{aligned}$$

Note that

$$T_\tau^{-\alpha} A T_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha} = \sum X_{ij}.$$

It is easy to see that $\mathbb{E}[X_{ij}] = 0$, and

$$\|X_{ij} - E(X_{ij})\| = \|X_{ij}\| \leq \frac{\|(a_{ij} - p_{ij}) E_{ij}\|}{(\delta + \tau)^{2\alpha}} \leq \frac{1}{(\delta + \tau)^{2\alpha}}.$$

Additionally,

$$\begin{aligned} \mathbb{E}[X_{ij}^2] &\leq \frac{\text{Var}(a_{ij} - p_{ij})}{(\delta + \tau)^{4\alpha}} \\ &= \frac{\text{Var}(a_{ij})}{(\delta + \tau)^{4\alpha}} \\ &\leq \frac{\theta_i \theta_j \rho_n (1 - \theta_i \theta_j \rho_n)}{(\delta + \tau)^{4\alpha}} \\ &\leq \frac{\theta_i \theta_j \rho_n}{(\delta + \tau)^{4\alpha}} \end{aligned}$$

so

$$\begin{aligned}
\sigma^2 &= \left\| \sum \text{Var}(X_i) \right\| \\
&= \left\| \sum_i \sum_j \mathbb{E} [X_{ij} X_{ij}^\top] \right\| \\
&\leq \left\| \sum_i \sum_{j=i} \mathbb{E} [X_{ij}^2] \right\| \\
&\leq \left\| \sum_i \sum_{j=i} \frac{\theta_i \theta_j \rho_n}{(\delta + \tau)^{4\alpha}} \right\| \\
&= \max_i \theta_i \sum_{j=i} \frac{\theta_j \rho_n}{(\delta + \tau)^{4\alpha}} \\
&\leq \frac{\theta_{\max} \|\theta\|_1 \rho_n}{(\delta + \tau)^{4\alpha}} \\
&= \frac{\|\theta\|_1 \rho_n}{(\delta + \tau)^{4\alpha}}
\end{aligned}$$

We now apply the lemma to $X = \sum X_{ij}$. Choose

$$s = \frac{2c \log n}{3(\delta + \tau)^{2\alpha}} + \sqrt{\frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}}}$$

so that

$$s \geq \frac{c \log n}{3(\delta + \tau)^{2\alpha}} + \sqrt{\frac{c^2 \log^2 n}{9(\delta + \tau)^{4\alpha}} + \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}}}.$$

This implies

$$\begin{aligned}
s - \frac{c \log n}{3(\delta + \tau)^{2\alpha}} &\geq \sqrt{\frac{c^2 \log^2 n}{9(\delta + \tau)^{4\alpha}} + \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}}} \\
\left(s - \frac{c \log n}{3(\delta + \tau)^{2\alpha}}\right)^2 &\geq \frac{c^2 \log^2 n}{9(\delta + \tau)^{4\alpha}} + \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}} \\
s^2 - \frac{2ac \log n}{3(\delta + \tau)^{2\alpha}} + \frac{c^2 \log^2 n}{9(\delta + \tau)^{4\alpha}} &\geq \frac{c^2 \log^2 n}{9(\delta + \tau)^{4\alpha}} + \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}} \\
s^2 - \frac{2ac \log n}{3(\delta + \tau)^{2\alpha}} &\geq \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}} \\
s^2 &\geq \frac{2c \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{4\alpha}} + \frac{2ac \log n}{3(\delta + \tau)^{2\alpha}} \\
s^2 &\geq 2(\sigma^2 + Ma/3)(c \log n) \\
\frac{s^2}{2(\sigma^2 + Ma/3)} &\geq c \log n.
\end{aligned}$$

Choosing $c = 3$ gives

$$\begin{aligned}
P\left(\|T_\tau^{-\alpha} A T_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha}\| \geq s\right) &\leq 2n \exp\left(-\frac{s^2}{2(v^2 + Ms/3)}\right) \\
&\leq 2n \exp(-3 \log n) \\
&\leq 2n^{-2}
\end{aligned}$$

and

$$\|T_\tau^{-\alpha} A T_\tau^{-\alpha} - T_\tau^{-\alpha} P T_\tau^{-\alpha}\| = O\left(\frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{2\alpha}}\right) \tag{C.3}$$

with probability at least $1 - 2n^{-2}$. Combining Equations (C.2) and (C.3) completes the proof.

C.2 Proof of Lemma 11

We begin by bounding the distance between the rows of $\hat{X}_{\alpha,\tau}$ and $X_{\alpha,\tau}$.

$$\begin{aligned}
\left\| \hat{X}_{\alpha,\tau} - X_{\alpha,\tau} \right\|_F &= \left\| \hat{U}_{\alpha,\tau} \hat{\Lambda}_{\alpha,\tau}^{1/2} - U_{\alpha,\tau} \Lambda_{\alpha,\tau}^{1/2} \right\|_F \\
&\leq \left\| \hat{U}_{\alpha,\tau} \hat{\Lambda}_{\alpha,\tau}^{1/2} - \hat{U}_{\alpha,\tau} \Lambda_{\alpha,\tau}^{1/2} \right\|_F - \left\| \hat{U}_{\alpha,\tau} \Lambda_{\alpha,\tau}^{1/2} - U_{\alpha,\tau} \Lambda_{\alpha,\tau}^{1/2} \right\|_F \\
&\leq \left\| \hat{\Lambda}_{\alpha,\tau}^{1/2} - \Lambda_{\alpha,\tau}^{1/2} \right\|_F - \left\| \hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right\|_F \left\| \Lambda_{\alpha,\tau}^{1/2} \right\|_F
\end{aligned} \tag{C.4}$$

By Lemma 7, $\left\| \hat{L}_{\alpha,\tau} - L_{\alpha,\tau} \right\| < \eta_{\alpha,\tau}$. Since both matrices are symmetric, Weyl's inequality implies

$$\max_{1 \leq k \leq K} \left| \sqrt{\hat{\lambda}_k} - \sqrt{\lambda_k} \right| \leq \max_{1 \leq k \leq K} \sqrt{|\hat{\lambda}_k - \lambda_k|} \leq \sqrt{\eta_{\alpha,\tau}}$$

and

$$\left\| \hat{\Lambda}_{\alpha,\tau}^{1/2} - \Lambda_{\alpha,\tau}^{1/2} \right\|_F \leq \sqrt{\eta_{\alpha,\tau} K}. \tag{C.5}$$

We may use the Davis-Kahan $\sin \theta$ theorem to bound $\left\| \hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right\|_F$. This was provided in Section 3.4 and reproduced here for convenience.

Lemma 19. *Let P be a rank K $n \times n$ symmetric matrix with smallest eigenvalue λ_K . Let A be any symmetric matrix and \hat{U}, U be the $n \times K$ matrix whose columns are the K top eigenvectors of A and P , respectively. Then there exists a $K \times K$ orthogonal matrix Q such that*

$$\left\| \hat{U} - UQ \right\|_F \leq \frac{2\sqrt{2K}}{\lambda_K} \|A - P\|$$

We may apply the previous lemma to $L_{\alpha,\tau}$ and $\hat{L}_{\alpha,\tau}$. For notational convenience, we work with the assumption that Q is the orthogonal matrix I_K , noting that the proof could be modified by replacing $U_{\alpha,\tau}$ with $U_{\alpha,\tau}Q$.

$$\left\| \hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right\|_F \leq \frac{2\sqrt{2K}}{\lambda_K} \left\| \hat{L}_{\alpha,\tau} - L_{\alpha,\tau} \right\| \leq \frac{2\sqrt{2K}\eta_{\alpha,\tau}}{\lambda_K}. \tag{C.6}$$

We now state a brief lemma that bounds the operator norm of $L_{\alpha,\tau}$.

Lemma 20. $\|L_{\alpha,\tau}\| \leq \Delta_{\alpha,\tau}^{1-2\alpha}$.

