COMMUNITY DETECTION FROM LOW-RANK EXCITATION OF GRAPH FILTERS

Hoi-To Wai[†], *Santiago Segarra*[‡], *Asuman E. Ozdaglar*[⋆], *Anna Scaglione*[†]

[†]School of ECEE, Arizona State University, Tempe, AZ, USA.
[‡]IDSS, MIT, Cambridge, MA, USA. [⋆]LIDS, MIT, Cambridge, MA, USA.
Emails: htwai@asu.edu, segarra@mit.edu, asuman@mit.edu, anna.scaglione@asu.edu

ABSTRACT

This paper considers the problem of inferring the topology of a graph from noisy outputs of an unknown graph filter excited by low-rank signals. Limited by this low-rank structure, we focus on the *community detection* problem, whose aim is to partition the node set of the unknown graph into subsets with high edge densities. We propose to detect the communities by applying spectral clustering on the low-rank output covariance matrix. To analyze the performance, we show that the low-rank covariance yields a sketch of the eigenvectors of the unknown graph. Importantly, we provide theoretical bounds on the error introduced by this sketching procedure based on spectral features of the graph filter involved. Finally, our theoretical findings are validated via numerical experiments in both synthetic and real-world graphs.

***Index Terms*—** graph signal processing, graph filter, topology identification, low rank excitation, community detection

1. INTRODUCTION

The emergence of new fields of knowledge such as Network Science and Big Data has generated a pressing need to broaden the scope beyond classical signal processing, to also accommodate signals defined on graphs [1–3]. Under the assumption that the signal properties are related to the topology of the graph where they are supported, the goal of graph signal processing (GSP) is to develop algorithms that leverage this relational structure, and can make inferences about these relationships when they are only partially observed [3]. A suitable way to accomplish these objectives is to rely on the so-called graph-shift operator (GSO), which is a matrix that reflects the local connectivity of the graph [2]. Most GSP works assume that the GSO (hence the graph) is known, and then analyze how the algebraic and spectral characteristics of the GSO impact the properties of the signals and filters defined on such a graph. This approach has been successfully implemented in the extension of classical problems to the realm of graphs such as sampling [4, 5] and reconstruction [6, 7]. By contrast, here we take the reverse path and investigate how to use information available from graph signals to infer aspects of the underlying topology. In particular, in this paper we focus on inferring the community structure [8] of the underlying graph.

Network topology inference from a set of (graph-signal) observations is a prominent problem in Network Science [3] with classical approaches for its solution including partial correlations [9], Gaussian graphical models [10], and structural equation models [11, 12], among others. Differently, recent GSP-based network inference frameworks postulate that the network exists as a latent underlying structure, and that observations are generated as a result of a network process defined in such a graph [13–21]. However, in order to infer the graph topology from the observed graph signals, the mentioned approaches rely on different assumptions such as smoothness of the observed signals [13–15], richness of the input signals [16–18], or partial knowledge of the involved filter or dynamic on the graph [19–21].

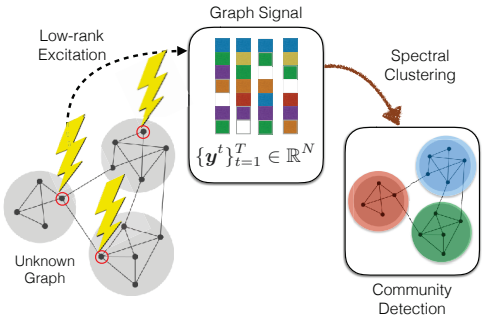

Fig. 1. Overview of the proposed approach. The observed graph signals are the outputs of a low-pass graph filter subject to a low-rank excitation, such as sparse inputs or inputs residing in a low-dimensional subspace. We show that spectral clustering can still recover the communities of the *unknown* underlying graph from these limited observations.

In the current paper, we model the observed graph signals as the outputs of an unknown graph filter subject to low-rank excitation, such as inputs that belong to a low-dimensional space or that are sparse. Under this weaker set of assumptions, exact inference of the underlying graph is in general infeasible. Nonetheless, we show that the *community structure* of this unknown graph can be recovered in this laxer setting. To show this result, we rely on the concept of *sketching* and, more specifically, on the theory behind recent applications of it for speeding up spectral clustering algorithms [22, 23]. Our contributions are twofold. First, we show that the covariance matrix of the observed graph signals retains a sketch of the principal eigenvectors of the unknown underlying GSO. Second and more importantly, we provide a theoretical characterization of the performance of spectral clustering when applied to the output covariance matrix. Precisely, using as a benchmark the communities that spectral clustering would output when perfect knowledge of the underlying graph is available, we bound the error in the detection of communities in terms of the spectral properties of the graph filter involved. Consequently, this bound can be used to identify linear network dynamics – modeled as graph filters – that are amenable to the problem of community detection. Fig. 1 overviews the proposed approach.

Notation. We use boldfaced lower-case (*resp.* upper-cased) letters to denote vectors (*resp.* matrices). For a vector $\boldsymbol{x}$, the notation x_i denotes its ith element. For a matrix $\boldsymbol{X}$, the notation X_{ij} denotes its (i,j)th element whereas $[\boldsymbol{X}]_{i,:}$ denotes its ith row vector. Also, $\mathcal{R}(\boldsymbol{X}) \subseteq \mathbb{R}^N$ denotes the range space of $\boldsymbol{X} \in \mathbb{R}^{N\times M}$. For a symmetric matrix $\boldsymbol{E}$, $\beta_i(\boldsymbol{E})$ denotes its ith largest eigenvalue. For a ma-

This work is supported by NSF CCF-1011811.

trix $\boldsymbol{M} \in \mathbb{R}^{P\times N}$, it admits the partition $\boldsymbol{M} = [\boldsymbol{M}_K \ \boldsymbol{M}_{N-K}]$ with $\boldsymbol{M}_K$ (*resp.* $\boldsymbol{M}_{N-K}$) denotes the matrix consisting of the *left-most* K (*resp. right-most* $N-K$) columns of $\boldsymbol{M}$. Similarly, $\boldsymbol{m} \in \mathbb{R}^N$ is partitioned into $\boldsymbol{m} = [\boldsymbol{m}_K; \boldsymbol{m}_{N-K}]$, where $\boldsymbol{m}_K$ (*resp.* $\boldsymbol{m}_{N-K}$) consists of its *top* K (*resp. bottom* $N-K$) elements.

2. GRAPH SIGNALS AND COMMUNITY DETECTION

We formally introduce GSP terminology that will be used throughout the paper and we briefly present the classical approach of spectral clustering for community detection.

