An Improved and Generalised Analysis for Spectral Clustering

George Tyler
Department of Mathematical Sciences
University of Bath
grlt20@bath.ac.uk

Luca Zanetti
Department of Mathematical Sciences
University of Bath
lz2040@bath.ac.uk

Abstract

We revisit the theoretical performances of Spectral Clustering, a classical algorithm for graph partitioning that relies on the eigenvectors of a matrix representation of the graph. Informally, we show that Spectral Clustering works well as long as the smallest eigenvalues appear in groups well separated from the rest of the matrix representation's spectrum. This arises, for example, whenever there exists a hierarchy of clusters at different scales, a regime not captured by previous analyses. Our results are very general and can be applied beyond the traditional graph Laplacian. In particular, we study Hermitian representations of digraphs and show Spectral Clustering can recover partitions where edges between clusters are oriented mostly in the same direction. This has applications in, for example, the analysis of trophic levels in ecological networks. We demonstrate that our results accurately predict the performances of Spectral Clustering on synthetic and real-world data sets.

1 Introduction

Spectral Clustering [1] is one of the most popular algorithms for clustering a graph. It exploits the eigenvectors of a matrix representing the graph to compute a low-dimensional Euclidean embedding of the vertices, which is then partitioned using geometric clustering algorithms, such as k-means.

While Spectral Clustering has been shown to work well in practice in a wide variety of settings, it still lacks a complete theoretical understanding. Typically, analyses of Spectral Clustering work by showing that the chosen eigenvectors are close to linear combinations of the indicator vectors of the clusters. For example, the structure theorem of Peng et al. [2] informally says that the bottom k eigenvectors of the graph Laplacian are close to linear combinations of the k "best" clusters whenever two conditions are satisfied: (1) the k-way expansion constant[1] of the optimal partition is *small*; (2) the $(k+1)$-st smallest eigenvalue of the Laplacian is *large*. However, the first condition is often not satisfied in practice and, we argue, is actually not necessary for Spectral Clustering to perform well.

Our main contribution is a strengthening of the structure theorem: we show Spectral Clustering works well whenever the informative eigenvalues are well-separated from the rest of the spectrum. This not only allows us to recover known bounds achieved by either the structure theorem or by classical perturbation arguments such as the Davis-Kahan theorem [5], but also to obtain new improved bounds in many scenarios. In particular, our techniques are well-suited to deal with situations where there exists a hierarchy of clusters at different scales. We believe our results are the first to correctly predict the excellent performances of Spectral Clustering for a large class of real-world networks.

An additional advantage of our analysis is that it applies to any Hermitian positive semidefinite representation of a graph, beyond the standard graph Laplacian. We showcase this strength by analysing recently proposed Hermitian representations of digraphs [6]. In particular, we consider a clustering problem in which we aim to partition the vertices of a digraph into subsets $S_1, \dots, S_k$ so that most of the edges are from S_i to S_{i+1} for any $i \in \{1, \dots, k-1\}$. In other words, we would like most of the edges to follow a directed path between the clusters. This has applications, for example,

[1] The k-way expansion constant is a measure of quality of a partition [3] similar to the normalised cut [4].

G. Tyler & L. Zanetti, An Improved and Generalised Analysis for Spectral Clustering. *Proceedings of the Fourth Learning on Graphs Conference (LoG 2025)*, PMLR 338, Hybrid Event, December 10–12, 2025.

in uncovering trophic levels in food chains or detecting patterns in trade networks [7]. We provide a new cost function for this task and apply our structure theorem to obtain bounds that accurately predict the performances of Spectral Clustering on both synthetic and real-world data sets.

1.1 Related work

There is a wealth of literature on Spectral Clustering; in this section we discuss only the most relevant work on the subject and we refer the reader to the classical surveys [8, 9] for additional background.

In the case of random graphs, and specifically stochastic block models, analyses of Spectral Clustering typically rely on the Davis-Kahan theorem [5] or similar tools [10]. Classical perturbation arguments, however, are poorly suited to deal with non-random graphs in which "noise" might be localised in relatively small regions of the graph [11]. For this reason, Peng et al. [2] introduced their structure theorem, which states that the bottom k-eigenvectors of the normalised Laplacian of an undirected graph are close to the indicator vectors of the "optimal" k clusters whenever the ratio $k^2 \cdot \rho(k)/\lambda_{k+1}$ is small, where $\rho(k)$ is the k-way expansion constant [3], and λ_{k+1} is the $(k+1)$-th smallest eigenvalue of the Laplacian. A series of recent results [12, 13], culminating with [14], further simplified the proof of the structure theorem, while improving the dependency on the number of clusters k.

All of these results, however, require $\rho(k)$ to be small: this is often not satisfied in real-world networks nor in stochastic block models, even for a range of parameters where Spectral Clustering is known to work well. Consider for example a graph consisting of two cliques of 50 vertices connected by a perfect matching in which each edge has weight 20. Spectral clustering perfectly partitions the graph by dividing the two cliques. If we apply the structure theorem of Macgregor and Sun [14], however, we obtain a bound on the distance between the eigenvector of the Laplacian associated with λ_2 and the closest linear combination of indicator vectors of the clusters equals to 0.28. This is because, even though there exists a large gap in the Laplacian spectrum, the value of $\rho(2)$ is relatively large (≈ 0.28). In contrast, our Corollary 4 will show that this distance is actually zero, correctly predicting the exact recovery of the clusters by spectral clustering.

Furthermore, our analysis is not restricted to the traditional (normalised) graph Laplacian and undirected graphs, but it can handle any Hermitian and positive semidefinite representation of undirected or directed graphs. Hermitian representations of digraphs were recently investigated for clustering purposes by Cucuringu et al. [6], who proposed their use to recover clusters characterised by a strong imbalance in the direction of the inter-cluster edges. They define a directed stochastic block model that captures this problem and show Spectral Clustering is able to recover its communities. Laenen and Sun [7] consider the specific instance of this clustering task in which the direction of most edges is required to follow a directed path between the clusters. They apply the techniques of Peng et al. [2] to obtain a structure theorem and an analysis of Spectral Clustering for appositely-constructed Hermitian representations of digraphs. In our work, we revisit this task and argue that the results of Laenen and Sun do not capture the practical performances of Spectral Clustering. Indeed, we demonstrate simple examples where Spectral Clustering works very well, but Laenen and Sun's results are completely uninformative. Applying our improved structure theorem together with a new cost function allows us to prove bounds that correctly predict Spectral Clustering's practical performances, vastly outperforming Laenen and Sun's results.

1.2 Organisation

The paper is organised as follows: in Section 2, we cover the necessary background. In Section 3, we state our main result, which is a generalised and improved structure theorem. We show how our results allow us to obtain meaningful bounds for Spectral Clustering, even for graphs where the k-way expansion constant is relatively large. In Section 4, we apply our generalised structure theorem to analyse Spectral Clustering on Hermitian Laplacians for digraphs. Proofs, together with additional experiments, are included in the Appendix. Code is available in a GitHub repository[2].

2 Background

Let $\mathcal{G} = (V, E, w)$ be a (possibly directed) graph with N vertices , M edges and weight function $w : V \times V \to \mathbb{R}_{\geq 0}$. For any edge $e = (u, v) \in E$, we write the weight of e as w_{uv} or w_e. For an

[2] https://github.com/GeorgeRLTyler/Improved-and-Generalised-Analysis-for-Spectral-Clustering

undirected graph $\mathcal{G}$, we denote the degree of vertex u by $d(u) = \sum_{v \in V} w_{uv}$. If $\mathcal{G}$ is directed, we define the in-degree and out-degree of vertex u as $d_{\text{in}}(u) = \sum_{v \in V} w_{vu}$ and $d_{\text{out}}(u) = \sum_{v \in V} w_{uv}$. The degree of vertex u in this case is $d(u) = d_{\text{in}}(u) + d_{\text{out}}(u)$. For any two sets $S, T \in V$, we define the cut value $w(S,T) = \sum_{\substack{(u,v) \in E \\ u \in S, v \in T}} w_{uv}$ and the volume of $S \subset V$ as $\text{vol}(S) = \sum_{u \in S} d(u)$.

We define the conductance of $\emptyset \neq S \subset V$ as

$$\Phi(S) = \frac{w(S, V - S)}{\text{vol}(S)}.$$

A k-way partition of V is a collection of k subsets $S_1, \ldots, S_k \subset V$ where $S_i \cap S_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{k} S_i = V$. For undirected graphs, we measure the quality of a partition by the k-way expansion:

$$\Phi(S_1, \ldots, S_k) = \max_{1 \leq i \leq k} \Phi(S_i).$$

The k-way expansion constant of $\mathcal{G}$ [3] is defined as

$$\rho(k) = \min_{\text{partition } S_1, \ldots S_k} \Phi(S_1, \ldots, S_k).$$

Spectral clustering leverages eigenvalues and eigenvectors of Hermitian matrices associated with the graph. The adjacency matrix $A \in \mathbb{R}^{N \times N}$ of an undirected graph $\mathcal{G} = (V, E, w)$ is defined as $A_{uv} = w_{uv}$ if $(u, v) \in E$ and zero otherwise. When $\mathcal{G}$ is directed, we will use the Hermitian adjacency matrix as defined in [6]:

$$A_{uv} = \begin{cases} w_{uv} \exp\left(2\pi \mathrm{i}/\tilde{k}\right) & \text{if } u \to v, \\ w_{uv} \exp\left(-2\pi \mathrm{i}/\tilde{k}\right) & \text{if } u \leftarrow v, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where i is the imaginary unit and the value $\tilde{k}$ must be prescribed. Throughout our paper we will assume for simplicity that, for a directed graph, $(u, v) \in E \implies (v, u) \notin E$. The following definitions all apply to both undirected and directed graphs using each respective adjacency matrix.

The degree matrix $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix where each diagonal entry equals the degree of a vertex: $D_{uu} = d(u)$. The Laplacian matrix $L \in \mathbb{R}^{N \times N}$ is defined as $L = D - A$, while the normalized Laplacian matrix $\mathcal{L} \in \mathbb{R}^{N \times N}$ is defined as $\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$.

We will use $M \in \mathbb{C}^{N \times N}$ to denote any Hermitian positive semidefinite matrix representation of a graph. We will denote its eigenvectors by $f_1, \ldots, f_N \in \mathbb{C}^N$ with corresponding eigenvalues $\lambda_1 \leq \ldots \leq \lambda_N$. We will write the eigendecomposition of M as $M = F \Delta F^*$ with $F = (f_1\, f_2 \cdots f_N)$ and $\Delta = \text{diag}(\lambda_1, \ldots, \lambda_N)$.

Given a basis of orthonormal vectors $g_1, \ldots, g_N \in \mathbb{C}^N$, we denote their Rayleigh quotients by $\gamma_i = g_i^* M g_i$, with $\gamma_1 \leq \ldots \leq \gamma_N$. We assemble these vectors in a matrix $G \in \mathbb{C}^{N \times N}$.

There are many variants of Spectral Clustering in the literature. Our results are quite general and will apply to most of these variants. For simplicity, we will consider the variant defined in Algorithm 1, which is the one considered in [2]. For undirected graphs, we typically choose $\tilde{k} = k$ and $M = \mathcal{L}$, which has real eigenvectors. The third step of Algorithm 1 is optional and typically depends on the matrix representation used; for example, it is needed when M is the normalised Laplacian $\mathcal{L}$, but not when M is the combinatorial Laplacian L.

