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 topological 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

A N 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. 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

1053-587X © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
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.

- **(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. Bold-face (capital) letters are used to denote vectors (matrices), Diag($\lambda$) is a diagonal matrix whose diagonal elements are taken from the vector $\lambda$ in order. For a vector $x$, $\|x\|$ is its Euclidean norm. For a matrix $Z \in \mathbb{R}^{m \times n}$, we denote $z_i^{row} \in \mathbb{R}^n$ as its $i$th row vector, while $\|Z\|_F$, $\|Z\|_2$ 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 $G = (V, E, A)$, where $V = \{1, \ldots, N\}$ is the node set, $E \subseteq V \times V$ is the edge set and $A \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix which is non-negative, symmetric and satisfies $A_{ij} = 0$ if and only if $(i, j) \notin E$. Let $K = \{1, \ldots, N\}$ be given and $C_1, \ldots, C_K$ be a non-overlapping, non-trivial partition of $V$, i.e., with $C_i \neq \emptyset$. We consider the normalized cut measure:

$$\text{NCut}(C_1, \ldots, C_K) = \sum_{k=1}^{K} \sum_{i \in C_k} \sum_{j \notin C_k} \frac{A_{ij}}{\sum_{l \in C_k} \sum_{m \in C_l} A_{lm}}. \quad (1)$$

Consider a partition $C_1, \ldots, C_K$. If the induced subgraph $G[C_k]$ is densely connected, for all $k$, and the number of edges between $C_k, C_k$ is small, then $\text{NCut}(C_1, \ldots, C_K) \approx 0$ 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 $D := \text{Diag}(A1)$ and the normalized adjacency matrix be $A_{\text{norm}} := D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. The normalized Laplacian is $L_{\text{sym}} := I - A_{\text{norm}}$ with eigen decomposition $L_{\text{sym}} = \Sigma \Lambda \Sigma^T$, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_N)$, $0 = \lambda_1 < \cdots < \lambda_N \leq 2$ and $V$ is orthonormal. It is well known [37] that the NCut measure is linked to the eigenvectors of the normalized Laplacian $L_{\text{sym}}$. Indeed, it can be shown that

$$\min_{c_1, \ldots, c_K \subseteq V} \text{NCut}(C_1, \ldots, C_K)$$

$$\iff \min_{X \in \mathbb{R}^{N \times K}} \text{Tr}(X^T D^2 L_{\text{sym}} D^2 X)$$

$$\text{s.t. } X^T DX = I, X_{ij} \in \{0, W_j^{-1/2}\}, \quad (2)$$

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

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

$$k\text{means}(Z, K) \in \arg\min_{C_1, \ldots, C_K \subseteq [p]} F(C_1, \ldots, C_K; Z), \quad (3)$$

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

$$F(C_1, \ldots, C_K; Z) := \frac{1}{K} \sum_{k=1}^{K} \sum_{i \in C_k} \left\| \frac{z_i - \frac{1}{|C_k|} \sum_{j \in C_k} z_j}{\text{row normalization}} \right\|^2. \quad (4)$$

In particular, the output of the spectral clustering method (without row normalization) [37] can be written as $k\text{means}(V_K, K)$, which approximates the NCut minimization (2). The method requires perfect knowledge of $L_{\text{sym}}$.

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

$$(C_1^\text{true}, \ldots, C_K^\text{true}) := k\text{means}(V_K, K). \quad (5)$$

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 $C_1^\text{true}, C_2^\text{true}$. The edge set of SBM is parameterized by

$$a = p((i, j) \in E, i \in C_1^\text{true}, j \in C_2^\text{true}),$$

$$b = p((i, j) \in E, i \in C_1^\text{true}, j \in C_2^\text{true}), \quad (6)$$

and $A$ is the binary adjacency matrix. When $a \gg b$, the nodes in $C_1^\text{true}$ or $C_2^\text{true}$ induce a densely connected subgraph. In particular, if $a, b$ satisfy the spectral resolution criterion [29], [38], then $(C_1^\text{true}, C_2^\text{true}) = (C_1^\text{true}, C_2^\text{true})$ or $(C_1^\text{true}, C_2^\text{true}) = (C_2^\text{true}, C_1^\text{true})$ (with label ambiguity), i.e., applying the spectral clustering method on $V_2$ 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 $\ell$th graph signal, $y^{(\ell)} \in \mathbb{R}^N$, is a filtered graph signal described as:

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

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

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

is a $T$th 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 $R = \text{rank}(C_x) \leq N$. Without loss of generality, we concentrate on the model $C_x = BB^\top$ where the columns of $B \in \mathbb{R}^{N \times R}$, $R \leq N$, describe the subspace where the excitation signals $x^{(\ell)}$ lies in. For example, when the excitation graph signals at the set of $R$ nodes $\{i_1, \ldots, i_R\}$ are statistically related to $L_{\text{sym}}$ and assumed to be low-pass [30], [39], [40]:

**Assumption 1:** We assume that the operator $H(S)$ can be written as:

$$H(S) = V \Sigma U^\top, \quad (9)$$
where $\Sigma \in \mathbb{R}^{N \times N}$ is a diagonal matrix of singular values ordered with respect to $V$, whose left-to-right column vectors are the eigenvectors of the normalized Laplacian $L_{\text{sym}}$ with increasing eigenvalues, and $U \in \mathbb{R}^{N \times N}$ is a square matrix. Lastly, the diagonal elements of $\Sigma$ satisfy

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

Note that if $H(S)$ is a polynomial graph filter $\sum_{i=0}^{p-1} h_i S^i$, then $U = V$ and (10) is equivalent to the definition of a $(\eta_K, K)$-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 $\eta_K$ 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 $S = L_{\text{sym}}$. It is straightforward to verify that $H(L_{\text{sym}}) = VH(\Lambda)V^T$, 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 $H(L_{\text{sym}}) = e^{-\sigma L_{\text{sym}}}$ for some $\sigma > 0$ or its discretized version with $H(L_{\text{sym}}) = (I - \sigma L_{\text{sym}})^P$ with $P \in \mathbb{Z}_+$, $0 < \sigma < 1$; see [41].