Proof. We first show the result for the regularized symmetric Laplacian, $L_{\text{Sym},\tau}$. Note that $T^{-1}A$ and $T^{-1/2}AT^{-1/2}$ are similar and share eigenvalues. Since $T^{-1}A$ is row stochastic, the Perron-Frobenius Theorem implies the largest eigenvalue of L_{Sym} is 1. Thus,

$$\begin{aligned}\|L_{\text{Sym},\tau}\| &= \|T_\tau^{-1/2}T^{1/2}L_{\text{Sym}}T^{1/2}T_\tau^{-1/2}\| \\ &\leq \|T_\tau^{-1/2}T^{1/2}\| \|T^{1/2}T_\tau^{-1/2}\| \\ &\leq \|T_\tau^{-1/2}T^{1/2}\|^2 \\ &= \frac{\Delta}{\Delta + \tau} \\ &\leq 1\end{aligned}$$

Now define a matrix $B = T_\tau^{1/2-\alpha}$ so that $L_{\alpha,\tau} = BL_{\text{Sym},\tau}B$. Note that when $\alpha < 1/2$, $1/2 - \alpha > 0$, so $\|B\|$ is the largest element of B . The opposite is true for $\alpha > 1/2$. Then

$$\begin{aligned}\|L_{\alpha,\tau}\| &= \|BL_{\text{Sym},\tau}B\| \\ &\leq \|B\| \|L_{\text{Sym},\tau}\| \|B\| \\ &\leq \Delta_{\alpha,\tau}^{1-2\alpha}.\end{aligned}$$

□

Finally, by Lemma 20, $\|L_{\alpha,\tau}\| \leq \Delta_{\alpha,\tau}^{1-2\alpha}$, so

$$\|\Lambda_{\alpha,\tau}^{1/2}\|_F \leq \Delta_{\alpha,\tau}^{1/2-\alpha} \sqrt{K}. \quad (\text{C.7})$$

Plugging Equations (C.5), (C.6) and (C.7) into Equation (C.4) gives

$$\left\| \hat{X}_{\alpha,\tau} - X_{\alpha,\tau} \right\|_F \leq \sqrt{\eta_{\alpha,\tau} K} + \frac{2\sqrt{2}K\eta_{\alpha,\tau}}{\lambda_K \Delta_{\alpha,\tau}^{\alpha-1/2}}$$

We now bound the difference between the row-normalized matrices. To this end, we state a brief lemma that relates the distance between two vectors to the distance between their normalized counterparts.

Lemma 21. *Let u and v be two vectors in \mathbb{R}^n and let u' and v' be the normalized vectors, $u' = u/\|u\|$ and $v' = v/\|v\|$. Let $M = \max\{\|u\|_2, \|v\|_2\}$. Then*

$$\|u' - v'\|_2 \leq \frac{2\|u - v\|}{M}$$

Proof. Without loss of generality, assume that $\|v\| \geq \|u\|$. Using the triangle inequality,

$$\begin{aligned} \|u' - v'\| &\leq \left\| \frac{u}{\|u\|} - \frac{u}{\|v\|} \right\| + \left\| \frac{u}{\|v\|} - \frac{v}{\|v\|} \right\| \\ &\leq \|u\| \frac{\|v - u\|}{\|u\| \|v\|} + \frac{\|u - v\|}{\|v\|} \\ &= \frac{2\|u - v\|}{\|v\|}. \end{aligned}$$

□

Applying the previous lemma to $(\hat{X}_{\alpha,\tau}^*)_{i^*}$ and $(X_{\alpha,\tau}^*)_{i^*}$,

$$\left\| (\hat{X}_{\alpha,\tau}^*)_{i^*} - (X_{\alpha,\tau}^*)_{i^*} \right\|_2 \leq \frac{2 \left\| (\hat{X}_{\alpha,\tau})_{i^*} - (X_{\alpha,\tau})_{i^*} \right\|_2}{(X_{\alpha,\tau})_{i^*}}$$

for each i . Thus,

$$\left\| \hat{X}_{\alpha,\tau}^* - X_{\alpha,\tau}^* \right\|_F \leq \frac{2 \left\| \hat{X}_{\alpha,\tau} - X_{\alpha,\tau} \right\|_F}{m},$$

where m is the shortest row of $X_{\alpha,\tau}$.

We now bound m from below. Using Lemma 6,

$$\begin{aligned}
\|(X_{\alpha,\tau})_{i^*}\|_2 &= \|(U_{\alpha,\tau}\Lambda_{\alpha,\tau}^{1/2})_{i^*}\|_2 \\
&= \sqrt{\lambda_1 u_{1,i}^2 + \lambda_2 u_{2,i}^2 + \dots + \lambda_K u_{K,i}^2} \\
&\geq \sqrt{\lambda_K} \sqrt{\sum_{k=1}^K u_{k,i}^2} \\
&= \sqrt{\lambda_K} \sqrt{\sum_{k=1}^K \left(\frac{\theta_i^{1-\alpha}}{\|\psi_{z_i}\|} v_{k,z_i} \right)^2} \\
&\geq \sqrt{\lambda_K} \frac{\theta_{\min}^{1-\alpha}}{\sqrt{n_{\max}}} \sqrt{\sum_{k=1}^K v_{k,z_i}^2} \\
&= \sqrt{\lambda_K} \frac{\theta_{\min}^{1-\alpha}}{\sqrt{n_{\max}}}
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\hat{X}_{\alpha,\tau}^* - X_{\alpha,\tau}^*\|_F &\leq 2 \left(\sqrt{\eta_{\alpha,\tau} K} + \frac{2\sqrt{2}K\eta_{\alpha,\tau}}{\lambda_K \Delta_{\alpha,\tau}^{\alpha-1/2}} \right) \frac{\sqrt{n_{\max}}}{\sqrt{\lambda_K} \theta_{\min}^{1-\alpha}} \\
&= 2 \left(\sqrt{\frac{K}{\eta_{\alpha,\tau}}} (\lambda_K) + \frac{2\sqrt{2}K}{\Delta_{\alpha,\tau}^{\alpha-1/2}} \right) \frac{\eta_{\alpha,\tau} \sqrt{n_{\max}}}{\lambda_K^{3/2} \theta_{\min}^{1-\alpha}} \\
&\leq 2 \left(\sqrt{\frac{K}{\eta_{\alpha,\tau}}} \|L_{\alpha}\| + \frac{2\sqrt{2}K}{\Delta_{\alpha,\tau}^{\alpha-1/2}} \right) \frac{\eta_{\alpha,\tau} \sqrt{n_{\max}}}{\lambda_K^{3/2} \theta_{\min}^{1-\alpha}} \\
&\leq 2 \left(\frac{\sqrt{K}}{\sqrt{\eta_{\alpha,\tau} \Delta_{\alpha,\tau}^{2\alpha-1}}} + \frac{2\sqrt{2}K}{\Delta_{\alpha,\tau}^{\alpha-1/2}} \right) \frac{\eta_{\alpha,\tau} \sqrt{n_{\max}}}{\lambda_K^{3/2} \theta_{\min}^{1-\alpha}}. \tag{C.8}
\end{aligned}$$

As a final step, we relate the eigenvalues of L_{α} to the eigenvalues of P .

Lemma 22. *Let A be a positive-definite symmetric matrix of rank K and D be a strictly positive diagonal matrix. Define L_{α} to be $D^{-\alpha} A D^{-\alpha}$ for $\alpha \in [0, 1]$. Then*

$$\frac{\lambda_{K,A}}{d_{\max}^{2\alpha}} \leq \lambda_{K,L} \leq \lambda_{1,L} \leq \frac{\lambda_{1,A}}{d_{\min}^{2\alpha}}.$$

Proof. This follows from the submultiplicativity of the spectral norm, because

$$\lambda_{1,L} = \|L\| = \|D^{-\alpha}AD^{-\alpha}\| \leq \|D^{-\alpha}\|^2 \|A\| = \frac{\lambda_{1,A}}{d_{\min}^{2\alpha}}.$$

Since $\lambda_{K,L} = 1/\|L^{-1}\|$

$$\lambda_{K,L} = \|L^{-1}\|^{-1} = \|D^\alpha A^{-1}D^\alpha\|^{-1} \geq (\|D^\alpha\|^2 \|A^{-1}\|)^{-1} = \frac{\lambda_{K,A}}{d_{\max}^{2\alpha}}.$$

□

Applying Lemma 22 to Equation (C.8),

$$\begin{aligned} \left\| \hat{X}_{\alpha,\tau}^* - X_{\alpha,\tau}^* \right\|_F &\leq 2 \left(\frac{\sqrt{K}}{\sqrt{\eta_{\alpha,\tau}} \Delta_{\alpha,\tau}^{2\alpha-1}} + \frac{2\sqrt{2}K}{\Delta_{\alpha,\tau}^{\alpha-1/2}} \right) \frac{\eta_{\alpha,\tau} \sqrt{n_{\max}}}{\lambda_K^{3/2} \theta_{\min}^{1-\alpha}} \\ &\leq 2 \left(\frac{\sqrt{K}}{\sqrt{\eta_{\alpha,\tau}} \Delta_{\alpha,\tau}^{\alpha-1/2}} + 2\sqrt{2}K \right) \frac{(\Delta + \tau)^{3\alpha} \eta_{\alpha,\tau} \sqrt{n_{\max}}}{\Delta_{\alpha,\tau}^{\alpha-1/2} \lambda_{K,P}^{3/2} \theta_{\min}^{1-\alpha}} \end{aligned}$$

and the proof is complete.

