Community Inference From Partially Observed Graph Signals: Algorithms and Analysis

Hoi-To Wai, *Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, Asuman E. Ozdaglar, *Fellow, IEEE*, and Anna Scaglione, *Fellow, IEEE*

***Abstract*—This paper considers community inference methods for finding communities on a graph. We treat the setting where the edges are not fully observed. Instead, inference is based on partially observed filtered graph signals where observations from some nodes are missing. Under this setup, we treat two related tasks: A) *blind* inference which recovers the inherited communities on the sub-graph; B) *semi-blind* inference which recovers communities on the full graph with additional partial topology information. For task A, we suggest a spectral method which analyzes the principal components of the data covariance matrix. We prove that it succeeds in finding the 'true' communities if the graph filter is low-pass and the nodes are uniformly sampled. For task B, we propose a method using spectral interpolation with a Nyström extension. The latter approach is proven to succeed in finding the 'true' communities for modular graphs and low-pass graph filters. Numerical experiments on synthetic and real data corroborate our results.**

***Index Terms*—Community detection, graph signal processing, low pass graph filter, Nyström extension.**

I. Introduction

AN OVERARCHING goal of data science is to infer information about complex systems from data. When dealing with network data where signals are observed on nodes (cf. graph signals), the underlying system can be described by a latent graph [2] such as the social graph of individuals embedded in opinion data, or the stock market graph of businesses embedded in daily return records. Among others, a problem of practical interest is to *infer or detect communities* of nodes from these graphs. Communities are subsets of nodes with dense connections. Learning them provides a macroscopic representation of the graph topology [3]. The community information is useful, for example, in designing marketing strategies to maximize sales of a product in social networks [4], or to classify nodes with similar functionalities in biological networks [5].

A popular heuristic for community inference is to apply off-the-shelf unsupervised learning methods, which usually involve a dimensionality reduction step such as principal component analysis (PCA) on the data correlation matrix, followed by a standard clustering algorithm such as K-means [6]. These methods are simple to implement and are shown to provide meaningful results on benchmark datasets [7]. To understand these techniques, a common assumption is that the data correlation matrix (or its nonlinear transformation such as in the t-SNE method [8]) acts as a *surrogate to the graph adjacency matrix*; the PCA and K-means steps produce an approximate minimum cut solution of the surrogate graph.

However, it is difficult, if not impossible, to ensure that the data correlation matrix is a faithful representation of the true graph without defining a proper data model, as the latter depends on the data generation process in relation to the graph. Another challenge is with partially observed network data, as there may be *hidden nodes* whose signals are missing in the data. For instance, in opinion data from social networks, some individuals may not participate in the social networking platform monitored, yet they can influence their peers on a different platform. As another example, in stock data from financial networks the unlisted companies may affect the performance of the listed companies. It is challenging to tackle community inference under this setting since the impact of the hidden nodes depends on the graph topology and the process which generates the data. Relevant questions are whether a community inference technique will deliver reliable outcomes in the presence of hidden nodes, and whether there is a principled way to design algorithms to infer communities.

This work addresses the above challenges by considering a graph signal processing (GSP) approach to derive community inference methods. Our idea is to develop a GSP model [9], [10] to treat the network data with hidden nodes as a collection of *spatially sampled and low-pass filtered graph signals* supported on the observed nodes. As demonstrated in [9], [10], GSP models provide a unified framework to analyze various network processes such as diffusion and network games. For community inference, we treat the 'true' graph topology as a latent parameter and our task entails inferring the communities therein.

Manuscript received February 24, 2021; revised September 8, 2021 and January 31, 2022; accepted March 1, 2022. Date of publication March 17, 2022; date of current version May 16, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael B Wakin. This work was supported in part by RGC Project under Grant 24203520, in part by CUHK Direct under Grant 4055135, and in part by ARO Project under Grant W911NF1810407. This work was presented at ICASSP 2019, Brighton, U.K. [DOI: 10.1109/ICASSP.2019.8683001]. *(Corresponding author: Hoi-To Wai.)*

Hoi-To Wai is with the Department of SEEM, The Chinese University of Hong Kong Shatin, Hong Kong (e-mail: htwai@se.cuhk.edu.hk).

Yonina C. Eldar is with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Asuman E. Ozdaglar is with LIDS, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: asuman@mit.edu).

Anna Scaglione is with the School of ECE, Cornell University and Cornell Tech, New York, NY 10044 USA (e-mail: as337@cornell.edu).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TSP.2022.3160277

With the GSP model in mind, a natural idea is to infer communities of the latent graph in two steps: first we estimate the graph topology from network data, and then we perform community detection on the estimated graph. The first step is also known as graph learning which has been extensively studied in the GSP literature [11], [12]. To list a few approaches, [13], [14] exploited smoothness in estimating the graph Laplacian matrix, [15] used the spectral template to estimate the graph shift operator, [16] considered a structural equation model to accommodate nonlinear dynamics, [17] considers causal modeling, and [18], [19] focused on inferring both the graph and the nonlinear dynamics generating the data. To estimate graphs with special structures like overlapping communities, the recent works [20], [21] considered learning with structural constraints. Likewise, graph learning with hidden nodes has been treated in [22]–[24] which studied optimization models for recovering the graph topology. Some recent works [20], [25] suggested graph learning with partial connectivity information incorporated as constraints in the optimization model, and [26]–[28] have studied graph learning with missing data via data imputation or kernel method. Besides treating different observation models, the above works on graph learning typically require a generation model with full-rank/white excitation to the graph filter in order to ensure the estimate is reliable. This limits their applications since a full-rank excitation to a large graph requires the excitation to be controlled by as many latent parameters as the number of nodes. For the second step, once the graph is estimated, standard community detection methods such as spectral clustering can be applied [29]. While each of the two steps are demonstrated to work, it is not clear how well the combined procedure performs.

The current paper is closely related to recent works on *blind community inference* which aim at inferring communities directly from graph signals without explicitly estimating the graph structure. Examples such as [30] analyzed the spectral properties of low-pass graph signals for community inference, [31] considered dynamical observations, [32] studied exact community inference based on the stochastic block model, [33] proposed a Bayesian estimation method; and [34] focused on gossip dynamics; also see [35], [36] which estimate node centrality. However, none of the above works consider hidden nodes in the observed graph signals.

In this paper, we fill the gap in the literature by studying community inference methods with partially observed, or equivalently, spatially sampled graph signals generated from exciting a low pass graph filter by a possibly low rank input. We contribute to the methodology and analysis for two tasks.

- *(Task A)* Using only partially observed graph signals, we perform blind community inference for finding communities within the observed network inherited from the complete graph. To this end, we study the application of a spectral method on the covariance matrix of the observed graph signals analogous to the one in [30]. Despite that a similar spectral method is applied, our analysis in Section III is new as it explicitly analyzes the effects on community inference performance due to the sampling pattern of the observed nodes and the low pass property of graph filters.

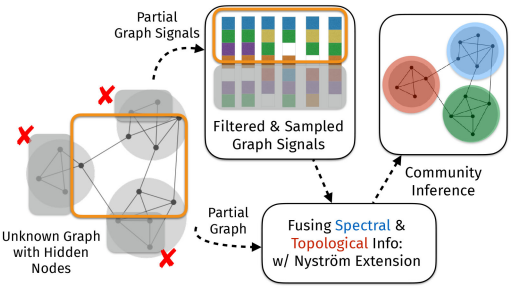

Fig. 1. Overview of the community inference tasks considered in this paper.

- *(Task B)* When additional information about the connectivity from hidden nodes to observable nodes is available, we consider a semi-blind method which finds the communities of the complete graph. Instead of interpolating the graph signals themselves, we suggest a novel application of the Nyström extension on the approximate eigenvectors. Our analysis in Section IV shows that the performance improves when more nodes are observed. In the special case when the graph signals are fully observed, the proposed method coincides with that of [30].

Our *Task A* is motivated by the practical scenario when a graph has a large number of nodes, which makes it impossible to observe the states on all of them. For example, we can only monitor the opinions of a subset of agents in a social network. *Task B* is motivated when crude estimates of the connectivity information between the observed nodes and the hidden nodes is available. For example, a firm may have access to the contacts of its employees that are external to the firm, but does not have further information about them; we may have partial information about the business ties between institutions on a stock market, but do not know the business (e.g., stock price) performance of all institutions.

An overview of the considered tasks is illustrated in Fig. 1. Compared to the conference version [1] whose results are mainly empirical, we provide performance analysis and include new numerical experiments. In the sequel, Section II introduces the graph and signal models. Section III discusses a strategy for tackling the blind community inference problem *(Task A)*; while Section IV treats the semi-blind inference problem *(Task B)*. Section V presents numerical experiments on synthetic and real data.

Throughout the paper, we use the following notation. Boldface (capital) letters are used to denote vectors (matrices), $\mathrm{Diag}(\boldsymbol{\lambda})$ is a diagonal matrix whose diagonal elements are taken from the vector $\boldsymbol{\lambda}$ in order. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ is its Euclidean norm. For a matrix $\boldsymbol{Z} \in \mathbb{R}^{p\times m}$, we denote $\boldsymbol{z}_i^{\mathrm{row}} \in \mathbb{R}^m$ as its ith row vector, while $\|\boldsymbol{Z}\|_{\mathrm{F}}$, $\|\boldsymbol{Z}\|_2$ are the Frobenius and spectral norm, respectively.

II. Signal and Graph Model

This section introduces the models of graphs, communities, and graph signal processing. We then pose the community inference problem with hidden nodes.

A. Graph and Communities

Consider an N-node undirected, connected graph described by $G = (V, E, \boldsymbol{A})$, where $V = \{1, \ldots, N\}$ is the node set, $E \subseteq V \times V$ is the edge set and $\boldsymbol{A} \in \mathbb{R}_+^{N \times N}$ is a weighted adjacency matrix which is non-negative, symmetric and satisfies $A_{ij} = 0$ if and only if $(i, j) \notin E$. Let $K \in \{1, \ldots, N\}$ be given and $\mathcal{C}_1, \ldots, \mathcal{C}_K$ be a non-overlapping, non-trivial partition of V, i.e., with $\mathcal{C}_i \neq \emptyset$. We consider the *normalized cut* measure:

$$\mathsf{NCut}(\mathcal{C}_1, \ldots, \mathcal{C}_K) = \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \sum_{j \notin \mathcal{C}_k} \frac{A_{ij}}{\sum_{\ell \in \mathcal{C}_k} \sum_{m \in V} A_{\ell m}}. \quad (1)$$

Consider a partition $\mathcal{C}_1, \ldots, \mathcal{C}_K$. If the induced subgraph $G[\mathcal{C}_k]$ is densely connected, for all k, and the number of edges between $\mathcal{C}_k, \mathcal{C}_{k'}$ is small, then $\mathsf{NCut}(\mathcal{C}_1, \ldots, \mathcal{C}_K) \approx 0$ and the partition corresponds to communities of the graph. In the latter case, we say that the graph is K-*modular* [3].

The community detection problem amounts to finding a non-trivial partition that minimizes the normalized cut (1). Let $\boldsymbol{D} := \mathrm{Diag}(\boldsymbol{A}\boldsymbol{1})$ and the normalized adjacency matrix be $\boldsymbol{A}_{\mathsf{norm}} := \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{D}^{-\frac{1}{2}}$. The normalized Laplacian is $\boldsymbol{L}_{\mathsf{sym}} := \boldsymbol{I} - \boldsymbol{A}_{\mathsf{norm}}$ with eigen decomposition $\boldsymbol{L}_{\mathsf{sym}} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^\top$, where $\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_N)$, $0 = \lambda_1 < \cdots < \lambda_N \leq 2$ and[1] $\boldsymbol{V}$ is orthonormal. It is well known [37] that the NCut measure is linked to the eigenvectors of the normalized Laplacian $\boldsymbol{L}_{\mathsf{sym}}$. Indeed, it can be shown that

$$\begin{aligned} &\min_{\mathcal{C}_1, \ldots, \mathcal{C}_K \subseteq V} \mathsf{NCut}(\mathcal{C}_1, \ldots, \mathcal{C}_K) \\ &\iff \min_{\boldsymbol{X} \in \mathbb{R}^{N \times K}} \mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{L}_{\mathsf{sym}} \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{X}) \\ &\qquad \text{s.t.} \boldsymbol{X}^\top \boldsymbol{D} \boldsymbol{X} = \boldsymbol{I}, X_{ij} \in \left\{0, \mathcal{W}_j^{-1/2}\right\}, \quad (2) \end{aligned}$$

where $\mathcal{W}_j := \sum_{\ell \in \mathcal{C}_j} D_{\ell\ell}$. Problem (2) is a combinatorial problem which is NP-hard in general. To obtain an approximate solution, we relax the combinatorial constraint $X_{ij} \in \{0, \mathcal{W}_j^{-1/2}\}$. The relaxed optimal solution is then $\boldsymbol{V}_K$, the left-most K column vectors of $\boldsymbol{V}$. Subsequently the communities can be estimated by applying a K-means subroutine on these eigenvectors, as described next.

For a given matrix $\boldsymbol{Z} \in \mathbb{R}^{p \times m}$, we define $\texttt{kmeans}(\boldsymbol{Z}, K)$ as the K-means subroutine which returns a non-overlapping K-partition of $[p] = \{1, \ldots, p\}$ satisfying

$$\texttt{kmeans}(\boldsymbol{Z}, K) \in \operatorname*{arg\,min}_{\mathcal{C}_1, \ldots, \mathcal{C}_K \subseteq [p]} F(\mathcal{C}_1, \ldots, \mathcal{C}_K; \boldsymbol{Z}), \quad (3)$$

where $F(\cdot; \boldsymbol{Z})$ is the K-means objective function:

$$F(\mathcal{C}_1, \ldots, \mathcal{C}_K; \boldsymbol{Z}) := \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \left\| \boldsymbol{z}_i^{\mathrm{row}} - \frac{1}{|\mathcal{C}_k|} \sum_{j \in \mathcal{C}_k} \boldsymbol{z}_j^{\mathrm{row}} \right\|^2. \quad (4)$$

In particular, the output of the spectral clustering method (without row normalization) [37] can be written as $\texttt{kmeans}(\boldsymbol{V}_K, K)$, which approximates the NCut minimization (2). The method requires perfect knowledge of $\boldsymbol{L}_{\mathsf{sym}}$.

In this paper, we define the 'true' communities on the graph according to this spectral clustering method, as follows:

$$(\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star) := \texttt{kmeans}(\boldsymbol{V}_K, K). \quad (5)$$

To justify our choice of 'true' community, consider the setting in which the graph is generated from the planted partition stochastic block model (SBM) with $K = 2$ communities given by $\mathcal{C}_1^{\mathsf{true}}, \mathcal{C}_2^{\mathsf{true}}$. The edge set of SBM is parameterized by

$$\begin{aligned} a &= \mathbb{P}((i,j) \in E, i, j \in \mathcal{C}_1^{\mathsf{true}} \text{or} i, j \in \mathcal{C}_2^{\mathsf{true}}), \\ b &= \mathbb{P}((i,j) \in E, i \in \mathcal{C}_1^{\mathsf{true}}, j \in \mathcal{C}_2^{\mathsf{true}}), \quad (6) \end{aligned}$$

and $\boldsymbol{A}$ is the binary adjacency matrix. When $a \gg b$, the nodes in $\mathcal{C}_1^{\mathsf{true}}$ or $\mathcal{C}_2^{\mathsf{true}}$ induce a densely connected subgraph. In particular, if a, b satisfy the spectral resolution criterion [29], [38], then $(\mathcal{C}_1^\star, \mathcal{C}_2^\star) = (\mathcal{C}_1^{\mathsf{true}}, \mathcal{C}_2^{\mathsf{true}})$ or $(\mathcal{C}_1^\star, \mathcal{C}_2^\star) = (\mathcal{C}_2^{\mathsf{true}}, \mathcal{C}_1^{\mathsf{true}})$ (with label ambiguity), i.e., applying the spectral clustering method on $\boldsymbol{V}_2$ gives the ground truth communities of G. We note that there are alternative definitions of 'true' communities, e.g., via eigenvectors of the unnormalized Laplacian matrix.