**Case 2. Markov Matrix**. Let the GSO be the Markov matrix $S = A_{\text{markov}} = D^{-1}A$, which is not symmetric. Using [42, Proposition 1], we observe that $A_{\text{markov}} = D^{-\frac{1}{2}} V (I - \Lambda) V^T D^{-\frac{1}{2}}$. Using the fact that $(D^{-\frac{1}{2}} V)^{-1} = V^T D^{\frac{1}{2}}$, $D^{\frac{1}{2}} H(A_{\text{markov}}) = VH(\Lambda)V^T D^{\frac{1}{2}}$. (11)

Therefore, the pre-multiplied graph filter $D^{\frac{1}{2}} H(A_{\text{markov}})$ also satisfies the requirements (9), (10) when the shifted polynomial $h(I - \lambda)$ is decreasing with respect to $\lambda \in [0, 2]$. Note that the rows of $A_{\text{markov}}$ sum to one and the matrix is non-negative. As such, applying $A_{\text{markov}}$ 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 $H(A_{\text{markov}}) = A_{P_{\text{markov}}}$ for some $P \in \mathbb{Z}_+$, which model $T$ rounds of gossiping averages; and the IIR graph filter $H(A_{\text{markov}}) = (I - \alpha A_{\text{markov}})^{-1}$, $\alpha \in (0, 1)$ 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 $V_{\text{obs}} - V_{\text{hid}}$ such that $V_{\text{obs}}$ is the set of $n$ observable nodes, and $V_{\text{hid}}$ is the set of hidden nodes with $V_{\text{hid}} = V \setminus V_{\text{obs}}$. We set $V_{\text{obs}} = \{1, \ldots, n\}$, $V_{\text{hid}} = \{n + 1, \ldots, N\}$ such that the adjacency and Laplacian matrices can be described respectively as:

$$A_{\text{norm}} = \begin{pmatrix} A_{0,0} & A_{0,h} & A_{h,h} \\ A_{h,0} & A_{h,h} & A_{h,0} \\ A_{h,0} & A_{h,h} & A_{h,0} \end{pmatrix}, \\ L_{\text{sym}} = \begin{pmatrix} L_{0,0} & L_{0,h} & L_{h,h} \\ L_{h,0} & L_{h,h} & L_{h,0} \\ L_{h,0} & L_{h,h} & L_{h,0} \end{pmatrix}. (12)$$

Here, $A_{0,0}$ (resp. $A_{h,h}$) represents the edges among the nodes in $V_{\text{obs}}$ (resp. $V_{\text{hid}}$) and $A_{h,0}$ (resp. $A_{h,h}$) represents the edges between $V_{\text{obs}}$ and $V_{\text{hid}}$. The observed signal corresponds to a spatially sampled version of $y(t)$ with entries in $V_{\text{obs}}$, i.e.,

$$y_{\text{obs}}(t) = (I_{n \times n}0_{n \times (N-n)}) y(t). (13)$$

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

Community inference aims to infer the ‘true’ community $C_1^*, \ldots, C_K^*$ defined in (5) based on the spatially sampled graph signals $\{y_{\text{obs}}(t)\}_{t=1}^{m}$. Consider the following two tasks:

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

$$V_{\text{obs}} \cap C_k^*, k = 1, \ldots, K.$$ (14)

This corresponds to the communities of $V_{\text{obs}}$ 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 $V_{\text{obs}}$ and $V_{\text{hid}}$, represented by the adjacency sub-matrix $A_{\text{norm}}$, is known. In this case, we have to combine two types of information: the graph signals on $V_{\text{obs}}$, and the connectivity information between $V_{\text{hid}} - V_{\text{obs}}$. Naturally, our aim is to recover the ‘true’ communities $C_1^*, \ldots, C_K^*$. Notice that $A_{\text{norm}}$ is not required in this task.

In **Task B**, $A_{\text{norm}}$ 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, $V^{(i)}_{\text{obs}} \cap C^*_k$, $k \in \{1, \ldots, K\}$, from the partially observed graph signals $\{y^{(f)}_{\text{obs}}\}_{f=1}^m$.

A. Blind Inference Method

We recall from (5) that $C_1, \ldots, C_K$ is defined through the $K$-means subroutine $\kmeans(V_K, K)$. Ideally, Task A aims at finding (14) is solved if $V_K$ 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 $y^{(i)}_{\text{obs}}$. Using (7), (9), (13), the covariance matrix can be evaluated as:

$$
C_{\text{obs}} = \mathbb{E}[y^{(i)}_{\text{obs}}(y^{(i)}_{\text{obs}})^\top] = V_0 \Sigma U^\top B^B U \Sigma V^\top_0 + C_c, \tag{15}
$$

where $V_0 := E_{\text{obs}} V$ and we noted that $C_c = BB^\top$. Also, we denote $\mathcal{C}_{\text{obs}} := C_{\text{obs}} - C_c$ 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 $\Sigma$ is dominated by its top-$K$ principal submatrix. This observation supports following approximation for the first term in (15):\(^2\)

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

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

$$
\mathcal{C}_{\text{obs}} \approx V_0 K C_K V_0^\top, \tag{17}
$$

such that $V_{0,K}$ is the leftmost $K$ column vectors of $V_0$ if $\eta_K \ll 1$. The above shows that the largest $K$ eigenvectors of $\mathcal{C}_{\text{obs}}$ roughly span the same subspace as that of $V_0 V_{0,K}$.

Now, suppose that we have an estimate of $V_0 K$, e.g., by As

Fig. 3. Scatter plots of the rows of the top-3 eigenvectors $\{\psi_{\text{row}}^{(i)}\}_{i=1}^N$ 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 $\psi_{\text{row}}^{(i)}$.

sumption 1, and we aim at inferring the communities $V_{0,K}$ from the partially observed graph signals $\{V^{(i)}_{\text{obs}}\}_{i=1}^N$. For the special case when $V_{0,K}$ is some positive semidefinite (PSD) ma-

and we highlighted the uniform samples taken from them. For

\(^2\)For the approximation to hold, we require a mild condition on $B$ such that $\min_{i=1,\ldots,K} \|B^\top u_i\| \geq \max_{j=k+1,\ldots,N} \|B^\top u_j\|$ and $u_i$ is the $i$th column of $U$. For the special case when $U = V$, this condition states that the energy of $BB^\top$ in the low frequency component is at least comparable to those in the high frequency components.

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).

The above shows that the largest $K$ eigenvectors of $\mathcal{C}_{\text{obs}}$ roughly span the same subspace as that of $V_0 V_{0,K}$. This observation is supported by Fig. 3 which shows the scatter plots of the rows of $V_K$ generated from two SBM graphs and we highlighted the uniform samples taken from them.

