COMMUNITY INFERENCE FROM GRAPH SIGNALS WITH HIDDEN NODES

Hoi-To Wai[a], Yonina C. Eldar[b], Asuman E. Ozdaglar[c], Anna Scaglione[d]

[a]Dept. of SEEM, The Chinese Univ. of Hong Kong, HKSAR. [b]Faculty of EE, Technion, Haifa, Israel, [c]LIDS, MIT, MA, USA, [d]School of ECEE, ASU, AZ, USA.

E-mails: htwai@se.cuhk.edu.hk, yonina@ee.technion.ac.il, Anna.Scaglione@asu.edu, asuman@mit.edu

ABSTRACT

Many recent works on inference of graph structure assume that the graph signals are *fully observable*. For large graphs with thousands or millions of nodes, this entails high complexity on the data collection and processing steps. Here, we study a community inference problem on partially observed (sub-sampled) graph signals which sidesteps topology inference, while revealing the coarse structure of the graph directly. Two variants of the inference task are studied: (i) a blind method that infers the communities that the observable nodes belong to; and (ii) a semi-blind method that infers the communities of all nodes using, in addition, side information about the sub-graph between observable and hidden nodes. These techniques for community inference are shown to be efficient and suitable for large graphs analytically and empirically.

***Index Terms*—** topology inference, graph signal processing, community inference, hidden nodes

1. INTRODUCTION

With recent advents in network science, graph signal processing (GSP), which is a formal extension of signal processing to irregularly structured data, has emerged to be an effective technique for data analytics. Many types of data such as actions/opinions of individuals, activity levels of regions in human brain, expression levels of genes, are examples of *graph signals* where each record is a vector of samples taken from nodes on a graph [1].

Graph signal data often arise from physical systems such as social and biological networks, where prediction and control are often desired. In these problems, inferring *structural properties* of the graph, e.g., topology or community structure, from the available graph signals is an important first step. A popular approach is to focus on the topology inference problem and tackle the inference task as an *inverse problem*. Compared to prior methods based on conditional independence [2] or dynamics equations [3–6], the GSP-inspired topology inference methods provide a unified way to infer graph topology using data collected from complex dynamics. Prior works show that it is possible to infer graph topology under conditions on smoothness of observed signals [7–9], or full-rank input [10–12]. However, it is assumed that the graph signals are *fully observable*, namely each node on the graph has to be sampled. For large graphs with thousands of nodes, collecting and processing these graph signals can be difficult due to the prohibitive computation complexity or the overwhelming efforts spent on data collection. It is therefore natural to consider the inference task with *partially observed* graph signals, where only a subset of nodes are sampled. Here, we consider an inference task with *hidden nodes*. While the signals on these hidden nodes are not observed, their presence affects the inference performance since they are connected to the observed nodes. Recently, some progress has been made in characterizing the impact of hidden nodes on the topology inference task [13], however the proposed algorithms are either without theoretical guarantee [14], or requires time series data to evaluate the time lag correlation matrix [15].

A common feature of social/biological networks is that they admit a *community structure* where the graph can be partitioned into densely connected clusters that are isolated from each other [16]. Departing from topology inference, we focus on inferring communities from graph signals, a problem which was first studied by the authors in [17, 18]. The goal is to infer a partition of such clusters while sidestepping topology inference. The intuition is that community structure offers a compact description of the graph, and therefore requires less information for its inference. As such, [18] shows that accurate inference can be performed when all nodes are observed even under the relaxed condition of low-rank graph signals, a situation where previous methods are poised to fail.

Here we extend [18] by considering the community inference problem under the impact of hidden nodes. In particular, we show that community inference using partially observed graph signals can be performed reliably while offering substantial reduction in computation complexity. The contributions of this paper are two-fold. First, we derive conditions under which the communities inherited from the complete graph can be inferred accurately from partially observed graph signals, leading to a *blind community inference* method for the observed nodes. Essentially, we prove that influences from the hidden nodes are absorbed into a noise subspace which our inference algorithm is agnostic to. Second, provided that side information on the topology is given, we derive a low complexity technique using Nyström method [19, 20] for community inference of all nodes, leading to a *semi-blind community inference* method.

Notations. We use boldfaced lower-case (upper-case) characters to denote vectors (matrices). For a vector $\boldsymbol{x} \in \mathbb{R}^n$, $[\boldsymbol{x}]_i$ denotes its ith element. $\mathrm{Diag}(\cdot)$ is the diagonalization operator that acts on a vector, and $(\cdot)^\top$ denotes the vector/matrix transpose.

2. SYSTEM MODEL

Consider a network specified by an undirected graph $G = (V, E)$ where the node set is $V = \{1, ..., N\}$, and the edge set is $E \subset V \times V$. The graph is defined by a symmetric adjacency matrix $\boldsymbol{A} \in \mathbb{R}_+^{N\times N}$ where $A_{ij} > 0$ if and only if $(i, j) \in E$. We focus on the scenario in which the network contains K communities, which we define as follows. Let $\mathcal{C}_1, ..., \mathcal{C}_K$ be a partition of V into K non-overlapping subsets. Consider the following *modularity* measure:

$$\mathsf{RatioCut}(\mathcal{C}_1, ..., \mathcal{C}_K) := \sum_{k=1}^K \frac{1}{|\mathcal{C}_k|} \sum_{i\in\mathcal{C}_k} \sum_{j\notin\mathcal{C}_k} A_{ij} \ , \quad (1)$$

which is the total weight of edges across subsets. We define

$$\mathcal{C}_1^\star, \ldots, \mathcal{C}_K^\star \in \arg\min_{\mathcal{C}_1,\ldots,\mathcal{C}_K} \ \mathsf{RatioCut}(\mathcal{C}_1, ..., \mathcal{C}_K), \quad (2)$$

as the optimal partition of V. The graph G is said to have a modularity of $\delta^\star$ when $\mathsf{RatioCut}(\mathcal{C}_1^\star, ..., \mathcal{C}_K^\star) = \delta^\star$. Throughout the paper, we assume $\delta^\star \approx 0$ and each of $\mathcal{C}_k^\star$ has similar size, such that the graph has K (nearly) isolated, equal size components / communities[1].