3 A generalised Structure Theorem

Our first main result is an improved and generalised structure theorem. It shows that, for any matrix M, as long as $\lambda_{k+1} \gg \lambda_1$, any set of k orthonormal vectors with Rayleigh quotient close to the smallest eigenvalue of M must be close to linear combinations of the bottom k eigenvectors of M. Furthermore, the bottom k eigenvectors of M will be close to linear combinations of these vectors.

Algorithm 1 Spectral Clustering

1: **Input:** $\mathcal{G} = (V, E, w), M \in \mathbb{C}^{|V|\times|V|}, k \geq 2, \tilde{k} \leq k$
2: Compute the matrix $F \in \mathbb{C}^{|V|\times\tilde{k}}$ whose columns are the orthonormal eigenvectors $f_1, \dots, f_{\tilde{k}}$ associated to $\lambda_1 \leq \dots \leq \lambda_{\tilde{k}}$.
3: $\tilde{F} \leftarrow D^{1/2}F$
4: Minimise the following k-means objective:

$$\min_{c_1,\dots,c_k \in \mathbb{C}^{\tilde{k}}} \sum_{u\in V} d(u)\|\tilde{F}_{u,:} - c_i\|_2^2$$

5: **Output:** partition $A_1, \dots, A_k$ corresponding to the solution of the above k-means problem.

Theorem 1. *Let $M \in \mathbb{C}^{N\times N}$ be Hermitian and positive semidefinite with eigenvalues $0 \leq \lambda_1 \leq \dots \leq \lambda_N$ and corresponding orthonormal basis of eigenvectors $f_1, \dots, f_N$. Let $g_1, \dots, g_k \in \mathbb{C}^N$ be orthonormal and let $\gamma_i := g_i^* M g_i$ $(1 \leq i \leq k < N)$. Then, if $\lambda_{k+1} > \lambda_1$,there exist $\hat{g}_1, \dots, \hat{g}_k \in \operatorname{span}\{g_1, \dots, g_k\}$, such that*

$$\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2 \leq \frac{\sum_{i=1}^{k}\gamma_i - k\lambda_1}{\lambda_{k+1} - \lambda_1}.$$

We observe we can simply recover the structure theorem - Theorem 1 of Macgregor and Sun [14] - by choosing $M = \mathcal{L}$, the normalised Laplacian of an undirected graph $\mathcal{G} = (V, E, w)$, and letting $g_1, \dots, g_k$ be the (normalised) indicator vectors of the clusters achieving the minimum $\rho(k)$.

Corollary 2. *Let $\mathcal{G}$ be undirected and connected with normalized Laplacian $\mathcal{L}$. Let $\{S_i\}_{i=1}^k$ be any optimal k-way partition that achieves $\rho(k)$. For any $1 \leq i \leq k$, define $\chi_i \in \mathbb{R}^N$ as $\chi_i(u) = 1$ if $u \in S_i$ and $\chi_i(u) = 0$ otherwise. Let $g_i = \frac{D^{1/2}\chi_i}{\|D^{1/2}\chi_i\|}$. Then, There exist $\hat{g}_1, \dots, \hat{g}_k \in \operatorname{span}\{g_1, \dots, g_k\}$, such that*

$$\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2 \leq \frac{k\rho(k)}{\lambda_{k+1}}.$$

Informally, Theorem 1 states that if we choose a matrix representation M of a graph and $g_1, \dots, g_k$ indicator vectors of some clusters, then the bottom k eigenvectors of M must be close to linear combinations of these indicator vectors as long as the Rayleigh quotients of the indicator vectors are significantly smaller than λ_{k+1}. This is crucial to show that Spectral Clustering works well: in Appendix A, we show how Theorem 1 can be used to obtain a bound on the volume of the symmetric difference between the partition output by spectral clustering and the one achieving $\rho(k)$.

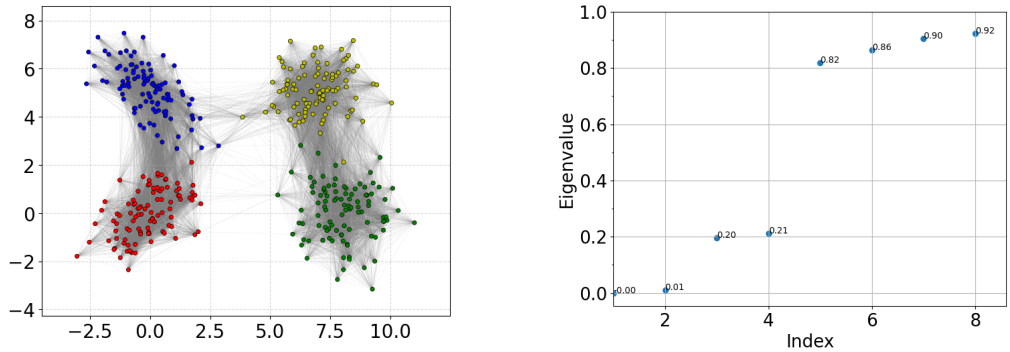
Besides recovering the original structure theorem, Theorem 1 is not confined to the normalised Laplacian of undirected graphs. In Section 4 we will show how to apply Theorem 1 to Hermitian representations of digraphs. In particular, we do not necessarily need each g_i to be a classical indicator vector: in certain cases (as in digraph clustering) it might be beneficial to choose a different set of "indicator vectors". Moreover, in digraphs, λ_1 might not necessarily be zero: subtracting λ_1 from both numerator and denominator on the RHS of our bounds can improve them substantially.

The main limitation of Theorem 1 and Corollary 2 is that they both rely on the gap $\gamma_i - \lambda_1$ being significantly smaller than $\lambda_{k+1} - \lambda_1$. This, in practice, is not always satisfied. For example, when using $M = \mathcal{L}$, we often see multiple gaps in the eigenvalues (which might result in some γ_i being considerably larger than $\lambda_1 = 0$), with Spectral Clustering still working effectively (see, e.g., Figure 1). To overcome this obstacle, we can recursively apply the following result to show that, as long as these gaps in the spectrum are large enough, Spectral Clustering can still be effective.

Theorem 3 (Recursive Structure Theorem)**.** *Let $q < k$ and $\hat{g}_1, \dots, \hat{g}_q \in \operatorname{span}\{g_1, \dots, g_q\}$. Then, there exist $\hat{g}_{q+1}, \dots, \hat{g}_k \in \operatorname{span}\{g_1, \dots, g_k\}$ and $\hat{f}_{q+1}, \dots, \hat{f}_k \in \operatorname{span}\{f_1, \dots, f_k\}$ such that*

$$\sum_{i=q+1}^{k} \|f_i - \hat{g}_i\|^2 \leq \frac{\sum_{i=q+1}^{k}(\gamma_i - \lambda_{q+1}) + \lambda_{k+1}\sum_{i=1}^{q}\|f_i - \hat{g}_i\|^2}{\lambda_{k+1} - \lambda_{q+1}}$$

The same bound also applies to $\sum_{i=q+1}^{k} \|\hat{f}_i - g_i\|^2$



(a) Geometric graph

(b) Smallest eigenvalues of the normalised Laplacian

Figure 1: Four clusters generated by sampling points from a mixture of 4 Gaussians and the corresponding geometric graph (1a). Notice how the smallest four eigenvalues come in pairs (1b).

The idea behind Theorem 3 is that, if the indicator vectors can be expressed *mostly* by the eigenvectors $f_{q+1}, \dots, f_k$, then by the orthonormality of both sets of vectors, the indicator vectors of $g_1, \dots, g_q$ can be expressed *mostly* by $f_1, \dots, f_q$, so the error term $\lambda_{k+1} \sum_{i=1}^{q} \|f_i - \hat{g}_i\|^2$ will be small and in exchange we have an improved ratio to that of Theorem 1 provided $\gamma_i < \lambda_{k+1}$ for $i = q+1, \dots, k$.

Theorem 3 is most appropriate when there are multiple gaps in the eigenvalues of M: we split the spectrum into groups of eigenvalues of similar magnitude, with a large gap between each group; we then recursively reapply Theorem 3 for each group $\lambda_{q+1}, \dots, \lambda_k$ as long as $\gamma_i - \lambda_{q+1} \ll \lambda_{k+1} - \lambda_{q+1}$.

Sketch Proof. The optimal choice of $\hat{g}_i$ for $i = q, \dots, k$ is the projection of the indicator vectors onto the first k eigenvectors. Extending the normalised indicator vectors to an orthonormal basis forming orthogonal matrix G, we obtain a orthogonal projection matrix Q where $F = GQ^*$. The rows/columns of Q have unit length so $\sum_{i=q+1}^{k} \|f_i - \hat{g}_i\|^2$ can be bounded by $(k-q) - \sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2$. The sum in this expression can then be bounded below as follows where we have again exploited the unit length rows of Q:

$$\gamma_i = g_i^* M g_i = \sum_{j=1}^{N} \lambda_j |Q_{ij}|^2 \geq (\lambda_{q+1} - \lambda_{k+1}) \sum_{j=q+1}^{k} |Q_{ij}|^2 + \lambda_{k+1} \left(1 - \sum_{j=1}^{q} |Q_{ij}|^2 \right)$$

Rearranging for $\sum_{j=q+1}^{k} |Q_{ij}|^2$ and summing over $i = q+1, \dots, k$ provides us with with a lower bound on $\sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2$. Consequently, this gives an upper bound on $(k-q) - \sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2$ as required. □

Typically, the choice for the indicator vectors $g_1, \dots, g_k$ is $g_i = \frac{D^{1/2}\chi_i}{\|D^{1/2}\chi_i\|}$ where $\chi_i(u) = 1$ if $u \in S_i$ and $\chi_i(u) = 0$ otherwise. The only real restrictions on our choice of indicator vectors, however, are that they are orthonormal and that $\{D^{-1/2} g_i\}_{i=1}^{k}$ are constant on the clusters. Since the first eigenvector f_1 of $\mathcal{L}$ (or L) is the uninformative vector $f_1 = \frac{D^{1/2}1}{\|D^{1/2}1\|}$ (or $f_1 = \frac{1}{\|1\|}$), we can choose $g_1 = f_1$ and then re-orthogonalise $g_2, \dots, g_k$. This allows us to prove the following corollary.

Corollary 4. *Let $\mathcal{G}$ be an undirected graph and $M = \mathcal{L}$. Let $g_1, \dots, g_k \in \mathbb{R}^N$ be orthonormal such that $g_1 = f_1$. Let $\gamma_i = g_i^* M g_i$ $(i = 1, \dots, k)$. Suppose $\lambda_{k+1} > \lambda_2$. Then, for $i = 1, \dots, k$, there exists $\hat{g}_i \in \operatorname{span}\{g_1, \dots, g_k\}$ such that*

$$\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2 \leq \frac{\sum_{i=2}^{k} (\gamma_i - \lambda_2)}{\lambda_{k+1} - \lambda_2}.$$

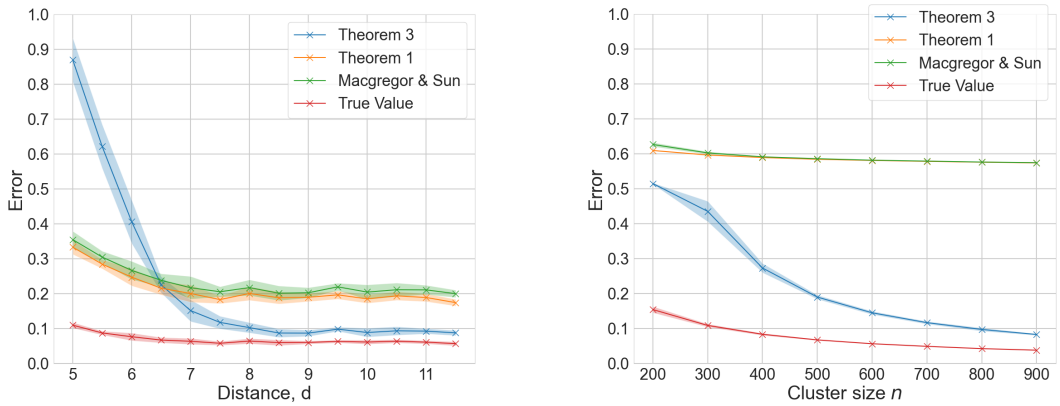
Corollary 4 essentially states that if the indicator vectors have Rayleigh quotient close to $\lambda_2 \ll \lambda_{k+1}$, then the bottom k eigenvectors of M must be very close to linear combinations of the indicator vectors. In contrast with the original structure theorem, this does not require λ_2 to be small.