Graph signals and graph filters. Let $G = (V, E)$ denote an undirected graph with a set of nodes $V := \{1, ..., N\}$ and a set of links $E \subseteq V \times V$, such that if node i is connected to j, then both (i, j) and (j, i) belong to E. We define the adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ as a matrix with non-zero elements $A_{ij} = A_{ji}$ if and only if $(i, j) \in E$. The adjacency matrix can be used to define the degree matrix $\boldsymbol{D} := \mathrm{diag}(\boldsymbol{A}\mathbf{1})$ and the (combinatorial) Laplacian matrix as $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$. Graph signals defined on the nodes of G are functions $f : V \to \mathbb{R}$, equivalently represented as vectors $\boldsymbol{x} = [x_1, ..., x_N]^\top \in \mathbb{R}^N$, where x_i denotes the signal value at node i. The graph G is endowed with the so-called graph shift operator $\boldsymbol{S} \in \mathbb{R}^{N\times N}$, a matrix whose entry S_{ij} can be non-zero only if $i = j$ or if $(i, j) \in E$. In this paper, we consider the case in which the GSO is given by the graph Laplacian, i.e. $\boldsymbol{S} = \boldsymbol{L}$, and has an eigenvalue decomposition given by $\boldsymbol{S} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ and λ_i is sorted in an *ascending order* $0 = \lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$.

The shift $\boldsymbol{S}$ can be used to define linear graph-signal operators denominated *graph filters* that have the form

$$\mathcal{H}(\boldsymbol{S}) := \textstyle\sum_{\ell=0}^{L-1} h_\ell \boldsymbol{S}^\ell = \boldsymbol{V}\big(\sum_{\ell=0}^{L-1} h_\ell \boldsymbol{\Lambda}^\ell\big)\boldsymbol{V}^\top . \qquad (1)$$

For a given input $\boldsymbol{z}$, the output of the filter is simply $\boldsymbol{x} = \mathcal{H}(\boldsymbol{S})\boldsymbol{z}$. The coefficients of the filter are collected into $\boldsymbol{h} := [h_0, \ldots, h_{L-1}]^\top$. Graph filters are of particular interest because they represent linear transformations that can be implemented locally and, thus, can be used to model linear network dynamics. Leveraging the spectral decomposition of $\boldsymbol{S}$, graph filters and signals can be represented in the frequency domain. To be precise, let us use the eigenvalues of $\boldsymbol{S}$ to define the $N \times L$ Vandermonde matrix $\boldsymbol{\Psi}$, where $\Psi_{ij} := (\lambda_i)^{j-1}$. Then, the frequency representations of a *signal* $\boldsymbol{z}$ and of a *filter* $\boldsymbol{h}$ are defined as $\tilde{\boldsymbol{z}} := \boldsymbol{V}^\top\boldsymbol{z}$ and $\tilde{\boldsymbol{h}} := \boldsymbol{\Psi}\boldsymbol{h}$. Exploiting such representations and with $\odot$ denoting the elementwise product, the filter's output $\boldsymbol{x} = \mathcal{H}(\boldsymbol{S})\boldsymbol{z}$ in the frequency domain is given by

$$\tilde{\boldsymbol{x}} = \mathrm{diag}\big(\boldsymbol{\Psi}\boldsymbol{h}\big)\boldsymbol{V}^\top\boldsymbol{z} = \tilde{\boldsymbol{h}} \odot \tilde{\boldsymbol{z}} . \qquad (2)$$

This is analogous to the convolution theorem for temporal signals.

Low-pass graph filters. Out of all possible filters, low-pass graph filters play a fundamental role for community detection. Formally, we say that a graph filter is *low-pass* and (K, η)*-separable* if

$$\tilde{h}_{K+1}/\tilde{h}_K \le \eta < 1, \quad \tilde{h}_N \ge 0 \quad \text{and} \quad \tilde{h}_i \ge \tilde{h}_{i+1} , \qquad (3)$$

for all $i = 1, \ldots, N-1$. Notice that since $\tilde{h}_i \ge \tilde{h}_{i+1}$, the eigenvectors of $\mathcal{H}(\boldsymbol{S})$ corresponding to the largest K eigenvalues are the left-most K vectors in $\boldsymbol{V}$, denoted by $\boldsymbol{V}_K$. Common examples of low-pass and (K, η)-separable graph filters (for all K) [15, 16] include $\mathcal{H}_1(\boldsymbol{S}) = (\boldsymbol{I} + \boldsymbol{S})^{-1}$, $\mathcal{H}_2(\boldsymbol{S}) = \exp(-\boldsymbol{S})$, and $\mathcal{H}_3(\boldsymbol{S}) = (\boldsymbol{I} - \alpha\boldsymbol{S})^{L-1}$ for $0 < \alpha < 1/\lambda_N$.

Observe that if $\eta \approx 0$, *i.e.,* the frequency response declines sharply from the Kth to the $(K+1)$th frequency, then $\mathcal{H}(\boldsymbol{S})$ is approximately rank-K. In fact, Section 3 reveals that a small value of η is essential to obtain a tight theoretical guarantee on detecting K communities from the unknown graph. We remark that a small value of η can be achieved by increasing the graph filter order L.

Community detection via spectral clustering. For an undirected graph $G = (V, E)$, a *community* is intuitively a set of nodes with higher edge density among themselves than to nodes outside the set. Thus, the problem of K-community detection amounts to finding K such disjoint communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$ such that $V = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_K$. A well-known way to establish a sense of optimal partition of the node sets into communities is that of minimizing the *ratio cut* [24]. Formally, defining the cut weight between two node sets as the sum of edge weights between them, *i.e.,* $\boldsymbol{A}_{\mathcal{C}_i,\mathcal{C}_j} := \sum_{k\in\mathcal{C}_i, k'\in\mathcal{C}_j} A_{kk'}$. The ratio cut of a given partition is defined as

$$\mathrm{RatioCut}(\mathcal{C}_1, \ldots, \mathcal{C}_K) := (1/2)\textstyle\sum_{i=1}^K \boldsymbol{A}_{\mathcal{C}_i,\overline{\mathcal{C}}_i}/|\mathcal{C}_i| , \qquad (4)$$

where $\overline{\mathcal{C}}_i$ is the complement of $\mathcal{C}_i$. Unfortunately, the ratio-cut minimization problem is hard due to its combinatorial nature. As a remedy, a widely used heuristic is the *spectral clustering* method [25] which can be seen as a convex relaxation of the ratio-cut minimization problem. Spectral clustering can be described as a two-step process[1]. First, every node is projected onto $\mathbb{R}^K$, where the coordinates of node i are given by $[V_{i1}, \ldots, V_{iK}]$, *i.e.,* the ith row vector of the left-most K columns of $\boldsymbol{V}$. Secondly, the K-means clustering method [26] is applied to the resulting projected space to obtain the sought K communities.