We are interested in the *community inference* problem, whose aim is to infer the partition (2) from observations on nodes of G. This is different from the *community detection* problem [16] which tackles (2) provided that the graph structure is *completely known*. Before describing our observation model, let $\boldsymbol{L} := \mathrm{Diag}(\boldsymbol{A}\mathbf{1}) - \boldsymbol{A}$ be the Laplacian of G with the eigenvalue decomposition (EVD) $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$. Then, a square matrix $\boldsymbol{\mathcal{H}}_G \in \mathbb{R}^{N\times N}$ is said to be a graph related operator if it admits a singular value decomposition (SVD) as:

$$\boldsymbol{\mathcal{H}}_G = \boldsymbol{V}\mathrm{Diag}(\boldsymbol{h})\boldsymbol{U}^\top, \tag{3}$$

where $\boldsymbol{h} \in \mathbb{R}_+^N$ is a vector controlling the *spectrum* of $\boldsymbol{\mathcal{H}}_G$ (whose order can be arbitrary), $\boldsymbol{V}$ is the orthogonal matrix containing the eigenvectors of $\boldsymbol{L}$, and $\boldsymbol{U}$ is an arbitrary orthogonal matrix. We shall fix the order of the eigenvectors in $\boldsymbol{V}$ as $\boldsymbol{V} := (\boldsymbol{v}_1 \cdots \boldsymbol{v}_N)$ where $\boldsymbol{v}_i$ is the eigenvector with the ith *smallest* eigenvalue in $\boldsymbol{L}$. Examples of this operator include graph filters [22] in which $\boldsymbol{\mathcal{H}}_G$ is a matrix polynomial of a graph shift operator (GSO) $\boldsymbol{S}$ such that $\boldsymbol{\mathcal{H}}_G = \sum_{t=0}^{T-1} h_t \boldsymbol{S}^t$ with the GSO taken as the Laplacian or Markov matrix [18, 22–24] since if $\boldsymbol{S} = \boldsymbol{L}$, then $\boldsymbol{\mathcal{H}}_G = \boldsymbol{V}\sum_{t=0}^{T-1} h_t \boldsymbol{\Lambda}^t \boldsymbol{V}^\top$.

Consider a set of m *graph signals*, *i.e.,* N-dimensional vectors whose elements correspond to nodes of G, given by:

$$\boldsymbol{y}^\ell = \overline{\boldsymbol{y}}^\ell + \boldsymbol{w}^\ell \;\text{ where }\; \overline{\boldsymbol{y}}^\ell = \boldsymbol{\mathcal{H}}_G \boldsymbol{x}^\ell,\; \ell = 1, ..., m, \tag{4}$$

where $\boldsymbol{x}^\ell$ is an excitation graph signal and $\boldsymbol{w}^\ell$ is a zero-mean, sub-Gaussian independent noise vector. As noted by [18, 22–24], a wide range of processes on networks, such as diffusion, opinion dynamics, equilibrium of quadratic games, etc., can be described as the mapping from $\boldsymbol{x}^\ell$ to $\overline{\boldsymbol{y}}^\ell$ in (4) with an appropriate form for $\boldsymbol{\mathcal{H}}_G$ satisfying (3). When $N \gg 1$, often times we only get to observe the graph signals on a subset of nodes. Partition the node set into an *observable* and a *hidden* part such that $V = V_{\mathsf{obs}} \cup V_{\mathsf{hid}}$ where $V_{\mathsf{obs}} = \{1, ..., n\}$ and $V_{\mathsf{hid}} = \{n+1, ..., N\}$. Accordingly, we partition $\boldsymbol{A}$ and $\boldsymbol{y}^\ell$ as:

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{A}_{\mathsf{o,o}} & \boldsymbol{A}_{\mathsf{o,h}} \\ \boldsymbol{A}_{\mathsf{h,o}} & \boldsymbol{A}_{\mathsf{h,h}} \end{pmatrix},\; \boldsymbol{y}^\ell := \begin{pmatrix} \boldsymbol{y}^\ell_{\mathsf{obs}} \\ \boldsymbol{y}^\ell_{\mathsf{hid}} \end{pmatrix} \tag{5}$$

such that $\boldsymbol{A}_{\mathsf{o,o}}$ (*resp.* $\boldsymbol{A}_{\mathsf{h,h}}$) correspond to the sub-graph induced by V_{obs} (*resp.* V_{hid}), and $\boldsymbol{A}_{\mathsf{o,h}} = \boldsymbol{A}_{\mathsf{h,o}}^\top$ corresponds to the bipartite graph between V_{obs} and V_{hid}. The signal $\boldsymbol{y}^\ell_{\mathsf{obs}}$ (*resp.* $\boldsymbol{y}^\ell_{\mathsf{hid}}$) is observed from V_{obs} (*resp.* V_{hid}). See the illustration in Fig. 1.