C.3 Proof of Theorem 10

The final step of Algorithm 3 is to apply k -means to the rows of \hat{X}_α^* , where each point is a vector in \mathbb{R}^K . Recalling that $\mathbb{M}_{n,K}$ is the set of all possible community membership matrices, the solution to the k -means problem is given by

$$\left(\hat{Z}, \hat{C} \right) = \min_{Z \in \mathbb{M}_{n,K}, C \in \mathbb{R}^{K \times K}} \left\| ZC - \hat{X}_\alpha^* \right\|_F^2.$$

Let

$$d = \min_{i,j,z_i \neq z_j} \left\| (X_\alpha^*)_{i*} - (X_\alpha^*)_{j*} \right\|_2,$$

the minimum distance between two rows of X_α^* whose corresponding nodes that do not share the same community.

Define a set of nodes in a community k by

$$\mathcal{M}_k = \left\{ i \mid z_i = k, \left\| (X_\alpha^*)_{i*} - (\hat{Z}\hat{C})_{i*} \right\|_2 > \frac{d}{2} \right\}$$

By the triangle inequality, we have

$$\begin{aligned} \left\| \hat{Z}\hat{C} - X_\alpha^* \right\|_F &\leq \left\| \hat{Z}\hat{C} - \hat{X}_\alpha^* \right\|_F + \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F \\ &\leq 2 \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F \end{aligned}$$

Then

$$\sum_{k=1}^K |\mathcal{M}_k| \left(\frac{d}{2} \right)^2 \leq \left\| \hat{Z}\hat{C} - X_\alpha^* \right\|_F^2 \leq 4 \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F^2. \quad (\text{C.9})$$

Let \mathcal{Z}_k be the set of nodes such that $\hat{z}_i = k$ and $\mathcal{C}_k = \mathcal{Z}_k \setminus \mathcal{M}_k$. We now show that \mathcal{C}_k may be thought of as the set of correctly classified nodes and \mathcal{M}_k may be thought of as the set of misclassified nodes.

Suppose $i, j \notin \mathcal{M}_k$ and $\hat{z}_i \neq \hat{z}_j$. This implies that $\hat{C}_{z_i*} \neq \hat{C}_{z_j*}$. Otherwise,

$$d \leq \left\| (X_\alpha^*)_{i*} - (X_\alpha^*)_{j*} \right\|_2 \leq \left\| (X_\alpha^*)_{i*} - \hat{C}_{\hat{z}_i*} \right\|_2 + \left\| (X_\alpha^*)_{j*} - \hat{C}_{\hat{z}_j*} \right\|_2 < \frac{d}{2} + \frac{d}{2} = d,$$

which is impossible. This also implies that \hat{C} must have exactly K distinct rows. On the other hand, if $i, j \notin \mathcal{M}_k$ and $\hat{z}_i = \hat{z}_j$, then $\hat{C}_{\hat{z}_i*} = \hat{C}_{\hat{z}_j*}$, since there are $K - 1$ distinct rows taken by nodes in \mathcal{Z}_l for $l \neq \hat{z}_i$. Thus, for $i, j \notin \mathcal{M}_K$, $(\hat{Z}\hat{C})_{i*} = (\hat{Z}\hat{C})_{j*}$ if and only if $Z_i = Z_j$, and \mathcal{M}_k is the set of misclassified nodes.

We now bound d from below using the following lemma, which follows from a straightforward application of Wielandt's inequality (see 7.4.12.6, (Horn and Johnson, 2013)).

Lemma 23. *Let u and v be two orthogonal vectors and let D be a diagonal matrix. Define*

$$\kappa = \frac{d_{max}}{d_{min}}.$$

Then

$$\left\| \frac{Du}{\|Du\|} - \frac{Dv}{\|Dv\|} \right\| \geq \sqrt{2 - 2 \left(\frac{\kappa^2 - 1}{\kappa^2 + 1} \right)} = \frac{2}{\sqrt{\kappa^2 + 1}}$$

Applying the previous lemma to the rows of X_α^* , as long as $n \geq 1$,

$$U_\alpha \Lambda_\alpha^{1/2} \geq \frac{1}{\sqrt{\kappa^2 + 1}} = \sqrt{\frac{\lambda_K}{\lambda_1 + \lambda_K}} \geq \sqrt{\frac{\lambda_K}{2\lambda_1}}.$$

We may apply this to Equation (C.9) to bound total number of misclassified nodes by

$$\sum_{k=1}^K |\mathcal{M}_k| \leq \frac{16}{d^2} \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F^2 = 32 \frac{\lambda_1}{\lambda_K} \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F^2.$$

Thus, we may write

$$\begin{aligned} L(\hat{Z}, Z) &\leq \frac{32\lambda_1 \left\| \hat{X}_\alpha^* - X_\alpha^* \right\|_F^2}{\lambda_K n} \\ &= 32 \left(\frac{K}{\eta_{\alpha,\tau} \Delta_{\alpha,\tau}^{2\alpha-1}} + 8K^2 \right) \frac{\lambda_1 (\Delta + \tau)^{6\alpha} (\eta_{\alpha,\tau})^2 n_{max}}{\Delta_{\alpha,\tau}^{2\alpha-1} (\lambda_K)^3 \theta_{min}^{2-2\alpha} n} \\ &= 32 \left(\frac{K}{\eta_{\alpha,\tau} \Delta_{\alpha,\tau}^{2\alpha-1}} + 8K^2 \right) \frac{\lambda_{1,P} (\Delta + \tau)^{8\alpha} (\eta_{\alpha,\tau})^2 n_{max}}{\Delta_{\alpha,\tau}^{2\alpha-1} (\delta + \tau)^{2\alpha} (\lambda_{K,P})^4 \theta_{min}^{2-2\alpha} n} \end{aligned}$$

where in the last step we used Lemma 22.

We have

$$\eta_{\alpha,\tau} = \frac{\Delta \|\theta\|_1 \rho_n \log n}{(\delta + \tau)^{2+2\alpha}} + \frac{\Delta \sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{1+2\alpha}} + \frac{\sqrt{\|\theta\|_1 \rho_n \log n}}{(\delta + \tau)^{2\alpha}}.$$

When $\alpha > 1/2$, $\Delta_{\alpha,\tau} = \delta + \tau$. If the first term dominates, then

$$\frac{\Delta \sqrt{\rho_n \|\theta\|_1 \log n}}{(\delta + \tau)^2} > 1 \quad \text{and} \quad \frac{\sqrt{\rho_n \|\theta\|_1 \log n}}{\delta + \tau} > 1,$$

which implies

$$\frac{\Delta \rho_n \|\theta\|_1 \log n}{(\delta + \tau)^3} > 1.$$

Thus, with high probability

$$\begin{aligned}
L(\hat{Z}, Z) &= O \left[\left(\frac{K(\delta + \tau)^3}{\Delta \|\theta\|_1 \rho_n \log n} + 8K^2 \right) \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha} \|\theta\|_1^2 \rho_n^2 (\log n)^2 n_{max}}{(\delta + \tau)^{8\alpha+3} (\lambda_{K,P})^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\left(\frac{K(\delta + \tau)^3}{\Delta \|\theta\|_1 \rho_n \log n} + 8K^2 \right) \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha} \|\theta\|_1^2 \rho_n^2 (\log n)^2 n_{max}}{(\delta + \tau)^{8\alpha+3} (\lambda_{K,P})^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[K^2 \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha+1} \|\theta\|_1^2 \rho_n^2 (\log n)^2 n_{max}}{(\delta + \tau)^{8\alpha} (\lambda_{K,P})^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{8\alpha} \sqrt{\Delta \|\theta\|_1 \rho_n \log n} n_{max}}{(\delta + \tau)^{8\alpha-3/2} (\lambda_{K,P})^4 \theta_{min}^{2-2\alpha} n} \right] \tag{C.10}
\end{aligned}$$

using the fact that $K \leq K^2$.

If the second term dominates, then

$$\frac{\Delta}{\delta + \tau} > 1 \quad \text{and} \quad 1 > \frac{\sqrt{\rho_n \|\theta\|_1 \log n}}{\delta + \tau}.$$

Thus, with high probability

$$\begin{aligned}
L(\hat{Z}, Z) &= O \left[\left(\frac{K(\delta + \tau)^2}{\Delta \sqrt{\|\theta\|_1 \rho_n \log n}} + 8K^2 \right) \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha} \|\theta\|_1 \rho_n (\log n) n_{max}}{(\delta + \tau)^{8\alpha+1} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\left(\frac{K(\delta + \tau)}{\sqrt{\|\theta\|_1 \rho_n \log n}} + 8K^2 \right) \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha} \|\theta\|_1 \rho_n (\log n) n_{max}}{(\delta + \tau)^{8\alpha+1} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\left(\frac{K^2 (\delta + \tau)}{\sqrt{\|\theta\|_1 \rho_n \log n}} \right) \left(\frac{\lambda_{1,P} \Delta^2 (\Delta + \tau)^{8\alpha} \|\theta\|_1 \rho_n (\log n) n_{max}}{(\delta + \tau)^{8\alpha+1} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\frac{\lambda_{1,P} K^2 \Delta^2 (\Delta + \tau)^{8\alpha} \sqrt{\|\theta\|_1 \rho_n \log n} n_{max}}{(\delta + \tau)^{8\alpha} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right] \tag{C.11}
\end{aligned}$$

using the fact that $K \leq K^2$.