3. COMMUNITY DETECTION FROM LOW-RANK DATA

Our goal is to detect communities in a graph when we do not have access to the graph itself – nor, equivalently, the GSO $\boldsymbol{S}$ – but rather to a set of graph signals defined on it. Formally, we observe a set of graph signals $\boldsymbol{y}^t \in \mathbb{R}^N$, $t = 1, ..., T$. Each graph signal is a noisy observation of the output of a low-pass and (K, η)-separable graph filter (1):

$$\boldsymbol{y}^t = \boldsymbol{x}^t + \boldsymbol{w}^t \ \text{ and } \ \boldsymbol{x}^t := \mathcal{H}(\boldsymbol{S})\boldsymbol{z}^t , \qquad (5)$$

where the observation noise is $\boldsymbol{w}^t \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$ and the excitation signal $\boldsymbol{z}^t \in \mathbb{R}^N$ is zero mean.

Unlike previous works [16] that tackle the problem of identifying $\boldsymbol{S}$ from a *full-rank excitation* model where $\mathbb{E}[\boldsymbol{z}^t(\boldsymbol{z}^t)^\top] = \boldsymbol{I}$, this paper focuses on low-rank excitation in which

$$\boldsymbol{C}_z = \mathbb{E}[\boldsymbol{z}^t(\boldsymbol{z}^t)^\top] = \boldsymbol{B}\boldsymbol{B}^\top, \qquad (6)$$

where $\boldsymbol{B} \in \mathbb{R}^{N\times R}$ with $R < N$. Practical scenarios that lead to (6) include the cases where the input signal $\boldsymbol{z}^t$ is applied on only a subset of R nodes and where the number of variations in the excitation signal is limited to R modes. Under this low-rank excitation model, the covariance matrix of a general output signal $\boldsymbol{x}^t$ is given by

$$\boldsymbol{C}_x = \mathbb{E}[\boldsymbol{x}^t(\boldsymbol{x}^t)^\top] = \mathcal{H}(\boldsymbol{S})\boldsymbol{B}\boldsymbol{B}^\top\mathcal{H}(\boldsymbol{S})^\top, \qquad (7)$$

whose rank is at most R. Inferring $\boldsymbol{S}$ from $\boldsymbol{C}_x$ is challenging for two reasons: i) we do not assume perfect knowledge of the specific filter coefficients $\boldsymbol{h}$ defining $\mathcal{H}$, and ii) information is in general lost due to the low-rank nature of $\boldsymbol{B}\boldsymbol{B}^\top$. Nevertheless, we show that we can recover the community structure in $\boldsymbol{S}$ by applying spectral clustering directly on $\boldsymbol{C}_x$, or an estimated version $\hat{\boldsymbol{C}}_x$ of it.

[1] Minor variations to the method presented here co-exist in the literature.

Algorithm 1 Community detection from low-rank excitation.

1: **Input**: Graph signals $\{\boldsymbol{y}^t\}_{t=1}^T$; desired number of clusters K.
2: Use $\{\boldsymbol{y}^t\}_{t=1}^T$ to compute the sample covariance $\hat{\boldsymbol{C}}_x$ as in (8).
3: Find the K eigenvectors to $\hat{\boldsymbol{C}}_x$ associated with the largest K eigenvalues. Denote the set of eigenvectors as $\hat{\boldsymbol{P}}_K \in \mathbb{R}^{N\times K}$.
4: Perform K-means clustering (e.g., [26]), which optimizes:

$$\min_{\mathcal{C}_1,\ldots,\mathcal{C}_K} \sum_{i=1}^{K}\sum_{j\in\mathcal{C}_i}\left\|\hat{\boldsymbol{p}}_j - \frac{1}{|\mathcal{C}_i|}\sum_{q\in\mathcal{C}_i}\hat{\boldsymbol{p}}_q\right\|_2^2 \text{ s.t. } \mathcal{C}_i\subseteq V\ , \quad (9)$$

where $\hat{\boldsymbol{p}}_i := [\hat{\boldsymbol{P}}_K]_{i,:} \in \mathbb{R}^K$. Let the solution be $\hat{\mathcal{C}}_1,...,\hat{\mathcal{C}}_K$.
5: **Output**: Partition of V into K communities, $\hat{\mathcal{C}}_1,...,\hat{\mathcal{C}}_K$.

3.1. Community detection algorithm

From the observed graph signals $\{\boldsymbol{y}^t\}_{t=1}^T$ [cf. (5)] we construct the empirical sample covariance

$$\hat{\boldsymbol{C}}_x = (1/T)\sum_{t=1}^T \boldsymbol{y}^t(\boldsymbol{y}^t)^\top, \quad (8)$$

and apply the spectral clustering procedure presented in Section 2 but based on the eigenvectors of $\hat{\boldsymbol{C}}_x$ associated to the K *largest* eigenvalues. This algorithm is summarized in Algorithm 1.

We first discuss the intuitive motivation of the proposed algorithm. First notice that the filter $\mathcal{H}(\boldsymbol{S})$, being a matrix polynomial on the symmetric matrix $\boldsymbol{S}$ [cf. (1)], preserves its eigenvectors. Moreover, since $R < N$, the matrix $\mathcal{H}(\boldsymbol{S})\boldsymbol{B} \in \mathbb{R}^{N\times R}$ is a '*sketch*' of $\mathcal{H}(\boldsymbol{S})$ with $\boldsymbol{B}$ being the *sketching matrix*. On the other hand, if $R \geq K$, the information of the top-K eigenvectors of $\mathcal{H}(\boldsymbol{S})$ can be preserved in the sketched version. Indeed, when $\mathcal{R}(\boldsymbol{B}) = \mathcal{R}(\boldsymbol{V}_K)$, it is easy to observe that $\mathcal{R}(\mathcal{H}(\boldsymbol{S})\boldsymbol{B}) = \mathcal{R}(\boldsymbol{V}_K)$. In this case, applying Algorithm 1 based on the covariance $\boldsymbol{C}_x = \mathcal{H}(\boldsymbol{S})\boldsymbol{B}\boldsymbol{B}^\top\mathcal{H}(\boldsymbol{S})^\top$ yields the same output as applying spectral clustering on $\boldsymbol{S}$.