\(^3\)To simplify notations, we assume that the partition $(C_{\text{obs},1}^*, \ldots, C_{\text{obs},K}^*)$ has been permuted to match with $(C_1^*, \ldots, C_K^*)$ to give the smallest $\delta_{cb}$.

\(^4\)That is, when the normalized cut defined in (2) is small.

\(^5\)where $\kmeans(V_{0,K}, K)$ and

\[\delta_{cb} := \max_{k \in [K]} \|v_{k}^* - v_{\text{obs},k}\|, \tag{19a}\]

\[R_1 := \max_{k \in [K]} \max_{i \in C_{\text{obs}}^*} \|v_{i,\text{row}}^\top - v_{\text{obs},k}\|, \tag{19b}\]

\[R_2 := \max_{k \in [K]} \max_{i \in C_{\text{obs}}^* \setminus V_{\text{obs}}} \|v_{i,\text{row}}^\top - v_{\text{obs},k}\|, \tag{19c}\]

where we defined the sampled centroid vectors $v_{\text{obs},k}^* := \frac{1}{|C_{\text{obs}}^*|} \sum_{j \in C_{\text{obs}}^*} v_{j,\text{row}}^\top$ for each $k$.

The constant $\delta_{cb}$ 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, 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 $v_i$ is taken as an eigenvector of a population version of $L_{\text{sym}}$, we have $\delta_{cb} = 0$; see [38, Definition 2] and [32, Proposition 3].

Now, applying [38, Theorem 2.1] shows that the eigenvectors of
Algorithm 1: Blind Community Inference.

1: Input: set of partially observed graph signals \(\{y_{\text{obs}}^{(t)}\}_{t=1}^{m}\); number of desirable clusters \(K\).
2: Compute the sampled covariance:
\[
\widehat{C}_{\text{obs}} = (1/m) \sum_{t=1}^{m} y_{\text{obs}}^{(t)} (y_{\text{obs}}^{(t)})^\top.
\]
3: Find the largest \(K\) eigenvectors of \(\widehat{C}_{\text{obs}}\), \(\hat{Q}_K \in \mathbb{R}^{n \times K}\).
4: Set \((\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}) = \text{kmeans}(\hat{Q}_K, K)\).
5: Output: estimated partition \(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}\).

\(L_{\text{syn}}\) and its population version coincide as \(N \to \infty\), implying that \(\delta_{\text{ce}} \to 0\). 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 \(V_{\text{obs},K}\) through \(\text{kmeans}(V_{\text{obs},K}, K)\), we can tackle Task A if \(\delta_{\text{ce}}, R_1, R_2 \ll 1\). To see this, we define the surrogate \(K\)-means objective function:
\[
\tilde{F}_O(C_{\text{obs},1}, \ldots, C_{\text{obs},K}) := \frac{1}{K} \sum_{k=1}^{K} \|v_{k}^{\text{row}} - \hat{v}_{k}\|^2\tag{20}
\]

such that \(C_{\text{obs},1}, \ldots, C_{\text{obs},K}\) is any partition of \(V_{\text{obs}}\) and \(v_k\) was defined in (18). Note that unlike \(F(\cdot; V_{\text{obs}})\) [cf. (4)], the centroid vector for each community is fixed at \(v_k\).

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

Lemma 1: Let \((C^*_{\text{obs},1}, \ldots, C^*_{\text{obs},K}) = \text{kmeans}(V_{\text{obs},K}, K)\). Then,
\[
|\tilde{F}_O(V_{\text{obs}} \cap C^*_{1}, V_{\text{obs}} \cap C^*_{2}) - \tilde{F}_O(C_{\text{obs},1}, \ldots, C_{\text{obs},K})| \leq 2N(R_1 + R_2 + \delta_{\text{ce}}) \delta_{\text{ce}}\tag{21}
\]

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 \(\hat{Q}_K\) 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 \(O(n^2m)\) floating point operations (flips), and the top-\(K\) eigenvectors are found in \(O(n^2K)\) flips.

The \(K\)-means clustering step is performed in flips \(O(2^{K+\log(K)})\) using [44], producing a \((1 + \epsilon)\) optimal solution to \(K\)-means, i.e., it finds a partition \(C^*_{1}, \ldots, C^*_{K}\) with \(\tilde{F}(C^*_{1}, \ldots, C^*_{K}; \hat{Q}_K) \leq (1 + \epsilon) \min_{C_1, \ldots, C^K} \tilde{F}(C_1, \ldots, C^K; \hat{Q}_K)\) [44]. The overall complexity is thus \(O(n^2(K + m))\) flips.

B. Performance Analysis

This section analyzes the performance of Algorithm 1 conditioned on the inputs \(\{y_{\text{obs}}^{(t)}\}_{t=1}^{m}\) generated by (13). Define the economy QR decomposition of \(V_{\text{obs},K}\) as:
\[
V_{\text{obs},K} = Q_K R_K,\tag{22}
\]

such that \(Q_K \in \mathbb{R}^{n \times K}\) is an orthogonal matrix which spans the range space of \(V_{\text{obs},K}\) and \(R_K\) is a upper triangular matrix. Define the following constant to be used in our analysis:
\[
\rho_{\text{gap}} := \lambda_{n-K} - \lambda_{n-K}(\hat{C}_{\text{obs}}) - \|\hat{C}^m_{\text{obs}} - \hat{C}_{\text{obs}}\|_2,\tag{23}
\]

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

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

1) \(\text{rank}(U_k^T C_2) = K, U_K\) is the K leftmost-columns of \(U\), defined in Assumption 1.

2) the constant \(\rho_{\text{gap}}\), 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 \(U_k^T\). For \(2\), we observe that typically \(\rho_{\text{gap}} > 0\) as the number of samples \(m\) grows and the noise is small, i.e., \(\|C_i\|_2 \ll 1\).

Our first analytical result is a bound on the difference between the range spaces of \(\hat{Q}_K\) and \(Q_K\):

Proposition 1: Suppose Assumption 1, 2 hold. Then
\[
\|\hat{Q}_K^\top Q_K^\top - Q_K^\top Q_K^\top\|_F \leq \sqrt{2K} \left(\sqrt{(2\|U_k^T B\|_2^2) + \|U_{N-K}^T B\|_2^2} + \lambda_{\text{gap}}\right) / \rho_{\text{gap}},\tag{24}
\]

where we have defined \(\gamma = \max\{\Sigma_{1}, \ldots, \Sigma_{N-K}\}\), and \(\hat{Q}_K\) 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 \(Q_K\) approximate \(Q_K = E_{\text{obs}} V_{\text{obs}} U_K^\top\), which spans the same subspace as the one spanned by the row-sampled version of \(V_K\). The approximation quality improves as the lowpass ratio \(\eta_K\) decreases and the number of samples \(m\) increases. Next, we benchmark the output \(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}\) of Algorithm 1 via the \(K\)-means objective function \(F(C_1, \ldots, C_K; V_{\text{obs}})\) [cf. (4)].

