JOINT CENTRALITY ESTIMATION AND GRAPH IDENTIFICATION FROM MIXTURE OF LOW PASS GRAPH SIGNALS

Yiran He, Hoi-To Wai

Department of SEEM, The Chinese University of Hong Kong, Shatin, Hong Kong SAR of China

ABSTRACT

This paper proposes a mixture model of low pass filtered graph signals. Our aim is to jointly estimate the eigen-centrality vectors for the underlying graphs and identify the graph signal samples with their corresponding graphs, without knowing the graph topology a-priori. The problem is challenging as the observed graph signals lack any obvious identity with their associated graphs. We leverage a low-rank plus sparse structure of the unknown parameters to derive a customized expectation-maximization (EM) algorithm for the joint problem. Our algorithm assumes general excitation and does not require prior knowledge of the graph topology. Numerical experiments show the efficacy of our customized EM algorithm.

***Index Terms*—** blind centrality estimation, graph identification, expectation maximization

1. INTRODUCTION

Graph signal processing (GSP) [1,2] has grown interest in social science, biology, and data science as the demand for analyzing graph-structured data rises in the recent years. Among others, an interesting problem is graph learning that infers the graph topology from observations on the nodes, i.e., graph signals, where popular algorithms based on smoothness [3], or spectral template [4] are proposed; also see [5, 6] and the references therein. On the other hand, the problem of learning features of graph topology such as centrality (e.g. [7–9]), communities (e.g. [10–14]) has received attention as they relax the strong requirements on data model found in works on graph learning, while achieving a similar end goal.

Many graph (feature) learning works take the default assumption that observed data is generated from a single underlying graph. A plenty of algorithms have been proposed based on this single graph assumption [5–8, 10–12]. However, it is natural that for some applications, the graph may vary significantly during the data collection process. For example, the structure of a social network may change when key opinion leaders start (un)following each other; the power network graph may vary between the daytime and nighttime operations, or when power outage occurs. To this end, algorithms for learning time varying graphs have been studied in [15–17]. These works require the graph topologies to be similar to each other and the timestamps of graph signal samples are known.

Another issue is that the correspondences between graphs and data samples can be unknown. For example, if we consider opinion data measured on different products, as people may attend to different peers when forming their opinions, this can form different graph topologies; yet the observed opinion data does not reveal which graph topology is associated with each sample of opinions. It results in a *graph identification* problem whose goal is to cluster data samples according to their associated graphs. Note that data clustering methods have been well studied, where efficient methods such as spectral clustering [18], DBSCAN [19] are widely used. However, filtered graph signals entail delicate structures which are ignored by these data clustering methods, leading to suboptimal performance.

In this paper, we consider a joint centrality estimation and graph identification problem which is a multi-graph learning problem. Our aim is to jointly estimate the eigen-centrality of nodes from multi-graph, while simultaneously identify the graph which is associated to each graph signal sample. Our approach features two innovations.

- We propose a mixture model of low pass filtered graph signals, which features possibly non-white excitation covariances.
- We design a customized expectation-maximization (EM) algorithm [20] for estimating the parameters in the mixture model under a low-rank plus sparse prior.

In detail, a closed form solution with easy implementation is provided for the E-step; it is shown that the M-step corresponds to an adaptive scaled robust PCA problem [21]. The parameters inferred from the EM algorithm can be exploited for graph identification which associates the graph signals according to the graph which generates them, and the centrality vectors of all the graphs involved in the mixture model can be derived. Moreover, we suggest a carefully crafted initialization for the EM algorithm.

Related works. To our best knowledge, this work is the first to consider the mixture model of graph signals for learning multiple graphs and their graph features. The setting for significantly different graph topologies and unknown graph identifiers are new to the literature. In the former case, a number of works on anomaly detection have considered related models. Most of them considered only two graphs and the graph topologies are assumed known. E.g., [22] considered a change point estimation for network topology, [23] proposed an approximate Neyman-Pearson detector to check which of the two graphs does the graph signal fit the best, [24] identified the edge disconnections for more than two graphs. In the latter case, prior works [18, 19] do not exploit structure in the mixture model of signals. Studies on mixture models of special signal structure are scarce except for simple linear models [25], and a graph Laplacian mixture model is considered in [26] which requires white noise excitation and known graph filter.