In most cases, however, Algorithm 1 *will not* return the same communities as those obtained by spectral clustering on the GSO $\boldsymbol{S}$. There are two sources for this discrepancy. First, the sketching matrix $\boldsymbol{B}$ does not share the range of $\boldsymbol{V}_K$ in general, thus distorting the eigenvectors in $\mathcal{H}(\boldsymbol{S})$. Second, having access to a finite set of graph signals, we do not get to observe $\boldsymbol{C}_x$ perfectly but rather an estimate $\hat{\boldsymbol{C}}_x$ of it. In the next section we bound the error introduced by these two sources of uncertainty.

3.2. Main Theoretical Guarantees

Denote by $\{\hat{\mathcal{C}}_1,\ldots,\hat{\mathcal{C}}_K\}$ the community structure obtained from Algorithm 1. Our goal is to find out how close these communities are to the ones obtained when $\boldsymbol{S}$ is available. To do so, we measure the optimality of the clustering result by comparing the following objective — let $\mathcal{C}_1,...,\mathcal{C}_K$ be any partition of V,

$$F(\mathcal{C}_1,...,\mathcal{C}_K) := \sum_{i=1}^{K}\sum_{j\in\mathcal{C}_i}\left\|\boldsymbol{v}_j - \frac{1}{|\mathcal{C}_i|}\sum_{q\in\mathcal{C}_i}\boldsymbol{v}_q\right\|_2^2, \quad (10)$$

where $\boldsymbol{v}_i := [\boldsymbol{V}_K]_{i,:}, i = 1,...,N$. Notice that (10) is the same objective function used in step 4 of Algorithm 1 but applied to the eigenvectors of the GSO $\boldsymbol{S}$ instead. In general, $F(\mathcal{C}_1,...,\mathcal{C}_K)$ can only be minimized when $\boldsymbol{S}$ is known perfectly.

Let us introduce the following definitions. The singular value decomposition of $\mathcal{H}(\boldsymbol{S})\boldsymbol{B}$ is given by $\mathcal{H}(\boldsymbol{S})\boldsymbol{B} := \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^\top$ with $\sigma_1 \geq \sigma_2 \geq \ldots$, the difference between the covariance and the sample covariance is $\boldsymbol{\Delta}_x := \hat{\boldsymbol{C}}_x - \boldsymbol{C}_x$. We have:

Theorem 1 *If the following conditions hold:*

1. $\mathcal{H}(\boldsymbol{S})$ *is a low-pass and* (K,η)*-separable filter [cf.* (3)*],*
2. *Step 4 of Algorithm 1 finds a solution* $\hat{\mathcal{C}}_1,...,\hat{\mathcal{C}}_K$ *that exactly minimizes the problem* (9)*,*
3. $\text{rank}(\boldsymbol{V}_K\text{diag}(\boldsymbol{\lambda}_K)\boldsymbol{V}_K^\top\boldsymbol{B}\boldsymbol{Q}_K) = K$*,*
4. $\sigma_K > 0$*,*
5. $\exists\,\delta > 0$ *such that* $\beta_1(\boldsymbol{\Delta}_x) + \delta \leq \beta_K(\boldsymbol{C}_x) - \beta_{K+1}(\boldsymbol{C}_x)$*,*

then, denoting by $F^\star := \min_{\mathcal{C}_1,...,\mathcal{C}_K\subseteq V} F(\mathcal{C}_1,...,\mathcal{C}_K)$ *as the optimal cost associated with spectral clustering with perfect knowledge of* $\boldsymbol{S}$*, we have that*

$$F(\hat{\mathcal{C}}_1,...,\hat{\mathcal{C}}_K) - F^\star \leq \sqrt{8K}\left(\sqrt{\frac{\gamma^2}{1+\gamma^2}} + \frac{|\beta_1(\boldsymbol{\Delta}_x)|}{\delta}\right)\ , \quad (11)$$

where γ *is bounded by*

$$\gamma \ \leq\ \eta\cdot\|\boldsymbol{V}_{N-K}^\top\boldsymbol{B}\boldsymbol{Q}_K\|_2\|(\boldsymbol{V}_K^\top\boldsymbol{B}\boldsymbol{Q}_K)^{-1}\|_2\ . \quad (12)$$

The proof of Theorem 1 – partially inspired by [22] – is omitted here due to space limitations but can be found in an online appendix[2].

As wanted, (11) bounds the difference between the optimal cost $F^\star$ associated with running spectral clustering when perfect knowledge of the graph is available and the cost achieved by the communities $\{\hat{\mathcal{C}}_1,...,\hat{\mathcal{C}}_K\}$ obtained from our algorithm. This optimality gap consists of the two summands on the right-hand side of (11). The first summand, which depends on γ, captures the distortion introduced by the sketching operation. From (12) it follows that this optimality gap depends on the parameter η of the low pass graph filter, which might be improved for higher filter order L, as discussed in Section 2. Furthermore, expression (12) reveals that γ depends on the angle between $\mathcal{R}(\boldsymbol{B}\boldsymbol{Q}_K)$ and $\mathcal{R}(\boldsymbol{V}_K)$. In particular, if $\mathcal{R}(\boldsymbol{B}\boldsymbol{Q}_K) = \mathcal{R}(\boldsymbol{V}_K)$, then $\|\boldsymbol{V}_{N-K}^\top\boldsymbol{B}\boldsymbol{Q}_K\|_2 = 0$ and $\gamma = 0$.

The second summand in the right-hand side of (11) bounds the error resulting from the fact that we have access to a finite number of signals. This bound follows from applying the classical Weyl's inequality combined with Davis-Kahan theorem [27]. Theorem 1 shows that when the largest eigenvalue of the error $\boldsymbol{\Delta}_x$ is small compared to the relevant *spectral gap of* $\boldsymbol{C}_x$ [cf. Condition 5], then the additive error introduced will simply be proportional to the largest eigenvalue of $\boldsymbol{\Delta}_x$. Moreover, using standard results from concentration inequalities and under mild assumptions on the statistics of $\boldsymbol{z}^t$, the error $|\beta_1(\boldsymbol{\Delta}_x)|$ decays at the rate of $\mathcal{O}(1/\sqrt{T})$ with high probability. Overall, Theorem 1 shows that when the graph filter has a favorable frequency response, and the number of collected samples is sufficient, we expect Algorithm 1 to produce communities comparable to those obtained by applying spectral clustering on $\boldsymbol{S}$.