In this paper, we observe a set of n-dimensional signals $\{\boldsymbol{y}^\ell_{\mathsf{obs}}\}_{\ell=1}^m$. In this setting performing community inference of the graph G is difficult as observations on the $N-n$ hidden nodes are unknown. Under the premise that the number of communities, K, in G is known, we consider two community inference tasks:

T1. *(Blind community inference).* We partition V_{obs} into $\mathcal{C}^1_{\mathsf{obs}}, ..., \mathcal{C}^K_{\mathsf{obs}}$ *inherited* from the original community of G, *i.e.,* with $\mathcal{C}^k_{\mathsf{obs}} = V_{\mathsf{obs}} \cap \mathcal{C}_k^\star$ for all k, using only the observed signals $\{\boldsymbol{y}^\ell_{\mathsf{obs}}\}_{\ell=1}^m$. This is the closest community inference of G one could obtain in the absence of knowledge about the hidden nodes.

T2. *(Semi-blind community inference).* We recover the full partition (2) of V with the observed signals $\{\boldsymbol{y}^\ell_{\mathsf{obs}}\}_{\ell=1}^m$ *and* an estimate on the bipartite graph between hidden and observable nodes, $\boldsymbol{A}_{\mathsf{h,o}}$. For example, this side information is provided by an external source who can estimate the sub-graph's topology, e.g., via monitoring the tweets between users on Twitter.

[1] One may also define the community structure with other forms of modularity measures such as [21].

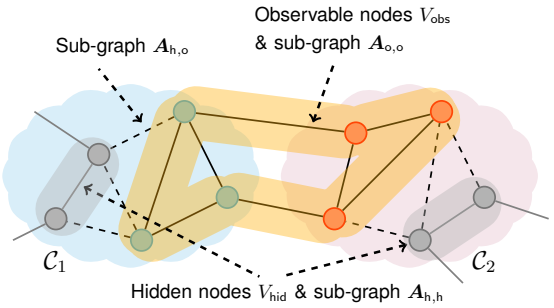

Fig. 1. Illustration of the configuration of G with $K = 2$. The nodes highlighted in orange are observable (V_{obs}), while the rest are hidden (V_{hid}). Note that the sub-graphs $\boldsymbol{A}_{\mathsf{o,o}}$, $\boldsymbol{A}_{\mathsf{h,h}}$ need not be connected.

If V_{obs} does not contain nodes from one of the communities, then it is impossible to tackle T1 and/or T2 since only $K-1$ communities could be formed from V_{obs}. To avoid such situations, we require that V_{obs} roughly samples an *equal* portion of nodes from all the K communities, *i.e.,* $|\mathcal{C}_k^\star \cap V_{\mathsf{obs}}|/|V_{\mathsf{obs}}| \approx |\mathcal{C}_k^\star|/|V|$ for $k = 1, ..., K$, e.g., with uniform sampling.

3. BLIND AND SEMI-BLIND COMMUNITY INFERENCE

To derive an inference method for the communities in G, let us borrow some insights from *community detection* methods for tackling (2). In particular, we observe that Problem (2) is equivalent to the discrete optimization problem [25, Section 5.2]:

$$\min_{\boldsymbol{P}\in\mathbb{R}^{N\times K}} \mathrm{Tr}(\boldsymbol{P}^\top \boldsymbol{L}\boldsymbol{P}) \;\text{ s.t. }\; \boldsymbol{P}^\top\boldsymbol{P} = \boldsymbol{I},\; P_{ij} \in \{0, \frac{1}{\sqrt{|\mathcal{C}_j|}}\}, \tag{6}$$

The binary constraint on P_{ij} can be relaxed to yield a tractable problem. Moreover, the optimal solution to the relaxed problem is precisely the smallest K eigenvectors of the Laplacian, $\boldsymbol{V}_K = (\boldsymbol{v}_1 \cdots \boldsymbol{v}_K)$. To retrieve an approximate solution to (2), one applies a K-means procedure on the row vectors of $\boldsymbol{V}_K$. This is the popular *spectral clustering* method [25].

In our model (4), the Laplacian matrix $\boldsymbol{L}$ is unknown. Let $\boldsymbol{x}^\ell$ be zero mean and its covariance matrix be given by $\boldsymbol{C}_x$. The covariance matrix of the graph signal $\overline{\boldsymbol{y}}^\ell$ is then

$$\boldsymbol{C}^{\mathsf{nl}} = \mathbb{E}[\overline{\boldsymbol{y}}^\ell(\overline{\boldsymbol{y}}^\ell)^\top] = \boldsymbol{V}\mathrm{Diag}(\boldsymbol{h})\boldsymbol{U}^\top\boldsymbol{C}_x\boldsymbol{U}\mathrm{Diag}(\boldsymbol{h})\boldsymbol{V}^\top. \tag{7}$$

By inspection, if $\min\{h_1, ..., h_K\} \gg \max\{h_{K+1}, ..., h_N\}$, then the top-$K$ eigenvectors of $\boldsymbol{C}^{\mathsf{nl}}$ are approximately $\boldsymbol{V}_K$ (up to a rotation), regardless of $\boldsymbol{U}$ and $\boldsymbol{C}_x$. Using this intuition, our previous work [18] proposed to infer communities from $\{\boldsymbol{y}^\ell\}_{\ell=1}^m$ by applying K-means on the top K eigenvectors of its sampled covariance matrix $\widehat{\boldsymbol{C}}_y$. In particular, [18] shows that this procedure successfully infers the partition in (2) when $\boldsymbol{\mathcal{H}}_G$ is taken as a *low-pass* graph filter defined on the Laplacian matrix [cf. Definition 1].