B. Graph Signals

We next discuss the signal model for the data that our community inference tasks depend on. We consider network data as the output *graph signals* of an unknown process on the graph G. For $\ell \in \{1, \ldots, m\}$, the ℓth graph signal, $\boldsymbol{y}^{(\ell)} \in \mathbb{R}^N$, is a *filtered graph signal* described as:

$$\boldsymbol{y}^{(\ell)} = \mathcal{H}(\boldsymbol{S}) \boldsymbol{x}^{(\ell)} + \boldsymbol{e}^{(\ell)}, \quad (7)$$

where $\boldsymbol{x}^{(\ell)} \in \mathbb{R}^N$ is a random and unknown excitation signal, and $\boldsymbol{e}^{(\ell)} \in \mathbb{R}^N$ is modeling noise. The random variables are independent and zero-mean with covariances $\mathbb{E}[\boldsymbol{x}^{(\ell)} (\boldsymbol{x}^{(\ell)})^\top] = \boldsymbol{C}_x$, $\mathbb{E}[\boldsymbol{e}^{(\ell)} (\boldsymbol{e}^{(\ell)})^\top] = \boldsymbol{C}_e$. The matrix $\boldsymbol{S} \in \mathbb{R}^{N \times N}$ is a graph shift operator (GSO) satisfying $S_{ij} = 0$ if and only if $(i, j) \notin E$ [9]. The linear *graph filter*

$$\mathcal{H}(\boldsymbol{S}) = \sum_{t=0}^{T} h_t \boldsymbol{S}^t \in \mathbb{R}^{N \times N} \quad (8)$$

is a Tth order polynomial of the GSO with coefficients $\{h_t\}_{t=0}^T$, where $T \in \mathbb{Z}_+ \cup \{\infty\}$. For convenience, we define the generating function as $h(\lambda) := \sum_{t=0}^{T} h_t \lambda^t$.

We consider a general model where the input covariance can be non-white and of low rank such that $R = \mathrm{rank}(\boldsymbol{C}_x) \leq N$. Without loss of generality, we concentrate on the model $\boldsymbol{C}_x = \boldsymbol{B}\boldsymbol{B}^\top$ where the columns of $\boldsymbol{B} \in \mathbb{R}^{N \times R}$, $R \leq N$, describe the subspace that the excitation signals $\boldsymbol{x}^{(\ell)}$ lies in. For example, when the excitation graph signals at the set of R nodes $\{i_1, \ldots, i_R\}$ are statistically independent, $\boldsymbol{B} = [\boldsymbol{e}_{i_1} \cdots \boldsymbol{e}_{i_R}]$ where $\boldsymbol{e}_{i_j}$ is the i_jth canonical basis vector. The graph filter $\mathcal{H}(\boldsymbol{S})$ is spectrally related to $\boldsymbol{L}_{\mathsf{sym}}$ and assumed to be low-pass [30], [39], [40]:

Assumption 1: We assume that the operator $\mathcal{H}(\boldsymbol{S})$ can be written as:

$$\mathcal{H}(\boldsymbol{S}) = \boldsymbol{V} \boldsymbol{\Sigma} \boldsymbol{U}^\top, \quad (9)$$

[1] We follow a descending order for the eigenvalues such that the left-most column vector of $\boldsymbol{V}$ corresponds to the smallest eigenvalue $\lambda_1 = 0$. For simplicity, we also assume that eigenvalues have multiplicity one.

where $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ *is a diagonal matrix of singular values ordered with respect to* $\boldsymbol{V}$*, whose left-to-right column vectors are the eigenvectors of the normalized Laplacian* $\boldsymbol{L}_{sym}$ *with increasing eigenvalues, and* $\boldsymbol{U} \in \mathbb{R}^{N\times N}$ *is a square matrix. Lastly, the diagonal elements of* $\boldsymbol{\Sigma}$ *satisfy*

$$\eta_K := \frac{|\max\{\Sigma_{(K+1),(K+1)}, \ldots, \Sigma_{N,N}\}|}{|\min\{\Sigma_{1,1}, \ldots, \Sigma_{K,K}\}|} < 1. \tag{10}$$

Note that if $\mathcal{H}(\boldsymbol{S})$ is a polynomial graph filter $\sum_{t=0}^{P-1} h_t \boldsymbol{S}^t$, then $\boldsymbol{U} = \boldsymbol{V}$ and (10) is equivalent to the definition of a (η_K, K)-low pass graph filter in [40].

As we will see, the proposed community inference algorithms return an accurate estimate of the communities in (5) when the parameter η_K is small. Examples of compatible GSO-graph filter pairs satisfying Assumption 1 are common in applications, see [40]. Two examples are given below:

Case 1. Normalized Laplacian Matrix . Suppose that the GSO is the normalized Laplacian matrix $\boldsymbol{S} = \boldsymbol{L}_{\mathsf{sym}}$. It is straightforward to verify that $\mathcal{H}(\boldsymbol{L}_{\mathsf{sym}}) = \boldsymbol{V} h(\boldsymbol{\Lambda}) \boldsymbol{V}^\top$, where $h(\cdot)$ is applied in an element-wise fashion. In this case, the graph filter satisfies (9), (10) when the polynomial $h(\lambda)$ is a *decreasing* function for $\lambda \in [0, 2]$.

Examples of graph filters satisfying Assumption 1 include the heat diffusion kernel where $\mathcal{H}(\boldsymbol{L}_{\mathsf{sym}}) = e^{-\sigma \boldsymbol{L}_{\mathsf{sym}}}$ for some $\sigma > 0$ or its discretized version with $\mathcal{H}(\boldsymbol{L}_{\mathsf{sym}}) = (\boldsymbol{I} - \sigma \boldsymbol{L}_{\mathsf{sym}})^P$ with $P \in \mathbb{Z}_+$, $0 < \sigma < 1$; see [41].

Case 2. Markov Matrix . Let the GSO be the Markov matrix $\boldsymbol{S} = \boldsymbol{A}_{\mathsf{markov}} = \boldsymbol{D}^{-1}\boldsymbol{A}$, which is not symmetric. Using [42, Proposition 1], we observe that $\boldsymbol{A}_{\mathsf{markov}} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{V} (\boldsymbol{I} - \boldsymbol{\Lambda}) \boldsymbol{V}^\top \boldsymbol{D}^{\frac{1}{2}}$. Using the fact that $(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{V})^{-1} = \boldsymbol{V}^\top \boldsymbol{D}^{\frac{1}{2}}$,

$$\boldsymbol{D}^{\frac{1}{2}} \mathcal{H}(\boldsymbol{A}_{\mathsf{markov}}) = \boldsymbol{V} h(\boldsymbol{I} - \boldsymbol{\Lambda}) \boldsymbol{V}^\top \boldsymbol{D}^{\frac{1}{2}}. \tag{11}$$

Therefore, the pre-multiplied graph filter $\boldsymbol{D}^{\frac{1}{2}} \mathcal{H}(\boldsymbol{A}_{\mathsf{markov}})$ also satisfies the requirements (9), (10) when the shifted polynomial $h(1-\lambda)$ is *decreasing* with respect to $\lambda \in [0, 2]$. Note that the rows of $\boldsymbol{A}_{\mathsf{markov}}$ sum to one and the matrix is non-negative. As such, applying $\boldsymbol{A}_{\mathsf{markov}}$ on a graph signal is equivalent to performing neighborhood mixing, where the shifted graph signal contains the weighted average of the values on neighboring nodes. The local averaging property is common for modeling linear opinion dynamics on social networks, where each node is taken as an individual and a graph shift models one step of opinion exchange.

Examples of graph filters satisfying Assumption 1 include $\mathcal{H}(\boldsymbol{A}_{\mathsf{markov}}) = \boldsymbol{A}^P_{\mathsf{markov}}$ for some $P \in \mathbb{Z}_+$, which model T rounds of gossiping averages; and the IIR graph filter $\mathcal{H}(\boldsymbol{A}_{\mathsf{markov}}) = (\boldsymbol{I} - \alpha \boldsymbol{A}_{\mathsf{markov}})^{-1}$, $\alpha \in (0, 1)$ which models the steady state of an opinion dynamics [43].

C. Hidden Nodes and Community Inference Problems

We consider situations where the signals on a subset of nodes are not observed during the data collection process (7). To describe the setup, consider an observation model with hidden nodes as depicted in Fig. 2. We partition V into $V_{\mathsf{obs}}, V_{\mathsf{hid}}$ such that V_{obs} is the set of n observable nodes, and V_{hid} is the set of hidden nodes with $V_{\mathsf{hid}} = V \setminus V_{\mathsf{obs}}$. We set $V_{\mathsf{obs}} = \{1, \ldots, n\}$, $V_{\mathsf{hid}} = \{n+1, \ldots, N\}$ such that the adjacency and Laplacian matrices can be described respectively as:

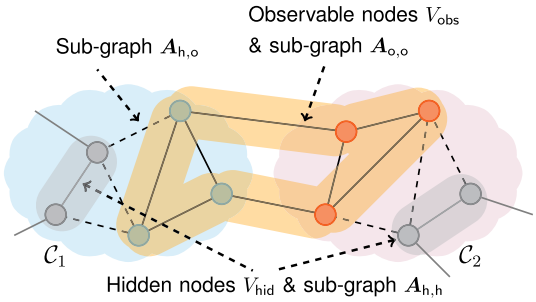

Fig. 2. Notations used for the hidden node setting. Only the graph signal values on the yellow shaded nodes V_{obs} are observed and used for community inference in *Task 1*; the sub-matrix $\boldsymbol{A}_{\mathsf{h,o}}$ is used in *Task 2*, in addition to the graph signals observed on V_{obs}. For the above graph, the colored nodes are the nodes we seek to cluster in *Task 1*.

$$\boldsymbol{A}^{\mathsf{norm}} = \begin{pmatrix} \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{o,o}} & \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{o,h}} \\ \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,o}} & \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,h}} \end{pmatrix}, \boldsymbol{L}_{\mathsf{sym}} = \begin{pmatrix} \boldsymbol{L}_{\mathsf{o,o}} & \boldsymbol{L}_{\mathsf{o,h}} \\ \boldsymbol{L}_{\mathsf{h,o}} & \boldsymbol{L}_{\mathsf{h,h}} \end{pmatrix}. \tag{12}$$

Here, $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{o,o}}$ (*resp.* $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,h}}$) represents the edges among the nodes in V_{obs} (*resp.* V_{hid}) and $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,o}}$ represents the edges between V_{obs} and V_{hid}. The observed signal corresponds to a *spatially sampled* version of $\boldsymbol{y}^{(\ell)}$ with entries in V_{obs}, i.e.,

$$\boldsymbol{y}^{(\ell)}_{\mathsf{obs}} = \big(\boldsymbol{I}_{n\times n} \boldsymbol{0}_{n\times(N-n)}\big) \boldsymbol{y}^{(\ell)}. \tag{13}$$

For convenience, let us denote $\boldsymbol{E}_{\mathsf{obs}} := (\boldsymbol{I}_{n\times n} \boldsymbol{0}_{n\times(N-n)})$. We note that the prior work [27] considered graph learning with a slightly different model with missing data where $V^{(\ell)}_{\mathsf{obs}}$ varies from sample to sample.

Community inference aims to infer the 'true' community $\mathcal{C}^\star_1, \ldots, \mathcal{C}^\star_K$ defined in (5) based on the spatially sampled graph signals $\{\boldsymbol{y}^{(\ell)}_{\mathsf{obs}}\}_{\ell=1}^m$. Consider the following two tasks:

- *Task A: Blind Community Inference* — As we do not know if the nodes in V_{hid} exist or not, we propose to only partition the nodes in V_{obs}. In this task, based on $\{\boldsymbol{y}^{(\ell)}_{\mathsf{obs}}\}_{\ell=1}^m$, we aim to find the communities given by

$$V_{\mathsf{obs}} \cap \mathcal{C}^\star_k, k = 1, \ldots, K. \tag{14}$$

  This corresponds to the communities of V_{obs} *inherited* from the complete graph. Obtaining (14) contributes to providing a macroscopic view of the complete graph.
- *Task B: Semi-blind Community Inference* — We consider the setting when, in addition, the sub-graph between V_{obs} and V_{hid}, represented by the adjacency sub-matrix $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,o}}$, is known. In this case, we have to combine two types of information: the graph signals on V_{obs}, and the connectivity information between $V_{\mathsf{hid}}, V_{\mathsf{obs}}$. Naturally, our aim is to recover the 'true' communities $\mathcal{C}^\star_1, \ldots, \mathcal{C}^\star_K$. Notice that $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,h}}$ is not required in this task.

In *Task B*, $\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h,o}}$ contains the partial graph topology whose exact form may be difficult to obtain. In practice, an estimate of the graph topology may be used for this task. For example, side information can be provided by an external source who estimates the sub-graph's topology.

III. Blind Community Inference

In this section, we focus on *Task A* whose aim is to retrieve the partition, $V_{\mathsf{obs}} \cap \mathcal{C}_k^\star$, $k \in \{1, \ldots, K\}$, from the partially observed graph signals $\{\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}\}_{\ell=1}^m$.

A. Blind Inference Method

We recall from (5) that $\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star$ is defined through the K-means subroutine $\texttt{kmeans}(\boldsymbol{V}_K, K)$. Ideally, *Task A* which aims at finding (14) is solved if $\boldsymbol{V}_K$ is available. However, the latter would be impossible since we only have the partially observed graph signals.