4. NUMERICAL EXAMPLES

This section presents numerical results to illustrate the performance of our low-rank community detection approach. We first focus on a controlled setting where the GSO and the associated graph signals are generated synthetically. More precisely, we consider an undirected graph G generated according to a stochastic block model (SBM) with K communities, i.e. $G \sim \mathsf{SBM}(N,K,a,b)$ such that a and b are the probabilities that an edge is formed within a community and between the communities, respectively, and we have that

[2] See `http://xxx` (For convenience, I am including the proof in the last page for now.)

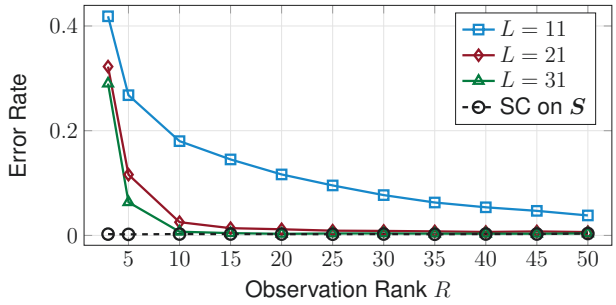

Fig. 2. **(Effect of filter order L and observation rank R)**. Matrix $\boldsymbol{S}$ is adjacency matrix of SBM graphs with $N = 150$ nodes and $K = 3$ communities, and $\mathcal{H}(\boldsymbol{S}) = (\boldsymbol{I} - \alpha \boldsymbol{S})^{L-1}$, under a noiseless observation setting with $\hat{\boldsymbol{C}}_x = \boldsymbol{C}_x$.

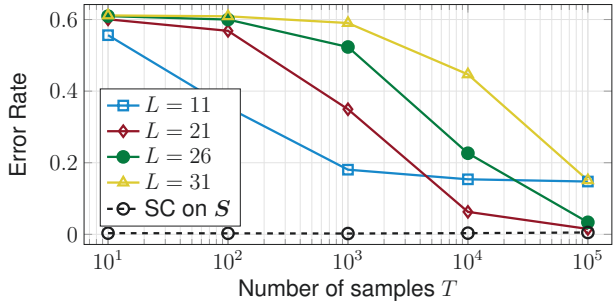

Fig. 3. **(Effect of sample size T)**. Similar settings as in Fig. 2 with $R = 15$. We consider noisy observations and only a finite number of samples are observed T.

$a > b$. The graph filter takes the form $\mathcal{H}(\boldsymbol{S}) = (\boldsymbol{I} - \alpha \boldsymbol{S})^{L-1}$ for $\alpha = 1/(2d_{\text{max}})$ where $d_{\text{max}} = \max_i d_i$ and d_i is the degree of node i. The filter order is given by L. It is immediate to check that the graph filter is low-pass and (K, η)-separable. The corresponding η is given as:

$$\eta = \left(\frac{1 - \alpha \lambda_{K+1}}{1 - \alpha \lambda_K} \right)^{L-1} . \tag{13}$$

Since $\alpha < 1/\lambda_N$, we have $1 - \alpha\lambda_{K+1} < 1 - \alpha\lambda_K$ if $\lambda_{K+1} > \lambda_K$. Therefore, we anticipate η to decrease to zero as L increases. Furthermore, the low-rank excitation signal is generated as $\boldsymbol{z}^t = \boldsymbol{B}\tilde{\boldsymbol{z}}^t$ where $\tilde{\boldsymbol{z}}^t \in \mathbb{R}^R$ satisfies $\tilde{\boldsymbol{z}}^t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{B}$ is a *row-sparse* binary matrix with only R non-zero rows, chosen at random, and each row has only $\lceil (R/N) d_i \rceil$ ones, also chosen at random. This corresponds to the setting when the graph is excited only on R nodes and each is driven by a few independent sources. The results reported here correspond to the average of 1000 Monte-Carlo simulations. To compute step 4 of Algorithm 1 we use the `kmeans` function in MATLAB.

The first example considers a noiseless setting where the algorithm has access to $\hat{\boldsymbol{C}}_x = \boldsymbol{C}_x$. We compare the average error rate (compared to the ground truth communities of the SBM) when applying Algorithm 1 for different filter orders L and ranks R of the sketching matrix $\boldsymbol{B}$. The SBM graph has $N = 150$ nodes, $K = 3$ communities and the parameters are set as $a = 8 \log N/N$ and $b = \log N/N$; see Fig. 2. We observe that the error rate decays as we increase the filter order L. This is consistent with our result in Theorem 1. Indeed, for larger values of L, we have that η in (12) decreases, thus leading to better performance. Furthermore, the performance improves with the observation rank R and approaches that of applying spectral clustering on the original $\boldsymbol{S}$. With a filter order of $L = 21$ and observation rank $R \approx 15$, the proposed method achieves a performance comparable to that of having perfect knowledge of the graph.

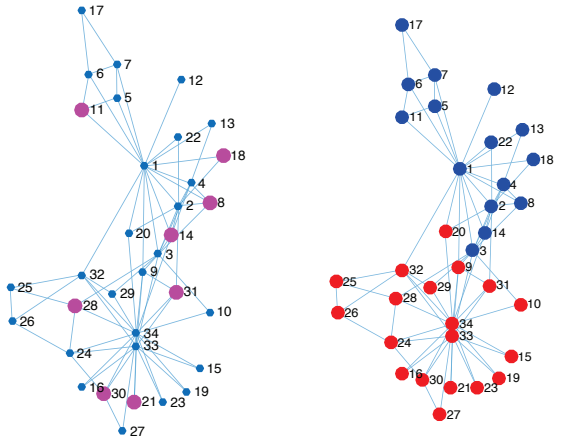

Fig. 4. **(Community Detection on Real Network)**. The Zachary Karate Club network with $N = 34$ nodes. (Left) Highlighted nodes are the locations of the non-zero rows of $\boldsymbol{B}$. (Right) Snapshot of the detected communities by proposed method with rank $R = 8$ observations and filter order $L = 8$.

The second example considers the finite sample setting and compares the error rate against the number of samples observed T. We focus on the same simulation settings as in the previous example and we fix $R = 15$, yet only a sample covariance $\hat{\boldsymbol{C}}_x$ is observed and the observation noise has a standard deviation of $\sigma = 10^{-1}$ [cf. (5) and (8)]. From Fig. 3, we observe that the error rate generally decreases as the number of samples increases. This is as predicted by Theorem 1, and the performance is ultimately limited by the filter order L. Interestingly, we observe that to achieve the same error rate, more samples are also required when the filter order L is high. This is because for higher filter orders, the magnitude of the eigenvalue $\beta_K(\boldsymbol{C}_x)$ decreases, which forces δ in Theorem 1 to decrease [cf. Condition 5].