Corollary 4 can be used to analyse Spectral Clustering on stochastic block models. While such an analysis can be obtained using standard perturbation arguments, stochastic block models are an egregious instance where the structure theorem of Corollary 2 fails. We illustrate the performance of Corollary 4 on SBMs in the appendix.

3.1 Experimental results

To illustrate the impact of Theorem 3, we consider synthetic graphs with a *hierarchical* structure and show the improved performance the theorem provides. We then consider some real-world networks to show that this hierarchical structure is naturally occurring and can be exploited to get a better bound on the performance of Spectral Clustering.

Geometric random graphs. We sample 100 points each from four two-dimensional Gaussians centred, respectively, $(0,0), (0,5), (d,0), (d,5)$ where d is a parameter we vary. A graph is constructed by assigning an edge between any two points if the Euclidean distance between them is less than a threshold, which we choose in this case to be 4. In Figure 2a we compare our bounds on the distance between indicator vectors of the clusters and the eigenvectors of the Laplacian with the results of Macgregor and Sun [14] and the actual distances. Results are averaged over 10 realisations.



(a) (b)

Figure 2: (a) Geometric random graph from Gaussian mixture model with varying distance between centres. (b) Stochastic block model with 4 blocks where two pairs have a higher affinity to each other. In both cases, the Error refers the bound given by Theorem 1 and 3, and by Macgregor and Sun [14] on $\frac{1}{k}\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2$, together to its actual value. Standard deviation is included as filled error bars.

Notice that, as the distance parameter d increases, drawing two pairs of clusters further apart from each other, Theorem 3 drastically outperforms Corollary 2 and Theorem 1.

Stochastic block models. We consider an SBM with 4 equal-sized clusters $S_1, \ldots, S_4$. Let P_{ij} be the probability that, for any $u \in S_i, v \in S_j$ there exists an edge between u and v. We set $P_{ii} = 0.5\,(i = 1, \ldots, 4)$ and $P_{12} = P_{21} = P_{34} = P_{43} = 0.4$. All other values are set equal to 0.1. We effectively divide the vertices into two pairs of clusters which are more strongly connected to one another. In Figure 2b, we compare our results with [14] and the true distances between indicator vectors of the optimal clusters and the eigenvectors of the Laplacian. Notice that Theorem 3 greatly outperforms Corollary 2 and Theorem 1. Further experiments (such as when the parameters of the stochastic block model are close to the detectability threshold) are available in the Appendix.

Real-world networks. In Table 1, we illustrate the improvement Theorem 3 provides over the results of Macgregor and Sun [14] on a variety of real-world networks. For MNIST [15], Fashion MNIST [16] and the Air Quality dataset [17], we constructed a graph from the data by computing a correlation matrix from its data points and assigning edges based on whether the correlation exceeded

a pre-defined threshold. Twitch [18], LastFM [19], Athletes [20] and CA-CondMat [21] are all network datasets available at SNAP [22]. N (resp. M) refers to the number of nodes (resp. edges). The number of clusters k has been chosen so that a relatively large gap between λ_k and λ_{k+1} exists. We include The improvements given by Theorem 1 over [14] are due to the fact that we sum over the Rayleigh quotients of the k indicator vectors, rather than simply upper-bound this value by $k\rho(k)$ as done in [14]. More interesting is, arguably, the improvement of Theorem 3, which is due to many networks having clustered eigenvalues and/or large λ_2. Notice how Theorem 3 is able to certify that, in these examples, the output of spectral clustering is well-correlated with the partition of minimum conductance, something not achievable with previous structure theorems.

Table 1: Comparison on the bounds on $\frac{1}{k}\sum_{i=1}^{k}\|f_i - \hat{g}_i\|^2$ given by Theorem 1 and 3, and by Macgregor and Sun [14] on real-world networks. We take a sample for MNIST and Fashion MNIST.

Dataset	**k**	**N**	**M**	**[14]**	**Theorem 1**	**Theorem 3**	$\tilde{\rho}(k)$	λ_{k+1}
MNIST*	5	2348	140018	0.4395	0.1915	0.1744	0.021	0.048
Fash. MNIST*	6	2872	1711206	0.5614	0.3761	0.3225	0.200	0.35
Air Quality	3	4942	2784780	0.6234	0.2883	0.2350	0.332	0.533
Twitch	2	168114	6797557	0.9567	0.4793	0.3937	0.128	0.133
LastFM	2	7622	27800	0.8373	0.4221	0.3123	0.013	0.016
Athletes	3	13866	86852	0.6951	0.4033	0.2427	0.053	0.037
CA-CondMat	3	21363	91342	0.7806	0.4818	0.3370	0.012	0.016

4 Digraph clustering

We now apply our techniques to Hermitian representations of digraphs, where the adjacency matrix is specified as in Equation 1. Cucuringu et al. [6] have shown experimentally that Spectral Clustering on these matrix representations is able to recover clusters characterised by large imbalances in the direction of inter-cluster edges, in the sense that most edges between two clusters S_i and S_j follow the same direction. This is a *higher-order* clustering problem, since clusters are defined according to their inter-cluster relations [23]. This is in contrast with traditional undirected graph clustering, in which a cluster is usually defined only according to its inner and outer density.

Cucuringu et al. [6] have proposed an analysis of Spectral Clustering on a directed analogue of the classical stochastic block model, while Laenen and Sun [7] have attempted to provide an analysis for more general graphs. In Section 4.1, however, we will argue that Laenen and Sun's results fail to explain the practical performances of Spectral Clustering on digraphs. Here we attempt to remedy this gap in the literature. We begin by defining a cost function for the task.

Definition 1. *Let $\mathcal{G} = (V, E, w)$ be a digraph and let $k \geq 2$. Let $S_1, \ldots, S_k$ be a k-way partition of V. We define the* cyclic expansion *of $S_1, \ldots, S_k$ as*

$$\Psi(S_1, \ldots, S_k) = \frac{1}{\mathrm{vol}(V)} \sum_{(i,j) \notin C_k} w(S_i, S_j),$$

where $C_k = \{(i,j) \mid j \equiv i+1 \mod k,\ 1 \leq i, j \leq k\}$, and the cyclic k-way expansion *of $\mathcal{G}$ as*

$$\Psi_k(\mathcal{G}) = \min_{\mathrm{partition}\{S_i\}_{i=1}^{k}} \Psi(S_1, \ldots, S_k).$$

We want to find clusters $S_1, \ldots, S_k$ so that most of the out-going edges from S_i are connected to vertices in $S_{i+1 \bmod k}$. When $k = 2$, our problem simply becomes finding *disassortative* clusters. Indeed, the Hermitian Laplacian for $k = 2$ is just the signless Laplacian, whose bottom eigenvectors are known to contain information about disassortative clusters [24, 25]. The next lemma clarifies the connection between this cost function and Hermitian Laplacians for digraphs.

Lemma 5. *Let $\mathcal{G}$ be a connected digraph. Then, $\lambda_1(\mathcal{L}) = 0$ if and only if $\Psi_k(\mathcal{G}) = 0$.*

As observed in [26], the bottom eigenvector of a Hermitian Laplacian already contains information about all k clusters. For this reason, given a k-way partition $\mathcal{S} = \{S_1, \ldots, S_k\}$, we define $\chi_{\mathcal{S}} \in \mathbb{C}^N$

as follows. For any $j = 1, \dots, k$,

$$\chi_{\mathcal{S}}(u) = \sqrt{\frac{d(u)}{\mathrm{vol}(V)}} \cdot \mathrm{e}^{\frac{2\pi \mathrm{i} j}{k}} \text{ for } u \in S_j,$$

where i is the imaginary unit. Fundamentally, $\chi_{\mathcal{S}}$ maps each cluster to a power of the k-th root of unity. The following simple but crucial lemma relates the Rayleigh quotient of $\chi_{\mathcal{S}}$ to $\Psi(S_1, \dots, S_k)$.

Lemma 6. *Let $k \geq 2$. It holds that $16k^{-2}\Psi(S_1, \dots, S_k) \leq \chi_{\mathcal{S}}^* \mathcal{L} \chi_{\mathcal{S}} \leq 4\Psi(S_1, \dots, S_k)$.*

We are now ready to state our structure theorem for digraphs.

Theorem 7 (Structure Theorem for Digraphs)**.** *Let $\mathcal{G}$ be a digraph with Hermitian Laplacian $\mathcal{L}$. Assume $\lambda_2 > \lambda_1$. Given any k-way partition $\mathcal{S} = \{S_1, \dots, S_k\}$, there exists $\alpha \in \mathbb{C}$ such that*

$$\|f_1 - \alpha \chi_{\mathcal{S}}\|^2 \leq \frac{\chi_{\mathcal{S}}^* \mathcal{L} \chi_{\mathcal{S}} - \lambda_1}{\lambda_2 - \lambda_1}. \tag{2}$$

Furthermore, if $\mathcal{S}$ achieves $\Psi_k(\mathcal{G})$, then

$$\|f_1 - \alpha \chi_{\mathcal{S}}\|^2 \leq \frac{4\Psi_k(\mathcal{G}) - \lambda_1}{\lambda_2 - \lambda_1}. \tag{3}$$

We observe that $\chi_{\mathcal{S}}^* \mathcal{L} \chi_{\mathcal{S}} = \dfrac{4}{\mathrm{vol}(V)} \displaystyle\sum_{i=1}^{k} \sum_{j=1}^{k} w(S_i, S_j) \sin^2\left(\frac{(i-(j+1))\pi}{k}\right)$. Therefore, inequality (2) is generally stronger than (3) because it assigns a smaller penalty to edges from S_i to S_j when $i - (j+1)$ is small.

4.1 Comparison with previous work

We now compare our structure theorem for digraphs to the one of Laenen and Sun [7]. They consider the following alternative to our cyclic expansion:

$$\theta_k(\mathcal{G}) \triangleq \max_{\{S_i\}_{i=1}^k \text{ partition}} \sum_{i=1}^{k-1} \frac{w(S_i, S_{i+1})}{\mathrm{vol}(S_i) + \mathrm{vol}(S_{i+1})}.$$

There are two differences compared with our definition of cyclic expansion. First, θ_k penalises edges from S_k to S_1, i.e. it tries to fit a path rather than a directed cycle between clusters. While reasonable, we argue this is not what Spectral Clustering on Hermitian Laplacians is actually doing. Secondly, and more importantly, the two cost functions differ in the normalisation based on the volume of the graph. We believe the normalisation chosen by [7] is poorly suited to analyse Spectral Clustering.