As a remedy, we examine what spectral information can be extracted from the observed partial graph signals by studying the covariance of $\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}$. Using (7), (9), (13), the covariance matrix can be evaluated as:

$$\boldsymbol{C}_{\mathsf{obs}} = \mathbb{E}[\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}(\boldsymbol{y}_{\mathsf{obs}}^{(\ell)})^\top] = \boldsymbol{V}_{\mathsf{o}}\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{B}\boldsymbol{B}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}_{\mathsf{o}}^\top + \boldsymbol{C}_e, \tag{15}$$

where $\boldsymbol{V}_{\mathsf{o}} := \boldsymbol{E}_{\mathsf{obs}}\boldsymbol{V}$ and we noted that $\boldsymbol{C}_x = \boldsymbol{B}\boldsymbol{B}^\top$. Also, we denote $\overline{\boldsymbol{C}}_{\mathsf{obs}} := \boldsymbol{C}_{\mathsf{obs}} - \boldsymbol{C}_e$ as the noiseless part of the covariance matrix. Under Assumption 1 that the graph filter is low-pass with $\eta_K \ll 1$ [cf. (10)], the diagonal matrix $\boldsymbol{\Sigma}$ is dominated by its top-K principal submatrix. This observation supports following approximation for the first term in (15):[2]

$$\boldsymbol{\Sigma}\boldsymbol{U}^\top\boldsymbol{B}\boldsymbol{B}^\top\boldsymbol{U}\boldsymbol{\Sigma} \approx \begin{pmatrix} \boldsymbol{C}_K & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \tag{16}$$

where $\boldsymbol{C}_K \in \mathbb{R}^{K\times K}$ is some positive semidefinite (PSD) matrix. Using the block matrix structure in (16), one obtains

$$\overline{\boldsymbol{C}}_{\mathsf{obs}} \approx \boldsymbol{V}_{\mathsf{o,K}}\boldsymbol{C}_K\boldsymbol{V}_{\mathsf{o,K}}^\top, \tag{17}$$

such that $\boldsymbol{V}_{\mathsf{o,K}}$ is the leftmost K column vectors of $\boldsymbol{V}_{\mathsf{o}}$ if $\eta_K \ll 1$. The above shows that the largest K eigenvectors of $\overline{\boldsymbol{C}}_{\mathsf{obs}}$ roughly span the same subspace as that of $\boldsymbol{V}_{\mathsf{o,K}}\boldsymbol{V}_{\mathsf{o,K}}^\top$.

Now, suppose that we have an estimate of $\boldsymbol{V}_{\mathsf{o,K}}$, e.g., by Assumption 1, and we aim at inferring the communities $V_{\mathsf{obs}} \cap \mathcal{C}_k^\star$ [cf. (14)] by clustering the rows of $\boldsymbol{V}_{\mathsf{o,K}}$ using K-means. Despite the close relationship between $\boldsymbol{V}_{\mathsf{o,K}}$ and $\boldsymbol{V}_K$, it remains unclear whether the procedure $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$ recovers (14) as the latter is not a minimizer of any K-means like objective function in the form (4), unless further assumptions are made on V_{obs}. We next investigate the geometry of the rows of $\boldsymbol{V}_{\mathsf{o,K}}$ to derive a set of conditions for tackling *Task A*.

First, observe that in order to treat *Task A*, it would be necessary for each community to have *at least one representation in* V_{obs}, i.e., $|V_{\mathsf{obs}} \cap \mathcal{C}_k^\star| \geq 1$, $k = 1, \ldots, K$. Otherwise, it is impossible to find all the K communities from V_{obs}. In fact, V_{obs} should contain a sizable representation from each community $\mathcal{C}_k^\star$. This observation is supported by Fig. 3 which shows the scatter plots of the rows of $\boldsymbol{V}_K$ generated from two SBM graphs and we highlighted the uniform samples taken from them. For

[2]For the approximation to hold, we require a mild condition on $\boldsymbol{B}$ such that $\min_{i=1,\ldots,K} \|\boldsymbol{B}^\top\boldsymbol{u}_i\| \gtrsim \max_{j=K+1,\ldots,N} \|\boldsymbol{B}^\top\boldsymbol{u}_j\|$ and $\boldsymbol{u}_i$ is the ith column of $\boldsymbol{U}$. For the special case when $\boldsymbol{U} = \boldsymbol{V}$, this condition states that the energy of $\boldsymbol{B}\boldsymbol{B}^\top$ in the low frequency component is at least comparable to those in the high frequency components.

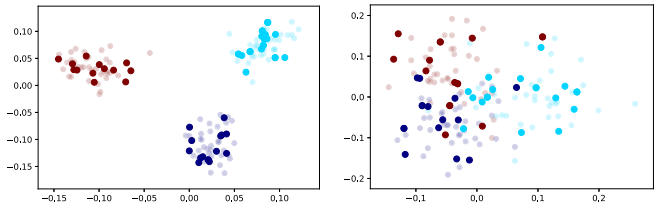

Fig. 3. Scatter plots of the rows of the top-3 eigenvectors $\{\boldsymbol{v}_i^{\mathrm{row}}\}_{i=1}^N$ of normalized Laplacian matrix when G is generated from an SBM with $N = 150$ nodes and $K = 3$ clusters and: (Left) $a = 8\log N/N$, $b = \log N/N$. (Right) $a = 8\log n/n$, $b = 4\log n/n$, where a (*resp.* b) is the intra (*resp.* inter)-cluster connectivity. The highlighted points correspond to $n = 50$ sampled rows of $\boldsymbol{v}_i^{\mathrm{row}}$.

modular graphs, e.g., SBMs with $a \gg b$ as seen in the left plot in Fig. 3 [cf. (6)], clustering the sampled points leads to the desired partition (14) since the points are well separated. On the other hand, for non-modular graphs, e.g., with $a \approx b$ for SBMs as seen in the right plot in Fig. 3, clustering the sampled points may not lead to (14).

Fig. 3 indicates that the desired partition (14) may be found by applying $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$ when the graph is K-modular [c.f. when $\mathsf{NCut}(\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star) \approx 0$] and a sufficient number of nodes are sampled from each community. We capture these effects quantitatively by comparing the clusters of rows in $\boldsymbol{V}_{\mathsf{o,K}}$ grouped by $V_{\mathsf{obs}} \cap \mathcal{C}_1^\star, \ldots, V_{\mathsf{obs}} \cap \mathcal{C}_K^\star$ to those induced by $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$. Define

$$\boldsymbol{v}_k^\star := \frac{1}{|\mathcal{C}_k^\star|}\sum_{j\in\mathcal{C}_k^\star} \boldsymbol{v}_j^{\mathrm{row}} \tag{18}$$

as the kth 'original' centroid vector computed from the desired partition. Our analysis depends explicitly on the following constants pertaining to $\boldsymbol{V}_{\mathsf{o,K}}$, V_{obs}:

Definition 1: Let $(\mathcal{C}_{obs,1}^\star, \ldots, \mathcal{C}_{obs,K}^\star) = \texttt{kmeans}(\boldsymbol{V}_{o,K}, K)$ *and*[3]

$$\delta_{\mathsf{ce}} := \max_{k\in[K]} \left\|\boldsymbol{v}_k^\star - \boldsymbol{v}_{\mathsf{obs,k}}^\star\right\|, \tag{19a}$$

$$R_1 := \max_{k\in[K]} \max_{i\in\mathcal{C}_{\mathsf{obs,k}}^\star} \|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_{\mathsf{obs,k}}^\star\|, \tag{19b}$$

$$R_2 := \max_{k\in[K]} \max_{i\in\mathcal{C}_k^\star\cap V_{\mathsf{obs}}} \|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_k^\star\|, \tag{19c}$$

where we defined the sampled centroid vectors $\boldsymbol{v}_{obs,k}^\star := \frac{1}{|\mathcal{C}_{obs,k}^\star|}\sum_{j\in\mathcal{C}_{obs,k}^\star} \boldsymbol{v}_j^{\mathrm{row}}$ *for each* k.

The constant δ_{ce} measures the drift in the sampled centroid vectors, and R_1, R_2 are the radius of the clusters in the K-dimensional space. These constants depend on the size, the modularity of the graph[4], and the sampling pattern of nodes. To gain more intuition for these variables, consider an SBM graph with equal sized communities [cf. (6)] and the same number of nodes are sampled from each community. If $\boldsymbol{v}_i$ is taken as an eigenvector of a population version of $\boldsymbol{L}_{\mathsf{sym}}$, we have $\delta_{\mathsf{ce}} = 0$; see [38, Definition 2] and [32, Proposition 3]. Now, applying [38, Theorem 2.1] shows that the eigenvectors of

[3]To simplify notations, we assume that the partition $(\mathcal{C}_{\mathsf{obs},1}^\star, \ldots, \mathcal{C}_{\mathsf{obs},\mathsf{K}}^\star)$ has been permuted to match with $(\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star)$ to give the smallest δ_{ce}.

[4]That is, when the normalized cut defined in (2) is small.

Algorithm 1: Blind Community Inference.

1: **Input**: set of partially observed graph signals $\{\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}\}_{\ell=1}^m$, number of desirable clusters K.
2: Compute the sampled covariance:
$$\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m = (1/m)\sum_{\ell=1}^m \boldsymbol{y}_{\mathsf{obs}}^{(\ell)}(\boldsymbol{y}_{\mathsf{obs}}^{(\ell)})^\top.$$
3: Find the largest K eigenvectors of $\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m$, $\widehat{\boldsymbol{Q}}_K \in \mathbb{R}^{n\times K}$.
4: Set $(\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}}) = \texttt{kmeans}(\widehat{\boldsymbol{Q}}_K, K)$.
5: **Output**: estimated partition $\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}}$

$\boldsymbol{L}_{\mathsf{sym}}$ and its population version coincide as $N\to\infty$, implying that $\delta_{\mathsf{ce}}\to 0$. On the other hand, when the nodes are sampled unevenly, intuitively, we observe from Fig. 3 that these constants can be large.

By clustering the rows of $\boldsymbol{V}_{\mathsf{o},\mathsf{K}}$ through $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o},\mathsf{K}}, K)$, we can tackle *Task A* if $\delta_{\mathsf{ce}}, R_1, R_2 \ll 1$. To see this, we define the surrogate K-means objective function:

$$\widetilde{F}_{\mathsf{o}}(\mathcal{C}_{\mathsf{obs},1},\ldots,\mathcal{C}_{\mathsf{obs},\mathsf{K}}) := \sum_{k=1}^K \sum_{i\in\mathcal{C}_{\mathsf{obs},\mathsf{k}}} \left\|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_k^\star\right\|^2 \quad (20)$$

such that $\mathcal{C}_{\mathsf{obs},1},\ldots,\mathcal{C}_{\mathsf{obs},\mathsf{K}}$ is any partition of V_{obs} and $\boldsymbol{v}_k^\star$ was defined in (18). Note that unlike $F(\cdot;\boldsymbol{V}_{\mathsf{o},\mathsf{K}})$ [cf. (4)], the centroid vector for each community is fixed at $\boldsymbol{v}_k^\star$.

Instead of (4), we use $\widetilde{F}_{\mathsf{o}}(\cdot)$ to compare the partition found by $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o},\mathsf{K}}, K)$ to the desired one, $V_{\mathsf{obs}}\cap\mathcal{C}_k^\star$, $k=1,\ldots,K$. If the two partitions share a similar function value, then the partitions formed shall be close to each other. The following lemma confirms that $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o},\mathsf{K}}, K)$ outputs a favorable solution when $\delta_{\mathsf{ce}}, R_1, R_2 \ll 1$:

Lemma 1: Let $(\mathcal{C}_{obs,1}^\star,\ldots,\mathcal{C}_{obs,K}^\star) = \texttt{kmeans}(\boldsymbol{V}_{o,K}, K)$. *Then,*

$$\begin{aligned}&|\widetilde{F}_{\mathsf{o}}(V_{\mathsf{obs}}\cap\mathcal{C}_1^\star,\ldots,V_{\mathsf{obs}}\cap\mathcal{C}_K^\star) - \widetilde{F}_{\mathsf{o}}(\mathcal{C}_{\mathsf{obs},1}^\star,\ldots,\mathcal{C}_{\mathsf{obs},\mathsf{K}}^\star)|\\ &\le 2N\left(R_1+R_2+\delta_{\mathsf{ce}}\right)\delta_{\mathsf{ce}}. \end{aligned}\quad (21)$$

The proof can be found in Appendix A.

Combining the observations in (17) and Lemma 1 motivates a spectral method to treat *Task A*, as summarized in Algorithm 1. The proposed method employs K-means on the rows of $\widehat{\boldsymbol{Q}}_K$ and is akin to the unsupervised learning heuristic via combining PCA and K-means, e.g., [6]. The main difference is that Algorithm 1 is not a heuristic as the latter is derived from the low pass property of the GSP model. We will show in the next section that the algorithm is able to detect the true communities. For the computation complexity, note that forming the sampled covariance matrix requires $\mathcal{O}(n^2m)$ floating point operations (flops), and the top-K eigenvectors are found in $\mathcal{O}(n^2K)$ flops. The K-means clustering step is performed in flops $\mathcal{O}(2^{(K/\epsilon)^{\mathcal{O}(1)}}Kn)$ using [44], producing a $(1+\epsilon)$ optimal solution to K-means, i.e., it finds a partition $\overline{\mathcal{C}}^1,\ldots,\overline{\mathcal{C}}^K$ with $F(\overline{\mathcal{C}}^1,\ldots,\overline{\mathcal{C}}^K;\widehat{\boldsymbol{Q}}_K)\le(1+\epsilon)\min_{\mathcal{C}^1,\ldots,\mathcal{C}^K}F(\mathcal{C}^1,\ldots,\mathcal{C}^K;\widehat{\boldsymbol{Q}}_K)$ [44]. The overall complexity is thus $\mathcal{O}(n^2(K+m))$ flops.

B. Performance Analysis

This section analyzes the performance of Algorithm 1 conditioned on the inputs $\{\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}\}_{\ell=1}^m$ generated by (13). Define the economy QR decomposition of $\boldsymbol{V}_{\mathsf{o},\mathsf{K}}$ as:

$$\boldsymbol{V}_{\mathsf{o},\mathsf{K}} = \boldsymbol{Q}_K\boldsymbol{R}_K, \quad (22)$$

such that $\boldsymbol{Q}_K\in\mathbb{R}^{n\times K}$ is an orthogonal matrix which spans the range space of $\boldsymbol{V}_{\mathsf{o},\mathsf{K}}$ and $\boldsymbol{R}_K$ is a upper triangular matrix. Define the following constant to be used in our analysis:

$$\rho_{\mathsf{gap}} := \lambda_{n-K-1}(\overline{\boldsymbol{C}}_{\mathsf{obs}}) - \lambda_{n-K}(\overline{\boldsymbol{C}}_{\mathsf{obs}}) - \|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2, \quad (23)$$

where $\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m$ is the sampled covariance matrix (cf. Algorithm 1), and $\lambda_i(\boldsymbol{X})$ denotes the ith smallest eigenvalue of a square, symmetric matrix $\boldsymbol{X}$. We consider the following assumption:

Assumption 2: We assume that the operator $\mathcal{H}(\boldsymbol{S})$ *satisfies*

1) $\mathrm{rank}(\boldsymbol{U}_K^\top\boldsymbol{C}_x) = K$, $\boldsymbol{U}_K$ *is the* K *leftmost-columns of* $\boldsymbol{U}$, *defined in Assumption 1.*
2) *the constant* ρ_{gap}, *defined in (23), is strictly positive.*

Condition 1) requires the excitation signal's covariance to be at least rank K and that it does not lie in the null space of $\boldsymbol{U}_K^\top$. For 2), we observe that typically $\rho_{\mathsf{gap}}>0$ as the number of samples m grows and the noise is small, i.e., $\|\boldsymbol{C}_e\|_2 \ll 1$.