Notations. We use boldfaced character (resp. boldfaced capital letter) to denote vector (resp. matrix). For a vector $\boldsymbol{x} \in \Re^n$, $\|\boldsymbol{x}\|$, $\|\boldsymbol{x}\|_1$ denote the Euclidean, ℓ_1 norm, respectively; for a matrix $\boldsymbol{X} \in \Re^{m\times n}$, $[\boldsymbol{X}]_{i,j}$ denotes its (i,j)th entry, and $\|\boldsymbol{X}\|$, $\|\boldsymbol{X}\|_1$, $\|\boldsymbol{X}\|_\star$ denote the spectral norm, 1-norm (of the vectorized matrix), nuclear norm, respectively. For a square, symmetric matrix $\boldsymbol{A} \in \Re^{n\times n}$, $\lambda_i(\boldsymbol{A})$ denotes its ith largest eigenvalue.

2. MIXTURE MODEL OF FILTERED GRAPH SIGNALS

Consider C undirected, connected graphs $G^{(c)} = (\mathcal{V}, \mathcal{E}^{(c)}, \boldsymbol{A}^{(c)})$, $c = 1, \ldots, C$, with a common node set $\mathcal{V} := \{1, \ldots, n\}$. Each

This work is supported by RGC Project #24203520. Emails: {yrhe,htwai}@se.cuhk.edu.hk

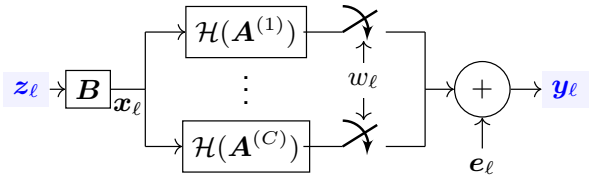

Fig. 1. Illustrating the generation process for the mixture model of graph signals (1). Black color denotes unknown variables while blue color denotes known variables [cf. Problem 1]. The graph identifier variable $\boldsymbol{w}_\ell$ chooses which graph filter is selected.

graph $G^{(c)}$ has a different edge set $\mathcal{E}^{(c)} \subseteq \mathcal{V} \times \mathcal{V}$ and the weighted adjacency matrix $\boldsymbol{A}^{(c)} \in \Re_+^{n \times n}$ satisfies $A_{ij}^{(c)} = A_{ji}^{(c)} > 0$ if and only if $(i,j) \in \mathcal{E}^{(c)}$; otherwise, $A_{ij}^{(c)} = 0$.

We study a mixture model of filtered graph signals. For each $\ell \in \mathbb{N}$, the observed sample $\boldsymbol{y}_\ell \in \Re^n$ is the output of a graph filter $\mathcal{H}(\cdot)$ defined using the adjacency matrix of one of the C graphs. Let $w_\ell \in \{1, \ldots, C\}$ be the graph identifier variable modeled as a multinomial random variable (r.v.) with the probability mass function $\mathbb{P}(w_\ell = c) = P_c$. We have

$$\boldsymbol{y}_\ell = \textstyle\sum_{c=1}^{C} \mathbb{1}(w_\ell = c) \mathcal{H}(\boldsymbol{A}^{(c)}) \boldsymbol{B} \boldsymbol{z}_\ell + \boldsymbol{e}_\ell, \tag{1}$$

where $\boldsymbol{e}_\ell \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is the observation noise and $\mathbb{1}(\cdot)$ is the $\{0,1\}$ indicator function. Furthermore, the excitation to the graph filter is given by the product $\boldsymbol{x}_\ell = \boldsymbol{B}\boldsymbol{z}_\ell$, where $\boldsymbol{B} \in \Re^{n \times k}$, $k \leq n$, controls the k-dimensional subspace of excitation and $\boldsymbol{z}_\ell \in \Re^k$ contains the excitation parameters. Note that this is without loss of generality as $\boldsymbol{B}$ can be arbitrary. Fig. 1 summarizes the generative process for the mixture model of filtered graph signals.