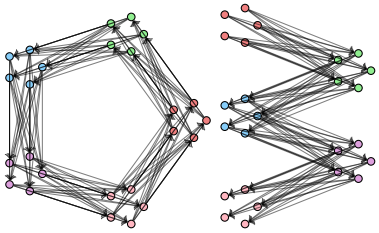
To motivate our assertions, let us delve deeper into their results. We first remark they choose a Hermitian Laplacian constructed with the $\lceil 2\pi k \rceil$-th root of unity. They also construct a different "indicator" vector $\tilde{\chi}_{\mathcal{S}}$ for a k-way partition $\mathcal{S}$. This choice is not overly important, so we refer to their paper for further details. Their main result is as follows.

Theorem 8 ([7])**.** *Let f_1 be the bottom eigenvector of the Hermitian Laplacian constructed with the $\lceil 2\pi k \rceil$-th root of unity. Let $\eta_k(\mathcal{G}) \triangleq \frac{\lambda_2}{1-(4/k)\theta_k(\mathcal{G})}$ and assume $\eta_k(\mathcal{G}) > 1$. Let $\mathcal{S}$ be a partition maximising $\theta_k(\mathcal{G})$. There exists $\beta \in \mathbb{C}$ such that $\|f_1 - \beta \tilde{\chi}_{\mathcal{S}}\|^2 \leq (\eta_k(\mathcal{G}) - 1)^{-1}$.*

We now compare Theorem 7 with Theorem 8 on the two very simple digraphs of Figure 3,

in which we have five clusters perfectly arranged on a directed cycle (Figure 3a) and on a directed path (Figure 3b). Clearly, in both cases, $\Psi_k = 0$. Applying our structure theorem, this implies that, in both cases, the first eigenvector is exactly a multiple of the indicator vector $\chi_{\mathcal{S}}$. Therefore, the clusters are embedded in k distinct and well-separated points, which are just the k powers of the k-th root of unity rotated by some $\alpha \in \mathbb{C}$. Our structure theorem correctly predicts that spectral clustering will recover the clusters perfectly.

If we apply Theorem 8 instead, we obtain that, for the digraph of Figure 3a, $\|f_1 - \beta \tilde{\chi}_{\mathcal{S}}\|^2 \leq (\eta_k(\mathcal{G}) - 1)^{-1} \approx 0.642$ which is not informative at all. If we consider the digraph of Figure 3b, we obtain a slightly better bound: $\|f_1 - \beta \tilde{\chi}_{\mathcal{S}}\|^2 \leq (\eta_k(\mathcal{G}) - 1)^{-1} \approx 0.294$. This is still far from the true value, which is equal to zero.



(a) Cycle. **(b)** Path.

Figure 3: Examples of directed cluster structures.

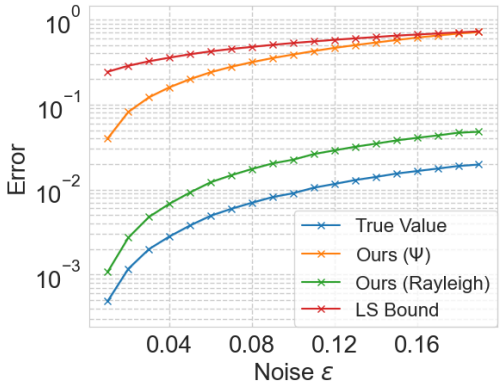


Figure 4: Comparison of the results given by Theorem 7 (green for (2) and orange for (3)) and by [7] (red) for a cyclic DSBM at varying level of noise. The actual values are reported in blue. Averaged over 10 realisations.

4.2 Experimental results

Directed stochastic block models. We now apply our results to the Directed Stochastic Block Model of [6]. Given parameters $k \geq 2, n \geq 1, P \in [0,1]^{k\times k}, F \in [0,1]^{k\times k}$, a directed stochastic block model $\mathcal{G} \sim \text{DSBM}(k, n, P, F)$ is a random graph of $N = kn$ vertices constructed as follows: each vertex belongs to one of k communities $S_1, \ldots, S_k$ of n vertices each. We place an edge independently at random between any two vertices $u \in S_i, v \in S_j$ with probability $P_{ij} = P_{ji}$. Furthermore, we orient the edge between u and v from u to v with probability F_{ij}, from v to u with probability $F_{ji} = 1 - F_{ij}$.

We consider a model $\mathcal{G} \sim \text{DSBM}(k, n, P, F)$ for $k = 4$ and $n = 100$ with P, F specified as follows.

$$F = \begin{pmatrix} .5 & 1 & .5 & .5 \\ 0 & .5 & 1 & .5 \\ .5 & 0 & .5 & 1 \\ .5 & .5 & 0 & .5 \end{pmatrix}, \quad P = \begin{pmatrix} \epsilon & 1 & \epsilon & \epsilon \\ 1 & \epsilon & 1 & \epsilon \\ \epsilon & 1 & \epsilon & 1 \\ \epsilon & \epsilon & 1 & \epsilon \end{pmatrix}$$

F represents a path structure, while this choice of P means the path structure is very pronounced. ϵ is a noise parameter: the smaller ϵ is, the closer the graph will be to having a perfect path cluster-structure.

In Figure 4, we compare the bounds of Theorem 7 with the results of [7], and with the true distance between the bottom eigenvector of the Hermitian Laplacian and the indicator vector of the clusters. Our results predict the true values exceptionally well and correctly imply Spectral Clustering will work almost perfectly for all noise levels considered. On the contrary, Laenen and Sun's result becomes non-informative even for small noise parameters. We provide a similar experiment for a DSBM with a cyclic cluster structure in the Appendix.

Table 2: Comparison on the bounds on $\|f_1 - \alpha\chi_S\|^2$ given by Theorem 7 (2) (Ours) and Laenen and Sun [7] (LS) on real-world networks.

Network	k	N	M	Ψ	Ours	LS
Yellowstone	4	15	37	0.027	0.086	0.662
COVID-19	4	67	66	0.000	0.000	N/D
St. Marks	5	49	226	0.154	0.324	N/D
St. Martin	4	45	224	0.118	0.352	N/D
Ythan	4	135	601	0.101	0.399	N/D

Real-world directed networks. We apply our results to real-world directed networks and summarise our findings in Table 2: for each network considered, we compare our bounds from Theorem 7 with Laenen and Sun's Theorem 8.

We first consider the directed graph analysed in [7] made from the Data Science for COVID-19 Dataset [27], where an edge from u to v exists if u has infected v. This graph has many disconnected components so we consider its largest weakly connected component. Spectral Clustering finds clusters fitting perfectly to a directed 4-cycle; therefore, $\Psi_4 = 0$ and our bound in Theorem 7 is $\|f_1 - \beta\chi_{\mathcal{S}}\|^2 = 0$. On the other hand, since $\eta_4 < 1$, Laenen and Sun's result is not applicable. Both this and the previous data set are characterised by unbalanced clusters, which results in a small θ_4 and makes Laenen and Sun's results uninformative.

The Yellowstone [28] data set a small digraph representing a food web for Yellowstone National Park [28], where vertices represent animal species and there is an edge from u to v if u is predated by v. As shown in Figure 5, this network can be partitioned into four clusters exhibiting an almost perfect directed path structure: only the outgoing edges for *Mule deer* are not consistent with this structure. Indeed, $\Psi_4 = 0.027$ and Spectral Clustering using the Hermitian Laplacian perfectly recovers these clusters. Theorem 7 suggests an error bound less than 0.086, which is close to the actual bound $\|f_1 - \beta\chi_{\mathcal{S}}\|^2 = 0.039$. On the other hand, Laenen and Sun's result can only obtain an upper bound of 0.662, which is not indicative of the actual performance of Spectral Clustering.

St. Marks Seagrass, St. Martin Island, and Ythan Estuary data sets [29] are networks representing other food webs. While these data sets do not present a cluster structure as obvious as the COVID-19 or Yellowstone data sets, our results imply there exists a nontrivial correlation between the indicator vector of the clusters and the bottom eigenvector of the Hermitian Laplacian. Notice that in all these three cases, $\eta < 1$, making Laenen and Sun's bounds uninformative.

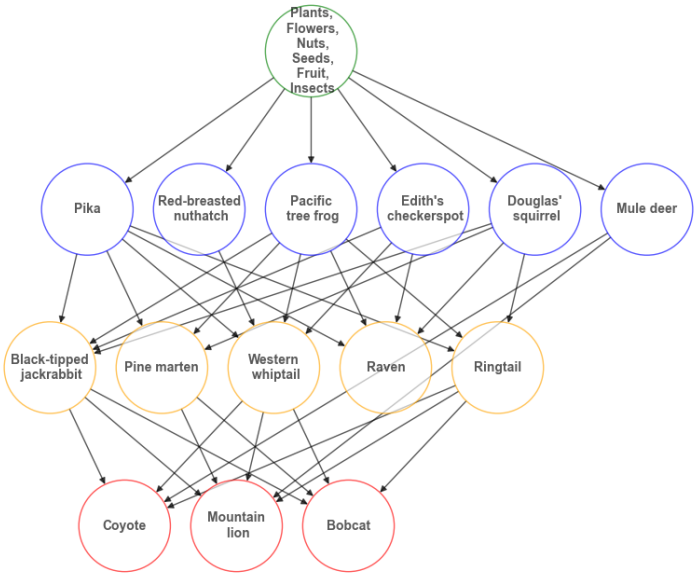


Figure 5: A directed graph illustrating a food web for Yellowstone National Park, United States.

5 Conclusion

We have presented a generalised framework for analysing the performance of Spectral Clustering. Our work generalises and improve upon the line of work based on the structure theorem of Peng et al. [2]. In particular, we have shown that spectral clustering works well as long as the bottom eigenvalues of the Laplacian matrix of an undirected graph can be divided in well-separated groups. Experimental results show that our bounds accurately predict the performance of spectral clustering on a wide range of both synthetic and real-world networks where previous analyses failed.

Furthermore, we have applied our techniques to analyse spectral clustering on Hermitian Laplacians for digraphs, as defined by Cucuringu et al. [6]. By exploiting a new cost function, which we call the cyclic k-way expansion, we are able to analyse the performances of spectral clustering on digraphs, vastly outperforming previous analyses.