Theorem 1: Assuming that the \(K\)-means in line 4 of Algorithm 1 outputs an \((1 + \epsilon)\) optimal solution. Then
\[
\sqrt{F(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}; V_{\text{obs}})} - \sqrt{(1 + \epsilon) F_0^*} \leq \|R_K\|_2 (2 + \epsilon) \|Q_K Q_K^\top - \hat{Q}_K \tilde{Q}_K^\top\|_F,\tag{25}
\]

where \(F_0^* = \min F(C_{\text{obs},1}, \ldots, C_{\text{obs},K}; V_{\text{obs}}), i.e., the optimal objective value attained by \text{kmeans}(V_{\text{obs}}, K)\).

The proof is provided in Appendix C. Notice that \(\|Q_K Q_K^\top - \hat{Q}_K \tilde{Q}_K^\top\|_F\) 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 $k\text{means}(V_{o,K}, K)$. This bound depends on two factors as seen in the right hand side of (24): the first factor is bounded as $O(\eta_K)$ 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 $(y^{(l)}_t)_{t=1}^m$) [45, Ch. 4] as

$$
\|C_{\text{obs}} - C\|_2 = O(\sqrt{\log(1/\delta)/m + \|C_e\|_2}).
$$

(26)

This bound decreases with the number of observed graph signals $m$ and depends on noise variance and the sampling pattern of nodes. We approximate the task $B$ by its surrogate

$$
\hat{A}_{o,K} \approx V_{o,K}\sqrt{\lambda_{\text{obs},1}} + \epsilon,
$$

(27)

and observe that $\lambda_{\text{obs},1}$ is a natural method for interpolating its eigenvectors through approximately solving the eigenvalue problem $\hat{A}_{o,K}v = \lambda_{\text{obs},1}v + \epsilon$. Specifically, take $c_L \in [\lambda_N, 2]$ be a user-designed constant and consider a flipped Laplacian matrix as

$$
\tilde{L}_{\text{sym}} = c_L I - \tilde{L}_{\text{sym}}.
$$

(28)

The column vectors of $\tilde{V}_K$ are the eigenvectors of $\tilde{L}_{\text{sym}}$ with the largest $K$ eigenvalues; the eigenvalues are $\{\hat{\lambda}_i\}_{i=1}^N = \{c_L - \lambda_i\}_{i=1}^N$. Now, suppose we are given the row sampled version of $\tilde{V}_K$, i.e., $V_{o,K} = E_{\text{obs}}\tilde{V}_K$. Let $k \in \{1, \ldots, K\}$ and $v_k$ be the kth column of $V_K$. Then,

$$
\tilde{L}_{\text{sym}}v_k = (c_L - \lambda_k)v_k.
$$

(29)

By expanding (29), the Nyström extension estimates the elements in $v_k$ for the hidden nodes, $i \in \text{hid}$, by

$$
v_{i,k} = \frac{1}{c_L - \lambda_k} \left( \sum_{j \in V_{\text{obs}}} [L_{\text{sym}}]_{ij} v_{j,k} + \sum_{j \in \text{hid}} [L_{\text{sym}}]_{ij} v_{j,k} \right),
$$

(30)

where the Nyström approximation is obtained by discarding the second summation. In the above, we have used the definition for $L_{\text{sym}}, \tilde{L}_{\text{sym}}$, as we recognized that $[-L_{\text{sym}}]_{ij}$ with $i \in \text{hid}, j \in V_{\text{obs}}$ can be mapped to the submatrix $A_{o,0}^{\text{norm}}$. An alternative approximation can be obtained by exploiting the structure of $L_{\text{sym}}$. Specifically, take $c_L = 2$ and observe that as $\tilde{L}_{\text{sym}} = I + A_{o,0}^{\text{norm}}$, we can retain the diagonal component of the second summation in (30) by the following approximation:

$$
v_{i,k} \approx \frac{1}{2 - \lambda_k} \left( \sum_{j \in V_{\text{obs}}} [A_{o,0}^{\text{norm}}]_{ij} v_{j,k} + v_{i,k} \right),
$$

(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 $V_{o,K}$ and the eigenvalues $\lambda_k$ are unknown, we replace $V_{o,K}$ by its surrogate $\hat{Q}_K$ used in Task $A$. Observe that $\lambda_k \approx 0$ for $1 \leq k \leq K$ when
the graph is \( K \)-modular. Applying \( \lambda_k \approx 0 \) to (30), (31) yields the following estimate of \( V_K \):

\[
\tilde{V}_K \approx \begin{pmatrix} Q_K \\ A_{h,o}^{\text{norm}} \tilde{Q}_K \end{pmatrix}, \quad \tilde{V}_K^{\text{imp}} := \begin{pmatrix} Q_K \\ A_{h,o}^{\text{norm}} \tilde{Q}_K \end{pmatrix}.
\]

(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 \( V_K \) achieves reasonable performance with \( \tilde{V}_K^{\text{imp}} \) 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 \( \tilde{V}_K^{\text{imp}} \) or \( \tilde{V}_K^{\text{sym}} \). The final clustering step is performed on the orthogonalized eigenmatrix \( \tilde{V}_K \) of \( V_K \).

Related algorithms have been proposed in the data clustering literature [47–49], 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 \( O(n^2 (K + m)) \) flops to obtain \( \hat{Q}_K \). In addition, the Nyström step in (32) takes \( O(n \min(\mathcal{A}_{h,o}) K) \) flops where \( n \min(\mathcal{A}_{h,o}) \) is the number of non-zeros in the matrix \( \mathcal{A}_{h,o}^{\text{norm}} \). Finally, the \( K \)-means step requires \( O(2^K / c^{(K)} (KN) \) flops. Overall, the algorithm requires \( O(Nn \min(\mathcal{A}_{h,o}) K + n^2 (K + m)) \) flops. In contrast, inferring partitions from fully observed graph signals involve a complexity of \( O(N^2 K) \) 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 \( V_K \). 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 \( V_{\text{obs}} \) 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 \( \{y_t\}_{t=1}^{m} \) given in (7) as fixed. We first define the following constants for facilitating our analysis: for any \( \Delta > 0 \),

\[
\delta_{\text{eig}}(\Delta) = \inf \left\{ \delta_{\text{eig}} : P(\max_{k \in [K]} |\lambda_k(L_{o,o})| \leq \delta_{\text{eig}}) \geq 1 - \Delta \right\}, \quad \rho_{\text{gap}}(\Delta) = \sup \{ \rho_{\text{gap}} : P(\rho_{\text{gap}} \geq \rho_{\text{gap}}) \geq 1 - \Delta \},
\]

(33)

where \( L_{o,o}, \rho_{\text{gap}} \) are defined in (12), (23), respectively, in which they are treated as random variables due to the selection of the observable node set \( V_{\text{obs}} \).