For large networks with $N \gg 1$, applying the method in [18] can be unfavorable. For instance, estimating the (approximately) rank K covariance matrix $\boldsymbol{C}_y = \mathbb{E}[\boldsymbol{y}^\ell(\boldsymbol{y}^\ell)^\top]$ requires $L = \mathcal{O}(K)$ samples for a fixed accuracy [26], each corresponding to acquiring the entire N dimensional graph signal $\boldsymbol{y}^\ell$. This data acquisition procedure can be difficult as one needs to sample the signal data on *every node*. Moreover, estimating the covariance matrix and computing the top K eigenvectors from it requires $\mathcal{O}(N^2K)$ time for a fixed EVD accuracy. It is undesirable as the runtime scales quadratically with N. In the following, we show that these shortcomings can be overcome through focusing on the partial graph signals $\{\boldsymbol{y}^\ell_{\mathsf{obs}}\}_{\ell=1}^m$.

3.1. Blind Community Inference [cf. T1]

Denote the partitions of the eigenvectors of the Laplacian $\boldsymbol{L}$ as:

$$\boldsymbol{V} := \begin{pmatrix} \boldsymbol{V}_{\mathsf{o}} \\ \boldsymbol{V}_{\mathsf{h}} \end{pmatrix} = (\boldsymbol{V}_K \;\; \boldsymbol{V}_{N-K}) = \begin{pmatrix} \boldsymbol{V}_{\mathsf{o},K} & \boldsymbol{V}_{\mathsf{o},N-K} \\ \boldsymbol{V}_{\mathsf{h},K} & \boldsymbol{V}_{\mathsf{h},N-K} \end{pmatrix}, \quad (8)$$

where $\boldsymbol{V}_{\mathsf{o},K}$ corresponds to the top n (observable) rows in the left most K eigenvectors. As observed by [25], when the RatioCut of the graph, $\delta^\star$, is close to zero, $\boldsymbol{V}_{\mathsf{o},K}$ contains columns of indicator vectors (with ± 1 elements) of the communities for nodes in V_{obs}. Under the regularity condition that $|\mathcal{C}_k^\star \cap V_{\mathsf{obs}}|/|V_{\mathsf{obs}}| \approx |\mathcal{C}_k^\star|/|V|$, the columns of $\boldsymbol{V}_{\mathsf{o},K}$ will be roughly orthogonal. In particular, let $\boldsymbol{V}_{\mathsf{o},K} = \boldsymbol{Q}_{\mathsf{o},K}\boldsymbol{R}_{\mathsf{o},K}$ denote the QR decomposition of $\boldsymbol{V}_{\mathsf{o},K}$ with $\boldsymbol{Q}_{\mathsf{o},K} \in \mathbb{R}^{n\times K}$, we observe that $\boldsymbol{R}_{\mathsf{o},K} \in \mathbb{R}^{K\times K}$ is close to diagonal. Therefore, clustering the row vectors of $\boldsymbol{Q}_{\mathsf{o},K}$ returns a partition of V_{obs} inherited from the original communities (2).

Motivated by the above discussions, we estimate $\boldsymbol{Q}_{\mathsf{o},K}$ from the partially observed graph signals $\{\boldsymbol{y}_{\mathsf{obs}}^\ell\}_{\ell=1}^m$. To do so, we propose to analyze the principal components of the sampled covariance matrix. The latter is defined along with its EVD as:

$$\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} := \tfrac{1}{m}\textstyle\sum_{\ell=1}^m \boldsymbol{y}_{\mathsf{obs}}^\ell(\boldsymbol{y}_{\mathsf{obs}}^\ell)^\top = \widehat{\boldsymbol{Q}}_{\mathsf{o}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{Q}}_{\mathsf{o}}^\top, \quad (9)$$

where $\widehat{\boldsymbol{\Sigma}} := \mathrm{Diag}([\widehat{\sigma}_1, ..., \widehat{\sigma}_n])$, with $\widehat{\sigma}_1 \geq \cdots \geq \widehat{\sigma}_n \geq 0$. To relate $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}}$ to $\boldsymbol{Q}_{\mathsf{o},K}$, from (7), the noiseless covariance matrix is:

$$\boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}} := \mathbb{E}[\overline{\boldsymbol{y}}_{\mathsf{obs}}^\ell(\overline{\boldsymbol{y}}_{\mathsf{obs}}^\ell)^\top] = \boldsymbol{V}_{\mathsf{o}}\mathrm{Diag}(\boldsymbol{h})\boldsymbol{U}^\top\boldsymbol{C}_x\boldsymbol{U}\mathrm{Diag}(\boldsymbol{h})\boldsymbol{V}_{\mathsf{o}}^\top, \quad (10)$$

where $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} \approx \boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}$ when the noise $\boldsymbol{w}^\ell$ is small. Consider the following low-pass property for $\boldsymbol{\mathcal{H}}_G$:

Definition 1 *The graph-related operator $\boldsymbol{\mathcal{H}}_G$ is said to be (K, η) low pass if its singular values satisfy*

$$\eta := \tfrac{\max\{h_{K+1},...,h_N\}}{\min\{h_1,...,h_K\}} < 1\,. \quad (11)$$

The above property is borrowed from the low-pass graph filter property studied in [18] and adapted to our setting of graph related operator. In the special case when $\boldsymbol{\mathcal{H}}_G$ is a graph filter, different cases for achieving $\eta \approx 0$ has been discussed in [18]. Intuitively, as G has K communities, the condition requires the social/biological network specified with G to *diffuse* information effectively among nodes that belong to the same community.