References

[1] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001. 1

[2] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In *Conference on learning theory*, pages 1423–1455. PMLR, 2015. 1, 2, 3, 10

[3] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):1–30, 2014. 1, 2, 3

[4] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000. 1

[5] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. 1, 2, 17

[6] Mihai Cucuringu, Huan Li, He Sun, and Luca Zanetti. Hermitian matrices for clustering directed graphs: insights and applications. In *International Conference on Artificial Intelligence and Statistics*, pages 983–992. PMLR, 2020. 1, 2, 3, 7, 9, 10, 12

[7] Steinar Laenen and He Sun. Higher-order spectral clustering of directed graphs. *Advances in neural information processing systems*, 33:941–951, 2020. 2, 7, 8, 9, 10, 22

[8] Santo Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010. 2

[9] Ulrike von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007. 2

[10] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, 39(4):1878–1915, 2011. 2

[11] Luca Zanetti. *Algorithms for partitioning well-clustered graphs*. PhD thesis, University of Bristol, 2017. 2

[12] Pavel Kolev and Kurt Mehlhorn. A note on spectral clustering. In *24th Annual European Symposium on Algorithms*, volume 57 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 57, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016. ISBN 978-3-95977-015-6. 2

[13] Tomohiko Mizutani. Improved analysis of spectral algorithm for clustering. *Optim. Lett.*, 15(4): 1303–1325, 2021. ISSN 1862-4472,1862-4480. 2

[14] Peter Macgregor and He Sun. A tighter analysis of spectral clustering, and beyond. In *International Conference on Machine Learning*, pages 14717–14742. PMLR, 2022. 2, 4, 6, 7, 13, 15

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791. 6

[16] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 6

[17] Mujtaba Mateen. Air quality and pollution assessment, 2024. URL https://www.kaggle.com/ds/6197184. 6

[18] Benedek Rozemberczki and Rik Sarkar. Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings, 2021. 7

[19] Benedek Rozemberczki and Rik Sarkar. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, page 1325–1334. ACM, 2020. 7

[20] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019. 7

[21] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007. 7

[22] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014. 7

[23] James Martin, Tim Rogers, and Luca Zanetti. An iterative spectral algorithm for digraph clustering. *Journal of Complex Networks*, 12(2):cnae016, 03 2024. ISSN 2051-1329. doi: 10.1093/comnet/cnae016. URL https://doi.org/10.1093/comnet/cnae016. 7

[24] Luca Trevisan. Max cut and the smallest eigenvalue. *SIAM J. Comput.*, 41(6):1769–1786, 2012. 7

[25] Shiping Liu. Multi-way dual Cheeger constants and spectral bounds of graphs. *Adv. Math.*, 268: 306–338, 2015. 7

[26] Huan Li, He Sun, and Luca Zanetti. Hermitian Laplacian and a Cheeger inequality for the Max-2-Lin problem. In *27th Annual European Symposium on Algorithms*, volume 144 of *LIPIcs. Leibniz Int. Proc. Inform.* Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019. 7

[27] Korea Centers for Disease Control & Prevention. COVID-19 data set. https://www.kaggle.com/kimjihoo/coronavirusdataset, 2020. Accessed: 2024-09-28. 10

[28] Yellowstone Food Web. *Yellowstone Food Web*, n.d. URL https://yellowstoneinfo.weebly.com/food-web.html. Accessed: 2024-09-28. 10

[29] COSIN project. Food web data sets, 2004. URL http://cosinproject.eu/extra/data/foodwebs/WEB.html. Accessed: 2024-10-05. 10

[30] Fan Chung and Mary Radcliffe. On the spectra of general random graphs. *the electronic journal of combinatorics*, pages P215–P215, 2011. 17

[31] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012. 17

[32] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016. 21

A Theoretical guarantees for spectral clustering

In this appendix we discuss how we can turn a structure theorem, like Theorem 1 and Corollary 2, into bounds on the symmetric difference between the clusters output by spectral clustering and the optimal ones. This is done by applying the following Lemma, whose proof is essentially adapted from Cucuringu et al. [6].

Lemma 9. *Assume the partition $A_1, \dots, A_k$ output by Algorithm 1 is computed by a $(1+\alpha)$-approximation algorithm for k-means, with $\tilde{k} = k$. Let $\{S_1, \dots, S_k\}$ be any k-way partition of V achieving $\rho(k)$. Let $G \in \mathbb{C}^{|V| \times \tilde{k}}$ be such that $(D^{-1/2}G)_{u,:} = (D^{-1/2}G)_{v,:}$ if $u, v \in S_i$ for some i. Let $F, \tilde{F}$ be defined as in Algorithm 1. For any $i = 1, \dots, k$, let $\mu_i = \mathrm{vol}(S_i)^{-1} \cdot \sum_{u \in S_i} d(u) \tilde{F}_{u,:}$. Define $\mathcal{D} := \min_{i,j} \|\mu_i - \mu_j\|$ and $U := \sum_{u \in V} \|F_{u,:} - G_{u,:}\|^2$. Assume $U \leq (1/5)\mathcal{D}^{-1}(2+\alpha)^{-1} \cdot \min_{i=1,\dots,k} \mathrm{vol}(S_i)$. Then, the volume of the symmetric difference between $\{S_1, \dots, S_k\}$ and $\{A_1, \dots, A_k\}$ (up to a permutation of the indices) is at most $\mathcal{O}(\frac{(1+\alpha)U}{\mathcal{D}^2})$.*

Proof. Let $c_i = \sum_{u \in A_i} d(u) \tilde{F}(u,.)$ be the centroid of A_i for $i = 1, \dots, k$. Let $c \colon V \to \mathbb{R}^k$ be a map from a vertex to its corresponding centroid, i.e., $c(u) = c_i$ if $u \in A_i$.

Since $A_1, \dots, A_k$ is a $(1+\alpha)$ approximation of the optimal k-means partition, we have that

$$\begin{aligned}
\sum_{j=1}^{k} \sum_{u \in A_j} d(u) \|\tilde{F}(u,.) - c_j\|^2 &= \sum_{j=1}^{k} \sum_{u \in S_j} d(u) \|\tilde{F}(u,.) - c(u)\|^2 \\
&\leq (1+\alpha) \sum_{j=1}^{k} \sum_{u \in S_j} d(u) \|\tilde{F}(u,.) - \mu_j\|^2 \\
&\leq (1+\alpha) \sum_{j=1}^{k} \sum_{u \in S_j} \|F(u,.) - G(u,.)\|^2 \leq (1+\alpha)U. \qquad (4)
\end{aligned}$$

By applying the triangle inequality and the simple inequality $(x-y)^2 \geq \frac{x^2}{2} - y^2$, we obtain the following:

$$\begin{aligned}\sum_{j=1}^{k}\sum_{u\in S_j} d(u)\|\tilde{F}(u,.)-c(u)\|^2 &\geq \sum_{j=1}^{k}\sum_{u\in S_j} d(u)(\|\mu_j - c(u)\| - \|\tilde{F}(u,.)-\mu_j\|)^2 \\ &\geq \sum_{j=1}^{k}\sum_{u\in S_j} \frac{1}{2}d(u)\|\mu_j - c(u)\|^2 - \sum_{j=1}^{k}\sum_{u\in S_j} d(u)\|\tilde{F}(u,.)-\mu_j\|^2 \\ &\geq \sum_{j=1}^{k}\sum_{u\in S_j} \frac{1}{2}d(u)\|\mu_j - c(u)\|^2 - U,\end{aligned}$$

where the last inequality follows from Equation 4.

By the assumption $\|\mu_i - \mu_j\| \geq \mathcal{D}$ for all $i \neq j$, it holds that $\|c_i - c_j\| \geq \mathcal{D}/2$. Otherwise, there would be μ_ℓ such that $\|\mu_\ell - c_i\| \geq \mathcal{D}/2$ for any i, but this would violate our assumption on U. Indeed, assume by contradiction $\|\mu_l - c_i\| \geq \mathcal{D}/2$ for any i. Then,

$$\begin{aligned}(1+\alpha)U &\geq \sum_{j=1}^{k}\sum_{u\in S_j} d(u)\|\tilde{F}(u,.)-c(u)\|^2 \\ &\geq \frac{1}{2}\sum_{u\in S_\ell} d(u)\|\mu_\ell - c(u)\|^2 - U \\ &\geq \frac{1}{4}\cdot \mathcal{D}\cdot \mathrm{vol}(S_\ell) - U.\end{aligned}$$

Therefore, $U \geq \frac{\mathcal{D}\cdot \mathrm{vol}(S_\ell)}{4(2+\alpha)}$, contradicting our assumption on U.

We say that a vertex $u \in S_j$ is *misclassified* if $\|c(u) - \mu_j\| \geq \mathcal{D}/2$. Let $\sigma\colon [k] \to [k]$ be the permutation of the indices minimising the volume of the symmetric difference $\sum_{i=1}^{k} \mathrm{vol}(S_i \setminus A_{\sigma(i)} \cup A_{\sigma(i)} \setminus S_i)$. Notice that we can obtain an upper bound to this volume by bounding the volume of misclassified vertices. We can upper bound the latter by noticing that, whenever we misclassify a vertex, we pay a price of $\Omega(D^2)$ in the k-means cost. Noting this, we obtain the following bound:

$$\begin{aligned}\sum_{j=1}^{k}\sum_{u\in S_j} d(u)\|\tilde{F}(u,.)-c(u)\|^2 &\geq \sum_{j=1}^{k}\sum_{u\in S_j} \frac{1}{2}d(u)\|\mu_j - c(u)\|^2 - U \\ &\geq \frac{\mathcal{D}^2}{8}\mathrm{vol}(\text{misclassified vertices}) - U,\end{aligned}$$

Together with Equation 4, this implies that

$$\min_{\sigma}\sum_{i=1}^{k}\mathrm{vol}(S_i \setminus A_{\sigma(i)} \cup A_{\sigma(i)} \setminus S_i) \leq \mathrm{vol}(\text{misclassified vertices}) \leq \frac{8(2+\alpha)U}{\mathcal{D}^2}.$$

□

Notice that, by choosing $G_{:,i} = \hat{g}_i$ for $i = 1,\dots,\tilde{k}$, we have $U = \sum_{i=1}^{\tilde{k}} \|f_i - \hat{g}_i\|^2$. Thus $\mathcal{D}$ is instead dependent on the choice of the indicator vectors. In the case of undirected graph clustering with $M = \mathcal{L}$ and the traditional choice of indicator vectors, as long as $U \ll 1$ (e.g., by Corollary 2, whenever $k\rho(k) \ll \lambda_{k+1}$), we have that $\mathcal{D} = \Omega\left(\min_{i=1,\dots,k}\mathrm{vol}(S_i)^{-1}\right)$[14]. Assuming a constant-factor approximation algorithm for k-means is used, Lemma 9 guarantees that the symmetric difference between the partition output by spectral clustering and the optimal partition is small.

Theorem 1 together with Lemma 9 give us a computable way to certify the quality of the partition $A_1,\dots,A_k$ output by spectral clustering compared to the optimal partition $S_1,\dots,S_k$. More precisely, let $g_1,\dots,g_k$ be a set of orthonormal indicator vectors for $A_1,\dots,A_k$ with Rayleigh

quotients $\gamma_i := g_i^* M g_i$. Compute the value $\frac{1}{k}\frac{\sum_{i=1}^k \gamma_i - k\lambda_1}{\lambda_{k+1}-\lambda_1}$. Theorem 1 and Lemma 9 ensure that, if this value is small, then $A_1, \ldots, A_k$ are close to $S_1, \ldots, S_k$ (up to permutation of the indices). Lemma 9 can also be used in conjunction with Theorem 3 by repeatedly applying Theorem 3 until a bound on $\sum_{i=1}^k \|f_i - \hat{g}_i\|^2$ is obtained.

B Omitted proofs and results

B.1 Section 3

In this section we prove the results from Section 3. For the following proofs, we must first define the matrices $Q, R \in \mathbb{C}^{N\times N}$ which map between F and G. Precisely,

$$G = FQ$$
$$F = GR.$$

In the following lemma we prove some basic properties of Q and R.

Lemma 10. *We have the following basic properties of Q:*

1. *Q is orthogonal.*
2. *$R = Q^*$.*

Proof. As F is orthogonal, we have that $Q = F^*G$. Noting that G is also orthogonal, we have that

$$Q^*Q = (F^*G)^*(F^*G) = G^*FF^*G = I$$

and similarly, $QQ^* = I$ so Q is orthogonal.
For 2., Clearly,

$$R = G^*F = Q^*$$

as G is orthogonal. □

B.1.1 Proof of Theorem 1 and Corollary 2

Below we use Lemma 10 to prove our Theorem 1.