The last example deals with detecting communities in a real network, namely Zachary's Karate Club [28]. The graph consists of $N = 34$ nodes and is depicted in Fig. 4 (Left). For the proposed method, we set $K = 2$ and the graph filter is the same as in the previous examples with $L = 8$. The low-rank excitation matrix $\boldsymbol{B}$ has rank $R = 8$ with a row-sparsity pattern given by the active nodes in Fig. 4 (Left). We consider the noisy observation setting with $T = 10^3$ and $\sigma = 10^{-1}$. For this case, we have an average error rate $\approx 8.8\%$ when compared to the communities recovered using spectral clustering on the original $\boldsymbol{S}$. An example of the recovered communities is depicted in Fig. 4 (Right).

Conclusions. We studied a practical method for recovering community structures from the outputs of a low-pass graph filter subject to low-rank excitation. The proposed method relies on applying spectral clustering on the output (sample) covariance matrix. We characterized the error of such procedure compared to the case where the true graph shift operator is available. Our analysis shows that the performance hinges on the spectral gap property of the graph filter involved. Numerical experiments were performed on synthetic and real networks to verify our results.

5. REFERENCES

[1] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.

[3] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. New York, NY: Springer, 2009.

[4] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, April 2016.

[5] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec 2015.

[6] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4363–4378, Aug 2016.

[7] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 764–778, Feb 2017.

[8] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3, pp. 75–174, 2010.

[9] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[10] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Shanghai, China, Mar. 20-25, 2016.

[11] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations," *PLoS, Computational Biology*, Jun. 2013.

[12] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2503–2516, 2017.

[13] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec 2016.

[14] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, March 2017, pp. 6508–6512.

[15] V. Kalofolias, "How to learn a graph from smooth signals," in *Intl. Conf. Artif. Intel. Stat. (AISTATS)*. J Mach. Learn. Res., 2016, pp. 920–929.

[16] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[17] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, March 2017, pp. 5870–5874.

[18] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Trans. Signal and Info. Process. over Networks*, 2017.

[19] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 2, no. 3, pp. 406–419, 2016.

[20] J. Mei and J. M. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.

[21] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 3, no. 3, pp. 484–499, Sept 2017.

[22] C. Boutsidis, P. Kambadur, and A. Gittens, "Spectral clustering via the power method-provably," in *International Conference on Machine Learning*, 2015, pp. 40–48.

[23] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *International Conference on Machine Learning*, 2016, pp. 1002–1011.

[24] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Trans. Comp. Aided Des. Integ. Cir. Sys.*, vol. 11, no. 9, pp. 1074–1085, Nov 2006.

[25] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[26] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[27] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. iii," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.

[28] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of anthropological research*, vol. 33, no. 4, pp. 452–473, 1977.

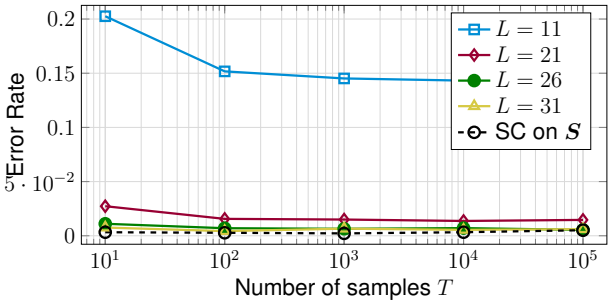

Fig. 5. **(Effect of sample size T without noise)**. Similar settings as in Fig. 2 with $R = 15$ and the exception of having noisy observations and only a finite number of samples observed T.

6. PROOF OF THEOREM 1

To simplify the notations while proving the theorem, let us define the following indicator matrices for the communities structures found. Firstly, the matrix $\hat{\boldsymbol{X}} \in \mathbb{R}^{N\times K}$ is associated with the communities $\{\hat{\mathcal{C}}_1, ..., \hat{\mathcal{C}}_K\}$ found with Algorithm 1 such that

$$\hat{X}_{ij} := \begin{cases} 1/|\hat{\mathcal{C}}_j|, & \text{if } i \in \hat{\mathcal{C}}_j \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Moreover, let $\upsilon^\star = F(\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star)$ be the optimal set of community found by minimizing $F(\mathcal{C}_1, ..., \mathcal{C}_K)$ [cf. (10)], similarly we can define $\boldsymbol{X}^\star \in \mathbb{R}^{N\times R}$ such that:

$$X_{ij}^\star := \begin{cases} 1/|\mathcal{C}_j^\star|, & \text{if } i \in \mathcal{C}_j^\star \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

If we define by $\mathcal{X}$ the set of all possible indicator matrices over all the combinations of communities, it can be verified that:

$$\begin{aligned} &\|\hat{\boldsymbol{P}}_K - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top \hat{\boldsymbol{P}}_K\|_{\mathrm{F}} = \min_{\boldsymbol{X}\in\mathcal{X}} \|\hat{\boldsymbol{P}}_K - \boldsymbol{X}(\boldsymbol{X})^\top \hat{\boldsymbol{P}}_K\|_{\mathrm{F}} \\ &= \min_{\mathcal{C}_1,\ldots,\mathcal{C}_K} \sum_{i=1}^K \sum_{j\in\mathcal{C}_i} \Big\|\hat{\boldsymbol{p}}_j - \frac{1}{|\mathcal{C}_i|}\sum_{q\in\mathcal{C}_i} \hat{\boldsymbol{p}}_q\Big\|_2^2 \text{ s.t. } \mathcal{C}_i \subseteq V \ , \end{aligned} \tag{16}$$

and similarly

$$\begin{aligned} \|\boldsymbol{V}_K - \boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top \boldsymbol{V}_K\|_{\mathrm{F}} &= \min_{\boldsymbol{X}\in\mathcal{X}} \|\boldsymbol{V}_K - \boldsymbol{X}(\boldsymbol{X})^\top \boldsymbol{V}_K\|_{\mathrm{F}} \\ &= \min_{\mathcal{C}_1,\ldots,\mathcal{C}_K} F(\mathcal{C}_1, ..., \mathcal{C}_K) = F^\star \ , \end{aligned} \tag{17}$$

furthermore, we have

$$\|\boldsymbol{V}_K - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top \boldsymbol{V}_K\|_{\mathrm{F}} = F(\hat{\mathcal{C}}_1, ..., \hat{\mathcal{C}}_K) \ . \tag{18}$$