Suppose that $\eta \approx 0$ and using the decomposition $\boldsymbol{V}_{\mathsf{o},K} = \boldsymbol{Q}_{\mathsf{o},K}\boldsymbol{R}_{\mathsf{o},K}$, then $\boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}} \approx \boldsymbol{Q}_{\mathsf{o},K}\boldsymbol{M}\boldsymbol{Q}_{\mathsf{o},K}^\top$ for some $K \times K$ matrix $\boldsymbol{M}$, *i.e.,* the top K eigenvectors of $\boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}$ is approximately $\boldsymbol{Q}_{\mathsf{o},K}$ (up to a rotation). The above intuition is made precise by letting the top K eigenvectors of $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}}$ be $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$, and we observe:

Proposition 1 *Suppose that the graph-related operator $\boldsymbol{\mathcal{H}}_G$ is (K, η) low-pass, $\mathrm{rank}(\boldsymbol{U}_K^\top\boldsymbol{C}_x) = K$, and we have*

$$\delta := \lambda_K(\boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}) - \lambda_{K+1}(\boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}) - \|\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} - \boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}\|_2 > 0, \quad (12)$$

where $\boldsymbol{U}_K$ is the first K columns of $\boldsymbol{U}$, $\lambda_K(\boldsymbol{X})$ denotes the Kth largest eigenvalue of a matrix $\boldsymbol{X}$. Then $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$ satisfies

$$\begin{aligned} &\|\widehat{\boldsymbol{Q}}_{\mathsf{o},K}\widehat{\boldsymbol{Q}}_{\mathsf{o},K}^\top - \boldsymbol{Q}_{\mathsf{o},K}\boldsymbol{Q}_{\mathsf{o},K}^\top\|_{\mathrm{F}} \\ &= \mathcal{O}\left(\sqrt{K}\left(\tfrac{\max\{h_1,...,h_K\}}{\min\{h_1,...,h_K\}}\eta + \delta^{-1}\|\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} - \boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}\|_2\right)\right), \end{aligned} \quad (13)$$

where $\mathcal{O}(\cdot)$ hides the constant factors that are due to $\boldsymbol{U}_K, \boldsymbol{C}_x$.

The proofs in this section are omitted in the interest of space[2]. The conditions in Proposition 1 can be justified as follows. First, it holds in general that $\mathrm{rank}(\boldsymbol{U}_K^\top\boldsymbol{C}_x) = K$ as long as $\mathrm{rank}(\boldsymbol{C}_x) \geq K$. Second, (12) holds if the noise variance σ is small and the number of samples L is sufficient. Precisely, when the noise vector $\boldsymbol{w}^\ell$ is sub-Gaussian with a covariance satisfying $\|\mathbb{E}[\boldsymbol{w}^\ell(\boldsymbol{w}^\ell)^\top]\|_2 \leq \sigma$, then $\|\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} - \boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}\|_2$ converges to σ at a rate of $\mathcal{O}(\sqrt{K/m})$ [26]. Proposition 1 thus shows that when $\boldsymbol{\mathcal{H}}_G$ is low-pass with $\eta \approx 0$, $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$ approximates $\boldsymbol{Q}_{\mathsf{o},K}$ up to a rotation. Our result implies that influences from the hidden nodes are insignificant on $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$ as they are absorbed into the non-principal spectral components of $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}}$.

As a final step, we apply K-means on the row vectors of $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$, a procedure akin to applying spectral clustering on $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}}$, to obtain a partition of V_{obs} as $\widehat{\mathcal{C}}_{\mathsf{obs}}^1, ..., \widehat{\mathcal{C}}_{\mathsf{obs}}^K$. The solution quality of the proposed procedure can be characterized through:

$$F_{\mathsf{obs}}(\mathcal{C}_1, ..., \mathcal{C}_K) := \sqrt{\textstyle\sum_{k=1}^K \sum_{i\in\mathcal{C}_k} \left\|\boldsymbol{q}_i^{\mathrm{row}} - \frac{1}{|\mathcal{C}_k|}\sum_{j\in\mathcal{C}_k}\boldsymbol{q}_j^{\mathrm{row}}\right\|_2^2},$$

where $\boldsymbol{q}_i^{\mathrm{row}}$ is the ith row of the matrix $\boldsymbol{Q}_{\mathsf{o},K}$. We then have the following result:

Proposition 2 *Suppose the conditions in Proposition 1 hold, and the K-means algorithm outputs an $(1+\epsilon)$ optimal solution[3]. Then*

$$\begin{aligned} &F_{\mathsf{obs}}(\widehat{\mathcal{C}}_{\mathsf{obs}}^1, ..., \widehat{\mathcal{C}}_{\mathsf{obs}}^K) - \sqrt{1+\epsilon}\min_{\mathcal{C}^1,...,\mathcal{C}^K} F_{\mathsf{obs}}(\mathcal{C}^1, ..., \mathcal{C}^K) \\ &= \mathcal{O}\left((2+\epsilon)\sqrt{K}\left(\tfrac{\max\{h_1,...,h_K\}}{\min\{h_1,...,h_K\}}\eta + \tfrac{1}{\delta}\|\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}} - \boldsymbol{C}_{\mathsf{o},\mathsf{o}}^{\mathsf{nl}}\|_2\right)\right). \end{aligned}$$

The left hand side of the above measures the gap between the 'optimal' partition of V_{obs} based on $\boldsymbol{Q}_{\mathsf{o},K}$ and the partition inferred from $\{\boldsymbol{y}_{\mathsf{obs}}^\ell\}_{\ell=1}^L$. The gap is small under the same set of desirable conditions in Proposition 1. This shows that the inferred partition $\widehat{\mathcal{C}}_{\mathsf{obs}}^1, ..., \widehat{\mathcal{C}}_{\mathsf{obs}}^K$ approximates one that is based on the actual $\boldsymbol{Q}_{\mathsf{o},K}$.

Finally, we consider the special case when $V_{\mathsf{obs}} = V$. We note $\boldsymbol{Q}_{\mathsf{o},K} = \boldsymbol{V}_K$ and therefore the proposition bounds the sub-optimality of the partition $\widehat{\mathcal{C}}_{\mathsf{obs}}^1, ..., \widehat{\mathcal{C}}_{\mathsf{obs}}^K$ with respect to spectral clustering on the full Laplacian $\boldsymbol{L}$. In fact, in this scenario, our bound recovers [18, Theorem 1] with an extra factor of $\frac{\max\{h_1,...,h_K\}}{\min\{h_1,...,h_K\}}$.

3.2. Semi-blind Community Inference [cf. T2]

Our next task is to infer the partition (2) of the graph G using $\{\boldsymbol{y}_{\mathsf{obs}}^\ell\}_{\ell=1}^m$ and the partial adjacency matrix $\boldsymbol{A}_{\mathsf{h},\mathsf{o}}$. As argued before, for this task it suffices to estimate the K eigenvectors in $\boldsymbol{V}_K$. Observe from the previous section that one can readily obtain $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$ by performing a top-K EVD on $\widehat{\boldsymbol{C}}_{\mathsf{o},\mathsf{o}}$, which satisfies, up to a rotation and diagonal scaling, that $\widehat{\boldsymbol{Q}}_{\mathsf{o},K} \approx \boldsymbol{Q}_{\mathsf{o},K} \approx \boldsymbol{V}_{\mathsf{o},K}$ under the conditions in Proposition 1 and the regularity condition that $|\mathcal{C}_k^\star \cap V_{\mathsf{obs}}|/|V_{\mathsf{obs}}| \approx |\mathcal{C}_k^\star|/|V|$. Our challenge lies in how to incorporate the side information $\boldsymbol{A}_{\mathsf{h},\mathsf{o}}$ for estimating $\boldsymbol{V}_K$.

To this end, the Nyström method is effective for approximating the *top (largest)* eigenvectors of a large matrix [19, 20]. However, we note that the desired $\boldsymbol{V}_K$ corresponds to the smallest eigenvalues of $\boldsymbol{L}$ only. To apply the Nyström method, we consider $\widetilde{\boldsymbol{L}} := \lambda_N\boldsymbol{I} - \boldsymbol{L}$ with $\lambda_N = \lambda_{\max}(\boldsymbol{L})$, which is a positive semidefinite matrix where $\boldsymbol{V}_K$ are now the *top* K eigenvectors. In particular,

$$\textstyle\sum_{j\in V}\widetilde{L}_{ij}[\boldsymbol{v}_k]_j = (\lambda_N - \lambda_k)[\boldsymbol{v}_k]_i,\; i \in V,\; k \in [K]\,, \quad (14)$$

[2] See: www1.se.cuhk.edu.hk/~htwai/pdf/nystrom_app.pdf

[3] Using the data points $\boldsymbol{Q}_{\mathsf{o},K}$ as input, it finds a partition $\overline{\mathcal{C}}^1, ..., \overline{\mathcal{C}}^K$ with $F_{\mathsf{obs}}(\overline{\mathcal{C}}^1, ..., \overline{\mathcal{C}}^K) \leq (1+\epsilon)\min_{\mathcal{C}^1,...,\mathcal{C}^K} F_{\mathsf{obs}}(\mathcal{C}^1, ..., \mathcal{C}^K)$, see [27].

where $\widetilde{L}_{ij}$ is the (i,j)th element of $\widetilde{\boldsymbol{L}}$, $\boldsymbol{v}_k$ is the kth column vector of $\boldsymbol{V}_K$ with eigenvalue λ_k. As it holds that $\widehat{\boldsymbol{Q}}_{\mathsf{o},K} \approx \boldsymbol{V}_{\mathsf{o},K}$ where an approximation of $[\boldsymbol{v}_k]_i$ is available for all $i \in V_{\mathsf{obs}}$, substituting this into (14) gives the Nyström extension [19]:

$$[\widehat{\boldsymbol{v}}_k]_i \approx \tfrac{1}{\lambda_N - \lambda_k} \textstyle\sum_{j \in V_{\mathsf{obs}}} \widetilde{L}_{ij} [\widehat{\boldsymbol{q}}_k]_j, \; i \in V_{\mathsf{hid}} \; , \tag{15}$$