Proof of Theorem 1. As $G = FQ$, this implies that $g_i = \sum_{j=1}^N Q_{ij} f_j$. We choose $\hat{f}_i = \sum_{j=1}^k Q_{ij} f_j$. Now notice that,

$$\gamma_i = \bar{g}_i^* M \bar{g}_i = \sum_{j=1}^N |Q_{ij}|^2 \lambda_j.$$

As Q is orthogonal, $\sum_{j=1}^N |Q_{ij}|^2 = 1$. This coupled with the assumption that $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$ provides the following bound:

$$\gamma_i = \sum_{j=1}^N |Q_{ij}|^2 \lambda_j \ge \lambda_1 \sum_{j=1}^k |Q_{ij}|^2 + \lambda_{k+1}(1 - \sum_{j=1}^k |Q_{ij}|^2)$$
$$\Rightarrow \gamma_i \ge \sum_{j=1}^k |Q_{ij}|^2(\lambda_1 - \lambda_{k+1}) + \lambda_{k+1}$$

Rearranging for $\sum_{j=1}^k |Q_{ij}|^2$, we get the following:

$$\sum_{j=1}^k |Q_{ij}|^2 \ge \frac{\lambda_{k+1} - \gamma_i}{\lambda_{k+1} - \lambda_1}$$

It follows that

$$1 - \sum_{j=1}^{k} |Q_{ij}|^2 \leq \frac{\gamma_i - \lambda_1}{\lambda_{k+1} - \lambda_1}$$

. As $F = GR$, $f_i = \sum_{j=1}^{N} R_{ij} g_j$. We choose $\hat{g}_i = \sum_{j=1}^{k} R_{ij} g_j$. Now notice that,

$$\|f_i - \hat{g}_i\|^2 = 1 - \sum_{j=1}^{k} |R_{ij}|^2.$$

Summing over $i = 1, \ldots, k$ gives the following.

$$\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2 = k - \sum_{i=1}^{k} \sum_{j=1}^{k} |R_{ij}|^2$$

Recall from Lemma 10 that $Q^* = R$ so $\sum_{i=1}^{k} \sum_{j=1}^{k} |R_{ij}|^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} |Q_{ij}|^2$. So, it follows that:

$$\sum_{i=1}^{k} \|f_i - \hat{g}_i\|^2 \leq \sum_{i=1}^{k} \frac{\gamma_i - \lambda_1}{\lambda_{k+1} - \lambda_1}.$$

□

By a particular choice of $\{g_i\}_{i=1}^{k}$ and M in Theorem 1, we obtain the structure theorem from [14].

Proof of Corollary 2. With the choice of $g_i = \frac{D^{1/2}\chi_i}{\|D^{1/2}\chi_i\|}$ for $i = 1, \ldots, k$, we have the following:

$$\begin{aligned}
\gamma_i &= g_i^* \mathcal{L} g_i \\
&= \frac{\chi_i^*(D - A)\chi_i}{\chi_i^* D \chi_i} \\
&= \frac{E(S_i, V - S_i)}{\mathrm{vol}(S_i)} \\
&\leq \max_{i=1,\ldots,k} \frac{E(S_i, V - S_i)}{\mathrm{vol}(S_i)} = \rho(k)
\end{aligned}$$

Applying Theorem 1 with $M = \mathcal{L}$ and noting that $\lambda_1 = 0$ completes the proof. □

B.1.2 Proof of Theorem 3 and Corollary 4

Proof of Theorem 3. Since $f_1, \ldots, f_N$ form an orthonormal basis, we can write that $g_i = \sum_{j=1}^{N} Q_{ij} f_j$ and we choose $\hat{f}_i = \sum_{j=1}^{q} Q_{ij} f_j$ for $i = 1, \ldots, q$ (summing only up to q by assumption) and $\hat{f}_i = \sum_{j=1}^{k} Q_{ij} f_j$ for $i = q+1, \ldots, k$. Now let i be some index between $q+1$ and k,

$$\begin{aligned}
\gamma_i = g_i^* M g_i = \sum_{j=1}^{N} \lambda_j |Q_{ij}|^2 &\geq \lambda_1 \sum_{j=1}^{q} |Q_{ij}|^2 + \lambda_{q+1} \sum_{j=q+1}^{k} |Q_{ij}|^2 + \lambda_{k+1} \sum_{j=k+1}^{n} |Q_{ij}|^2 \\
&= \lambda_1 \sum_{j=1}^{q} |Q_{ij}|^2 + \lambda_{q+1} \sum_{j=q+1}^{k} |Q_{ij}|^2 + \lambda_{k+1} \left(1 - \sum_{j=q+1}^{k} |Q_{ij}|^2 - \sum_{j=1}^{q} |Q_{ij}|^2 \right)
\end{aligned}$$

Rearranging for $\sum_{j=q+1}^{k} |Q_{ij}|^2$, we get:

$$\sum_{j=q+1}^{k} |Q_{ij}|^2 \geq \frac{\lambda_{k+1}\left(1 - \sum_{j=1}^{q} |Q_{ij}|^2\right) - \gamma_i + \lambda_1 \sum_{j=1}^{q} |Q_{ij}|^2}{\lambda_{k+1} - \lambda_{q+1}}$$
$$\geq \frac{\lambda_{k+1}\left(1 - \sum_{j=1}^{q} |Q_{ij}|^2\right) - \gamma_i}{\lambda_{k+1} - \lambda_{q+1}}$$

Next, we sum over $i = q+1, \ldots, k$ so we have that:

$$\sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2 \geq \frac{\lambda_{k+1}\left((k-q) - \sum_{i=q+1}^{k} \sum_{j=1}^{q} |Q_{ij}|^2\right) - \sum_{i=q+1}^{k} \gamma_i}{\lambda_{k+1} - \lambda_{q+1}}.$$

Since $\hat{f}_i \in \text{span}\{f_1, \ldots, f_q\}$ for $i \in \{1, \ldots, q\}$, the sum in the brackets can be bounded as follows:

$$\sum_{i=q+1}^{k} \sum_{j=1}^{q} |Q_{ij}|^2 \leq \sum_{i=q+1}^{N} \sum_{j=1}^{q} |Q_{ij}|^2 = q - \sum_{i=1}^{q} \sum_{j=1}^{q} |Q_{ij}|^2 = \sum_{i=1}^{q} \|\hat{f}_i - g_i\|^2$$

So,

$$\sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2 \geq \frac{\lambda_{k+1}\left((k-q) - \sum_{i=1}^{q} \|f_i - \hat{g}_i\|^2\right) - \sum_{i=q+1}^{k} \gamma_i}{\lambda_{k+1} - \lambda_{q+1}}$$

Now, noting that Q is orthogonal and therefore its rows are unit length, it follows that:

$$\sum_{i=q+1}^{k} \|\hat{f}_i - g_i\|^2 = (k-q) - \sum_{i=q+1}^{k} \sum_{j=1}^{k} |Q_{ij}|^2 \leq (k-q) - \sum_{i=q+1}^{k} \sum_{j=q+1}^{k} |Q_{ij}|^2$$

Therefore,

$$\sum_{i=q+1}^{k} \|\hat{f}_i - g_i\|^2 \leq \frac{\sum_{i=q+1}^{k} \gamma_i - (k-q)\lambda_{q+1} - \lambda_{k+1} \sum_{i=1}^{q} \|\hat{f}_i - g_i\|^2}{\lambda_{k+1} - \lambda_{q+1}}$$

Now, we show the same bound for $\sum_{i=q+1}^{k} \|f_i - \hat{g}_i\|^2$. Choosing $\hat{g}_i = \sum_{j=1}^{q} R_{ij} f_j$ for $i = 1, \ldots, q$, Lemma 10 implies that

$$\sum_{i=1}^{q} \|\hat{f}_i - g_i\|^2 = \sum_{i=1}^{q} \|f_i - \hat{g}_i\|^2$$

as $Q = R^*$. Similarly, choosing $\hat{g}_i = \sum_{j=1}^{k} R_{ij} f_j$ for $i = q+1, \ldots, k$ implies that

$$\sum_{i=q+1}^{k} \|\hat{f}_i - g_i\|^2 = \sum_{i=q+1}^{k} \|f_i - \hat{g}_i\|^2$$

which completes the proof. □

Proof of Corollary 4. Choosing $g_1 = f_1$ and remaining indicator vectors $g_2, \ldots, g_k$ all orthonormal means the conditions for Theorem 3 are satisfied with $q = 1$. $\hat{g}_1 = g_1 = f_1$ means that:

1. $\|\hat{g}_1 - f_1\|^2 = 0$

2. $\sum_{i=2}^{k} \|f_i - \hat{g}_i\|^2 \leq \frac{\sum_{i=2}^{k}(\gamma_i - \lambda_2)}{\lambda_{k+1} - \lambda_2}$

which completes the proof. □

B.1.3 Spectral Clustering performance on SBMs

Using Corollary 4 we have a result regarding the performance of Spectral Clustering on SBMs.

We generate $\mathcal{G} \sim SBM(n, k, p, q)$ as follows: we divide kn vertices into k communities $S_1, \ldots, S_k$ of equal size n. For any $u \in S_i, v \in S_j$, we place an edge between u and v with probability p if $i = j$, with probability q otherwise. We recover the following well-known result.

Corollary 11. *Let $\mathcal{G} \sim SBM(n, k, p, q)$ for some $k \geq 2$. Choose $M = L$ and let $g_1, \ldots, g_k \in \mathbb{C}^{nk}$ be the first k eigenvectors of $\mathbb{E}L$. Assume $p - q \geq 40\sqrt{pk\log(kn)/n}$. Then, with high probability,*

$$\sum_{i=2}^{k} \|f_i - \hat{g}_i\|^2 = \mathcal{O}\left(\frac{k}{p-q}\sqrt{\frac{pk\log(kn)}{n}}\right).$$

Corollary 11 together with Lemma 9 implies Spectral Clustering misclassifies vertices with a volume of at most $\mathcal{O}\left(\frac{k^2}{p-q}\sqrt{p(kn)\log(kn)}\right)$. We note that whilst this is not an improvement on what can be achieved with results from perturbation analysis such as the Davis-Kahan theorem [5], the original structure theorem achieves merely a constant bound for SBMs.

In order to prove Corollary 11, we require some prerequisite results.

Theorem 12 ([30])**.** *Let $X_1, X_2, \ldots, X_m$ be independent random $d \times d$ Hermitian matrices. Moreover, assume that $\|X_j - \mathbb{E}(X_j)\| \leq M$ for all j, and let $\sigma^2 = \|\sum_{j=1}^{m} \mathbb{E}((X_j - \mathbb{E}(X_j))^2)\|$. Let $X = \sum_{j=1}^{m} X_j$. Then, for any $a > 0$, it holds that:*

$$\mathbb{P}(\|X - \mathbb{E}(X)\| > a) < 2d\exp\left(\frac{-a^2}{2\sigma^2 + 2Ma/3}\right)$$

Theorem 13 (Courant-Fischer [31])**.** *Let A be a Hermitian matrix in $\mathbb{C}^{N\times N}$. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ of A, arranged in non-increasing order ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$), can be characterized by the Courant-Fischer min-max principle as follows:*

For $k = 1, 2, \ldots, N$:

$$\lambda_k = \max_{\substack{S \subseteq \mathbb{C}^N \\ \dim(S)=k}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle x, Ax\rangle}{\langle x, x\rangle}$$

and

$$\lambda_k = \min_{\substack{T \subseteq \mathbb{C}^N \\ \dim(T)=N-k+1}} \max_{\substack{x \in T \\ x \neq 0}} \frac{\langle x, Ax\rangle}{\langle x, x\rangle}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{C}^N$.

Corollary 14 (Courant-Fischer Corollary)**.** *Let A and B be Hermitian matrices in $\mathbb{C}^{N\times N}$, and let $\lambda_k(A)$ and $\lambda_k(B)$ denote their k-th eigenvalues, respectively. Then*

$$|\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|.$$

In the following lemma, we apply Theorem 12 to obtain a bound on $\|L - \mathbb{E}(L)\|$ for SBMs.