Our first analytical result is a bound on the difference between the range spaces of $\widehat{\boldsymbol{Q}}_K$ and $\boldsymbol{Q}_K$:

Proposition 1: Suppose Assumption 1, 2 hold. Then

$$\begin{aligned}&\|\widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top - \boldsymbol{Q}_K\boldsymbol{Q}_K^\top\|_{\mathrm{F}} \le \sqrt{2K}\\ &\times\left(\frac{\sqrt{2}\gamma(2\|\boldsymbol{U}_K^\top\boldsymbol{B}\|_2 + \|\boldsymbol{U}_{N-K}^\top\boldsymbol{B}\|_2)}{\lambda_K(\boldsymbol{B}^\top\boldsymbol{U}_K\boldsymbol{V}_{\mathsf{o},K}^\top\boldsymbol{V}_{\mathsf{o},K}\boldsymbol{U}_K^\top\boldsymbol{B})}\,\eta_K + \frac{\|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2}{\rho_{\mathsf{gap}}}\right),\end{aligned}\quad (24)$$

where we have defined $\gamma = \frac{\max\{\Sigma_{1,1},\ldots,\Sigma_{K,K}\}}{\min\{\Sigma_{1,1},\ldots,\Sigma_{K,K}\}}$, *and* $\widehat{\boldsymbol{Q}}_K$ *is computed in line 3 of Algorithm 1.*

The proof is relegated to Appendix B. The above proposition gives a quantitative account for the discussions in (15), (16). In particular, the eigenvectors $\widehat{\boldsymbol{Q}}_K$ approximate $\boldsymbol{Q}_K = \boldsymbol{E}_{\mathsf{obs}}\boldsymbol{V}_K\boldsymbol{R}_K^{-1}$, which spans the same subspace as the one spanned by the row-sampled version of $\boldsymbol{V}_K$. The approximation quality improves as the lowpass ratio η_K decreases and the number of samples m increases. Next, we benchmark the output $\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}}$ of Algorithm 1 via the K-means objective function $F(\mathcal{C}_1,\ldots,\mathcal{C}_K;\boldsymbol{V}_{\mathsf{o},\mathsf{K}})$ [cf. (4)].

Theorem 1: Assuming that the K*-means in line 4 of Algorithm 1 outputs an* $(1+\epsilon)$ *optimal solution. Then*

$$\begin{aligned}&\sqrt{F(\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}};\boldsymbol{V}_{\mathsf{o},\mathsf{K}})} - \sqrt{(1+\epsilon)F_{\mathsf{o}}^\star}\\ &\qquad\le \|\boldsymbol{R}_K\|_2(2+\epsilon)\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}.\end{aligned}\quad (25)$$

where $F_{\mathsf{o}}^\star = \min F(\mathcal{C}_{obs,1},\ldots,\mathcal{C}_{obs,K};\boldsymbol{V}_{o,K})$, *i.e., the optimal objective value attained by* $\texttt{kmeans}(\boldsymbol{V}_{o,K}, K)$.

The proof is provided in Appendix C. Notice that $\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}$ may be upper bounded by (24).

Combining Proposition 1, Theorem 1 provides a bound on the difference between the output by Algorithm 1 and the surrogate

solution $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$. This bound depends on two factors as seen in the right hand side of (24): the first factor is bounded as $\mathcal{O}(\eta_K)$ which improves if the graph filter is sufficiently low-pass; it is known that the second factor is bounded with probability at least $1-\delta$ (with respect to the randomness in the generation of $\{\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}\}_{\ell=1}^m$) [45, Ch. 4] as

$$\|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2 = \mathcal{O}(\sqrt{\log(1/\delta)/m} + \|\boldsymbol{C}_e\|_2). \tag{26}$$

This bound decreases with the number of observed graph signals m and depends on noise variance $\|\boldsymbol{C}_e\|_2$. To summarize, with probability at least $1-\delta$, it holds that

$$\begin{aligned}&\sqrt{F(\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}};\boldsymbol{V}_{\mathsf{o,K}})} - \sqrt{(1+\epsilon)F_{\mathsf{o}}^{\star}}\\ &= \mathcal{O}\left(\eta_K + \rho_{\mathsf{gap}}^{-1}\left(\sqrt{\log(1/\delta)/m} + \|\boldsymbol{C}_e\|_2\right)\right).\end{aligned} \tag{27}$$

The big-$\mathcal{O}$ notation suppresses the factors $\|\boldsymbol{U}_K^\top \boldsymbol{B}\|_2$, $\|\boldsymbol{U}_{N-K}^\top \boldsymbol{B}\|_2$, $\lambda_{N-K-1}(\boldsymbol{B}^\top \boldsymbol{U}_K \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \boldsymbol{U}_K^\top \boldsymbol{B})$ which describe the effects of the correlation between $\boldsymbol{U}_K$ and the excitation covariance. In the case of $\boldsymbol{U}_K = \boldsymbol{V}_K$, i.e., the GSO is symmetric, the excitation graph signals at each community have to be independent for these factors to be small. For example, this would happen if there is an independent influencer at each community. Naturally, if the excitation signals only span the orthogonal subspace $\boldsymbol{U}_{N-K}$, these factors can be large.

Our results show that Algorithm 1 provides an approximate solution to $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$ when $m \gg 1$, $\eta_K \ll 1$. The derived bounds are comparable to [30], particularly they match each other (up to a constant factor that depends on $\|\boldsymbol{R}_K\|_2$, $\|\boldsymbol{U}_K^\top \boldsymbol{B}\|_2$) if all nodes are observed. To tackle *Task A*, we further recall from Lemma 1 that $\texttt{kmeans}(\boldsymbol{V}_{\mathsf{o,K}}, K)$ approximates the desired solution if $\delta_{\mathsf{ce}}, R_1, R_2 \ll 1$. Interestingly, the analysis reveals that the performance of Algorithm 1 is insensitive to the number of observed nodes, but instead it depends on the low-pass coefficient η_K and the sampling pattern of nodes. We verify these findings in Section V.

IV. Nyström Based Semi-Blind Community Inference

This section focuses on *Task B*. Our goal is to infer the partition, $\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star$, for the complete graph.

A. Semi-Blind Inference Method

As we recall from definition (5), the desired partition is found using the eigenvectors $\boldsymbol{V}_K$ of the normalized Laplacian. While $\boldsymbol{V}_K$ is not readily available, we utilize the spectral information obtained from the partial graph signals and then *interpolate* it using the partial topology information $\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}$. We begin by recalling from Proposition 1 that the eigenvectors $\widehat{\boldsymbol{Q}}_K$ computed in Algorithm 1 approximate a row-sampled version of $\boldsymbol{V}_K$, i.e., $\boldsymbol{Q}_K = \boldsymbol{E}_{\mathsf{obs}} \boldsymbol{V}_K \boldsymbol{R}_K^{-1}$. In this regard, the Nyström extension [46] is a natural method for *interpolating* $\widehat{\boldsymbol{Q}}_K$.

In a nutshell, the Nyström extension method treats the unknown positive semindefinite (PSD) matrix as a kernel, and interpolates its eigenvectors through approximately solving the eigenvector equations. To derive the Nyström extension method,

Algorithm 2: Semi-Blind Community Inference.

1: **Input**: partially observed graph signals $\{\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}\}_{\ell=1}^m$, number of desirable clusters K, partial network topology $\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}$.
2: Follow line 2–3 in Algorithm 1 to obtain $\widehat{\boldsymbol{Q}}_K$.
3: Compute (32) to obtain either $\widetilde{\boldsymbol{V}}_K^{\mathsf{nys}}$ (or $\widetilde{\boldsymbol{V}}_K^{\mathsf{imp}}$).
4: Perform economy QR decomposition to obtain $\widetilde{\boldsymbol{V}}_K = \widehat{\boldsymbol{V}}_K \widehat{\boldsymbol{\Delta}}_K$, where columns of $\widehat{\boldsymbol{V}}_K \in \mathbb{R}^{N\times K}$ are orthogonal and $\widehat{\boldsymbol{\Delta}}_K \in \mathbb{R}^{K\times K}$ is upper triangular.
5: Set $(\widehat{\mathcal{C}}_1, \ldots, \widehat{\mathcal{C}}_K) = \texttt{kmeans}(\widehat{\boldsymbol{V}}_K, K)$.
6: **Output**: estimated partition $\widehat{\mathcal{C}}_1, \ldots, \widehat{\mathcal{C}}_K$

we let $\mathrm{c}_L \in [\lambda_N, 2]$ be a user-designed constant and consider a flipped Laplacian matrix as

$$\widetilde{\boldsymbol{L}}_{\mathsf{sym}} = \mathrm{c}_L \boldsymbol{I} - \boldsymbol{L}_{\mathsf{sym}}. \tag{28}$$

The column vectors of $\boldsymbol{V}_K$ are the eigenvectors of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$ with the largest K eigenvalues; the eigenvalues are $\{\tilde{\lambda}_i\}_{i=1}^N = \{\mathrm{c}_L - \lambda_i\}_{i=1}^N$. Now, suppose we are given the row sampled version of $\boldsymbol{V}_K$, i.e., $\boldsymbol{V}_{\mathsf{o,K}} = \boldsymbol{E}_{\mathsf{obs}} \boldsymbol{V}_K$. Let $k \in \{1, \ldots, K\}$ and $\boldsymbol{v}_k$ be the kth column of $\boldsymbol{V}_K$. Then,

$$\widetilde{\boldsymbol{L}}_{\mathsf{sym}} \boldsymbol{v}_k = (\mathrm{c}_L - \lambda_k)\boldsymbol{v}_k. \tag{29}$$

By expanding (29), the Nyström extension estimates the elements in $\boldsymbol{v}_k$ for the hidden nodes, $i \in V_{\mathsf{hid}}$, by

$$\begin{aligned} v_{i,k} &= \frac{1}{\mathrm{c}_L - \lambda_k}\left(\sum_{j\in V_{\mathsf{obs}}} \left[\widetilde{\boldsymbol{L}}_{\mathsf{sym}}\right]_{ij} v_{j,k} + \sum_{j\in V_{\mathsf{hid}}} \left[\widetilde{\boldsymbol{L}}_{\mathsf{sym}}\right]_{ij} v_{j,k}\right)\\ &\approx \frac{1}{\mathrm{c}_L - \lambda_k}\left(\sum_{j\in V_{\mathsf{obs}}} [\boldsymbol{A}^{\mathsf{norm}}]_{ij}\, v_{j,k}\right),\end{aligned} \tag{30}$$

where the Nyström approximation is obtained by discarding the second summation. In the above, we have used the definition for $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$, $\boldsymbol{L}_{\mathsf{sym}}$, as we recognized that $[-\boldsymbol{L}_{\mathsf{sym}}]_{ij}$ with $i \in V_{\mathsf{hid}}, j \in V_{\mathsf{obs}}$ can be mapped to the submatrix $\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}$.

An alternative approximation can be obtained by exploiting the structure of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$. Specifically, take $\mathrm{c}_L = 2$ and observe that as $\widetilde{\boldsymbol{L}}_{\mathsf{sym}} = \boldsymbol{I} + \boldsymbol{A}^{\mathsf{norm}}$, we can retain the diagonal component of the second summation in (30) by the following approximation:

$$\begin{aligned} v_{i,k} &\approx \frac{1}{2-\lambda_k}\left(\sum_{j\in V_{\mathsf{obs}}} [\boldsymbol{A}^{\mathsf{norm}}]_{ij}\, v_{j,k} + v_{i,k}\right)\\ \implies v_{i,k} &\approx \frac{1}{1-\lambda_k}\sum_{j\in V_{\mathsf{obs}}} [\boldsymbol{A}^{\mathsf{norm}}]_{ij}\, v_{j,k}.\end{aligned} \tag{31}$$

We find that (30), (31) lead to two different Nyström extension models, where (31) enjoys better empirical performance when the number of observed nodes is small.

For *Task B*, since both the sub-sampled matrix $\boldsymbol{V}_{\mathsf{o,K}}$ and the eigenvalues λ_k are unknown, we replace $\boldsymbol{V}_{\mathsf{o,K}}$ by its surrogate $\widehat{\boldsymbol{Q}}_K$ used in *Task A*. Observe that $\lambda_k \approx 0$ for $1 \le k \le K$ when

the graph is K-modular. Applying $\lambda_k \approx 0$ to (30), (31) yields the following estimate of $\boldsymbol{V}_K$:

$$\widetilde{\boldsymbol{V}}_K^{\mathsf{nys}} := \begin{pmatrix} \widehat{\boldsymbol{Q}}_K \\ \frac{1}{\mathrm{c}_L} \boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}} \widehat{\boldsymbol{Q}}_K \end{pmatrix}, \widetilde{\boldsymbol{V}}_K^{\mathsf{imp}} := \begin{pmatrix} \widehat{\boldsymbol{Q}}_K \\ \boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}} \widehat{\boldsymbol{Q}}_K \end{pmatrix}. \quad (32)$$

Note that the two estimates differ by a factor of c_L in the lower block matrix. As we will demonstrate in Section V, Algorithm 2 using either one of the estimates of $\boldsymbol{V}_K$ achieves reasonable performance with $\widetilde{\boldsymbol{V}}_K^{\mathsf{imp}}$ offering slightly better performance. We summarize the proposed inference method for *Task B* in Algorithm 2 which performs K-means clustering on the rows of $\widetilde{\boldsymbol{V}}_K^{\mathsf{nys}}$ or $\widetilde{\boldsymbol{V}}_K^{\mathsf{imp}}$. The final clustering step is performed on the orthogonalized eigenmatrix $\widehat{\boldsymbol{V}}_K$ of $\widetilde{\boldsymbol{V}}_K$.

Related algorithms have been proposed in the data clustering literature [47]–[50], yet these works assume knowledge of the adjacency matrix between observable nodes, while in our settings, we do not observe the edges between observable nodes. The Nyström extension has also been applied to graph signal interpolation [42]. In fact, one may be tempted to first interpolate the partial graph signals, and then applying [30] to infer communities of the complete graph based on the interpolated graph signals. However, graph signal interpolation (such as [42]) also requires the Laplacian matrix between observable nodes, which is not available in our setting.

We comment on the computation complexity of Algorithm 2. As discussed in Section III it requires $\mathcal{O}(n^2(K+m))$ flops to obtain $\widehat{\boldsymbol{Q}}_K$. In addition, the Nyström step in (32) takes $\mathcal{O}(\mathrm{nnz}(\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}})K)$ flops where $\mathrm{nnz}(\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}})$ is the number of non-zeros in the matrix $\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}$. Finally, the K-means step requires $\mathcal{O}(2^{(K/\epsilon)^{\mathcal{O}(1)}}KN)$ flops. Overall, the algorithm requires flops $\mathcal{O}(N(\mathrm{nnz}(\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}})+K)+n^2(K+m))$ to run. In contrast, inferring partitions from fully observed graph signals involve a complexity of $\mathcal{O}(N^2K)$ flops (or through the formed similarity matrix). Algorithm 2 thus reduces the complexity of community inference when $n \ll N$ and partial topology information is available.

B. Performance Analysis

Algorithm 2 tackles the semi-blind inference *Task B* using estimates of the full eigenvectors $\widehat{\boldsymbol{V}}_K$. Unlike the analysis of Algorithm 1, we have to explicitly consider the effects of randomness in sampling nodes. For simplicity, we make the following assumption.

Assumption 3: The observable node set V_{obs} is sampled uniformly at random from $V = \{1, \ldots, N\}$.