where $\widehat{\boldsymbol{q}}_k$ is the kth column vector of $\widehat{\boldsymbol{Q}}_{\mathsf{o},K}$. Furthermore, the graph has K communities, we have $\lambda_k \ll \lambda_{K+1} \leq \lambda_N$ and $\lambda_N - \lambda_k \approx \lambda_N$ for all $k = 1, ..., K$ [25]. Therefore, together with the fact that $\widetilde{L}_{ij} = A_{ij}$ for $i \in V_{\mathsf{hid}}, j \in V_{\mathsf{obs}}$, an approximation of $\boldsymbol{V}_K$ is:

$$\widehat{\boldsymbol{V}}_K := \begin{pmatrix} \widehat{\boldsymbol{Q}}_{\mathsf{o},K} \\ \widehat{\lambda}_N^{-1} \boldsymbol{A}_{\mathsf{h,o}} \widehat{\boldsymbol{Q}}_{\mathsf{o},K} \end{pmatrix} , \; \widehat{\lambda}_N := \frac{\mathbf{1}^\top \boldsymbol{A}_{\mathsf{h,o}} \mathbf{1}}{N - n} \; . \tag{16}$$

To tackle T2, we simply infer the partition (2) by performing K-means clustering on the row vectors of $\widehat{\boldsymbol{V}}_K$.

Obviously, the accuracy of the Nyström method depends on the ratio between the number of observable nodes n and the network size N. In the simulations, we show that reasonable performance can be attained by observing a small fraction (20%–30%) of the nodes.

3.3. Computation Complexity

The proposed methods for T1 and T2 are based on the top-K EVD of the sampled covariance matrix in (9). As it requires $m = \mathcal{O}(K)$ to obtain a reliable estimation, with a naive implementation, the computation complexity for forming the estimate and the EVD step is $\mathcal{O}(n^2 K)$ for a fixed accuracy. For T1, using the K-means algorithm in [27] outputs an $(1+\epsilon)$ optimal solution in time $\mathcal{O}(2^{(K/\epsilon)^{\mathcal{O}(1)}} Kn)$, *i.e.,* it is dominated by the EVD step. With a moderate K, the overall complexity of tackling T1 is $\mathcal{O}(n^2 K)$. For T2, the Nyström step (16) takes an additional time $\mathcal{O}(\mathrm{nnz}(\boldsymbol{A}_{\mathsf{h,o}})K)$ where $\mathrm{nnz}(\boldsymbol{A}_{\mathsf{h,o}})$ is the number of non-zeros in the matrix $\boldsymbol{A}_{\mathsf{h,o}}$. This is negligible when the graph is sparse. On the other hand, the K-means step requires $\mathcal{O}(2^{(K/\epsilon)^{\mathcal{O}(1)}} KN)$. With a moderate K, the overall complexity is $\mathcal{O}(n^2 K + 2^{(K/\epsilon)^{\mathcal{O}(1)}} KN)$, *i.e.,* linear in N.

Overall, the proposed methods are efficient in terms of data acquisition and computation. Particularly, estimating the covariance matrix $\widehat{\boldsymbol{C}}_{\mathsf{o,o}}$ requires only samples of n-dimensional graph signals, and the overall computation complexity of T2 is significantly lower than that of [18] since it is only quadratic in n (with $n \ll N$).

4. NUMERICAL EXPERIMENTS

We verify the claims in this paper using synthetic data. Throughout this section, the undirected graphs are generated according to a stochastic block model (SBM) $G \sim \mathsf{SBM}(N, K, a, b)$ such that G has $N = 150$ nodes and $K = 3$ communities, with $0 \leq b < a \leq 1$ where $a = 8 \log N/N$ (*resp.* $b = \log N/(2N)$) is the connection probability between nodes in the same community (*resp.* different communities). Weights on the adjacency matrix $\boldsymbol{A}$ are assigned to be 1 if $(i,j) \in E$, or 0 if otherwise. For the data model (4), we focus on the diffusion process with $\boldsymbol{\mathcal{H}}_G = (\boldsymbol{I} - \alpha \boldsymbol{L})^M$ and $\alpha = (1/2)\|\boldsymbol{A}\mathbf{1}\|_\infty^{-1}$; the excitation signal satisfies $\mathbb{E}[\boldsymbol{x}^\ell (\boldsymbol{x}^\ell)^\top] = \boldsymbol{B}\boldsymbol{B}^\top$ where $\boldsymbol{B} \in \mathbb{R}^{N \times R}$ is set with $\mathcal{N}(0,1)$ elements and $R = 10$ to simulate a situation with low-rank excitations in which previous methods will fail even when all nodes are observable. Notice that the low pass parameter is $\eta = \left(\frac{1-\alpha\lambda_{K+1}}{1-\alpha\lambda_K}\right)^M$, *i.e.,* decreasing to zero exponentially with M. The set of observable nodes V_{obs} is selected uniformly from $[N]$. We observe $m = 50$ samples with the noise vector generated through $\boldsymbol{w}^\ell \sim \mathcal{N}(\mathbf{0}, 10^{-1}\boldsymbol{I})$. The results are averaged over 5,000 trials and we used the built in `kmeans` function of MATLAB for the K-means procedure required in our method.