Lemma 15. *Suppose that $G \sim SBM(n, k, p, q)$ for some $k \geq 2$. Let L be the Laplacian of G. Then, with high probability,*

$$\|L - \mathbb{E}(L)\| \leq 18\sqrt{pkn\log(kn)}.$$

Proof. Firstly, let

$$M_{uv} = (e_u e_u^T + e_v e_v^T) - (e_u e_v^T + e_v e_u^T),$$

which has precisely four non-zero entries: $M_{uv}(u,u) = M_{uv}(v,v) = 1$ and $M_{uv}(u,v) = M_{uv}(v,u) = -1$. Now we define the random matrix

$$X_{uv} = \begin{cases} M_{uv} & \text{if } u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $L = \sum_{(u,v)\in E} X_{uv}$. Moreover, we have that $\|X_{uv} - \mathbb{E}(X_{uv})\| \leq 2$.

By the identity

$$\mathbb{E}((X_{uv} - \mathbb{E}(X_{uv}))^2) = \mathbb{E}((X_{uv})^2) - (\mathbb{E}(X_{uv}))^2,$$

and the fact that $M_{uv}^2 = 2M_{uv}$, we obtain that

$$\mathbb{E}((X_{uv})^2) - (\mathbb{E}(X_{uv}))^2 = \begin{cases} 2p(1-p)M_{uv} & \text{if } u \text{ and } v \text{ are in the same block} \\ 2q(1-q)M_{uv} & \text{otherwise.} \end{cases}$$

Summing over all $u, v \in V$ and taking the norm, we obtain the following:

$$\sigma^2 = \| \sum_{\{u,v\}\in\binom{V}{2}} \mathbb{E}((X_{uv})^2) - (\mathbb{E}(X_{uv}))^2 \| \leq 4(p + (k-1)q)n.$$

Applying Theorem 12 with $a = 18\sqrt{pkn\log(kn)}$ yields the statement.

□

The following result states the eigenvalues of the expected Laplacian in stochastic block models. Its proof is by elementary calculations.

Lemma 16. *Suppose that $G \sim SBM(n,k,p,q)$ for some $k \geq 2$. Let $L = D - A$ be the Laplacian of G and let $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$ be the normalized Laplacian of G. Then, the first eigenvalue of $\mathbb{E}(L)$ and $\mathbb{E}(\mathcal{L})$ are both 0. The next $k-1$ eigenvalues of $\mathbb{E}(L)$ and $\mathbb{E}(\mathcal{L})$ are knq and $\frac{knq}{(n-1)p+(k-1)nq}$ respectively and the $(k+1)$st eigenvalues are $np + (k-1)nq$ and $\frac{np+(k-1)nq}{(n-1)p+(k-1)nq}$.*

We are now ready to prove Corollary 11.

Proof of Corollary 11. Firstly, let $\lambda_1(\mathbb{E}(L)) \leq \cdots \leq \lambda_N(\mathbb{E}(L))$ denote the eigenvalues of $\mathbb{E}(L)$ with corresponding orthonormal eigenvectors $g_1, \ldots, g_k$. Let $\gamma_i = \frac{g_i^* L g_i}{g_i^* g_i}$. Notice that $g_1 = f_1 = \frac{1}{\sqrt{N}}1$. By Corollary 4, we have that

$$\sum_{i=2}^{k} \|\hat{f}_i - g_i\|^2 \leq \frac{\sum_{i=2}^{k}(\gamma_i - \lambda_2)}{\lambda_{k+1} - \lambda_2}.$$

We will bound the numerator of the previous expression as follows:

$$\begin{aligned} |\gamma_i - \lambda_2| &= |\gamma_i - \lambda_i(\mathbb{E}(L)) + \lambda_i(\mathbb{E}(L)) - \lambda_2| \\ &\leq |\gamma_i - \lambda_i(\mathbb{E}(L))| + |\lambda_i(\mathbb{E}(L)) - \lambda_2|. \end{aligned}$$

The first term can be bounded above as follows:

$$|\gamma_i - \lambda_i(\mathbb{E}(L))| = \|g_i^*(L - \mathbb{E}(L))g_i\| \leq \|L - \mathbb{E}(L)\|.$$

By Lemma 15, with high probability, $\|L - \mathbb{E}(L)\| \leq 18\sqrt{(pkn)\log(kn)}$. By Corollary 14, for all i, it holds that:

$$|\lambda_i - \lambda_i(\mathbb{E}(L)| \leq \|L - \mathbb{E}(L)\| \leq 18\sqrt{(pkn)\log(kn)}. \tag{5}$$

Thus, by Lemma 16, for $i = 1, \ldots, k$,

$$\lambda_i \in \left[knq - 18\sqrt{(pkn)\log(kn)}, knq + 18\sqrt{(pkn)\log(kn)}\right]$$

which implies

$$|\lambda_i(\mathbb{E}(L)) - \lambda_2| = |knq - \lambda_2| \leq 18\sqrt{(pkn)\log(kn)}$$

and

$$\gamma_i - \lambda_2 \leq 36\sqrt{(pkn)\log(kn)}.$$

On the other hand, Lemma 16 and (5) imply that

$$\lambda_{k+1} \in \left[np + (k-1)nq - 18\sqrt{(pkn)\log(kn)}, np + (k-1)nq + 18\sqrt{(pkn)\log(kn)}\right]$$

and, therefore,

$$\begin{aligned}\lambda_{k+1} - \lambda_2 &\geq (np + (k-1)nq - 18\sqrt{(pkn)\log(kn)}) - (knq + 18\sqrt{(pkn)\log(kn)}) \\ &= n(p-q) - 36\sqrt{(pkn)\log(kn)}.\end{aligned}$$

Finally, by our assumption that $p - q \geq 40\sqrt{pk\log(kn)/n}$, it holds that

$$\begin{aligned}\sum_{i=2}^{k} \|\hat{f}_i - g_i\|^2 &\leq \frac{\sum_{i=2}^{k}(\gamma_i - \lambda_2)}{\lambda_{k+1} - \lambda_2} \\ &\leq \frac{36(k-1)\sqrt{(pkn)\log(kn)}}{n(p-q) - 36\sqrt{(pkn)\log(kn)}} \\ &= \mathcal{O}\left(\frac{k}{p-q}\sqrt{\frac{pk\log(kn)}{n}}\right).\end{aligned}$$

□

B.2 Section 4

B.2.1 Proof of Lemma 6

Proof. Consider the rayleigh quotient of $\mathcal{L}$ with some vector $x \in \mathbb{C}^N$:

$$\begin{aligned}\frac{x^*\mathcal{L}x}{x^*x} &= \frac{x^*(I - D^{-1/2}AD^{-1/2})x}{x^*x} \\ &= \frac{y^*(D-A)y}{y^*Dy} \text{ (where } y = D^{-1/2}x) \\ &= \frac{\sum_{u\to v}|y_u|^2 - \bar{y}_u y_v \omega_k - y_u \bar{y}_v \bar{\omega_k} + |y_v|^2}{\sum_{u\to v}|y_u|^2 + |y_v|^2} \\ &= \frac{\sum_{u\to v}|y_u - \omega_k y_v|^2}{\sum_{u\to v}|y_u|^2 + |y_v|^2}\end{aligned}$$

where $\omega_k = \exp(\frac{2\pi i}{k})$. Selecting $\chi_{\mathcal{S}}$

$$\chi_{\mathcal{S}}(u) = \begin{cases}\sqrt{\frac{d(u)}{\mathrm{vol}(V)}} \cdot \mathrm{e}^{\frac{2\pi \mathrm{i} j}{k}} & \text{for } u \in S_j, \\ 0 & \text{otherwise.}\end{cases}$$

Then the expression resolves to

$$\begin{aligned}\frac{\chi_{\mathcal{S}}^*\mathcal{L}\chi_{\mathcal{S}}}{\chi_{\mathcal{S}}^*\chi_{\mathcal{S}}} &= \frac{1}{\mathrm{vol}(V)}\sum_{i=1}^{k}\sum_{j=1}^{k} 2w(S_i, S_j)(1 - \cos\left(\frac{(i-(j+1))2\pi}{k}\right)) \\ &= \frac{4}{\mathrm{vol}(V)}\sum_{i=1}^{k}\sum_{j=1}^{k} w(S_i, S_j)\sin^2\left(\frac{(i-(j+1))\pi}{k}\right) \\ &= \frac{4}{\mathrm{vol}(V)}\sum_{(i,j)\notin C} w(S_i, S_j)\sin^2\left(\frac{(i-(j+1))\pi}{k}\right) \\ &\leq \frac{4}{\mathrm{vol}(V)}\sum_{(i,j)\notin C} w(S_i, S_j) \\ &= 4\Psi(S_1, \ldots, S_k)\end{aligned}$$

The second equality follows from the identity $1 - \cos(x) = 2\sin^2(x/2)$, the third equality follows as $\sin^2\left(\frac{(i-(j+1))\pi}{k}\right) = 0$ for $(i,j) \in C$, and the inequality holds because $\sin^2(x) \leq 1$.
Now to prove the lower bound, consider the following function:

$$y(x) = \begin{cases} \frac{2}{\pi}(-x-\pi) & \text{for } x \in [-\pi, -\pi/2) \\ \frac{2}{\pi}x & \text{for } x \in [-\pi/2, \pi/2] \\ \frac{2}{\pi}(\pi - x) & \text{for } x \in (\pi/2, \pi] \end{cases}$$

It can be easily verified that $\sin^2(x) \geq y(x)^2$. Considering only $(i,j) \notin C$, the smallest values of $\sin^2\left(\frac{(i-(j+1))\pi}{k}\right)$ are attained when $i - (j+1) = \pm 1, \pm(k-1)$. The resulting value of $y\left(\frac{(i-(j+1))\pi}{k}\right)^2$ in all these instances is $\frac{4}{k^2}$. The bound follows. □

B.2.2 Proof of Lemma 5

Proof. Let us first assume $\Psi_k(\mathcal{G}) = 0$. Let S be a k-way partitioning such that $\Psi(S_1, \ldots, S_k) = 0$. Then, by 6, $\chi_{\mathcal{S}}^* \mathcal{L} \chi_{\mathcal{S}} = 0$. This implies the if direction.

For the reverse direction, let y be an eigenvector of $\mathcal{L}$ with eigenvalue 0, i.e., $y^* \mathcal{L} y = 0$. Then, as shown in the proof of 6, $y^* \mathcal{L} y = \sum_{u \to v} |y_u - \omega_k y_v|^2 = 0$. By the weak connectivity of $\mathcal{G}$, this implies there exists $\alpha \in \mathbb{C} \setminus \{0\}$ such that, for any u, there exists $i \in \{0, \ldots, k-1\}$ such that $y_u = \alpha \omega_k^i$. Let us fix such α and let $S_1, \ldots, S_k$ be a partition of V defined as follows: $S_i = \{u \in V \colon y_u = \alpha \omega_k^{i-1}\}$ for $i = 1, \ldots, k$. We have that $\Psi(S_1, \ldots, S_k) = 0$, which implies $\Psi_k(\mathcal{G}) = 0$. □

B.2.3 Proof of Theorem 7

Proof. The first statement follows from Theorem 1 using $k = 1$ and $g_1 = \chi_{\mathcal{S}}$. The second statement then follows from directly applying Lemma 6 to the first statement. □

C Omitted experiments

C.1 SBMs: Additional structure

In addition to our SBM experiment in Section 3, we have a similar experiment for an SBM with 8 clusters where one pair of clusters have a higher affinity for each other. More precisely, P is an 8 by 8 matrix with 0.5 along its diagonal, $P_{12} = P_{21} = 0.3$ and every other value is set to 0.05. The bounds for this SBM for a varied cluster size n can be seen in Figure 6.