Let $T \in \mathbb{Z}_+ \cup \{\infty\}$, we consider the linear graph filter $\mathcal{H}(\cdot)$ given as a T-th order matrix polynomial, i.e.,

$$\mathcal{H}(\boldsymbol{A}^{(c)}) = \textstyle\sum_{t=0}^{T-1} h_t [\boldsymbol{A}^{(c)}]^t \tag{2}$$

where $\{h_t\}_{t=0}^{T-1}$ are the filter coefficients[1]. We assume that the graph filters satisfy the following condition regarding their spectrums:

H1 *Denote the extreme eigenvalues of the graph adjacency matrices by* $\underline{\lambda}_1 := \min_{c=1,\ldots,C} \lambda_1(\boldsymbol{A}^{(c)})$, $\overline{\lambda}_1 := \max_{c=1,\ldots,C} \lambda_1(\boldsymbol{A}^{(c)})$, $\overline{\lambda}_2 := \max_{c=1,\ldots,C} \lambda_2(\boldsymbol{A}^{(c)})$, *and* $\underline{\lambda}_n := \min_{c=1,\ldots,C} \lambda_n(\boldsymbol{A}^{(c)})$. *We assume that (i)* $\underline{\lambda}_1 > \overline{\lambda}_2$, *and (ii) the low pass ratio, defined as:*

$$\eta := \left(\min_{\lambda \in [\underline{\lambda}_1, \overline{\lambda}_1]} |\textstyle\sum_{t=0}^{T-1} h_t \lambda^t| \right)^{-1} \max_{\lambda \in [\underline{\lambda}_n, \overline{\lambda}_2]} |\textstyle\sum_{t=0}^{T-1} h_t \lambda^t|,$$

is strictly less than one.

Notice that H1(i) ensures that the graph adjacency matrices have similar spectrums in terms of the first two graph frequencies. Moreover, H1(ii) is modified from [27] to the multi-graph setting. Under H1(ii), it can be seen that each of the graph filter $\mathcal{H}(\boldsymbol{A}^{(c)})$ is 1-low pass with the low pass ratio not exceeding $\eta < 1$. The lowpass ratio η quantifies the strength of the low pass filter as it describes how fast the filter's frequency response drops over λ_1. When $\eta \ll 1$, the matrix $\mathcal{H}(\boldsymbol{A}^{(c)})$ is approximately rank-one for any c.

We aim to learn the unknown graphs $\{G^{(c)}\}_{c=1}^C$ from m samples of observed graph signals in (1). Recovering the complete adjacency matrices may not be possible since the observations are low rank in general (with $k \leq n$). We are thus concerned with:

Problem 1 *Given the filtered graph signals* $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m]$, *the excitation parameters* $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m]$, *and the number of graphs* C, *we estimate* (A) *graph identifiers* $w_\ell \in \{1, \ldots, C\}$ *for each* $\ell = 1, \ldots, m$ *(subject to permutation ambiguity), and* (B) *the eigen-centrality vector given by the top eigenvector of adjacency matrix* $\boldsymbol{c}_{\mathsf{eig}}^{(c)} := \mathsf{TopEV}(\boldsymbol{A}^{(c)})$ *for each* $c = 1, \ldots, C$.

[1] It is possible to extend our analysis to setting when the graph filter coefficients varies with the graph identifier.