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

\[
L_{o,o} = I - E_{\text{obs}} D^{-1/2} AD^{-1/2} E_{\text{obs}}^\top
= I - D_{\text{obs}}^{-1/2} E_{\text{obs}} A E_{\text{obs}}^\top D_{\text{obs}}^{-1/2},
\]

(34)

where \( D_{\text{obs}} \) is the diagonal matrix of the degrees of the nodes in \( V_{\text{obs}} \). Notice that \( E_{\text{obs}} A E_{\text{obs}}^\top \) is the adjacency matrix of the subgraph \( G[V_{\text{obs}}] \). Under Assumption 3 and assuming that \( G \) is \( K \)-modular, with high probability \( V_{\text{obs}} \) includes nodes from each cluster, forming a \( K \)-modular graph itself. Thus, the \( K \) largest eigenvalues of \( D_{\text{obs}}^{-1/2} E_{\text{obs}} A E_{\text{obs}}^\top D_{\text{obs}}^{-1/2} \) will be close to \( \frac{1}{n} \) as \( n \) nodes are selected to form \( G[V_{\text{obs}}] \). This leads to the observation that the \( K \) smallest eigenvalues of \( L_{o,o} \) are approximately \( \frac{N-n}{N} \), i.e., \( \delta_{\text{eig}}(\Delta) \approx \frac{N-n}{N} \). Furthermore, \( \rho_{\text{gap}}(\Delta) \) 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 sub-optimality 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 \( \rho_{\text{gap}}(\Delta) > 0 \), \( n \geq 4\mu K \log(3K/\delta) \), where \( \mu \equiv \frac{N}{K} \max_{j=1,\ldots,N} \|\tilde{V}_{1:j,k}\| \) such that \( \tilde{V}_{1:j,k} \) is the \( j \)th row of the matrix of first \( K \) eigenvectors of \( L_{\text{sym}} \), then with probability at least \( 1 - \delta \), Algorithm 2 with \( \tilde{V}_K \) [cf. (32)] satisfies

\[
\sqrt{F(\hat{C}_1, \ldots, \hat{C}_K; V_K)} - \sqrt{(1 + \epsilon)F^*} = \frac{2 + \epsilon}{c_L - \lambda_K} O \left( \delta_{\text{eig}}(\Delta) + \sqrt{K} \left( \eta_K + \frac{\|C_{\text{obs}} - C_{\text{sym}}\|_2}{\rho_{\text{gap}}(\Delta)} \right) \right) + \left( 1 + \sqrt{\frac{2N}{n}} \right)(c_L - \lambda_K + 1),
\]

(35)

where \( F^* \) is the minimum objective value of \( F(\cdot; V_K) \) in (4), and \( \lambda_{K+1} \) is the \( K + 1 \)th smallest eigenvalue of \( L_{\text{sym}} \).

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 \( L_{\text{sym}} \) and the error of approximating \( Q_K \) by \( \hat{Q}_K \) in line 2 of Algorithm 2. In our analysis, we adopt [48], [49] to show that the low-rank approximation of \( L_{\text{sym}} \) produced by the Nyström method admits an error of \( O(1/\sqrt{n}) \), where \( n \) is the number of observed nodes, and we apply Proposition 1 to treat the error between \( Q_K, \hat{Q}_K \).

Similar to the analysis for Task A, we recall that \( \|C_{\text{obs}} - C_{\text{sym}}\|_2 = O(\log(1/\delta)/m + \|C_{\text{e}}\|_2) \) with probability at least \( 1 - \delta \) [45, Ch. 4]. As such, when we take \( c_L = 1 \), (35) implies
that
\[
\sqrt{F(\hat{C}_1, \ldots, \hat{C}_K; V_K)} - \sqrt{(1 + \epsilon)F^*} = \mathcal{O} \left( \eta_K + \frac{\log(\delta^{-1})}{m} + \frac{\|C_e\|_2}{\rho_{gap}(\Delta)} + \frac{N - n}{N} + \sqrt{\frac{1}{n}} \right),
\]
(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 \(\delta_{\text{alg}}(\Delta) \approx \frac{N - n}{N}\), \(\lambda_K \approx 0\) and \(\lambda_{K+1} \approx 1\) for \(K\)-modular graphs. The latter follows from the intuition that \(A\) 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 \(c_L = 1\). The requirement \(n \geq 8\rho_K \log(3K/\Delta)\) 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 \(\mathcal{H}(S) = (I - 0.5S)^P\) and \(P\) is the order of diffusion. The input to the graph filter is \(z^{(t)} = Bz^{(t)}\) where \(B \in \mathbb{R}^{N \times R}\) is a random matrix with i.i.d. elements generated from \(\mathcal{N}(0, 1)\), and \(z^{(t)} \sim \mathcal{N}(0, I)\). The GSO is selected as the normalized Laplacian matrix \(S = L_{\text{sym}} = I - A^{\text{norm}}\). The graph is generated according to a planted partition stochastic block model denoted by \(\text{SBM}(N, K, a, b)\) 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 \(C_e = 10^{-4} I\). For benchmarking purpose, we compare the error rate of the detected communities/partitions with the SBM’s ground truth:

\[
P_c := \mathbb{E} \left[ \frac{1}{N} \min_{i : \pi([K] - [K]) \sum_{i=1}^N j \neq \pi(c_i), i \neq i_{\text{true}}} \right],
\]
(37)

where \(c_i\) is the output of Algorithm 1 or Algorithm 2, and \(c_i^{\text{true}}\) 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 \(L_{0,0}\) and full Laplacian \(L\); 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...
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 $L$ and applying [30] on the zero-padded graph signals. We fix $a = 8 \log N/N, b = a/16, I = 10$ 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 $V_{imp}^{\text{mp}}$ achieves a better performance than plain Nyström extension with $V_{nys}^{\text{nys}}$. The performance with $V_{imp}^{\text{mp}}$ 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 $A_{h,o}^{\text{norm}}$ used in Algorithm 2 is contaminated with noise. In particular, we consider using $\hat{A}_{h,o}^{\text{norm}} = A_{h,o}^{\text{norm}} + E_{h,o}$ where $E_{h,o} \in \mathbb{R}^{(N-n) \times n}$ 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 $\hat{A}_{h,o}^{\text{norm}}$.