In the following, we analyze the performance of Algorithm 2 under Assumption 3 while treating the *full graph signals* $\{\boldsymbol{y}_\ell\}_{\ell=1}^m$ given in (7) as fixed. We first define the following constants for facilitating our analysis: for any $\Delta > 0$,

$$\tilde{\delta}_{\mathsf{eig}}(\Delta) = \inf\left\{\hat{\delta}_{\mathsf{eig}} : \mathbb{P}(\max_{k\in[K]} |\lambda_k(\boldsymbol{L}_{\mathsf{o,o}})| \le \hat{\delta}_{\mathsf{eig}}) \ge 1-\Delta\right\},$$
$$\tilde{\rho}_{\mathsf{gap}}(\Delta) = \sup\{\hat{\rho}_{\mathsf{gap}} : \mathbb{P}(\rho_{\mathsf{gap}} \ge \hat{\rho}_{\mathsf{gap}}) \ge 1-\Delta\}, \quad (33)$$

where $\boldsymbol{L}_{\mathsf{o,o}}$, ρ_{gap} are defined in (12), (23), respectively, in which they are treated as random variables due to the selection of the observable node set V_{obs}.

Observe that $\tilde{\delta}_{\mathsf{eig}}(\Delta) \approx 0$ when G is K-modular and the number of observed nodes is large. To see this, denote $\boldsymbol{E}_{\mathsf{obs}} \in \{0,1\}^{n\times N}$ as the selection matrix induced by V_{obs}. Then

$$\begin{aligned}\boldsymbol{L}_{\mathsf{o,o}} &= \boldsymbol{I} - \boldsymbol{E}_{\mathsf{obs}}\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}\boldsymbol{E}_{\mathsf{obs}}^\top \\ &= \boldsymbol{I} - \boldsymbol{D}_{\mathsf{obs}}^{-1/2}\boldsymbol{E}_{\mathsf{obs}}\boldsymbol{A}\boldsymbol{E}_{\mathsf{obs}}^\top\boldsymbol{D}_{\mathsf{obs}}^{-1/2}, \end{aligned} \quad (34)$$

where $\boldsymbol{D}_{\mathsf{obs}}$ is the diagonal matrix of the degrees of the nodes in V_{obs}. Notice that $\boldsymbol{E}_{\mathsf{obs}}\boldsymbol{A}\boldsymbol{E}_{\mathsf{obs}}^\top$ is the adjacency matrix of the subgraph $G[V_{\mathsf{obs}}]$. Under Assumption 3 and assuming that G is K-modular, with high probability V_{obs} includes nodes from each cluster, forming a K-modular graph itself. Thus, the K largest eigenvalues of $\boldsymbol{D}_{\mathsf{obs}}^{-1/2}\boldsymbol{E}_{\mathsf{obs}}\boldsymbol{A}\boldsymbol{E}_{\mathsf{obs}}^\top\boldsymbol{D}_{\mathsf{obs}}^{-1/2}$ will be close to $\frac{n}{N}$ as n nodes are selected to form $G[V_{\mathsf{obs}}]$. This leads to the observation that the K smallest eigenvalues of $\boldsymbol{L}_{\mathsf{o,o}}$ are approximately $\frac{N-n}{N}$, i.e., $\tilde{\delta}_{\mathsf{eig}}(\Delta) \approx \frac{N-n}{N}$. Furthermore, $\tilde{\rho}_{\mathsf{gap}}(\Delta)$ is bounded away from zero when the number of observed graph signals, m, is moderate [cf. (26)]. We remark that these constants can be characterized when a specific (random) graph model is assumed.

The following theorem introduces a bound on the suboptimality of Algorithm 2:

Theorem 2: Let Assumption 1, 2, 3 hold and the K-means algorithm outputs an $(1+\epsilon)$ optimal solution. Fix any failure probability $\Delta > 0$. If $\tilde{\rho}_{gap}(\Delta) > 0$, $n \ge 8\mu K \log(3K/\Delta)$ where $\mu := \frac{N}{K}\max_{j=1,\ldots,N} \|\widetilde{\boldsymbol{V}}_{1:K,j}^{\mathrm{row}}\|$ such that $\widetilde{\boldsymbol{V}}_{1:K,j}^{\mathrm{row}}$ is the jth row of the matrix of first K eigenvectors of $\widetilde{\boldsymbol{L}}_{sym}$, then with probability at least $1-\Delta$, Algorithm 2 with $\widetilde{\boldsymbol{V}}_K^{nys}$ [cf. (32)] finds a partition of $V = \{1, \ldots, N\}$ satisfying

$$\begin{aligned}&\sqrt{F(\widehat{\mathcal{C}}_1, \ldots, \widehat{\mathcal{C}}_K; \boldsymbol{V}_K)} - \sqrt{(1+\epsilon)F^\star} \\ &= \frac{2+\epsilon}{\mathrm{c}_L - \lambda_K}\mathcal{O}\left(\tilde{\delta}_{\mathsf{eig}}(\Delta) + \sqrt{K}\left(\eta_K + \frac{\|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2}{\tilde{\rho}_{\mathsf{gap}}(\Delta)}\right)\right. \\ &\quad \left. + \left(1+\sqrt{\frac{2N}{n}}\right)(\mathrm{c}_L - \lambda_{K+1})\right), \end{aligned} \quad (35)$$

where $F^\star$ is the minimum objective value of $F(\cdot; \boldsymbol{V}_K)$ in (4), and λ_{K+1} is the $K+1$th smallest eigenvalue of $\boldsymbol{L}_{sym}$.

The proof can be found in Appendix D. Our result is achieved by decomposing the l.h.s. of (35) into the error of the Nyström extension for the eigenvectors of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$ and the error of approximating $\boldsymbol{Q}_K$ by $\widehat{\boldsymbol{Q}}_K$ in line 2 of Algorithm 2. In our analysis, we adopt [48], [49] to show that the low-rank approximation of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$ produced by the Nyström method admits an error of $\mathcal{O}(1/\sqrt{n})$, where n is the number of observed nodes, and we apply Proposition 1 to treat the error between $\boldsymbol{Q}_K, \widehat{\boldsymbol{Q}}_K$.

Similar to the analysis for Task A, we recall that $\|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2 = \mathcal{O}(\sqrt{\log(1/\delta)/m} + \|\boldsymbol{C}_e\|_2)$ with probability at least $1-\delta$ [45, Ch. 4]. As such, when we take $\mathrm{c}_L = 1$, (35) implies

that

$$\sqrt{F(\widehat{\mathcal{C}}_1,\ldots,\widehat{\mathcal{C}}_K;\boldsymbol{V}_K)}-\sqrt{(1+\epsilon)F^\star} = \mathcal{O}\left(\eta_K+\frac{\sqrt{\frac{\log(\delta^{-1})}{m}}+\|\boldsymbol{C}_e\|_2}{\tilde{\rho}_{\mathsf{gap}}(\Delta)}+\frac{N-n}{N}+\sqrt{\frac{1}{n}}\right) \quad (36)$$

with probability at least $1-\delta-\Delta$, where the randomness comes from the selection of n nodes in V_{obs} and the excitation/noise signals in (13). Note we have taken the approximations $\tilde{\delta}_{\mathsf{eig}}(\Delta)\approx\frac{N-n}{N}$, $\lambda_K\approx 0$ and $\lambda_{K+1}\approx 1$ for K-modular graphs. The latter follows from the intuition that $\boldsymbol{A}$ is approximately rank-K; see [51, Th. 10] for a precise bound for $K=2$.

The bound in (36) consists of two parts. The first two terms are small provided that the graph filter is low pass $\eta_K\ll 1$, the number of observed nodes is sufficiently large, and the number of graph signal samples m is large. The remaining terms decrease with the number of observed nodes n if we take $\mathrm{c}_L=1$. The requirement $n\geq 8\mu K\log(3K/\Delta)$ reveals that for Algorithm 2 to succeed, the number of observed nodes has to be at least a constant multiple of the number of existing communities. In conclusion, the performance of Algorithm 2 is sensitive to the number of observed nodes, as we will demonstrate in Section V.

V. Numerical Experiments

This section examines the efficiency of our proposed methods applied on synthetic and real data. We validate the theoretical results on generated graph signals and present an application on finding communities in the S&P100 stocks network through the daily returns of a subset of stocks.

A. Synthetic Data

We consider applying Algorithm 1 and 2 on tackling *Task A* and *Task B*, respectively. To generate the partial graph signals data, we adopt the data generation model from (7) where the graph related operator is given as the graph diffusion filter $\mathcal{H}(\boldsymbol{S})=(\boldsymbol{I}-0.5\boldsymbol{S})^P$ and P is the order of diffusion. The input to the graph filter is $\boldsymbol{x}^{(\ell)}=\boldsymbol{B}\boldsymbol{z}^{(\ell)}$ where $\boldsymbol{B}\in\mathbb{R}^{N\times R}$ is a random matrix with i.i.d. elements generated from $\mathcal{N}(0,1)$, and $\boldsymbol{z}^{(\ell)}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$. The GSO is selected as the normalized Laplacian matrix $\boldsymbol{S}=\boldsymbol{L}_{\mathsf{sym}}=\boldsymbol{I}-\boldsymbol{A}^{\mathsf{norm}}$. The graph is generated according to a planted partition stochastic block model denoted by $\mathsf{SBM}(N,K,a,b)$ – such that there are N nodes divided into K equal sized clusters, the intra-cluster (*resp.* inter-cluster) connectivity is a (*resp.* b). Throughout this section, we fix $N=150$ nodes, $K=3$ clusters in the randomly generated graphs, and the excitation rank is $R=15$. We observe $m=100$ samples from the output of the graph filter with an observation noise following $\boldsymbol{C}_e=10^{-4}\boldsymbol{I}$. For benchmarking purpose, we compare the *error rate* of the detected communities/partitions with the SBM's ground truth:

$$P_e:=\mathbb{E}\left[\tfrac{1}{N}\min_{\pi:[K]\to[K]}\textstyle\sum_{i=1}^{N}\mathbb{1}_{\pi(c_i)\neq c_i^{\mathsf{true}}}\right], \quad (37)$$

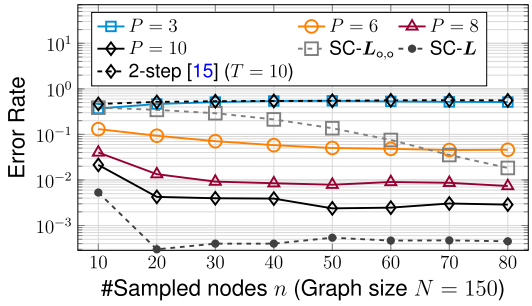

Fig. 4. *(Task A)* Error rate of inferred communities from graph signals against the number of sampled nodes n. Different plots represent different filter parameters for the filter $(\boldsymbol{I}-0.5\boldsymbol{L}_{\mathsf{sym}})^P$.

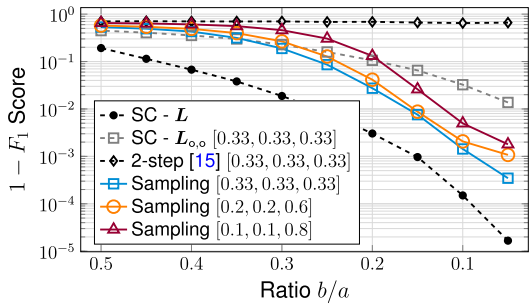

Fig. 5. *(Task A)* $1-F_1$ scores of inferred communities from graph signals against the ratio of inter/intra partition connectivity in the SBM model. In the legend, $[x,y,z]$ means that a fraction of x (*resp.* y, z) nodes are sampled from the 1st (*resp.* 2nd, 3 rd) cluster, e.g., [0.1,0.1,0.8] is an uneven sampling.

where c_i is the output of Algorithm 1 or Algorithm 2, and c_i^{true} is the ground truth used in generating the SBM graph. We evaluate the error rates using 10^4 Monte-carlo trials.

First, we focus on the *Task A* of blind inference which partitions only the observable nodes V_{obs}. For reference, we compare the performance of spectral clustering (SC) on the ground truth partial Laplacian $\boldsymbol{L}_{\mathrm{o,o}}$ and full Laplacian $\boldsymbol{L}$; and a procedure that first learns a graph topology via [15], [52], then applying SC. We begin by assessing the effect of graph filter and number of sampled nodes on the accuracy of partition inference. We fix the parameters at $a=8\log N/N$, $b=\log N/N$ and the observed nodes are sampled uniformly at random. The results are presented in Fig. 4. Observe that the performance is generally invariant with respect to the number of sampled nodes n; the performance improves significantly as we increase the order of diffusion T. Our observation is consistent with the analysis in Section III-B.

The second example evaluates the impact of node sampling scheme on the blind partition inference performance. Motivated by Lemma 1, we are interested in the combined effect of the SBM parameters and the sampling scheme. Here, we fix $a=8\log N/N$, and order of diffusion at $P=10$. There are $N=150$ nodes and $K=3$ clusters. Notice that due to the uneven sampling, the ground truth clusters are of different sizes. In this case, we compare the averaged F1 score in lieu of the error rate. The results are presented in Fig. 5, where we observe that when the ratio b/a is large, i.e., when the graph is not

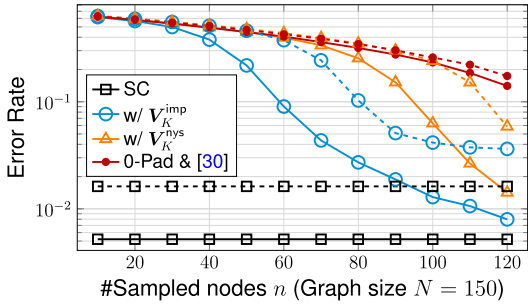

Fig. 6. *(Task B)* Error rate of inferred communities from graph signals against the number of sampled nodes n. The solid (resp. dashed) lines represent the error rates with $\frac{a}{b} = 16$ (resp. $\frac{a}{b} = 4$). We compare the error rate using different Nyström extension approximations in (32).

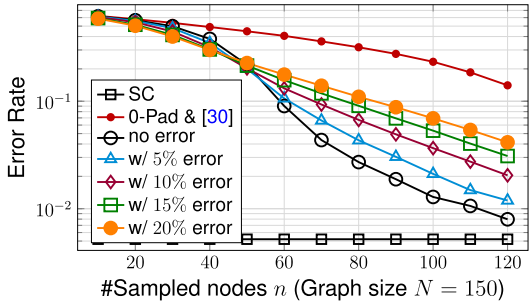

Fig. 7. *(Task B)* Error rate of inferred communities from graph signals against the number of sampled nodes n. We compare the error rate with different error level in the side information $\boldsymbol{A}_{\text{h,o}}^{\text{norm}}$ with $\boldsymbol{V}_K^{\text{imp}}$ in (32) for Algorithm 2.

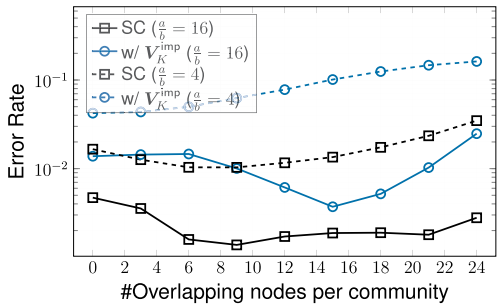

Fig. 8. *(Task B)* Error rate of inferred communities from graph signals against the number of overlapping nodes in each community. We compare the error rate with different SBM parameters a, b.

modular, the performance deteriorates if an uneven sampling scheme is adopted. The above observation is consistent with Lemma 1. Furthermore, for the two examples, the proposed approach outperforms the 2-step procedure. It demonstrates the benefits of direct community inference.