Let us begin the proof of Theorem 1. Define an error matrix as $\boldsymbol{E} = \boldsymbol{V}_K\boldsymbol{V}_K^\top - \hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top$. We observe the following chain:

$$\begin{aligned} &F(\hat{\mathcal{C}}_1, ..., \hat{\mathcal{C}}_K) = \|\boldsymbol{V}_K - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top \boldsymbol{V}_K\|_{\mathrm{F}} \\ &= \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)\boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}} = \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)(\hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top + \boldsymbol{E})\|_{\mathrm{F}} \\ &\le \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)\hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_{\mathrm{F}} + \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)\boldsymbol{E}\|_{\mathrm{F}} \\ &\le \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)\hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_{\mathrm{F}} + \|\boldsymbol{E}\|_{\mathrm{F}} \ , \end{aligned}$$

where the second equality is due to the orthogonality of $\boldsymbol{V}_K^\top$ and the last inequality is due to the fact that $(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)$ is a projection matrix. Using (16), we proceed as

$$\begin{aligned} &F(\hat{\mathcal{C}}_1, ..., \hat{\mathcal{C}}_K) \le \|(\boldsymbol{I} - \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^\top)\hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_{\mathrm{F}} + \|\boldsymbol{E}\|_{\mathrm{F}} \\ &\le \|(\boldsymbol{I} - \boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)\hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_{\mathrm{F}} + \|\boldsymbol{E}\|_{\mathrm{F}} \\ &= \|(\boldsymbol{I} - \boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)(\boldsymbol{V}_K\boldsymbol{V}_K^\top - \boldsymbol{E})\|_{\mathrm{F}} + \|\boldsymbol{E}\|_{\mathrm{F}} \\ &\le \|(\boldsymbol{I} - \boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)\boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}} + 2\|\boldsymbol{E}\|_{\mathrm{F}} \\ &= F^\star + 2\|\boldsymbol{E}\|_{\mathrm{F}} \ . \end{aligned} \tag{19}$$

To bound $\|\boldsymbol{E}\|_{\mathrm{F}} = \|\boldsymbol{V}_K\boldsymbol{V}_K^\top - \hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_{\mathrm{F}}$, we invoke the following lemma:

Lemma 1 *[22] For any $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ with $m \ge n$ and $\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{I}$, it holds that:*

$$\|\boldsymbol{A}\boldsymbol{A}^\top - \boldsymbol{B}\boldsymbol{B}^\top\|_{\mathrm{F}}^2 \le 2n\|\boldsymbol{A}\boldsymbol{A}^\top - \boldsymbol{B}\boldsymbol{B}^\top\|_2^2 \ . \tag{20}$$

As a result, we obtain the bound that:

$$\|\boldsymbol{E}\|_{\mathrm{F}} \le \sqrt{2K}\|\boldsymbol{V}_K\boldsymbol{V}_K^\top - \hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_2 \ , \tag{21}$$

which can be further bounded using the following propositions —

Proposition 1 *Under Condition 1, 3 and 4 in Theorem 1, we have*

$$\|\boldsymbol{P}_K\boldsymbol{P}_K^\top - \boldsymbol{V}_K\boldsymbol{V}_K^\top\|_2^2 = \frac{\gamma^2}{1+\gamma^2} \ , \tag{22}$$

where the columns of $\boldsymbol{P}_K$ is the top K eigenvectors of $\boldsymbol{C}_x$ and γ is bounded as stated in (12)*.*

Proposition 2 *Under Condition 5 in Theorem 1, it holds that*

$$\|\boldsymbol{P}_K\boldsymbol{P}_K^\top - \hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_2 \le \frac{|\beta_1(\boldsymbol{\Delta}_x)|}{\delta} \ . \tag{23}$$

Obviously, combining (22), (23) and using the triangular inequality yields the desired result in expression (11). The proof is thus concluded.

6.1. Proof of Proposition 1

Denote the rank-K approximation to $\boldsymbol{S}$ as $[\boldsymbol{S}]_K := \boldsymbol{V}_K\mathrm{diag}(\boldsymbol{\lambda}_K)\boldsymbol{V}_K^\top$ and define the shorthand notation $\tilde{\boldsymbol{B}} := \boldsymbol{B}\boldsymbol{Q}_K$, we observe that

$$\mathcal{R}([\boldsymbol{S}]_K) = \mathcal{R}([\mathcal{H}(\boldsymbol{S})]_K) = \mathcal{R}([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}) \ , \tag{24}$$

where the second equality is due to the assumption that $\mathcal{H}(\boldsymbol{S})$ is a low-pass and separable graph filter, the last equality is due to condition 3 of Theorem 1, such that the linear transformation on the right does not modify the range space of $[\mathcal{H}(\boldsymbol{S})]_K$ (or equivalently $[\boldsymbol{S}]_K$). If we denote the columns of $\tilde{\boldsymbol{V}}_K$ as the top K left singular vectors of $[\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}$, the above shows that the two products are equal $\boldsymbol{V}_K\boldsymbol{V}_K^\top = \tilde{\boldsymbol{V}}_K\tilde{\boldsymbol{V}}_K^\top$.

Similarly, we define $[\boldsymbol{C}_x]_K := \boldsymbol{P}_K\mathrm{diag}(\boldsymbol{\sigma}_K)^2\boldsymbol{P}_K^\top$ as the rank K approximation to $\boldsymbol{C}_x$ and observe that

$$\mathcal{R}([\boldsymbol{C}_x]_K) = \mathcal{R}([\mathcal{H}(\boldsymbol{S})\boldsymbol{B}]_K) = \mathcal{R}(\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}}) \tag{25}$$

where the last equality is due to the fact that $\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}} = \mathcal{H}(\boldsymbol{S})\boldsymbol{B}\boldsymbol{Q}_K = \boldsymbol{P}_K\mathrm{diag}(\boldsymbol{\sigma}_K)$ as the columns of $\boldsymbol{Q}_K$ are the top K right singular vectors. Likewise, if the columns of $\tilde{\boldsymbol{P}}_K$ are the top K left singular vectors of $\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}}$, then $\boldsymbol{P}_K\boldsymbol{P}_K^\top = \tilde{\boldsymbol{P}}_K\tilde{\boldsymbol{P}}_K^\top$.