If the third term dominates, then

$$1 > \frac{\Delta}{\delta + \tau} \quad \text{and} \quad 1 > \frac{\Delta \sqrt{\rho_n \|\theta\|_1 \log n}}{(\delta + \tau)^2}.$$

Thus, with high probability

$$\begin{aligned}
L(\hat{Z}, Z) &= O \left[\left(\frac{K(\delta + \tau)}{\sqrt{\|\theta\|_1 \rho_n \log n}} + 8K^2 \right) \left(\frac{\lambda_{1,P}(\Delta + \tau)^{8\alpha} \|\theta\|_1 \rho_n (\log n) n_{max}}{(\delta + \tau)^{8\alpha-1} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\left(\frac{K\Delta}{\delta + \tau} + 8K^2 \right) \left(\frac{\lambda_{1,P}(\Delta + \tau)^{8\alpha} \|\theta\|_1 \rho_n (\log n) n_{max}}{(\delta + \tau)^{8\alpha-1} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\left(\frac{K\Delta}{\delta + \tau} \right) \left(\frac{\lambda_{1,P}(\Delta + \tau)^{8\alpha} \sqrt{\|\theta\|_1 \rho_n (\log n) n_{max}}}{\Delta(\delta + \tau)^{8\alpha-3} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right) \right] \\
&= O \left[\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{8\alpha} \sqrt{\|\theta\|_1 \rho_n \log n n_{max}}}{(\delta + \tau)^{8\alpha-2} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right] \tag{C.12}
\end{aligned}$$

using the fact that $K \leq K^2$.

Thus, when $\alpha > 1/2$,

$$L(\hat{Z}, Z) = O \left(\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{8\alpha} \sqrt{\Delta \|\theta\|_1 \rho_n \log n n_{max}}}{(\delta + \tau)^{8\alpha-3/2} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right)$$

with high probability.

The analysis is similar when $\alpha < \frac{1}{2}$. In this case, with high probability,

$$L(\hat{Z}, Z) = O \left(\frac{\lambda_{1,P} K^2 (\Delta + \tau)^{6\alpha+1} \sqrt{\Delta \|\theta\|_1 \rho_n \log n n_{max}}}{(\delta + \tau)^{6\alpha-1/2} \lambda_{K,P}^4 \theta_{min}^{2-2\alpha} n} \right).$$

We omit the algebraic details.

We briefly remark that we have used the assumption that $K \leq K^2$, which is true because K is the number of communities. This proof technique could be used to get a slightly more transparent bound in terms of K by keeping track of the two terms involving K and K^2 explicitly. In practice, the number of communities is either known and fixed, or small compared to n . We thus choose not to pursue this bound in favor of clarity of the final result.

C.4 Investigation of Modifications to Spectral Clustering

C.4.1 Additional Eigenvectors

In this section, we review the numerical experiments conducted to investigate the effect of selecting additional eigenvectors for analysis in spectral clustering algorithms. A summary of the results is provided in Figure C.1. In each of these experiments, 50 random graphs are generated and either ASC ($\alpha = 0$), RSC ($\alpha = 1/2$), or DCSC ($\alpha = 1$) is applied. For each algorithm, either 3, 4 or 5 eigenvectors are selected corresponding to the top eigenvalues.

For all trials, $n = 350$, $K = 3$, and the communities are balanced. The degree-correction parameters are randomly selected from a $\text{Unif}(0.7, 1)$ distribution. The community connection matrix B changes between rows. All trials use a homogeneous stochastic block model, with changing α and λ values. For the top row $\alpha = 0.4$ and $\lambda = 0.3$. For the second row, $\alpha = 0.4$ and $\lambda = 0.35$. For the third row, $\alpha = 0.4$ and $\lambda = 0.4$. Since increasing values of λ correspond to increasing signal, community detection becomes easier from as we move down the figure.

From Figure C.1, we may conclude that including additional eigenvalues seems to have a decreasing effect as α increases and as the signal increases.

C.4.2 Eigenvalue Scaling

In this section, we review the numerical experiments conducted to investigate the effect of scaling the eigenvectors in spectral clustering algorithms. A summary of the results is provided in Figure C.1. In each of these experiments, 50 random graphs are generated and either ASC ($\alpha = 0$), RSC ($\alpha = 1/2$), or DCSC ($\alpha = 1$) is applied. For each algorithm, the columns of $\hat{U}_{\alpha,\tau}$ are scaled by $\hat{\Lambda}_{\alpha,\tau}^{\beta}$ for $\beta \in \{0, 0.5, 1\}$.

In the first set of experiments, a homogeneous balanced 4-community DCSCBM is studied. We set $\gamma = 0.5$, $\lambda = 0.4$, and the degree-correction parameters are randomly sampled from a $\text{Unif}(0.7, 1)$ distribution with increasing n . The values used were $n \in \{100, 150, \dots, 500\}$. These experiments correspond to the first row, Figures C.2a-C.2c. In this setup, the choice of scaling has almost no effect on the results of the clustering algorithms.

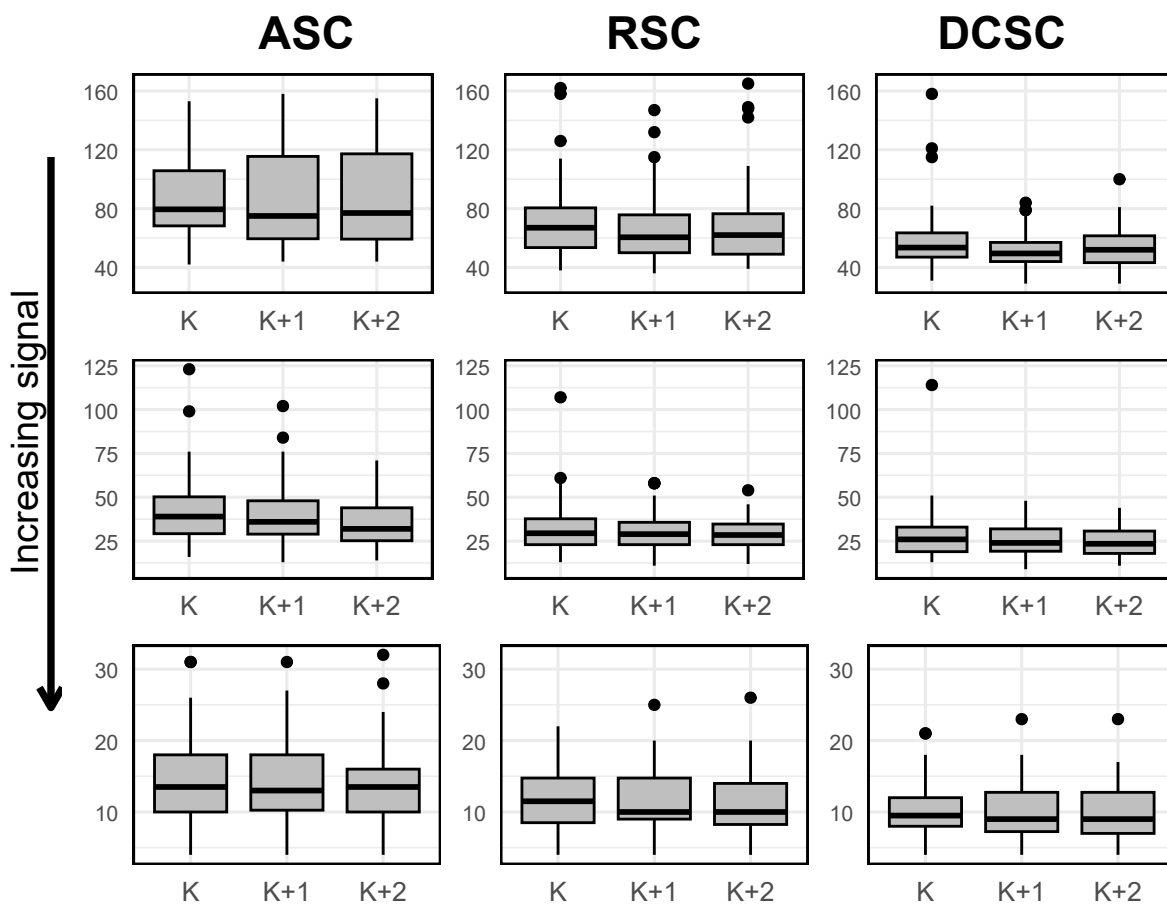


Figure C.1: The effect of including additional eigenvectors for various spectral clustering algorithms. Each column corresponds to a different algorithm, and each row corresponds to a different model, with increasing signal from the top row to the bottom row. Each graph shows boxplots of 50 replicates, using K , $K + 1$, or $K + 2$ eigenvectors for analysis.