Next, we focus on *Task B* for semi-blind community inference. We compare the performance of SC on the ground truth Laplacian $\boldsymbol{L}$ and applying [30] on the zero-padded graph signals. We fix $a = 8 \log N/N$, $b = a/16$, $P = 10$ for the order of diffusion, and the observed nodes are sampled uniformly at random. In Fig. 6, we present the error rate against the number of sampled nodes n in inferring the communities of the full graph G using Algorithm 2. From the figure, we see that the performance of Algorithm 2 improves when the number of sampled nodes increases. With approximately 70 out of 150 nodes sampled, the semi-blind inference procedure infers the communities correctly. Interestingly, we observe that when n is small, the modified Nyström extension with $\boldsymbol{V}_K^{\text{imp}}$ achieves a better performance than plain Nyström extension with $\boldsymbol{V}_K^{\text{nys}}$. The performance with $\boldsymbol{V}_K^{\text{nys}}$ remains consistent with the analysis in Theorem 2.

The next example studies the effects of erroneous side topology information on the semi-blind community inference. Again, we consider the same simulation setting as the previous example while we fix $\frac{a}{b} = 16$ for the SBM parameters. The partial network topology $\boldsymbol{A}_{\text{h,o}}^{\text{norm}}$ used in Algorithm 2 is contaminated with noise. In particular, we consider using $\widehat{\boldsymbol{A}}_{\text{h,o}}^{\text{norm}} = \boldsymbol{A}_{\text{h,o}}^{\text{norm}} + \boldsymbol{E}_{\text{h,o}}$ where $\boldsymbol{E}_{\text{h,o}} \in \mathbb{R}^{(N-n)\times n}$ is a sparse matrix with varying density and the non-zero elements with the uniform distribution $\mathcal{U}[0, 0.2]$. The error rate performance against the number of sampled nodes n is presented in Fig. 7. We observe that the performance of Algorithm 2 is robust to erroneous side information $\widehat{\boldsymbol{A}}_{\text{h,o}}^{\text{norm}}$.

In the last example, we study the effects of *overlapping* community structure on the semi-blind inference performance. We consider an overlapping SBM model where in each block, there are r nodes that belong to two communities simultaneously. The rest of the settings are similar to the previous examples. In Fig. 8, we compare the error rate against the number of overlapped nodes per community. For nodes that belong to two communities, we count the detection as erroneous only if it is not found in neither of the ground truth communities. From the result, for graphs that are not modular (when $\frac{a}{b}$ is small), we observe that the performance generally deteriorates as the amount of overlap increases. On the other hand, for modular graphs (when $\frac{a}{b}$ is large), the semi-blind inference method achieves reasonable performance.

B. Real Data

We consider applying the proposed community inference methods on a dataset of S&P100 stocks obtained from https://www.kaggle.com/camnugent/sandp500. We focus on a subset of the data with the stock prices of $N = 92$ stocks in S&P100 that are collected from a window of Feb. 2013 to Dec. 2016, where the opening and closing prices of 975 days are considered. We treat the daily returns on day ℓ, defined as the ratio between closing and opening price, of each stock as the ℓth graph signal. Denote the ℓth graph signal by $\boldsymbol{y}^{(\ell)} \in \mathbb{R}^N$ and consider $m = 875$ days of data and leave the remaining 100 days of data for later use. We also normalize the variances of the daily returns of each stock in the data. We postulate that these graph signals follow a model like (7) with a low pass graph filter satisfying $\eta_K < 1$ [cf. (10)] for any K. The graph signals on each day is the result of a latent excitation signal modeling the global market environment. Our goal is to discover communities of companies with close ties, where the latter forms a stock network. Ideally, the communities detected include companies from the same business sector.

TABLE I
FINDING $K = 10$ COMMUNITIES FROM S&P 100 DATA. RED COLOR INDICATES HIDDEN NODES WHOSE PRICE SIGNALS ARE NOT OBSERVED

1	2	3	4	5	6	7	8	9	10
BMY	DUK	CL	COP	AAPL	CMCSA	T	BAC	ABBV	BA
ORCL	EXC	COST	KMI	AMZN	CVS	VZ	C	ABT	CAT
	SO	KO	OXY	FB	DIS		JPM	AMGN	EMR
	SPG	MCD	SLB	INTC	HD		MS	BIIB	IBM
				MSFT	NKE		USB	JNJ	MMM
				V	SBUX			LLY	UPS
								PFE	

Task A – with $n = 46$ stocks

1	2	3	4	5	6	7	8	9	10
BA	ABBV	AAPL	BMY	COP	AMGN	CMCSA	T	DUK	BAC
CAT	ABT	AMZN	CL	KMI	BIIB	COST	VZ	EXC	C
EMR	JNJ	FB	KO	OXY	AGN	CVS		SO	JPM
IBM	LLY	INTC	ORCL	SLB	CELG	DIS		SPG	MS
MMM	PFE	MSFT	CVX		GILD	HD		NEE	USB
UPS	MO	V	MA		TWX	MCD			AIG
FDX	MRK	ACN	MDLZ			NKE			ALL
GOOG	PG	CSCO	PEP			SBUX			AXP
HON	PM	GD	TGT						BK
LMT	UNH	GOOGL	WBA						BLK
MDT		HAL							COF
RTN		MON							F
UTX		PCLN							GE
		QCOM							GM
		TXN							GS
									LOW
									MET
									UNP
									WFC

Task B – with $N = 92$ stocks

TABLE II
(TASK A) ACCURACY OF FINDING COMMUNITIES COMPARED TO THE GICS SECTOR CLASSIFICATION

	Algorithm 1	2-step with [15]
F_1 Score	**0.7988**	0.1598

To evaluate the proposed methods, we hand picked $n = 46$ stocks from $K = 10$ business sectors (classified by GICS) as the observed nodes. We consider *Task A* by applying Algorithm 1 to infer $K = 10$ communities using the covariance matrix of the observables stocks. The results can be found in Table I. We observe close match with the clustering as Community 5 are technology companies such as 'Apple,' 'Intel,' etc.; Community 8 are financial companies such as 'Bank of America,' 'JP Morgan,' etc.. In Table II, we compare the F_1 score of classification of the result from Algorithm 1 to that of a 2-step procedure with [15]. Note that the ground truth communities are unknown. We take the GICS classification of stocks as the ground truth. We observe that the F_1 score of Algorithm 1 is significantly higher than that of [15].

Next, we consider *Task B* where we infer communities by combining the observed graph signals with partial graph information using Algorithm 2. To estimate the connectivity from hidden nodes to observable nodes, i.e., the matrix $\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}$, we compute the correlation between these nodes using the daily return data from the *unused 100 days of data*. Our results can be found in Table I, where we have highlighted the stocks that were not observed in the partial graph signals $\boldsymbol{y}_{\mathsf{obs}}^{(\ell)}$. Determining their communities has to rely on the Nyström extension method. We observe that companies in related sectors are grouped together. For example, Community 3 are technology companies such as 'Cisco,' 'Alphabet,' 'Qualcomm,' etc..; Community 2 are health care companies including 'Biogen Inc.' and 'Eli Lilly and Co'; Community 1 are industrial companies such as '3 M,' 'Raytheon Co.,' 'Honeywell,' etc.. We notice that there are some mismatches such as 'General Motors' and 'Ford Motors' in Community 10 while the majority are financial companies.

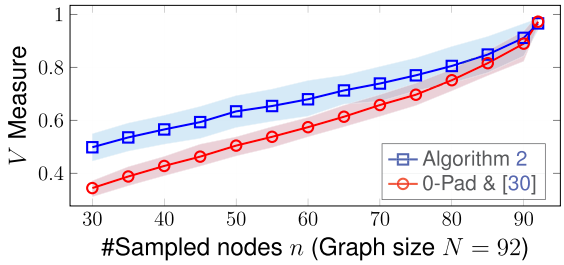

Fig. 9. *(Task B)* V-measure against number of sampled stocks/nodes in S&P100 dataset. Shaded area represents the 90% confidence interval.

Finally, we compare the consistency of our algorithm against the number of randomly (and uniformly) sampled stocks/nodes n in Fig. 9. We consider the V-measure[5] of the output of Algorithm 2 against the ground truth given by applying [30] on the full graph signals, averaged over 100 trials. The V-measure increases with n consistently.

VI. CONCLUSION

We introduced community inference methods for filtered graph signals with hidden nodes. Under the assumption of a low-pass graph filter and conditions on how the hidden nodes are selected, we analyzed the performance of spectral methods for blind and semi-blind inference. Our analysis reveals that the key factors determining the performance of the spectral methods are the strengths of the low-pass graph filter and the sampling pattern of the observable nodes. We then propose a Nyström extension based community inference method when connectivity information from hidden nodes to observable nodes is available. Numerical results on synthetic and real data justify our theoretical claims.

APPENDIX A
PROOF OF LEMMA 1

Under Assumption 1, expanding the squared norm and using the Cauchy-Schwarz inequality, we obtain for all $i \in \mathcal{C}_{\mathsf{obs,k}}^{\star}$ the following inequalities:

$$\begin{aligned}
&\|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_k^{\star}\|^2 - \|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_{\mathsf{obs,k}}^{\star}\|^2 \\
&\leq 2\|\boldsymbol{v}_i^{\mathrm{row}} - \boldsymbol{v}_{\mathsf{obs,k}}^{\star}\|\|\boldsymbol{v}_k^{\star} - \boldsymbol{v}_{\mathsf{obs,k}}^{\star}\| + \|\boldsymbol{v}_k^{\star} - \boldsymbol{v}_{\mathsf{obs,k}}^{\star}\|^2 \\
&\leq (2R_1 + \delta_{\mathsf{ce}})\,\delta_{\mathsf{ce}}. \qquad (38)
\end{aligned}$$

[5]The complexity of computing the F_1 score can be overwhelming in order to take care of label permutation with $K = 10$. As a remedy, we use the V-measure which is built in with `Clustering.jl`.

Similarly, for all $i \in \mathcal{C}_k^\star$, it holds that

$$\|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_{\mathsf{obs},\mathsf{k}}^\star\|^2 \le \|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_k^\star\|^2 + (2R_2 + \delta_{\mathsf{ce}})\,\delta_{\mathsf{ce}}. \quad (39)$$

Analyzing the surrogate objective function, we then have:

$$\begin{aligned}
\widetilde{F}_{\mathsf{o}}(\mathcal{C}_{\mathsf{obs},1}^\star, \ldots, \mathcal{C}_{\mathsf{obs},\mathsf{K}}^\star) &= \sum_{k=1}^K \sum_{i \in \mathcal{C}_{\mathsf{obs},\mathsf{k}}^\star} \left\|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_k^\star\right\|^2 \\
&\le \sum_{k=1}^K \sum_{i \in \mathcal{C}_{\mathsf{obs},\mathsf{k}}^\star} \left\{ \left\|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_{\mathsf{obs},\mathsf{k}}^\star\right\|^2 + (2R_1 + \delta_{\mathsf{ce}})\,\delta_{\mathsf{ce}} \right\} \\
&\overset{(a)}{\le} \sum_{k=1}^K \sum_{i \in \mathcal{C}_k^\star \cap V_{\mathsf{obs}}} \left\|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_{\mathsf{obs},\mathsf{k}}^\star\right\|^2 + N\,(2R_1 + \delta_{\mathsf{ce}})\,\delta_{\mathsf{ce}} \\
&\le \sum_{k=1}^K \sum_{i \in \mathcal{C}_k^\star \cap V_{\mathsf{obs}}} \left\|\boldsymbol{v}_i^{\rm row} - \boldsymbol{v}_k^\star\right\|^2 + 2N\,(R_1 + R_2 + \delta_{\mathsf{ce}})\,\delta_{\mathsf{ce}},
\end{aligned}$$

where (a) is due to that $\mathcal{C}_{\mathsf{obs},1}^\star, \ldots, \mathcal{C}_{\mathsf{obs},\mathsf{K}}^\star$ is a K-means optimal solution with the centroid vector $\boldsymbol{v}_{\mathsf{obs},\mathsf{k}}^\star$ and the fact that $\sum_{k=1}^K |\mathcal{C}_{\mathsf{obs},\mathsf{k}}^\star| = N$. This concludes the proof.

APPENDIX B
PROOF OF PROPOSITION 1

Observe the following bound:

$$\begin{aligned}
\|\widehat{\boldsymbol{Q}}_K \widehat{\boldsymbol{Q}}_K^\top - \boldsymbol{Q}_K \boldsymbol{Q}_K^\top\|_{\rm F} &\le \|\overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top - \boldsymbol{Q}_K \boldsymbol{Q}_K^\top\|_{\rm F} \\
&\quad + \|\widehat{\boldsymbol{Q}}_K \widehat{\boldsymbol{Q}}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_{\rm F}, \quad (40)
\end{aligned}$$

where we have defined $\overline{\boldsymbol{Q}}_K$ as the largest K eigenvectors of the noiseless covariance matrix $\overline{\boldsymbol{C}}_{\mathsf{obs}}$.

To bound the first term in r.h.s. of (40), observe that $\overline{\boldsymbol{Q}}_K$ can be obtained as the top-K left singular vectors of $\boldsymbol{E}_{\mathsf{obs}} \boldsymbol{V} \boldsymbol{\Sigma} \boldsymbol{U}^\top \boldsymbol{B}$ [cf. (15)], and recall that $\boldsymbol{C}_x = \boldsymbol{B}\boldsymbol{B}^\top$. The latter can be decomposed as:

$$\underbrace{\boldsymbol{V}_{\mathsf{o},K} \boldsymbol{\Sigma}_K \boldsymbol{U}_K^\top \boldsymbol{B}}_{=:\boldsymbol{P}} + \underbrace{\boldsymbol{V}_{\mathsf{o},N-K} \boldsymbol{\Sigma}_{N-K} \boldsymbol{U}_{N-K}^\top \boldsymbol{B}}_{=:\boldsymbol{T}}, \quad (41)$$

we have defined the partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{U}$ with $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_K \boldsymbol{0}; \boldsymbol{0} \boldsymbol{\Sigma}_{N-K})$ and $\boldsymbol{U} = (\boldsymbol{U}_K \boldsymbol{U}_{N-K})$, and we have $\boldsymbol{V}_{\mathsf{o},N-K} = \boldsymbol{E}_{\mathsf{obs}} \boldsymbol{V}_{N-K}$ with $\boldsymbol{V}_{N-K}$ corresponding to the last $N-K$ eigenvectors of $\boldsymbol{L}_{\mathsf{sym}}$. On the other hand, the range space of $\boldsymbol{P}$ is equivalent to that of $\boldsymbol{Q}_K$ provided that $\mathrm{rank}(\boldsymbol{U}_K^\top \boldsymbol{C}_x^{\frac{1}{2}}) = K$ [cf. Assumption 2]. If we let $\widetilde{\boldsymbol{Q}}_K$ be the largest K left singular vector of $\boldsymbol{P}$, then we have $\boldsymbol{Q}_K \boldsymbol{Q}_K^\top = \widetilde{\boldsymbol{Q}}_K \widetilde{\boldsymbol{Q}}_K^\top$. We invoke [53, Th. 3], which is a variant of the Wedin's theorem:

$$\begin{aligned}
&\|\boldsymbol{Q}_K \boldsymbol{Q}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_{\rm F} \\
&= \|\widetilde{\boldsymbol{Q}}_K \widetilde{\boldsymbol{Q}}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_{\rm F} = \|\sin\Theta(\widetilde{\boldsymbol{Q}}_K, \overline{\boldsymbol{Q}}_K)\|_{\rm F} \\
&\le \frac{2(2\|\boldsymbol{P}\|_2 + \|\boldsymbol{T}\|_2)}{\sigma_K^2(\boldsymbol{P})} \min\{\sqrt{K}\|\boldsymbol{T}\|_2, \|\boldsymbol{T}\|_{\rm F}\}\,, \quad (42)
\end{aligned}$$

where $\sigma_i(\boldsymbol{P})$ denotes the ith largest singular value of the matrix $\boldsymbol{P}$ and $\|\boldsymbol{T}\|_2$ denotes the spectral norm of the matrix $\boldsymbol{T}$. It is easy to obtain:

$$\begin{aligned}
&\|\boldsymbol{P}\|_2 \le \max\{s_1, \ldots, s_K\} \|\boldsymbol{U}_K^\top \boldsymbol{B}\|_2 \|\boldsymbol{V}_{\mathsf{o},K}\|_2\,, \\
&\|\boldsymbol{T}\|_2 \le \max\{s_{K+1}, \ldots, s_N\} \|\boldsymbol{U}_{N-K}^\top \boldsymbol{B}\|_2 \|\boldsymbol{V}_{\mathsf{o},N-K}\|_2, \quad (43)
\end{aligned}$$

where we defined $s_k := \Sigma_{k,k}$. Furthermore, we have

$$\begin{aligned}
\sigma_K(\boldsymbol{P}) &= \min\{s_1, \ldots, s_K\} \sigma_K\left(\boldsymbol{V}_{\mathsf{o},K} \mathrm{Diag}(\widehat{\boldsymbol{s}}) \boldsymbol{U}_K^\top \boldsymbol{B}\right) \\
&= \min\{s_1, \ldots, s_K\} \sqrt{\lambda_K\left(\overline{\boldsymbol{V}}_K^{\rm sym}\right)}, \quad (44)
\end{aligned}$$

where we scaled $\boldsymbol{s}_K := (s_1, \ldots, s_K)$ as $\widehat{\boldsymbol{s}} := \frac{1}{\min\{s_1, \ldots, s_K\}} \boldsymbol{s}_K \ge \boldsymbol{1}$. Moreover, we have

$$\begin{aligned}
\overline{\boldsymbol{V}}_K^{\rm sym} &:= \boldsymbol{B}^\top \boldsymbol{U}_K \mathrm{Diag}(\widehat{\boldsymbol{s}}) \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \mathrm{Diag}(\widehat{\boldsymbol{s}}) \boldsymbol{U}_K^\top \boldsymbol{B} \\
&\succeq \boldsymbol{B}^\top \boldsymbol{U}_K \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \boldsymbol{U}_K^\top \boldsymbol{B}\,, \quad (45)
\end{aligned}$$

where $\boldsymbol{A}_1 \succeq \boldsymbol{A}_2$ indicates that $\boldsymbol{A}_1 - \boldsymbol{A}_2$ is positive semidefinite, the above inequality is obtained by (i) rewriting $\mathrm{Diag}(\widehat{\boldsymbol{h}}) \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \mathrm{Diag}(\widehat{\boldsymbol{h}})$ as $\boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \odot (\widehat{\boldsymbol{h}}\widehat{\boldsymbol{h}}^\top)$ and (ii) applying Schur's theorem [54, Theorem 7.5.3] with the fact that $\widehat{\boldsymbol{h}}\widehat{\boldsymbol{h}}^\top - \boldsymbol{I} \succeq \boldsymbol{0}$. Consequently,

$$\sigma_K(\boldsymbol{P}) \ge \min\{s_1, \ldots, s_K\} \sqrt{\lambda_K(\boldsymbol{B}^\top \boldsymbol{U}_K \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \boldsymbol{U}_K^\top \boldsymbol{B})} \quad (46)$$

such that $\boldsymbol{U}_K$ contains the first K columns of $\boldsymbol{U}$. Combining (42), (43) and (46) gives the desired bound as

$$\begin{aligned}
&\|\boldsymbol{Q}_K \boldsymbol{Q}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_{\rm F} \\
&\le 2\gamma\sqrt{K} \frac{2\|\boldsymbol{U}_K^\top \boldsymbol{B}\|_2 + \|\boldsymbol{U}_{N-K}^\top \boldsymbol{B}\|_2}{\lambda_K(\boldsymbol{B}^\top \boldsymbol{U}_K \boldsymbol{V}_{\mathsf{o},K}^\top \boldsymbol{V}_{\mathsf{o},K} \boldsymbol{U}_K^\top \boldsymbol{B}_x)} \eta_K. \quad (47)
\end{aligned}$$

To bound the second term in (40), we follow a similar derivation as in [30, Proposition 2]. First,

$$\|\widehat{\boldsymbol{Q}}_K \widehat{\boldsymbol{Q}}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_{\rm F} \le \sqrt{2K} \|\widehat{\boldsymbol{Q}}_K \widehat{\boldsymbol{Q}}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_2.$$

As $\rho_{\mathsf{gap}} > 0$ [cf. (23)], applying the Weyl's inequality and Davis-Kahan theorem [55, Sec. VII.3] show that:

$$\|\widehat{\boldsymbol{Q}}_K \widehat{\boldsymbol{Q}}_K^\top - \overline{\boldsymbol{Q}}_K \overline{\boldsymbol{Q}}_K^\top\|_2 \le \frac{\|\widehat{\boldsymbol{C}}_{\mathsf{obs}}^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2}{\rho_{\mathsf{gap}}}\,. \quad (48)$$

Substituting the above back into (40) shows the desired bound.

APPENDIX C
PROOF OF THEOREM 1

We adopt the proof from [30, Theorem 1] (also see [56]). Define the indicator matrix $\widehat{\boldsymbol{X}} \in^{n \times K}$ associated with the partition

$\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}}$ found with the proposed method:

$$\widehat{X}_{ik} := \begin{cases} 1/\sqrt{|\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{k}}|}, & \text{if} i \in \widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{k}} \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

Similarly, we define the indicator matrix $\boldsymbol{X}^\star$ based on the partition $\mathcal{C}^\star_{\mathsf{obs},1},\ldots,\mathcal{C}^\star_{\mathsf{obs},\mathsf{K}}$. The latter partition is an *optimal solution* obtained by clustering the rows of $\boldsymbol{V}_{\mathsf{obs}}$, i.e., it minimizes the K-means objective function (4). In particular, using our definition it can be shown that

$$\begin{aligned} F(\mathcal{C}^\star_{\mathsf{obs},1},\ldots,\mathcal{C}^\star_{\mathsf{obs},\mathsf{K}};\boldsymbol{V}_{\mathsf{o},\mathsf{K}}) &= \|(\boldsymbol{I}-\boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)\boldsymbol{V}_{\mathsf{o},\mathsf{K}}\|_{\mathrm{F}}^2 \\ &\leq F(\mathcal{C}_1,\ldots,\mathcal{C}_K;\boldsymbol{V}_{\mathsf{o},\mathsf{K}}), \forall \mathcal{C}_1,\ldots,\mathcal{C}_K. \end{aligned} \tag{50}$$

We proceed by observing

$$\begin{aligned} &F(\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}};\boldsymbol{V}_{\mathsf{o},\mathsf{K}})^{\frac{1}{2}} = \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{V}_{\mathsf{o},\mathsf{K}}\|_{\mathrm{F}} \\ &\overset{(a)}{=} \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{Q}_K\boldsymbol{R}_K\|_{\mathrm{F}} \leq \|\boldsymbol{R}_K\|_2\|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{Q}_K\|_{\mathrm{F}} \end{aligned} \tag{51}$$

where (a) uses the QR factorization in (22). Moreover,

$$\begin{aligned} &\|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{Q}_K\|_{\mathrm{F}} \overset{(a)}{=} \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{Q}_K\boldsymbol{Q}_K^\top\|_{\mathrm{F}} \\ &\leq \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} \\ &\overset{(b)}{\leq} \sqrt{1+\epsilon}\|(\boldsymbol{I}-\boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)\widehat{\boldsymbol{Q}}_K\|_{\mathrm{F}} \\ &\quad + \|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}, \end{aligned} \tag{52}$$

where (a) is due to the orthogonal property of $\boldsymbol{Q}_K$, and (b) is due to the fact $\widehat{\boldsymbol{X}}$ is an $(1+\epsilon)$ optimal solution to the clustering problem based on the rows of $\widehat{\boldsymbol{Q}}_K$. By following a similar upper bounding techniques in (51), (52), and observe that $\sqrt{1+\epsilon} \leq 1+\epsilon$, we obtain that

$$\begin{aligned} &F(\widehat{\mathcal{C}}_{\mathsf{obs},1},\ldots,\widehat{\mathcal{C}}_{\mathsf{obs},\mathsf{K}};\boldsymbol{V}_{\mathsf{o},\mathsf{K}})^{\frac{1}{2}} \leq \\ &\left\{(1+\epsilon)\overline{F}^\star_{\mathsf{o}}\right\}^{\frac{1}{2}} + \|\boldsymbol{R}_K\|_2(2+\epsilon)\left\{\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}\right\}. \end{aligned}$$

Lastly, $\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}$ is upper bounded using Proposition 1. Combining terms conclude the proof.

Appendix D
Proof of Theorem 2

Similar to the proof of Theorem 1, we begin by defining the $N \times K$ indicator matrices $\widehat{\boldsymbol{X}}$, $\boldsymbol{X}^\star$, based on the partitions $(\widehat{\mathcal{C}}_1,\ldots,\widehat{\mathcal{C}}_K)$ and $(\mathcal{C}^\star_1,\ldots,\mathcal{C}^\star_K)$, respectively. For instance, with a slight abuse of notations, we define

$$\widehat{X}_{ik} := \begin{cases} 1/\sqrt{|\widehat{\mathcal{C}}_{\mathsf{k}}|}, & \text{if} i \in \widehat{\mathcal{C}}_{\mathsf{k}}, \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

Similar to (51), we observe that

$$\begin{aligned} &F(\widehat{\mathcal{C}}_1,\ldots,\widehat{\mathcal{C}}_K;\boldsymbol{V}_K)^{\frac{1}{2}} = \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\boldsymbol{V}_K\|_{\mathrm{F}} \\ &\leq \|(\boldsymbol{I}-\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top)\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top\|_{\mathrm{F}} + \|\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top - \boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}} \\ &\overset{(a)}{\leq} \sqrt{1+\epsilon}\|(\boldsymbol{I}-\boldsymbol{X}^\star(\boldsymbol{X}^\star)^\top)\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top\|_{\mathrm{F}} \\ &\quad + \|\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top - \boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}} \\ &\overset{(b)}{\leq} \sqrt{(1+\epsilon)F^\star} + (2+\epsilon)\|\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top - \boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}}. \end{aligned} \tag{54}$$

where (a) uses the fact that $\widehat{\boldsymbol{X}}$ is an $(1+\epsilon)$ optimal solution, and (b) uses $\sqrt{1+\epsilon} \leq 1+\epsilon$ with the triangular inequality.

Next, we denote the SVD of the partial Laplacian as $\boldsymbol{L}_{\mathsf{o},\mathsf{o}} = \boldsymbol{Q}\boldsymbol{\Lambda}_{\mathsf{obs}}\boldsymbol{Q}^\top$. The following matrix is the Nyström rank-K approximation of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$ [48]:

$$\boldsymbol{L}_{\mathsf{nym},\mathsf{K}} = \begin{pmatrix} \boldsymbol{Q}_K \\ \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}}\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}^{-1}_{\mathsf{obs},\mathsf{K}} \end{pmatrix} \widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}} \begin{pmatrix} \boldsymbol{Q}_K \\ \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}}\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}^{-1}_{\mathsf{obs},\mathsf{K}} \end{pmatrix}^\top, \tag{55}$$

where $\boldsymbol{Q}_K$ is the largest K eigenvectors of $2\boldsymbol{I} - \boldsymbol{L}_{\mathsf{o},\mathsf{o}}$, and $\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}} = 2\boldsymbol{I} - \boldsymbol{\Lambda}_{\mathsf{obs},\mathsf{K}}$ such that $\boldsymbol{\Lambda}_{\mathsf{obs},\mathsf{K}}$ is the principal K submatrix of $\boldsymbol{\Lambda}_{\mathsf{obs}}$.

On the other hand, $\widehat{\boldsymbol{V}}_K$ has the same range space as

$$\widehat{\boldsymbol{L}}_{\mathsf{nym},\mathsf{K}} = \begin{pmatrix} \widehat{\boldsymbol{Q}}_K \\ 2^{-1}\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}}\widehat{\boldsymbol{Q}}_K \end{pmatrix} \widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}} \begin{pmatrix} \widehat{\boldsymbol{Q}}_K \\ 2^{-1}\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}}\widehat{\boldsymbol{Q}}_K \end{pmatrix}^\top, \tag{56}$$

where we recall that $\widehat{\boldsymbol{Q}}_K$ is obtained from applying line 2–3 in Algorithm 1. To proceed with the bound in (54), we apply the Davis-Kahan theorem [55, Sec. VII.3] to obtain

$$\|\widehat{\boldsymbol{V}}_K\widehat{\boldsymbol{V}}_K^\top - \boldsymbol{V}_K\boldsymbol{V}_K^\top\|_{\mathrm{F}} \leq \frac{\sqrt{2K}}{\mathrm{c}_L - \lambda_K}\|\widehat{\boldsymbol{L}}_{\mathsf{nym},\mathsf{K}} - \widetilde{\boldsymbol{L}}_{\mathsf{sym}}\|_2, \tag{57}$$

where we recall that λ_K is the Kth smallest eigenvalue of $\boldsymbol{L}_{\mathsf{sym}}$. The triangular inequality gives

$$\begin{aligned} \|\widehat{\boldsymbol{L}}_{\mathsf{nym},\mathsf{K}} - \widetilde{\boldsymbol{L}}_{\mathsf{sym}}\|_2 &\leq \|\widehat{\boldsymbol{L}}_{\mathsf{nym},\mathsf{K}} - \boldsymbol{L}_{\mathsf{nym},\mathsf{K}}\|_2 \\ &+ \|\boldsymbol{L}_{\mathsf{nym},\mathsf{K}} - \widetilde{\boldsymbol{L}}_{\mathsf{sym}}\|_2. \end{aligned} \tag{58}$$

For the first term on the right hand side of (58), we further observe the following decomposition,

$$\begin{aligned} &\|\widehat{\boldsymbol{L}}_{\mathsf{nym},\mathsf{K}} - \boldsymbol{L}_{\mathsf{nym},\mathsf{K}}\|_2 \\ &\leq \|\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}}\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}}\widehat{\boldsymbol{Q}}_K^\top\|_2 \\ &+ 2\left\| \left(\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \tfrac{1}{2}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}}\widehat{\boldsymbol{Q}}_K^\top\right)(\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}})^\top \right\|_2 \\ &+ \left\| \boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}} \left(\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}^{-1}_{\mathsf{obs},\mathsf{K}}\boldsymbol{Q}_K^\top - \tfrac{1}{4}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs},\mathsf{K}}\widehat{\boldsymbol{Q}}_K^\top\right)(\boldsymbol{A}^{\mathsf{norm}}_{\mathsf{h},\mathsf{o}})^\top \right\|_2 \\ &\equiv T_1 + 2T_2 + T_3. \end{aligned}$$