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 $\ell$, defined as the ratio between closing and opening price, of each stock as the $\ell$th graph signal. Denote the $\ell$th graph signal by $y^{(\ell)} \in \mathbb{R}^N$ 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.
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. Under Assumption 1, expanding the squared norm and using the Cauchy-Schwarz inequality, we obtain for all $i$:

$$\|v_i^{\text{row}} - v_k^{*}\|^2 - \|v_i^{\text{row}} - v_{\text{obs,k}}^{*}\|^2 \leq 2\|v_i^{\text{row}} - v_{\text{obs,k}}^{*}\| \|v_k^{*} - v_{\text{obs,k}}^{*}\| + \|v_k^{*} - v_{\text{obs,k}}^{*}\|^2 \leq 2(R_1 + \delta_{ce})\delta_{ce}.\quad (38)$$

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

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 $i \in C_{\text{obs,k}}^{\text{row}}$ the following inequalities:

$$\|v_i^{\text{row}} - v_k^{*}\|^2 - \|v_i^{\text{row}} - v_{\text{obs,k}}^{*}\|^2 \leq 2\|v_i^{\text{row}} - v_{\text{obs,k}}^{*}\| \|v_k^{*} - v_{\text{obs,k}}^{*}\| + \|v_k^{*} - v_{\text{obs,k}}^{*}\|^2 \leq 2(R_1 + \delta_{ce})\delta_{ce}.\quad (38)$$

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.

### Table I

<table>
<thead>
<tr>
<th>$K$</th>
<th>GICS</th>
<th>DIS</th>
<th>HLD</th>
<th>SMI</th>
<th>EMT</th>
<th>PUR</th>
<th>EDU</th>
<th>COX</th>
<th>HSC</th>
<th>CAC</th>
<th>PIN</th>
<th>POM</th>
<th>SHT</th>
<th>TLT</th>
<th>VNY</th>
<th>HMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

### Table II

(Task A) Accuracy of finding communities compared to the GICS sector classification

<table>
<thead>
<tr>
<th></th>
<th>Algorithm 1</th>
<th>2-step with [15]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$ Score</td>
<td>0.7988</td>
<td>0.1598</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: Weizmann Institute of Science. Downloaded on May 19, 2022 at 11:10:12 UTC from IEEE Xplore. Restrictions apply.
Similarly, for all $i \in C^*_k$, it holds that
\[
\| v_{i}^{\text{row}} - v_{\text{obs}, k}^* \|^2 \leq \| v_{i}^{\text{row}} - v_k^* \|^2 + (2R_2 + \delta_{\text{ce}}) \delta_{\text{ce}}. \tag{39}
\]

Analyzing the surrogate objective function, we then have:
\[
\tilde{F}_c(C^{*}_{\text{obs},1}, \ldots, C^{*}_{\text{obs},K}) = \sum_{k=1}^{K} \sum_{i \in C^*_k} \| v_{i}^{\text{row}} - v_k^* \|^2 \
\leq \sum_{k=1}^{K} \sum_{i \in C^*_k} \left\{ \| v_{i}^{\text{row}} - v_{\text{obs}, k}^* \|^2 + (2R_1 + \delta_{\text{ce}}) \delta_{\text{ce}} \right\} \
\leq \sum_{k=1}^{K} \sum_{i \in C^*_k \cap \text{Obs}} \| v_{i}^{\text{row}} - v_k^* \|^2 + 2N (R_1 + \delta_{\text{ce}}) \delta_{\text{ce}},
\]
where (a) is due to that $C^{*}_{\text{obs},1}, \ldots, C^{*}_{\text{obs},K}$ is a $K$-means optimal solution with the centroid vector $v_{\text{obs}, k}^*$ and the fact that $\sum_{k=1}^{K} |C^*_k \cap \text{Obs}| = N$. This concludes the proof.

**APPENDIX B**

**PROOF OF PROPOSITION 1**

Observe the following bound:
\[
\| \tilde{Q}_K \tilde{Q}_K^T - Q_K Q_K^T \|_F \leq \| \overline{Q}_K \overline{Q}_K^T - Q_K Q_K^T \|_F + \| \tilde{Q}_K \tilde{Q}_K^T - \overline{Q}_K \overline{Q}_K^T \|_F, \tag{40}
\]
where we have defined $\overline{Q}_K$ as the largest $K$ eigenvectors of the noiseless covariance matrix $\Sigma_{\text{obs}}$.