In the second set of experiments, a homogeneous imbalanced 4-community DCSBM is studied. We set $\gamma = 0.5, \lambda = 0.3$, and the degree-correction parameters are randomly sampled from a $\text{Unif}(0.7, 1)$ distribution with increasing n . The values used were $n \in \{100, 150, \dots, 500\}$. These experiments correspond to the second row, Figures C.2d-C.2f. In this setup, the scaling weight appears to have a modest effect on the efficacy of the algorithms, with scaling by $\hat{\Lambda}_{\alpha, \tau}$ being preferred.

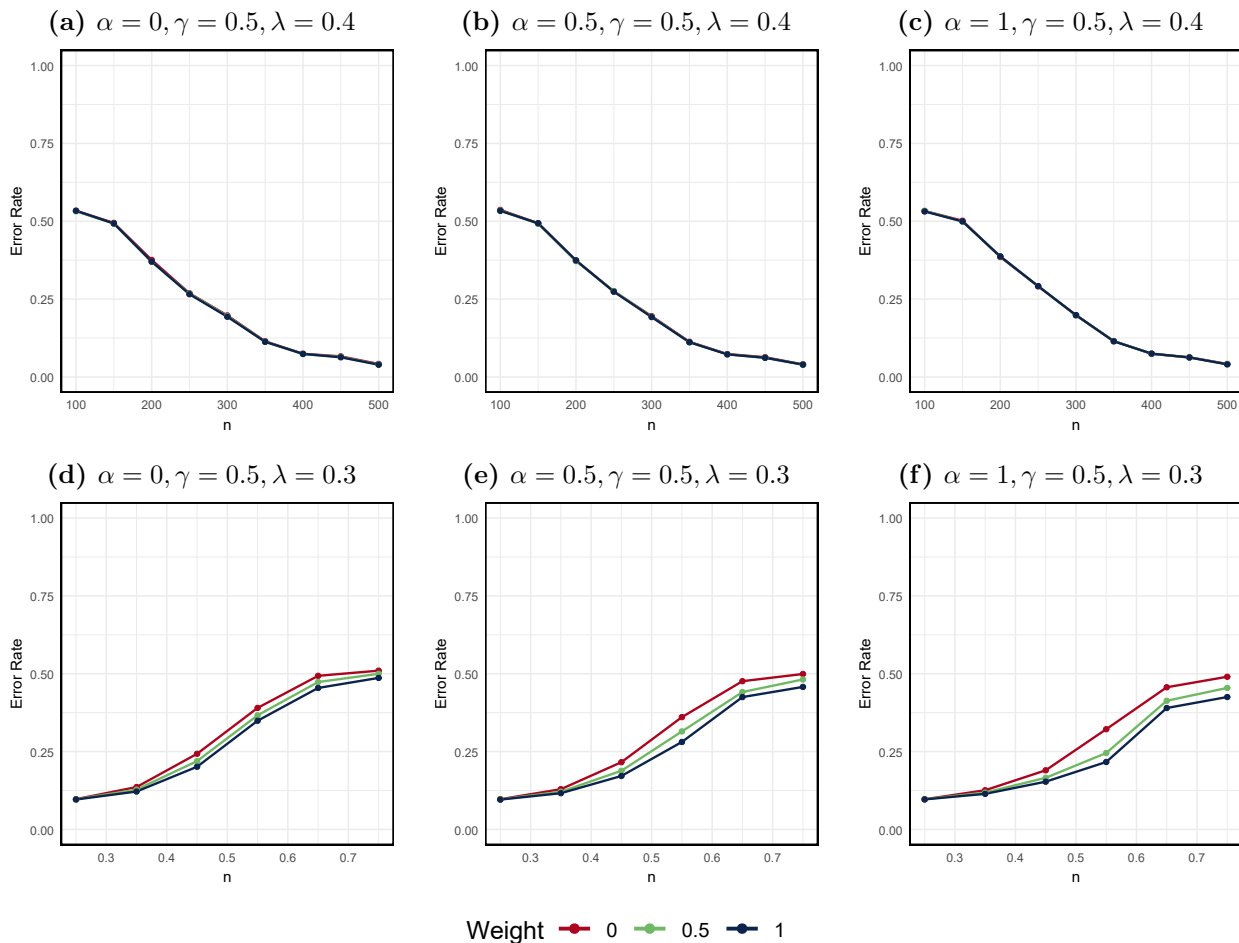


Figure C.2: The effect of scaling eigenvectors for various spectral clustering algorithms. Each column corresponds to a different algorithm, and each row corresponds to a different model, where the top row has a higher signal than the second. For each algorithm, the eigenvectors are scaled by $\hat{\Lambda}_{\alpha, \tau}^{\beta}$ for $\beta \in 0, 0.5, 1$.

C.5 Lemma 24 and Proof

We first define several matrices. Let $U_{\alpha,\tau,i}$ be the $n \times (K+i)$ matrix whose columns consist of the $K+i$ top unit eigenvectors of $L_{\alpha,\tau}$. Let $\Lambda_{\alpha,\tau,i} = \text{diag}(\lambda_1, \dots, \lambda_{K+i})$ be the diagonal matrix whose entries are the $K+i$ top eigenvalues of $L_{\alpha,\tau}$. Note that under this notation, $U_{\alpha,\tau,0}$ and $\Lambda_{\alpha,\tau,0}$ reduce to $U_{\alpha,\tau}$ and $\Lambda_{\alpha,\tau}$. Finally, define $X_{\alpha,\tau,\beta,i} = U_{\alpha,\tau,i} \Lambda_{\alpha,\tau,i}^\beta$ and let $X_{\alpha,\tau,\beta,i}^*$ denote the row-normalized version of $X_{\alpha,\tau,\beta,i}$. These definitions extend naturally to the observed quantities.

Lemma 24. *Under a DCSBM parameterized by (Z, Θ, B) satisfying Conditions 3 and 5, if $\|\theta\|_1 \rho_n \geq \delta + \tau$, the following bound holds with*

$$\left\| \hat{X}_{\alpha,\tau,\beta,i}^* - X_{\alpha,\tau,\beta,i}^* \right\|_F \leq 2 \left(\frac{\sqrt{K+i}}{\eta_{\alpha,\tau}^{1-\beta} \Delta_{\alpha,\tau}^{(2\alpha-1)\beta}} + 2\sqrt{2}K \right) \frac{\Delta^{(2+2\beta)\alpha} \eta_{\tau,\alpha} \sqrt{n_{\max}}}{\Delta_{\alpha,\tau}^{(2\alpha-1)\beta} \lambda_{K,P}^{1+\beta} \theta_{\min}^{1-\alpha}}$$

with high probability.

Proof. The proof of Lemma 11 begins by proving

$$\left\| \hat{X}_{\alpha,\tau} - X_{\alpha,\tau} \right\|_F \leq \left\| \hat{\Lambda}_{\alpha,\tau}^{1/2} - \Lambda_{\alpha,\tau}^{1/2} \right\|_F - \left\| \hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right\|_F \left\| \Lambda_{\alpha,\tau}^{1/2} \right\|_F.$$

We similarly show

$$\begin{aligned} \left\| \hat{X}_{\alpha,\tau,\beta,i} - X_{\alpha,\tau,\beta,i} \right\|_F &= \left\| \hat{U}_{\alpha,\tau,i} \hat{\Lambda}_{\alpha,\tau,i}^\beta - U_{\alpha,\tau,i} \Lambda_{\alpha,\tau,i}^\beta \right\|_F \\ &\leq \left\| \hat{U}_{\alpha,\tau,i} \hat{\Lambda}_{\alpha,\tau,i}^\beta - \hat{U}_{\alpha,\tau,i} \Lambda_{\alpha,\tau,i}^\beta \right\|_F + \left\| \hat{U}_{\alpha,\tau,i} \Lambda_{\alpha,\tau,i}^\beta - U_{\alpha,\tau,i} \Lambda_{\alpha,\tau,i}^\beta \right\|_F \\ &= \left\| \hat{U}_{\alpha,\tau,i} \left(\hat{\Lambda}_{\alpha,\tau,i}^\beta - \Lambda_{\alpha,\tau,i}^\beta \right) \right\|_F + \left\| \left(\hat{U}_{\alpha,\tau,i} - U_{\alpha,\tau,i} \right) \Lambda_{\alpha,\tau,i}^\beta \right\|_F \\ &\leq \left\| \hat{\Lambda}_{\alpha,\tau,i}^\beta - \Lambda_{\alpha,\tau,i}^\beta \right\|_F + \left\| \left(\hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right) \Lambda_{\alpha,\tau}^\beta \right\|_F \\ &= \left\| \hat{\Lambda}_{\alpha,\tau,i}^\beta - \Lambda_{\alpha,\tau,i}^\beta \right\|_F + \left\| \hat{U}_{\alpha,\tau} - U_{\alpha,\tau} \right\|_F \left\| \Lambda_{\alpha,\tau}^\beta \right\|_F. \end{aligned}$$