The following holds with probability at least $1-\Delta$,

$$T_1 \le 2\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_2 + 2\|2\boldsymbol{I} - \widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\|_2 \le 2\left(\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \tilde{\delta}_{\mathsf{eig}}(\Delta)\right), \quad (59)$$

where we have used the norm equivalence $\|\cdot\|_2 \le \|\cdot\|_{\mathrm{F}}$ for symmetric matrix. Moreover,

$$T_2 \le \|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2 \left\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \tfrac{1}{2}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\widehat{\boldsymbol{Q}}_K^\top\right\|_2. \quad (60)$$

Since

$$\left\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \tfrac{1}{2}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\widehat{\boldsymbol{Q}}_K^\top\right\|_2 \le \|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \|\widehat{\boldsymbol{Q}}_K(\boldsymbol{I} - \tfrac{1}{2}\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}})\widehat{\boldsymbol{Q}}_K^\top\|_2,$$

we have

$$T_2 \le \|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2 \left(\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \frac{\tilde{\delta}_{\mathsf{eig}}(\Delta)}{2}\right). \quad (61)$$

Finally,

$$T_3 \le \|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2^2 \left\|\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}^{-1}\boldsymbol{Q}_K^\top - \tfrac{1}{4}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\widehat{\boldsymbol{Q}}_K^\top\right\|_2, \quad (62)$$

with probability at least $1-\Delta$. Since

$$\left\|\boldsymbol{Q}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}^{-1}\boldsymbol{Q}_K^\top - \tfrac{1}{4}\widehat{\boldsymbol{Q}}_K\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\widehat{\boldsymbol{Q}}_K^\top\right\|_2 \le \tfrac{1}{2}\left(\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \|2\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}^{-1} - \boldsymbol{I}\|_2 + \|\boldsymbol{I} - \tfrac{1}{2}\widetilde{\boldsymbol{\Lambda}}_{\mathsf{obs,K}}\|_2\right), \quad (63)$$

we obtain

$$T_3 \le \tfrac{1}{2}\|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2^2\left(\|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}} + \tfrac{3}{2}\tilde{\delta}_{\mathsf{eig}}(\Delta)\right). \quad (64)$$

Combining the above observations, if we denote that $e_K := \|\boldsymbol{Q}_K\boldsymbol{Q}_K^\top - \widehat{\boldsymbol{Q}}_K\widehat{\boldsymbol{Q}}_K^\top\|_{\mathrm{F}}$, then

$$\|\widehat{\boldsymbol{L}}_{\mathsf{nym,K}} - \boldsymbol{L}_{\mathsf{nym,K}}\|_2 \le \left(2 + 2\|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2 + \|\boldsymbol{A}_{\mathsf{h,o}}^{\mathsf{norm}}\|_2^2\right)\left(e_K + \tilde{\delta}_{\mathsf{eig}}(\Delta)\right). \quad (65)$$

Note that e_K can be bounded using Proposition 1 as $\mathcal{O}(\sqrt{K}(\eta_K + \|\widehat{\boldsymbol{C}}_y^m - \overline{\boldsymbol{C}}_{\mathsf{obs}}\|_2/\tilde{\rho}_{\mathsf{gap}}(\Delta)))$.

For the last term in the r.h.s. of (58), under the uniform sampling assumption, we can invoke [49, Lemma 8] which shows that with probability at least $1-\Delta$,

$$\|\boldsymbol{L}_{\mathsf{nym,K}} - \widetilde{\boldsymbol{L}}_{\mathsf{sym}}\|_2 \le \left(1 + \sqrt{\frac{2N}{n}}\right)(\mathrm{c}_L - \lambda_{K+1}), \quad (66)$$

provided that $n \ge 8\mu K\log(3K/\Delta)$, where μ is a coherence factor for the top-K eigenvector of $\widetilde{\boldsymbol{L}}_{\mathsf{sym}}$. The bound is obtained from [49, Lemma 8] by setting $q=1$, $\epsilon = 1/2$ and observing the bound $\|\widetilde{\boldsymbol{L}}_{\mathsf{sym}} - \widetilde{\boldsymbol{L}}_{\mathsf{sym,K}}\|_2 \le \mathrm{c}_L - \lambda_{K+1}$.

Invoking Proposition 1 and Assumption 3-(1) gives an upper bound to e_K with probability at least $1-\Delta$. Lastly, combining (54), (57), (65), (66) yield the desired bound.

REFERENCES

[1] H.-T. Wai, Y. C. Eldar, A. E. Ozdaglar, and A. Scaglione, "Community inference from graph signals with hidden nodes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 4948–4952.

[2] E. D. Kolaczyk and G. Csárdi, *Statistical Analysis of Network Data With R*, vol. 65, Berlin, Germany: Springer, 2014.

[3] M. E. Newman, "Spectral methods for community detection and graph partitioning," *Phys. Rev. E*, vol. 88, no. 4, 2013, Art. no. 0 42822.

[4] E. Balkanski, N. Immorlica, and Y. Singer, "The importance of communities for learning to influence," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5862–5871.

[5] D. Marbach *et al.*, "Wisdom of crowds for robust gene network inference," *Nature Methods*, vol. 9, no. 8, pp. 796–804, 2012.

[6] H. Zha, X. He, C. Ding, M. Gu, and H. D. Simon, "Spectral relaxation for k-means clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 1057–1064.

[7] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 225–232.

[8] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no., pp. 2579–2605, Nov. 2008.

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

[10] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[11] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.

[12] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.

[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] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. Artif. Intell. Statist.*, 2016, pp. 920–929.

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

[16] 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, May 2017.

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

[18] H.-T. Wai, A. Scaglione, B. Barzel, and A. Leshem, "Joint network topology and dynamics recovery from perturbed stationary points," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4582–4596, Sep. 2019.

[19] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-blind inference of topologies and dynamical processes over dynamic graphs," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2263–2274, May 2019.

[20] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 825–841, Sep. 2017.

[21] Y. Yuan *et al.*, "Learning overlapping community-based networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 4, pp. 684–697, Dec. 2019.

[22] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," in *Proc. IEEE 48th Annu. Allerton Conf. Commun., Control, Comput.*, 2010, pp. 1610–1613.

[23] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 622–646, Jan. 2019.

[24] A. Buciulea, S. Rey, C. Cabrera, and A. G. Marques, "Network reconstruction from graph-stationary signals with hidden variables," in *Proc. IEEE 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 56–60.

[25] R. Shafipour and G. Mateos, "Online network topology inference with partial connectivity informatio," in *Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2019, pp. 226–230.

[26] A. Venkitaraman, S. Chatterjee, and P. Händel, "Predicting graph signals using kernel regression where the input signal is agnostic to a graph," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 4, pp. 698–710, Dec. 2019.

[27] P. Berger, G. Hannak, and G. Matz, "Efficient graph learning from noisy and incomplete data," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 105–119, 2020.

[28] X. Pu, S. L. Chau, X. Dong, and D. Sejdinovic, "Kernel-based graph learning from smooth signals: A functional viewpoint," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 7, pp. 192–207, 2021.
[29] E. Abbe, "Community detection and stochastic block models: Recent developments," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6446–6531, 2017.
[30] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind community detection from low-rank excitations of a graph filter," *IEEE Trans. Signal Process.*, vol. 68, pp. 436–451, 2020.
[31] M. T. Schaub, S. Segarra, and J. N. Tsitsiklis, "Blind identification of stochastic block models from dynamical observations," *SIAM J. Math. Data Sci.*, vol. 2, no. 2, pp. 335–367, 2020.
[32] T. M. Roddenberry, M. T. Schaub, H.-T. Wai, and S. Segarra, "Exact blind community detection from signals on multiple graphs," *IEEE Trans. Signal Process.*, vol. 68, pp. 5016–5030, 2020.
[33] T. Hoffmann, L. Peel, R. Lambiotte, and N. S. Jones, "Community detection in networks with unobserved edges," *Sci. Adv.*, vol. 6, no. 4, 2020.
[34] Y. Xing, X. He, H. Fang, and K. H. Johansson, "Community detection for gossip dynamics with stubborn agents," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 4915–4920.
[35] Y. He and H.-T. Wai, "Estimating centrality blindly from low-pass filtered graph signals," in *Proc. ICASSP*, 2020, pp. 5330–5334.
[36] T. M. Roddenberry and S. Segarra, "Blind inference of centrality rankings from graph signals," in *Proc. ICASSP*, 2020, pp. 5335–5339.
[37] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.
[38] K. Rohe *et al.*, "Spectral clustering and the high-dimensional stochastic blockmodel," *Ann. Statist.*, vol. 39, no. 4, pp. 1878–1915, 2011.
[39] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis.," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
[40] R. Ramakrishna, H.-T. Wai, and A. Scaglione, "A user guide to low-pass graph signal processing and its applications," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 74–85, Nov. 2020.
[41] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 484–499, Sep. 2017.
[42] A. Heimowitz and Y. C. Eldar, "A Markov variation approach to smooth graph signal interpolation," 2018, *arXiv:1806.03174*.
[43] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 406–419, Sep. 2016.
[44] A. Kumar, Y. Sabharwal, and S. Sen, "A simple linear time $(1+\epsilon)$+ approximation algorithm for K-Means clustering in any dimensions," in *Proc. IEEE FOCS*, 2004, pp. 454–462.
[45] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[46] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nystrom method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 214–225, Feb. 2004.
[47] A. Choromanska, T. Jebara, H. Kim, M. Mohan, and C. Monteleoni, "Fast spectral clustering via the Nyström method," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2013, pp. 367–381.
[48] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling methods for the nyström method," *J. Mach. Learn. Res.*, vol. 13, pp. 981–1006, 2012.
[49] A. Gittens and M. W. Mahoney, "Revisiting the nyström method for improved large-scale machine learning," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 3977–4041, 2016.
[50] N. Tremblay and A. Loukas, "Approximating spectral clustering via sampling: A review," in *Sampling Techniques for Supervised or Unsupervised Tasks*, Springer, 2020, pp. 129–183.
[51] S. Deng, S. Ling, and T. Strohmer, "Strong consistency, graph laplacians, and the stochastic block model," *J. Mach. Learn. Res.*, vol. 22, no. 117, pp. 1–44, 2021.
[52] R. Shafipour, A. Hashemi, G. Mateos, and H. Vikalo, "Online topology inference from streaming stationary graph signals.," in *Proc. DSW*, 2019, pp. 140–144.
[53] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the davis-kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2014.
[54] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[55] R. Bhatia, *Matrix Analysis*. vol. 169. Berlin, Germany: Springer Science & Business Media, 2013.
[56] C. Boutsidis, P. Kambadur, and A. Gittens, "Spectral clustering via the power method-provably," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 40–48.

Hoi-To Wai (Member, IEEE) received the B.Eng. (with first-class Hons.) and M.Phil. degrees in electronic engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2010 and 2012, respectively, and the Ph.D. degree in electrical engineering from Arizona State University (ASU), Tempe, AZ, USA, in 2017. He is currently an Assistant Professor with the Department of Systems Engineering and Engineering Management, CUHK. He has held research positions with ASU, University of California, Davis, Davis, CA, USA, Telecom ParisTech, Paris, France, Ecole Polytechnique, Palaiseau, France, and LIDS, MIT, Cambridge, MA, USA. His research interests include signal processing, machine learning and distributed optimization, with a focus of their applications to network science. His dissertation was the recipient of the 2017's Dean's Dissertation Award from the Ira A. Fulton Schools of Engineering of ASU, and he was also the recipient of a Best Student Paper Award at ICASSP 2018.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was a Professor with the Department of Electrical Engineering, Technion, Haifa, Israel. She is also a Visiting Professor with MIT, a Visiting Scientist with Broad Institute, and an Adjunct Professor with Duke University, Durham, NC, USA, She was a Visiting Professor with Stanford. She is the author of the book *Sampling Theory: Beyond Bandlimited Systems* and coauthor of five other books published by Cambridge University Press. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging and optics.

Dr. Eldar was the recipient of many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and IEEE Kiyo Tomiyasu Award (2016). She was a Horev Fellow of the Leaders in Science and Technology Program at the Technion and an Alon Fellow. She was the recipient of the Michael Bruno Memorial Award from the Rothschild Foundation, Weizmann Prize for Exact Sciences, Wolf Foundation Krill Prize for Excellence in Scientific Research, Henry Taub Prize for Excellence in Research (twice), Hershel Rich Innovation Award (three times), Award for Women with Distinguished Contributions, Andre and Bella Meyer Lectureship, Career Development Chair at the Technion, Muriel & David Jacknow Award for Excellence in Teaching, and Technionís Award for Excellence in Teaching (two times). She was also the recipient of several best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize, UFFC Outstanding Paper Award, Signal Processing Society Best Paper Award, and IET Circuits, Devices and Systems Premium Award, was selected as one of the 50 most influential women in Israel and in Asia, and is a highly cited Researcher.

She is a member of the Israel Academy of Sciences and Humanities was elected in 2017, and EURASIP Fellow. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing*, *SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*. She was the Co-Chair and Technical Co-Chair of several international conferences and workshops.

Asuman E. Ozdaglar (Fellow, IEEE) received the B.S. degree in electrical engineering from Middle East Technical University, Ankara, Turkey, in 1996, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1998 and 2003, respectively.

She is currently the Mathworks Professor of electrical engineering and computer science (EECS) with the Massachusetts Institute of Technology. She is also the Department Head of EECS and the Deputy Dean of academics of the Schwarzman College of Computing with MIT. She is the coauthor of the book entitled *Convex Analysis and Optimization*, Athena Scientific, 2003. Her expertise and research interests include optimization theory and algorithms, game theory, machine learning and network analysis with applications in social, economic and financial networks. Her current research focuses on the designing incentives and algorithms for data-driven online systems with many diverse human-machine participants.

Prof. Ozdaglar was the recipient of the Microsoft fellowship, MIT Graduate Student Council Teaching Award, NSF Career Award, 2008 Donald P. Eckman Award of the American Automatic Control Council, 2014 Spira teaching award, and Keithley, Distinguished School of Engineering and Mathworks professorships. She was selected as an Invited Speaker at the International Congress of Mathematicians. She was on the Board of Governors of the Control System Society in 2010 and was an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. She was the inaugural Area Co-Editor of the area entitled Games, Information, and Networks in the journal Operations Research.

Anna Scaglione (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in 1995 and 1999, respectively. She is currently a Professor in electrical and computer wih Cornell Tech, Cornell University, New York, NY, USA. Prior to that, she held faculty positions with Arizona State University, Tempe, AZ, USA, the University of California at Davis, Davis, CA, USA, Cornell University (the first time), and The University of New Mexico, Albuquerque, NM, USA.

Her expertise and research interests include theoretical and applied problems is in statistical signal processing, communications theory and cyber-physical systems.

She was the recipient of the 2013 IEEE Donald G. Fink Prize Paper Award, 2000 IEEE Signal Processing Transactions Best Paper Award the NSF CAREER Grant in 2002. She was the co-recipient with her students of several best student papers awards at conferences, such as the 2013 IEEE Signal Processing Society Young Author Best Paper Award with one of the Ph.D. students. She was also the recipient of the 2020 Technical Achievement Award from the IEEE Communication Society Technical Committee on Smart Grid Communications and was Distinguished Lecturer of the Signal Processing Society in 2019 and 2020.