To bound the first term in r.h.s. of (40), observe that $\overline{Q}_K$ can be obtained as the top-$K$ left singular vectors of $E_{\text{obs}} \Sigma U^T B$ [cf. (15)], and recall that $C_x = BB^T$. The latter can be decomposed as:
\[
\begin{align*}
V_{0,K} \Sigma_K U_K^T B + V_{0,N-K} \Sigma_{N-K} U_{N-K}^T B =: P \\
= T
\end{align*}
\]
we have defined the partitions of $\Sigma$, $U$ with $\Sigma = (\Sigma_K 0; 0 \Sigma_{N-K})$ and $U = (U_K U_{N-K})$, and we have $V_{0,N-K} = E_{\text{obs}} V_{N-K}$ with $V_{N-K}$ corresponding to the last $N - K$ eigenvectors of $L_{\text{sym}}$. On the other hand, the range space of $P$ is equivalent to that of $Q_K$ provided that $\text{rank}(U_K^T C_x^T) = K$ [cf. Assumption 2]. If we let $\tilde{Q}_K$ be the largest $K$ left singular vector of $P$, then we have $Q_K \tilde{Q}_K = \tilde{Q}_K Q_K$. We invoke [53, Th. 3], which is a variant of the Wedin’s theorem:
\[
\| Q_K Q_K^T - \tilde{Q}_K \tilde{Q}_K^T \|_F \\
= \| \tilde{Q}_K \tilde{Q}_K^T - Q_K Q_K^T \|_F = \| \sin(\Theta(Q_K, \tilde{Q}_K)) \|_F \\
\leq 2(\| P \|_2 + \| T \|_2) \min \left\{ \sqrt{\gamma K} \| T \|_2, \| T \|_F \right\}, \tag{42}
\]
where $\sigma_i(P)$ denotes the $i$th largest singular value of the matrix $P$ and $\| T \|_2$ denotes the spectral norm of the matrix $T$. It is easy to obtain:
\[
\| P \|_2 \leq \max \{ s_1, \ldots, s_K \} \| U_K^T B \|_2 \| V_{0,K} \|_2, \\
\| T \|_2 \leq \max \{ s_{K+1}, \ldots, s_N \} \| U_{N-K}^T B \|_2 \| V_{0,N-K} \|_2 \|, \tag{43}
\]
where we defined $s_k := \sigma_k \Sigma_{\text{obs}}$. Furthermore, we have
\[
\sigma_k(P) = \min \{ s_1, \ldots, s_K \} \sigma_k (V_{0,K} \text{Diag}(\hat{s}) U_K^T B) = \min \{ s_1, \ldots, s_K \} \sqrt{\lambda_k \left( V_{K}^{\text{sym}} \right)}, \tag{44}
\]
where we scaled $s_k := (s_1, \ldots, s_K)$ as $\hat{s} := \frac{1}{\min \{ s_1, \ldots, s_K \}} s_k \geq 1$. Moreover, we have
\[
V_{K}^{\text{sym}} := B^T U_K \text{Diag}(\hat{s}) V_{0,K} V_{0,K} \text{Diag}(\hat{s}) U_K^T B, \tag{45}
\]
where $A_1 \geq A_2$ indicates that $A_1 - A_2$ is positive semidefinite, the above inequality is obtained by (i) rewriting $\text{Diag}(\hat{h}) V_{0,K}^T V_{0,K} \text{Diag}(\hat{h})$ as $V_{0,K} V_{0,K} \cap (\hat{h} \hat{h}^T)$ and (ii) applying Schur’s theorem [54, Theorem 7.5.3] with the fact that $\hat{h} \hat{h}^T - I \succeq 0$. Consequently,
\[
\sigma_k(P) \geq \min \{ s_1, \ldots, s_K \} \sqrt{\lambda_k \left( B^T U_K V_{0,K} V_{0,K} U_K^T B \right)} \tag{46}
\]
such that $U_K$ contains the first $K$ columns of $U$. Combining (42), (43) and (46) gives the desired bound as
\[
\| Q_K Q_K^T - \tilde{Q}_K \tilde{Q}_K^T \|_F \leq 2\sqrt{K} \| U_K \|_2 + \| U_{N-K} \|_{2} \eta_K . \tag{47}
\]

To bound the second term in (40), we follow a similar derivation as in [30, Proposition 2]. First,
\[
\| \tilde{Q}_K \tilde{Q}_K^T - Q_K Q_K^T \|_F \leq \sqrt{2K} \| \tilde{Q}_K \tilde{Q}_K^T - Q_K Q_K^T \|_F . \tag{48}
\]
As $\rho_{\text{gap}} > 0$ [cf. (23)], applying the Weyl’s inequality and Davis-Kahan theorem [55, Sec. VII.3] show that:
\[
\| \tilde{Q}_K \tilde{Q}_K^T - Q_K Q_K^T \|_F \leq \| \tilde{C}_{\text{obs}} - \overline{C}_{\text{obs}} \|_2 . \tag{48}
\]
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 $\tilde{X} \in \mathbb{R}^{n \times K}$ associated with the partition
\( \hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K} \) found with the proposed method:

\[
\hat{X}_{ik} := \begin{cases} 
1/\sqrt{\hat{C}_{ik}}, & \text{if } i \in \hat{C}_{ik} \\
0, & \text{otherwise}.
\end{cases}
\] (49)

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

\[
F(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}; V_{\text{obs}}) = \| (I - X^* (X^*)^T )V_{\text{obs}} \|_F^2
\leq F(\hat{C}_1, \ldots, \hat{C}_K; V_{\text{obs}}), \forall \hat{C}_1, \ldots, \hat{C}_K.
\] (50)

We proceed by observing

\[
F(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}; V_{\text{obs}})^2 = \| (I - X^* (X^*)^T )V_{\text{obs}} \|_F^2
\leq \| (I - \hat{X}^* (\hat{X}^*)^T )V_{\text{obs}} \|_F^2
\leq \| (I - \hat{X}^* (\hat{X}^*)^T )Q_K V_{\text{obs}} \|_F^2
\leq \| Q_K V_{\text{obs}} - Q_K \hat{Q}_K \|_F^2
\leq \sqrt{1 + \epsilon} \| Q_K V_{\text{obs}} - Q_K \hat{Q}_K \|_F^2
\leq \| Q_K Q_K^T - Q_K \hat{Q}_K^T \|_F^2,
\] (51)

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

\[
\| (I - X^* (X^*)^T )Q_K \|_F = \| (I - \hat{X}^* (\hat{X}^*)^T )Q_K \|_F
\leq \| (I - \hat{X}^* (\hat{X}^*)^T )Q_K \|_F + \| Q_K Q_K^T - Q_K \hat{Q}_K^T \|_F
\leq \sqrt{1 + \epsilon} \| (I - X^* (X^*)^T )Q_K \|_F
\] (b)

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

\[
F(\hat{C}_{\text{obs},1}, \ldots, \hat{C}_{\text{obs},K}; V_{\text{obs}})^2 \leq \{ (1 + \epsilon)F_0 \}^{\frac{1}{2}} + \| R_K \|_2 (2 + \epsilon) \left\{ \| Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T \|_F \right\}.
\]

Lastly, \( \| Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T \|_F \) is upper bounded using Proposition 1. Combining terms conclude the proof.

\[ \] (53)

\textbf{APPENDIX D}

\textbf{PROOF OF THEOREM 2}

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

\[
\hat{X}_{ik} := \begin{cases} 
1/\sqrt{\hat{C}_ik}, & \text{if } i \in \hat{C}_k \\
0, & \text{otherwise}.
\end{cases}
\]