In the penultimate line, we have used the fact that the rows of $\hat{U}_{\alpha,\tau,i}$ are unit vectors and that $\lambda_{K+i} = 0$ for all $0 < i \leq (n-K)$. The remainder of the proof then proceeds as in proof of Lemma 8, mutatis mutandis. \square

Appendix D

Appendix to Chapter 5

D.1 Derivation of Equation (5.3)

We first compute the means and covariances given by Theorem 9. The means may be calculated using elementary linear algebra,

$$\mu_1 = \begin{bmatrix} \sqrt{\frac{n(p+q)}{2(p+q)^\alpha}} \\ \sqrt{\frac{n(p-q)}{2(p+q)^\alpha}} \end{bmatrix} \quad \mu_2 = \begin{bmatrix} \sqrt{\frac{n(p+q)}{2(p+q)^\alpha}} \\ -\sqrt{\frac{n(p-q)}{2(p+q)^\alpha}} \end{bmatrix}.$$

In this model, the random vector $\xi \in \mathbb{R}^2$ has the distribution

$$P(\xi = x_1) = P(\xi = x_2) = \frac{1}{2},$$

where $X_1 = \begin{bmatrix} \sqrt{\frac{p+q}{2}} \\ \sqrt{\frac{p-q}{2}} \end{bmatrix}$ and $X_2 = \begin{bmatrix} \sqrt{\frac{p+q}{2}} \\ -\sqrt{\frac{p-q}{2}} \end{bmatrix}$. For the covariances, we first calculate

$$\mu = \mathbb{E}[\xi] = \begin{bmatrix} \sqrt{\frac{p+q}{2}} \\ 0 \end{bmatrix} \quad \tilde{\Upsilon} = \mathbb{E} \begin{bmatrix} \xi \xi^\top \\ \langle \xi, \mu \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{p-q}{p+q} \end{bmatrix}.$$

The conditional block covariance matrix for each latent position can then be written

$$\Sigma_\alpha(x) = \tilde{\Upsilon}^{-1} \mathbb{E} \left[\frac{\langle x, \xi \rangle (1 - \langle x, \xi \rangle)}{\langle x, \mu \rangle^{2\alpha}} \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle} \right) \left(\frac{\xi}{\langle \xi, \mu \rangle} - \frac{\alpha \tilde{\Upsilon} x}{\langle x, \mu \rangle} \right)^\top \right] \tilde{\Upsilon}^{-1}.$$

We may calculate

$$\langle x_1, \mu \rangle = \langle x_2, \mu \rangle = \frac{p+q}{2}$$

This observation allows us to write

$$\Sigma_\alpha(x) = \left(\frac{2}{p+q} \right)^{2\alpha+2} \mathbb{E} \left[\langle x, \xi \rangle (1 - \langle x, \xi \rangle) \left(\tilde{\Upsilon}^{-1} \xi - \alpha x \right) \left(\tilde{\Upsilon}^{-1} \xi - \alpha x \right)^\top \right].$$

We may now compute the expectation explicitly at each latent position. When $x = X_1$,

$$\frac{1}{2} \left[p(1-p) \left(\tilde{\Upsilon}^{-1} X_1 - \alpha X_1 \right) \left(\tilde{\Upsilon}^{-1} X_1 - \alpha X_1 \right)^\top + q(1-q) \left(\tilde{\Upsilon}^{-1} X_2 - \alpha X_1 \right) \left(\tilde{\Upsilon}^{-1} X_2 - \alpha X_1 \right)^\top \right].$$

Again we may compute

$$\begin{aligned} \left(\tilde{\Upsilon}^{-1} X_1 - \alpha X_1 \right) \left(\tilde{\Upsilon}^{-1} X_1 - \alpha X_1 \right)^\top &= \begin{bmatrix} \frac{(p+q)(\alpha-1)^2}{2} & -\frac{\sqrt{p+q}(\alpha-1)(p-\alpha(p-q)+q)}{2\sqrt{p-q}} \\ -\frac{\sqrt{p+q}(\alpha-1)(p-\alpha(p-q)+q)}{2\sqrt{p-q}} & \frac{(p-\alpha(p-q)+q)^2}{2(p-q)} \end{bmatrix} \\ \left(\tilde{\Upsilon}^{-1} X_2 - \alpha X_1 \right) \left(\tilde{\Upsilon}^{-1} X_2 - \alpha X_1 \right)^\top &= \begin{bmatrix} \frac{(p+q)(\alpha-1)^2}{2} & \frac{\sqrt{p+q}(\alpha-1)(p+\alpha(p-q)+q)}{2\sqrt{a-b}} \\ \frac{\sqrt{p+q}(\alpha-1)(p+\alpha(p-q)+q)}{2\sqrt{a-b}} & \frac{(p+\alpha(p-q)-q)^2}{2(p-q)} \end{bmatrix}. \end{aligned}$$

Combining terms and simplifying gives

$$\Sigma_{\alpha,1} := \Sigma_\alpha(X_1) = \begin{bmatrix} \frac{4^\alpha(\alpha-1)^2(p(1-p)+q(1-q))}{(p+q)^{2\alpha+1}} & -\frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} \\ -\frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} & \frac{4^\alpha(p(1-p)\gamma_2^2+q(1-q)\gamma_1^2)}{(p-q)(p+q)^{2\alpha+2}} \end{bmatrix},$$

where $\gamma_1 = p - \alpha(p - q) + q$ and $\gamma_2 = p - \alpha(p - q) + q$.

A similar calculation gives

$$\Sigma_{\alpha,2} := \Sigma_\alpha(X_2) = \begin{bmatrix} \frac{4^\alpha(\alpha-1)^2(p(1-p)+q(1-q))}{(p+q)^{2\alpha+1}} & \frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} \\ \frac{4^\alpha(\alpha-1)(p(1-p)\gamma_2-q(1-q)\gamma_1)}{\sqrt{p-q}(p+q)^{2\alpha+\frac{3}{2}}} & \frac{4^\alpha(p(1-p)\gamma_2^2+q(1-q)\gamma_1^2)}{(p-q)(p+q)^{2\alpha+2}} \end{bmatrix}.$$

We may now compute $\rho_{\alpha, \alpha'}$. To do so, we need to compute Equation (5.2), which we restate below.

$$\begin{aligned} \rho_{\alpha}^* &\approx \sup_{t \in (0,1)} \left[t(1-t) \|\mu_1 - \mu_2\|_{\Sigma_{\alpha,12}^{-1}(t)}^2 \right] \\ &= \sup_{t \in (0,1)} \left[t(1-t) (\mu_1 - \mu_2)^{\top} (t\Sigma_{\alpha,1} + (1-t)\Sigma_{\alpha,2})^{-1} (\mu_1 - \mu_2) \right]. \end{aligned}$$

We compute this quantity in two steps. First, we show that the supremum is achieved at $t = 1/2$. Second, we derive an explicit formula for $(t\Sigma_{\alpha,1} + (1-t)\Sigma_{\alpha,2})^{-1}$.

We begin with the first claim. Let $x = \mu_1 - \mu_2$, $A = \Sigma_{\alpha,1}$ and $B = \Sigma_{\alpha,2}$. Define

$$f(t) = t(1-t)(\mu_1 - \mu_2)^{\top} (t(\Sigma_{\alpha,1} - \Sigma_{\alpha,2}) + \Sigma_{\alpha,2})^{-1} (\mu_1 - \mu_2).$$

Notice that $t(\Sigma_{\alpha,1} - \Sigma_{\alpha,2}) + \Sigma_{\alpha,2}$ has the form

$$\begin{bmatrix} c_1 & (1-2t)c_2 \\ (1-2t)c_2 & c_3 \end{bmatrix}$$

for constants c_1, c_2 and c_3 , so

$$(t(\Sigma_{\alpha,1} - \Sigma_{\alpha,2}) + \Sigma_{\alpha,2})^{-1} = \frac{1}{c_1 c_3 - (1-2t)^2 c_2^2} \begin{bmatrix} c_3 & -(1-2t)c_2 \\ -(1-2t)c_2 & c_1 \end{bmatrix}.$$

Additionally, $\mu_1 - \mu_2$ has the form $(0, c_4)$ for some constant c_4 . Therefore,

$$(\mu_1 - \mu_2)^{\top} (t(\Sigma_{\alpha,1} - \Sigma_{\alpha,2}) + \Sigma_{\alpha,2})^{-1} (\mu_1 - \mu_2) = \frac{c_1 c_4^2}{c_1 c_3 - (1-2t)^2 c_2^2}$$

and

$$f(t) = \frac{t(1-t)c_1 c_4^2}{c_1 c_3 - (1-2t)^2 c_2^2}.$$

Elementary calculus and the fact that $\Sigma_{\alpha,1}$ and $\Sigma_{\alpha,2}$ are positive semi-definite show that $t = 1/2$ is indeed the unique value at which the supremum in Equation (5.2) is achieved.