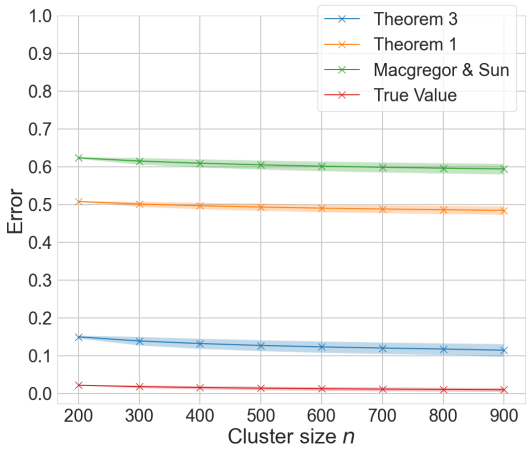


Figure 6: Stochastic block model with 8 blocks where one pair of blocks have a higher affinity to each other. Each point is an average over 10 realisations and the standard deviation is included as filled error bars.

We also considered an SBM with multiple layers in its hierarchy of clusters. We took an SBM with 12 blocks that can be split into three sets of four. In each quartet, two pairs of the blocks have a higher affinity for each other. Precisely, this was the 12×12 probability matrix P that was used:

$$P = \begin{bmatrix} Q & \mathbf{.05} & \mathbf{.05} \\ \mathbf{.05} & Q & \mathbf{.05} \\ \mathbf{.05} & \mathbf{.05} & Q \end{bmatrix}$$

Where

$$Q = \begin{bmatrix} .9 & .7 & .2 & .2 \\ .7 & .9 & .2 & .2 \\ .2 & .2 & .9 & .7 \\ .2 & .2 & .7 & .9 \end{bmatrix}$$

and $\mathbf{.05}$ is a 4×4 matrix of 0.05. The bounds for this SBM for varied cluster size n can be seen in Figure 7.

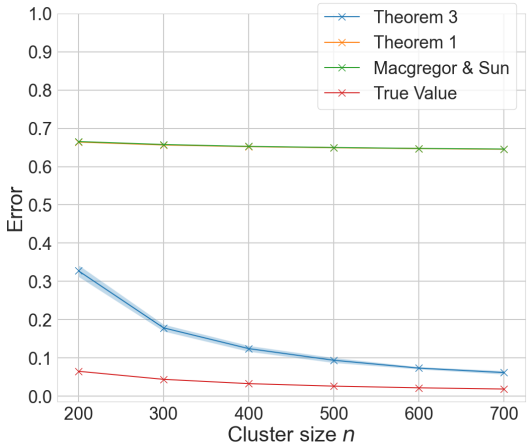


Figure 7: Stochastic block model with 12 blocks where three sets of four blocks have a higher affinity to blocks inside their quartet. Inside each quartet, two pairs of blocks have an even higher affinity to each other. Each point is an average over 5 realisations and the standard deviation is included as filled error bars.

In both experiments, we see that Theorem 3 greatly outperforms the other bounds.

C.2 SBMs: Close to the perfect recovery threshold

We now consider our bounds for an SBM with 2 blocks with probability matrix

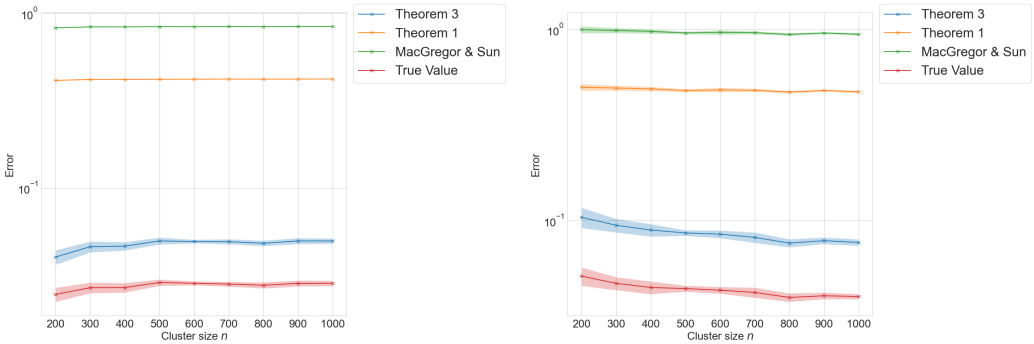
$$P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

where $p = \frac{\alpha \log(N)}{N}, q = \frac{\beta \log(N)}{N}$, with α and β taken at or close to the perfect recovery threshold $\sqrt{\alpha} - \sqrt{\beta} \geq \sqrt{2}$. Details of this threshold and its significance can be found in [32]. We study the results of several experiments for different values of α and β. In Figure 8a, we try $\beta = 20$ with α satisfying $\sqrt{\alpha} - \sqrt{\beta} = 2$ and study the bounds for a varied cluster size n. In Figure 8b, we try the same experiment with $\beta = 1$.

The bounds do not vary much for our choices of n, however we see that Theorem 3 provides a tighter bound on the true value than the other results. We also considered varying β with α satisfying $\sqrt{\alpha} - \sqrt{\beta} = 2$ and cluster size fixed at $n = 500$. The results can be seen in Figure 9.

We notice a similar behaviour in that the bounds do not change much at all as we vary our parameter β. Again, the bound from Theorem 3 is much closer to the true value. Finally, we consider varying β for a fixed choice of α ($\alpha = 35$). This will mean that as β is increased we will get closer to the threshold and eventually will be in violation of it. The results are pictured in Figure 10.

In this experiment, we see that the true distance of the eigenvectors from the indicator vectors of the clusters grows with β, and the bound from Theorem 3 stays relatively close to it. The other bounds are rather poor and increase slightly.



(a) $\beta = 20$ **(b)** $\beta = 1$

Figure 8: $\beta = 20$ (left) and $\beta = 1$ (right), $\alpha = (\sqrt{\beta} + \sqrt{2})^2$, varying cluster size n. An average is taken over 10 realisations for each value and the standard deviation is included as filled error bars.

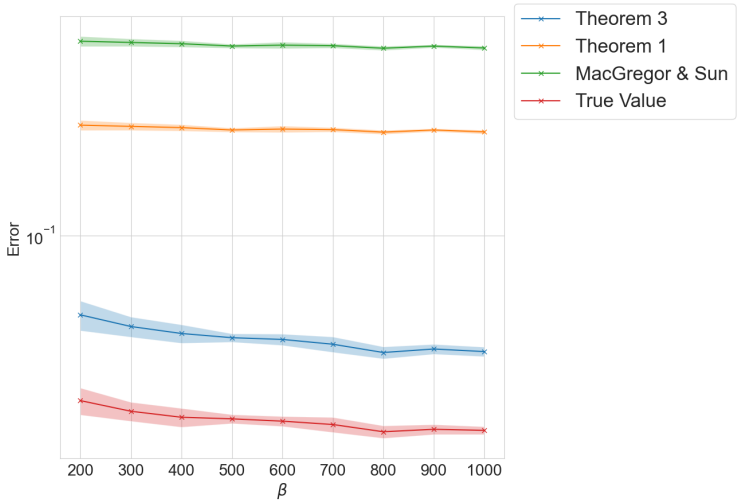


Figure 9: cluster size $n = 500$; β is varied and α is parametrised by β with $\alpha = (\sqrt{\beta} + \sqrt{2})^2$. Each value is an average over 10 realisations and the standard deviation is included as filled error bars.

C.3 DSBMs: Cyclic cluster structure

Similar to our experiment in Section 4, we consider a graph $\mathcal{G} \sim \text{DSBM}(k, n, P, F)$ where $k = 4$, $n = 100$ and P and F are defined as follows:

$$F = \begin{pmatrix} .5 & 1 & .5 & 0 \\ 0 & .5 & 1 & .5 \\ .5 & 0 & .5 & 1 \\ 1 & .5 & 0 & .5 \end{pmatrix}, \; P = \begin{pmatrix} \epsilon & 1 & \epsilon & 1 \\ 1 & \epsilon & 1 & \epsilon \\ \epsilon & 1 & \epsilon & 1 \\ 1 & \epsilon & 1 & \epsilon \end{pmatrix}.$$

As before, ϵ is a noise parameter that we vary. This results in a graph which has a distinct cyclic cluster structure for small ϵ. The results of this experiment can be seen in Figure 11.

As one can see, our bounds significantly outperform the result of [7]. In particular our bound from 2 hugs the true value very tightly for all noise levels.

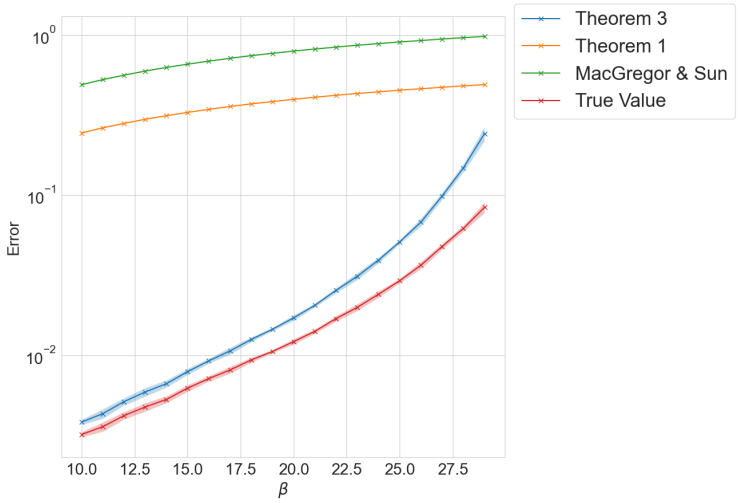


Figure 10: $n = 500$; β is varied and α is fixed at $\alpha = 35$. Each value is an average over 10 realisations and the standard deviation is included as filled error bars.

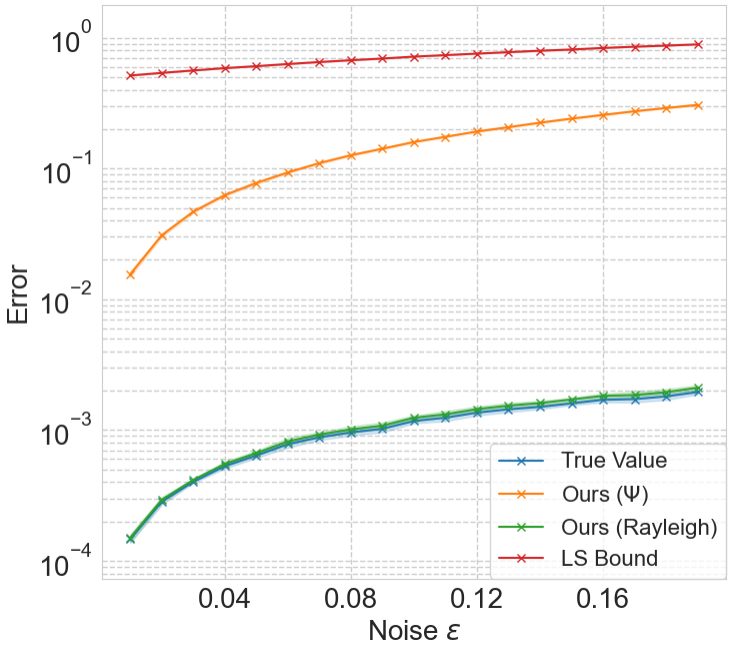


Figure 11: Comparison of the results given by the structure theorem of our paper (green for equation 2 and orange for equation 3) and by Laenen and Sun (red) for two DSBMs at varying level of noise. The actual values are reported in blue. Each value is an average over 10 trials.