Similar to (51), we observe that

\[
F(\hat{C}_1, \ldots, \hat{C}_K; V_K)^2 = \| (I - X^* (X^*)^T )Q_K \|_F^2
\leq \| (I - \hat{X}^* (\hat{X}^*)^T )Q_K \|_F^2
\leq \sqrt{1 + \epsilon} \| (I - X^* (X^*)^T )Q_K \|_F^2
\leq \| Q_K Q_K^T - Q_K \hat{Q}_K^T \|_F^2
\] (a)

where (a) uses the fact that \( \hat{X}^* \) 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 \( L_{o,0} = Q_A Q_A^T \). The following matrix is the Nyström rank-\( K \) approximation of \( \tilde{L}_{\text{sym}} \) [48]:

\[
L_{\text{sym},K} = \left( \begin{array}{ll}
Q_K & A_{h,0}^{-1} \tilde{A}_{obs,K} \end{array} \right) \Lambda_{obs,K} \left( \begin{array}{l}
Q_K \\
A_{h,0}^{-1} \tilde{A}_{obs,K}
\end{array} \right)^\top
\] (55)

where \( Q_K \) is the largest \( K \) eigenvectors of \( 2I - L_{o,0} \) and \( \tilde{A}_{obs,K} = 2I - A_{obs,K} \) such that \( \Lambda_{obs,K} \) is the principal \( K \) submatrix of \( A_{obs} \).

On the other hand, \( \tilde{V}_K \) has the same range space as

\[
\tilde{L}_{\text{sym},K} = \left( \begin{array}{l}
\tilde{Q}_K \\
2^{-1}A_{h,0}^{-1} \tilde{A}_{obs,K} \end{array} \right) \Lambda_{obs,K} \left( \begin{array}{l}
\tilde{Q}_K \\
2^{-1}A_{h,0}^{-1} \tilde{A}_{obs,K}
\end{array} \right)^\top
\] (56)

where we recall that \( \tilde{Q}_K \) 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

\[
\| \tilde{V}_K \tilde{V}_K^\top - V_K V_K^\top \|_F \leq \sqrt{2K \lambda_{\text{min}} \tilde{L}_{\text{sym},K} - \tilde{L}_{\text{sym}} \|_2},\] (57)

where we recall that \( \lambda_{\text{min}} \) is the \( K \)th smallest eigenvalue of \( L_{\text{sym}} \). The triangular inequality gives

\[
\| \tilde{L}_{\text{sym},K} - \tilde{L}_{\text{sym}} \|_2 \leq \| \tilde{L}_{\text{sym},K} - L_{\text{sym},K} \|_2
+ \| L_{\text{sym},K} - \tilde{L}_{\text{sym}} \|_2.
\] (58)

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

\[
\| \tilde{L}_{\text{sym},K} - L_{\text{sym},K} \|_2
\leq \| Q_K \Lambda_{obs,K} Q_K^\top - \hat{Q}_K \tilde{A}_{obs,K} \|_2
+ 2 \| \left( Q_K Q_K^\top - \hat{Q}_K \tilde{A}_{obs,K} \right) (A_{h,0}^{-1}) \|_2
+ \| A_{h,0}^{-1} \left( Q_K \Lambda_{obs,K} Q_K^\top - \hat{Q}_K \tilde{A}_{obs,K} \right) (A_{h,0}^{-1}) \|_2
\leq T_1 + T_2 + T_3.
\]
The following holds with probability at least $1 - \Delta$,

$$T_1 \leq 2\|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_2 + 2\|I - \tilde{A}_{obs,K}\|_2 \leq 2 \left(\|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F + \tilde{\delta}_{eig}(\Delta)\right),$$

(59)

where we have used the norm equivalence $\|\cdot\|_2 \leq \|\cdot\|_F$ for symmetric matrix. Moreover,

$$T_2 \leq \|A_{h,o}^{\text{norm}}\|_2 \|Q_K Q_K^T - \frac{1}{2} \hat{Q}_K \hat{A}_{obs,K} \hat{Q}_K^T\|_2.$$  

(60)

Since

$$\|Q_K Q_K^T - \frac{1}{2} \hat{Q}_K \hat{A}_{obs,K} \hat{Q}_K^T\|_2 \leq \|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F + \|\hat{Q}_K (I - \frac{1}{2} \hat{A}_{obs,K}) \hat{Q}_K^T\|_2,$$

we have

$$T_2 \leq \|A_{h,o}^{\text{norm}}\|_2 \left(\|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F + \frac{\tilde{\delta}_{eig}(\Delta)}{2}\right).$$

(61)

Finally,

$$T_3 \leq \|A_{h,o}^{\text{norm}}\|_2 \|Q_K \hat{A}_{obs,K} Q_K^T - \frac{1}{2} \hat{Q}_K \hat{A}_{obs,K} \hat{Q}_K^T\|_2,$$

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

$$\|Q_K \hat{A}_{obs,K} Q_K^T - \frac{1}{2} \hat{Q}_K \hat{A}_{obs,K} \hat{Q}_K^T\|_2 \leq \frac{1}{2} \left(\|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F + \|2 \hat{A}_{obs,K} - I\|_2 + \|I - \frac{1}{2} \hat{A}_{obs,K}\|_2\right),$$

(62)

we obtain

$$T_3 \leq \frac{1}{2} \|A_{h,o}^{\text{norm}}\|_2 \left(\|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F + \frac{3}{2} \tilde{\delta}_{eig}(\Delta)\right).$$

(63)

Combining the above observations, if we denote that $e_K := \|Q_K Q_K^T - \hat{Q}_K \hat{Q}_K^T\|_F$, then

$$L_{\text{sym}, K} - L_{\text{sym}, K} \leq (2 + \|A_{h,o}^{\text{norm}}\|_2 + \|A_{h,o}^{\text{norm}}\|_2^2) e_K + \tilde{\delta}_{eig}(\Delta).$$

(64)

Note that $e_K$ can be bounded using Proposition 1 as $O\left(\sqrt{n} \log(3K/\Delta)\right)$.

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 $1 - \Delta$,

$$\|L_{\text{sym}, K} - \tilde{L}\|_{\text{sym}} \leq 1 + \sqrt{\frac{2}{n}} \left(\sqrt{n} - \lambda_{K+1}\right),$$

(66)

provided that $n \geq 8K \log(3K/\Delta)$, where $\mu$ is a coherence factor for the top-$K$ eigenvector of $L_{\text{sym}}$. The bound is obtained from [49, Lemma 8] by setting $q = 1, \epsilon = 1/2$ and observing the bound $\|L_{\text{sym}} - \hat{L}_{\text{sym}}\|_2 \leq c_L - \lambda_{K+1}$.

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.


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 with 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 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.