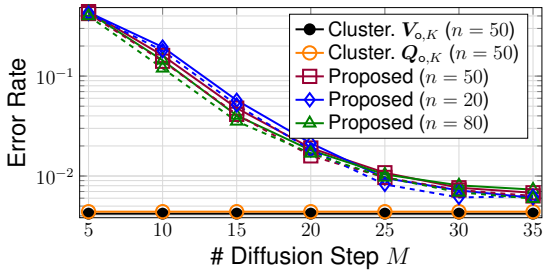

Fig. 2. Error rate performance comparison for task T1 against the number of diffusion steps M. Dashed lines refer to the noiseless setting with direct observation on $\boldsymbol{C}^{\mathsf{nl}}_{\mathsf{o,o}}$ [cf. (10)].

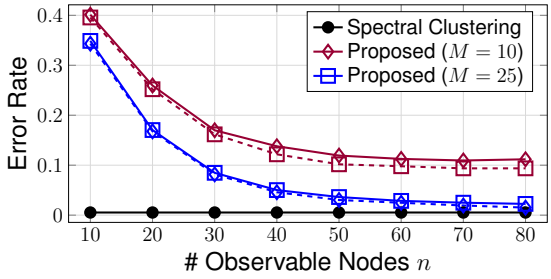

Fig. 3. Performance comparison for task T2 against $|V_{\mathsf{obs}}| = n$. Dashed lines refer to the noiseless setting with observation on $\boldsymbol{C}^{\mathsf{nl}}_{\mathsf{o,o}}$.

T1. Community inference of $G[V_{\mathsf{obs}}]$: In this example, we focus on the performance of tackling T1, where the inferred communities $\widehat{\mathcal{C}}^1_{\mathsf{obs}}, ..., \widehat{\mathcal{C}}^K_{\mathsf{obs}}$ are compared against the ground truth in generating the SBMs, e.g., $\mathcal{C}^\star_1 \cap V_{\mathsf{obs}}, ..., \mathcal{C}^\star_K \cap V_{\mathsf{obs}}$. The numerical results are reported in Fig. 2, which plots the error rate of the inferred community against the number of diffusion steps M elapsed before collecting each sample and we have fixed $|V_{\mathsf{obs}}| = n \in \{20, 50, 80\}$. As observed, the identical performance between clustering on $\boldsymbol{V}_{\mathsf{o},K}$ and $\boldsymbol{Q}_{\mathsf{o},K}$ justifies the observation made before Proposition 2 that it suffices to estimate $\boldsymbol{Q}_{\mathsf{o},K}$. This is further corroborated by the improved performance of the proposed methods with increasing M. We remark that the performance of our method for this task is insensitive to the size of observable nodes n as we observe similar performance over different values of n.

T2. Community inference of G: In this example, we consider tackling T2 with the Nystrom extension based method. Here, the goal is to infer the ground truth partition of nodes as generated from the SBM. We report the numerical results in Fig. 3, which shows the error rate against the number of observed nodes n. As expected, the error rate decreases as the number of observable nodes n increases. It improves to a reasonable level when $n \approx 40$, which is less than 30% of the nodes. Furthermore, the result suggests that the performance can be affected by the first stage in estimating $\boldsymbol{Q}_{\mathsf{o},K}$, as the error rate decreases with increased M.

5. CONCLUSION

In this paper, we have studied several techniques for community inference from observations of the graph signals on a subset of nodes, providing an estimate of the coarse network structure. Despite the influences from hidden nodes, we showed that the inference is accurate as long as the graph related operator is low-pass. The proposed methods are efficient for networks with a large number of nodes.

6. REFERENCES

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

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

[3] M. Timme, "Revealing network connectivity from response dynamics," *Physical review letters*, vol. 98, no. 22, p. 224101, 2007.

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

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

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

[7] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[8] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, 2017.

[9] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 6508–6512.

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

[11] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *arXiv preprint arXiv:1801.03862*, 2018.

[12] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 3, pp. 481–496, 2018.

[13] B. Ata, A. Belloni, and O. Candogan, "Latent agents in networks: Estimation and pricing," *arXiv preprint arXiv:1808.04878*, 2018.

[14] T. Xie, S. Liu, and A. O. Hero, "Semiblind subgraph reconstruction in gaussian graphical models," in *Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on*. IEEE, 2017, pp. 628–632.

[15] V. Matta and A. H. Sayed, "Consistent tomography under partial observations over adaptive networks," *IEEE Transactions on Information Theory*, 2018.

[16] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[17] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Community detection from low-rank excitations of a graph filter," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4044–4048.

[18] ——, "Blind community detection from low-rank excitations of a graph filter," *arXiv preprint arXiv:1809.01485*, 2018.

[19] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the nystrom method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 2, pp. 214–225, 2004.

[20] M.-A. Belabbas and P. J. Wolfe, "Spectral methods in machine learning and new strategies for very large datasets," *Proceedings of the National Academy of Sciences*, pp. pnas–0 810 600 105, 2009.

[21] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

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

[23] A. Heimowitz and Y. C. Eldar, "A unified view of diffusion maps and signal processing on graphs," in *SampTA*, July 2017.

[24] ——, "A unified view of diffusion maps and signal processing on graphs," *arXiv preprint arXiv:1806.03174*, 2018.

[25] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec 2007.

[26] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.

[27] A. Kumar, Y. Sabharwal, and S. Sen, "A simple linear time $(1+\epsilon)$-approximation algorithm for k-means clustering in any dimensions," in *FOCS*. IEEE, 2004, pp. 454–462.

[28] 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.