We now establish the second claim, which will allow us to compute $\rho_{\alpha,\alpha'}^*$, we begin with a lemma from [Cape et al. \(2019a\)](#), which follows from elementary matrix calculations.

Lemma 25. *Let $M_0, M_1 \in \mathbb{R}^{2 \times 2}$ be two invertible matrices. For each $t \in [0, 1]$, let $M_t = (1-t)M_0 + t(M_1)$. If M_t is invertible, then*

$$M_t^{-1} = \frac{(1-t)M_0^{-1} + |M_1M_0^{-1}|tM_1^{-1}}{|M_1M_0^{-1}|t^2 + \text{tr}(M_1M_0^{-1})t(1-t) + (1-t)^2}.$$

In particular, note that $|\Sigma_\alpha(X_1)| = |\Sigma_\alpha(X_2)|$. This implies that $|\Sigma_\alpha(X_1)\Sigma_\alpha(X_2)^{-1}| = 1$ and the expression from Lemma 25 becomes

$$M_t^{-1} = \frac{(1-t)M_0^{-1} + tM_1^{-1}}{t^2 + \text{tr}(M_1M_0^{-1})t(1-t) + (1-t)^2}.$$

At the value $t = 1/2$, if $\text{tr}(M_1M_0^{-1}) \neq -2$ (which is true because covariance matrices are positive semi-definite) the expression further simplifies to

$$M_{1/2}^{-1} = \frac{2(M_0^{-1} + M_1^{-1})}{2 + \text{tr}(M_1M_0^{-1})}.$$

Finally,

$$\begin{aligned} \rho_{\alpha,\alpha'}^* &= \frac{\|\mu_{\alpha,1} - \mu_{\alpha,2}\|_{\Sigma_{\alpha,1,2}^{-1}(1/2)}^2}{\|\mu_{\alpha',1} - \mu_{\alpha',2}\|_{\Sigma_{\alpha',1,2}^{-1}(1/2)}^2} \\ &= \frac{2(\mu_{\alpha,1} - \mu_{\alpha,2})^\top (\Sigma_{\alpha,1}^{-1} + \Sigma_{\alpha,2}^{-1})(\mu_{\alpha,1} - \mu_{\alpha,2})}{2 + \text{tr}(\Sigma_{\alpha,1}\Sigma_{\alpha,2}^{-1})} \\ &= \frac{2(\mu_{\alpha',1} - \mu_{\alpha',2})^\top (\Sigma_{\alpha',1}^{-1} + \Sigma_{\alpha',2}^{-1})(\mu_{\alpha',1} - \mu_{\alpha',2})}{2 + \text{tr}(\Sigma_{\alpha',1}\Sigma_{\alpha',2}^{-1})} \\ &= \frac{2 + \text{tr}(\Sigma_{\alpha',1}\Sigma_{\alpha',2}^{-1})}{2 + \text{tr}(\Sigma_{\alpha,1}\Sigma_{\alpha,2}^{-1})} \end{aligned}$$

This last term may be computed explicitly, which proves the lemma.

D.2 Description of Real-World Networks

1. **Karate ($K=2$):** This network consists of members of a karate club, where edges represent a friendship between two members. The author studied a fracture within the group, whereby each member split into one of two communities, led either by Mr. Hi or John. We use the leader of the group as the true labels. (Zachary, 1977)
2. **Football ($K=5$):** This network summarizes American football games between college teams during the regular season of Fall 2022, and the presence of an edge represents a regular-season game. The nodes are grouped into 11 conferences, which are used for the true labels. 5 of the nodes are labeled “independent”, and are removed from the graph. (Girvan and Newman, 2002)
3. **Polbooks ($K=2$):** This network consists of books about US politics published around 2004 and sold by Amazon. The presence of an edge represents frequent purchases of books by the same buyers. The nodes are labeled as either “conservative”, “liberal” or “independent”. The last group is removed from the graph. This network data is unpublished but may be downloaded at <https://websites.umich.edu/~mejn/netdata/>.
4. **Polblogs ($K=2$):** This network consists of blogs about US politics run around 200. The presence of an edge represents links between blogs. The nodes are labeled as either “conservative” or “liberal”. Only the largest connected component is considered in the graph (Adamic and Glance, 2005).
5. **UKFaculty ($K=2$):** This network consists of faculty at a UK university consisting of three separate schools. The presence of an edge represents friendship between the nodes. The smallest group only contains 2 nodes, which is removed from the graph (Nepusz et al., 2008).
6. **Caltech ($K=8$):** This network consists of students at Caltech, with an edge representing a friendship between nodes. The dormitory of the student is used as the true community.

7. **UA**: This network consists of all departments at the University of Arizona, and edges represent collaborations between academics. The data are provided by the KMAP group. To pre-process the data, all nodes of degree less than 5 are removed from analysis. Then, the largest connected component is selected for study. For the example in Section 2.3.1, a subgraph consisting of 4 departments is chosen for community detection (Hossain et al., 2025).

D.3 List of Departments and Grouping

This appendix lists each department and their group in the KMAP dataset analyzed in Section 5.3.1. The table also includes the agreement percentage, which is the number of members in that department assigned to the larger group, and the total number of members in each department. We remark although the graph and corresponding visualization consists of 166 communities, only 141 appear in the table below due to missing data in the cross-tabulation.

Department	Group	Agreement Pct.	Total Members
Agricultural Resource Economics - Research	1	1.00	8
Apache County Office	1	1.00	5
Applied Science	1	0.46	13
AZ Center for Judaic Studies	1	0.75	4
AZ Institute for Resilience	1	1.00	12
AZ Veterinary Diagnostic Lab	1	1.00	8
Biosystems Engineering Research	1	0.89	9
Cochise County Office	1	1.00	4
Disability Psychoeduc Studies	1	0.79	14
Ecology & Evolutionary Biology	1	0.57	14
Education - Dean's Office	1	1.00	3
Educational Policy Studies Practice	1	1.00	17
Educational Psychology	1	0.89	9
English	1	0.71	7
Entomology-Research	1	1.00	13
Environmental Science-Research	1	0.79	33
Environmental Science - Extension	1	1.00	4
Gila County Office	1	1.00	5
History	1	0.67	3
Information Science	1	0.55	22

Law Instruction	1	0.70	10
Maricopa County Office	1	1.00	8
Mexican American Studies	1	0.67	3
Nutritional Sciences and Wellness Extended Studies - Extension	1	1.00	4
Office of Societal Impacts	1	0.50	4
Other	1	0.59	90
Pima County Office	1	0.75	8
Pinal County Office	1	1.00	4
Planning Degree Program	1	0.86	7
Public & Applied Humanities	1	1.00	5
Radiation Oncology	1	0.67	15
Research Engagement	1	0.90	10
Research Innovation & Impact	1	0.67	3
Santa Cruz County Office	1	1.00	10
School of Anthropology	1	0.57	7
School of Architecture	1	0.57	7
School of Geography, Development and Environment	1	0.83	6
School of Global Studies	1	0.56	9
School of Health Professions	1	0.50	4
School of Human Ecology - Research	1	1.00	15
School of Landscape Architecture	1	0.79	14

School of Natural Resources and Environmental - Research	1	0.97	35
School of Plant Sciences - Research	1	0.84	19
Sociology	1	0.57	7
Spanish and Portuguese	1	0.50	4
Special Collections	1	1.00	5
Student Learning & Engagement	1	1.00	7
SW Institute for Research on Women	1	0.78	9
Teaching Learning and Sociocultural Studies	1	0.89	28
The Honors College	1	0.82	11
UCATT	1	0.50	6
Udall Center	1	1.00	6
Water Resources Research Center	1	1.00	4
Yavapai County Office	1	1.00	3
Yuma County Office	1	1.00	3
Aerospace and Mechanical Engineering	2	0.93	27
Arizona Geological Survey	2	1.00	21
Arizona State Museum	2	1.00	9
Chemical & Environmental Engineering	2	0.56	16
Civil and Architectural Engineering and Mechanics	2	0.75	16
Computer Science	2	0.90	20
Economics	2	0.60	5