Furthermore, we observe that

$$\mathcal{R}([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}) \perp \mathcal{R}((\mathcal{H}(\boldsymbol{S}) - [\mathcal{H}(\boldsymbol{S})]_K)\tilde{\boldsymbol{B}}) \ . \tag{26}$$

Invoking [22, Lemma 8] through setting $\mathbf{D} = \mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}}$, $\mathbf{C} = [\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}$ and $\mathbf{E} = [\mathcal{H}(\boldsymbol{S})]_{N-K}\tilde{\boldsymbol{B}}$ therein, shows the following:

$$\begin{aligned}&\|\tilde{\boldsymbol{V}}_K\tilde{\boldsymbol{V}}_K^\top - \tilde{\boldsymbol{P}}_K\tilde{\boldsymbol{P}}_K^\top\|_2^2 \\ &= 1 - \beta_K\Big([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}\big((\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}})^\top\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}}\big)^\dagger([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}})^\top\Big) \ .\end{aligned}$$

Denote the matrix in the middle of the expression above as $\boldsymbol{\Pi} := (\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}})^\top\mathcal{H}(\boldsymbol{S})\tilde{\boldsymbol{B}}$. Now, under condition 4 of Theorem 1 that $\sigma_K > 0$, the $K \times K$ matrix $\boldsymbol{\Pi}$ is non-singular. We can observe the following chain:

$$\begin{aligned}&\beta_K\Big([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}\boldsymbol{\Pi}^{-1}([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}})^\top\Big) \\ &= \beta_K\Big(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}}\boldsymbol{\Pi}^{-1}(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^\top\Big) \\ &= \frac{1}{\beta_1\Big((\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-\top}\boldsymbol{\Pi}(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-1}\Big)} \ ,\end{aligned} \tag{27}$$

where the first equality is due to $\beta_K(\boldsymbol{V}\boldsymbol{A}\boldsymbol{V}^\top) = \beta_K(\boldsymbol{A})$ for any $\boldsymbol{V} \in \mathbb{R}^{N\times K}$ with orthogonal columns. Moreover, observe that $\boldsymbol{\Pi}$ has the following decomposition:

$$\begin{aligned}\boldsymbol{\Pi} &= (\mathcal{H}(\boldsymbol{S})\boldsymbol{B}\boldsymbol{Q}_K)^\top\mathcal{H}(\boldsymbol{S})\boldsymbol{B}\boldsymbol{Q}_K \\ &= (\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\boldsymbol{B}\boldsymbol{Q}_K)^\top(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\boldsymbol{B}\boldsymbol{Q}_K) \\ &\qquad + \boldsymbol{Q}_K^\top\boldsymbol{B}^\top\boldsymbol{V}_{N-K}\mathrm{diag}(\tilde{\boldsymbol{h}}_{N-K})^2\boldsymbol{V}_{N-K}^\top\boldsymbol{B}\boldsymbol{Q}_K \ .\end{aligned} \tag{28}$$

This yields:

$$\begin{aligned}&\beta_K\Big([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}}\boldsymbol{\Pi}^{-1}([\mathcal{H}(\boldsymbol{S})]_K\tilde{\boldsymbol{B}})^\top\Big) \\ &= \Big(1 + \beta_1\Big((\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-\top}\tilde{\boldsymbol{B}}^\top\boldsymbol{V}_{N-K} \\ &\qquad \mathrm{diag}(\tilde{\boldsymbol{h}}_{N-K})^2\boldsymbol{V}_{N-K}^\top\tilde{\boldsymbol{B}}(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-1}\Big)\Big)^{-1} \\ &= \frac{1}{1 + \|\mathrm{diag}(\tilde{\boldsymbol{h}}_{N-K})\boldsymbol{V}_{N-K}^\top\tilde{\boldsymbol{B}}(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-1}\|_2^2} \\ &= \Big(1 + \gamma^2\Big)^{-1} \ ,\end{aligned} \tag{29}$$

where we have defined γ such that:

$$\begin{aligned}\gamma &:= \|\mathrm{diag}(\tilde{\boldsymbol{h}}_{N-K})\boldsymbol{V}_{N-K}^\top\tilde{\boldsymbol{B}}(\mathrm{diag}(\tilde{\boldsymbol{h}}_K)\boldsymbol{V}_K^\top\tilde{\boldsymbol{B}})^{-1}\|_2 \\ &\leq \left(\frac{\tilde{h}_{K+1}}{\tilde{h}_K}\right)\cdot\|\boldsymbol{V}_{N-K}^\top\boldsymbol{B}\boldsymbol{Q}_K\|_2\|(\boldsymbol{V}_K^\top\boldsymbol{B}\boldsymbol{Q}_K)^{-1}\|_2 \ ,\end{aligned} \tag{30}$$

as desired. This concludes the proof of our claim.

6.2. Proof of Proposition 2

Observe that the left hand side of (23) can be written as:

$$\|\boldsymbol{P}_K\boldsymbol{P}_K^\top - \hat{\boldsymbol{P}}_K\hat{\boldsymbol{P}}_K^\top\|_2 = \|\hat{\boldsymbol{P}}_{N-K}^\top\boldsymbol{P}_K\|_2 \ , \tag{31}$$

where the last equality is due to [**?**, Theorem 2.6.1].

Now, observe that Condition 5 of Theorem 1 implies that the largest eigenvalue in $\hat{\boldsymbol{\Sigma}}_{N-K}$ will never exceed $\beta_K(\boldsymbol{C}_x) - \delta$ since

$$\beta_{\mathsf{max}}(\hat{\boldsymbol{\Sigma}}_{N-K}) = \beta_{K+1}(\hat{\boldsymbol{C}}_x) \leq \beta_{K+1}(\boldsymbol{C}_x) + \beta_1(\boldsymbol{\Delta}_x) \ , \tag{32}$$

where the last inequality is due to the Weyl's inequality. The perturbed matrix $\hat{\boldsymbol{C}}_x$ thus satisfies the requirement of the the Davis-Kahan's $\sin(\Theta)$ theorem[3] [27], which gives:

$$\|\hat{\boldsymbol{P}}_{N-K}^\top\boldsymbol{P}_K\|_2 \leq \frac{\|\hat{\boldsymbol{P}}_{N-K}^\top\boldsymbol{\Delta}_x\boldsymbol{P}_K\|_2}{\delta} \ . \tag{33}$$

The last inequality in (23) is obtained by observing that both $\boldsymbol{P}_K$ and $\hat{\boldsymbol{P}}_{N-K}$ are orthogonal matrices.

[3] I have used the form given in `http://www.cs.columbia.edu/~djhsu/coms47`