Electrical and Computer Engineering	2	0.78	32
Geosciences	2	0.93	43
Hydrology & Atmospheric Sciences	2	0.56	25
James C. Wyant College of Optical Sciences	2	0.69	52
Linguistics	2	1.00	10
Management Information Systems	2	0.69	13
Materials Science and Engineering	2	1.00	19
Mathematics	2	0.54	28
Research & Discovery Tech	2	0.67	3
Sch Mining Engr Mineral Res	2	1.00	15
Systems and Industrial Engineering	2	0.80	15
Tech Launch Arizona	2	1.00	3
Tree Ring Laboratory	2	0.86	22
AHS Library	3	0.80	5
Andrew Weil Center for Integrative Medicine	3	0.88	8
Anesthesiology	3	1.00	11
Animal & Biomedical Sciences - Research	3	0.64	25
Arthritis Center-Research	3	1.00	5
Asthma and Airway Disease Research Center	3	1.00	31
BIO5 Institute	3	0.68	22
Biomedical Engineering	3	0.87	15

Cancer Center Division	3	1.00	46
Cellular & Molecular Medicine	3	1.00	34
Center for Rural Health	3	0.88	8
Center for Toxicology	3	1.00	4
Chemistry & Biochemistry - Science	3	0.50	48
Clinical Teaching	3	1.00	4
College of Medicine Phoenix Basic Medical Sciences	3	1.00	22
College Of Medicine Phoenix Child Health	3	1.00	11
College of Medicine Phoenix Emergency Medicine	3	1.00	3
College of Medicine Phoenix Family Community and Preventive Medicine	3	0.92	12
College of Medicine Phoenix Internal Medicine	3	1.00	89
College of Medicine Phoenix NanoBioscience & Medicine	3	1.00	5
College of Medicine Phoenix Obstetrics and Gynecology	3	1.00	6
College of Medicine Phoenix Orthopedics	3	1.00	6
College of Medicine Phoenix Psychiatry	3	1.00	3
College of Medicine Phoenix Radiology	3	1.00	5
College of Medicine Phoenix Surgery	3	1.00	11
College of Nursing	3	1.00	81
Communication	3	0.75	4
Community Environment & Policy	3	0.62	37
D7: AZ Data Science Initiative	3	0.60	5

Department of Emergency Medicine	3	1.00	55
Epidemiology and Biostatistics	3	0.74	34
Family and Community Medicine	3	0.92	61
Health Promotion Sciences	3	0.92	40
Immunobiology	3	1.00	19
Liver Research Institute	3	1.00	5
Medicine	3	1.00	142
Molecular and Cellular Biology	3	0.52	21
Neurology	3	1.00	23
Neuroscience	3	0.62	8
Neurosurgery	3	1.00	13
Nutritional Science Wellness - Research	3	0.59	17
Obstetrics and Gynecology	3	1.00	10
Ophthalmology & Vision Science	3	1.00	7
Orthopedic Surgery	3	1.00	11
Otolaryngology	3	1.00	6
Pathology and Lab Medicine	3	1.00	18
Pediatrics	3	1.00	56
Pharmacology	3	1.00	22
Pharmacology and Toxicology	3	1.00	32
Pharmacy Practice and Science	3	0.97	32

Physiology	3	0.97	35
Poison Control Center	3	1.00	6
Psychiatry	3	1.00	19
Psychology	3	0.91	35
Public Health Practice and Translational Research	3	1.00	18
Radiology & Imaging Sciences Department	3	0.97	31
Research Laboratory and Safety Services	3	1.00	4
RII Core Facilities	3	1.00	7
Sarver Heart Center	3	1.00	14
Senior VP Health Sciences	3	1.00	15
Speech Language & Hearing Sciences	3	0.92	12
Student Development & Career Advancement	3	1.00	3
Surgery	3	0.98	65
UAHS Brain Science	3	1.00	10
UAHS Research	3	0.91	11
University Animal Care	3	1.00	3
Urology	3	1.00	5
Valley Fever Center for Excellence	3	1.00	7
Astronomy	4	1.00	27
Lunar and Planetary Laboratory	4	0.97	36
Physics	4	0.55	20

Planetary Sciences	4	0.95	21
Steward Observatory	4	1.00	72

References

- E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452, 2020.
- L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.
- E. M. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Advances in Neural Information Processing Systems*, 21, 2008.
- A. A. Amini, A. Chen, P. Bickel, and L. Levina. Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, pages 1–27, 2013.
- A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, Y. Qin, and D. L. Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.
- V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10):P10008, 2008.
- J. Cape, M. Tang, and C. E. Priebe. On spectral embedding performance and elucidating network structure in stochastic blockmodel graphs. *Network Science*, 7(3):269–291, 2019a.

- J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, 2019b.
- K. Chaudhuri, F. Chung, and A. Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. *Journal of Machine Learning Research*, pages 1–23, 2012.
- K. Chen and J. Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251, 2018.
- Y. Chen, X. Li, and J. Xu. Convexified modularity maximization for degree-corrected stochastic block models. *The Annals of Statistics*, 46(4):1573–1602, 2018.
- H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.
- F. Chung and M. Radcliffe. On the spectra of general random graphs. *The Electronic Journal of Combinatorics*, pages P215–P215, 2011.
- F. R. Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.
- A. Clauset, M. E. Newman, and C. Moore. Finding community structure in very large networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 70(6):066111, 2004.
- M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- P. D’haeseleer. How does gene expression clustering work? *Nature Biotechnology*, 23(12):1499–1501, 2005.
- W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.
- P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.

- C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *Journal of Machine Learning Research*, 18(60):1–45, 2017.
- E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.
- M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- L. Hagen and A. B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE transactions on computer-aided design of integrated circuits and systems*, 11(9):1074–1085, 1992.
- M. S. Handcock, A. E. Raftery, and J. M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 170(2):301–354, 2007.
- P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.
- P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
- R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2013.
- I. Hossain, T. Harman, W. Nguyen, and R. Chadha. Understanding research dynamics at the university of arizona: An ai-driven metadata analysis. In *2025 ASEE Annual Conference & Exposition*, 2025.
- J. Jin. Fast community detection by score. *The Annals of Statistics*, 43(1):57–89, 2015.
- J. Jin, Z. T. Ke, and S. Luo. Improvements on score, especially for weak signals. *Sankhya A*, 84(1):127–162, 2022.
- B.-Y. Jing, T. Li, N. Ying, and X. Yu. Community detection in sparse networks using the symmetrized laplacian inverse matrix (slim). *Statistica Sinica*, 32(1):1–22, 2022.

- A. Joseph and B. Yu. Impact of regularization on spectral clustering. *The Annals of Statistics*, 44(4):1765–1791, 2016.
- B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.
- A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor. Compartments revealed in food-web structure. *Nature*, 426(6964):282–285, 2003.
- J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.
- F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006.
- L. Lu and F. Chung. *Complex graphs and Networks*. American Mathematical Society, 2006.
- L. Lu and X. Peng. Spectra of edge-independent random graphs. *The Electronic Journal of Combinatorics*, 20(4):P27, 2013.
- A. Modell and P. Rubin-Delanchy. Spectral clustering under degree heterogeneity: a case for the random walk laplacian, 2021. URL <https://arxiv.org/abs/2105.00987>.
- T. Nepusz, A. Petróczy, L. Négyessy, and F. Bacsó. Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E*, 77:016107, Jan 2008.
- J. Neyman and E. S. Pearson. Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.
- A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 2001.
- S. Pal and M. Coates. Scalable mcmc in degree corrected stochastic block model. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5461–5465. IEEE, 2019.

- J. Park, Y. Zhao, and N. Hao. A note on the identifiability of the degree-corrected stochastic block model. *Stat*, 14(2):e70067, 2025.
- T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- H. Qing and J. Wang. Applications of dual regularized Laplacian matrix for community detection. *Advances in Data Analysis and Classification*, 18(4):1001–1043, Dec. 2024. ISSN 1862-5347, 1862-5355.
- K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, pages 1878–1915, 2011.
- P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe. A statistical interpretation of spectral embedding: The generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1446–1473, 06 2022. ISSN 1369-7412.
- J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- M. Tang and C. E. Priebe. Limit theorems for eigenvectors of the normalized laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, 2018.
- R. Vershynin. *High-dimensional probability: An introduction with applications in Data Science*. Cambridge University Press, 2018.
- U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- Y. J. Wang and G. Y. Wong. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.

S. J. Young and E. R. Scheinerman. Random dot product graph models for social networks. In *International workshop on algorithms and models for the web-graph*, pages 138–149. Springer, 2007.

W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.