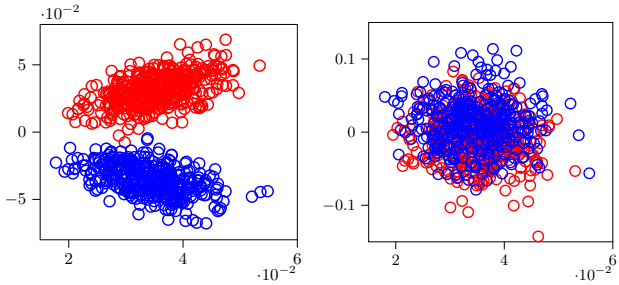

Fig. 2. Toy Example illustrating the data mixing on a core periphery graph: (Left) with strong low pass filter $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}) = (\boldsymbol{I} - \frac{1}{40}\boldsymbol{A})^{-1}$; (Right) with weak low pass filter $\mathcal{H}_{\mathsf{w}}(\boldsymbol{A}) = (\boldsymbol{I} - \frac{1}{80}\boldsymbol{A})^{-1}$.

The above problem addresses the *joint centrality inference and graph identification*. In particular, the goal (A) identifies the graph associated with a graph signal sample, while the goal (B) makes inference about the centrality of the graphs. Before discussing the proposed algorithm, we observe how the prior approaches can not be applied to tackle Problem 1.

Achieving Goal (A) Irrespective of Signal Structure. Observe that goal (A) is essentially a data clustering problem which groups m samples of graph signals into C clusters. To this end, one may apply the classical spectral clustering (SC) method on the observations $\boldsymbol{Y}$ [18]. Denote the correlation matrix of data $\boldsymbol{Y}$ as $\boldsymbol{C}_{\boldsymbol{y}} = \boldsymbol{Y}^\top \boldsymbol{Y}$, the SC method applies K-means on the top-C eigenvectors of $\boldsymbol{C}_{\boldsymbol{y}}$ to cluster the graph signals. Estimates of the graph identifiers can then be obtained from the clustering results.

To evaluate the possibility of the direct application of SC method to achieve goal (A), we perform a toy example simulation to visualize a typical instance of the data generated from the mixture model. We use a case with $C = 2$ graphs, where each underlying graph topology is a 100-nodes core-periphery graph generated from a stochastic block model with 10 fully connected central nodes and the excitation signals lie in a low-dimensional space with $k = 40$; see Section 4. Fig. 2 shows a scatter plot of the top-2 eigenvectors of $\boldsymbol{C}_{\boldsymbol{y}}$ that are colored according to the true graph identifiers $\{w_\ell\}_{\ell=1}^m$. Only when data points are separable, the SC method can accurately estimate the graph identifiers. We observe significant overlap between the data points in Fig. 2 (right) and deduce that the SC method does not achieve goal (A), especially when the underlying graph filter is weak low pass. A possible explanation is that the SC method does not make use of the graph signal structure (1) and the known information on excitation parameters $\boldsymbol{Z}$.

Achieving Goal (B) With Known Graph Identifiers. When the graph identifiers $\{w_\ell\}_{\ell=1}^m$ are available, i.e., goal (A) is achieved, the centrality inference problem from the mixture model (1) can be decomposed into C single graph learning problems. We can reorganize the observations $\boldsymbol{Y}$ into $\{\boldsymbol{Y}_c\}_{c=1}^C$ and the excitation parameters $\boldsymbol{Z}$ into $\{\boldsymbol{Z}_c\}_{c=1}^C$ based on their graph identifiers $\{w_\ell\}_{\ell=1}^m$. Estimating the centrality vector on a single graph model is a well-studied problem [7–9] when $\boldsymbol{Z}_c$ is known. For example, we can directly apply the algorithm in [7] on each data pair $(\boldsymbol{Y}_c, \boldsymbol{Z}_c)$. Such approach produces reliable estimate of the eigen-centrality.

From the above discussions, we observe that tackling Problem 1 requires utilizing structures in the mixture model (1). This motivates us to derive a customized expectation maximization algorithm

which utilizes a maximum-a-posterior (MAP) estimation framework to leverage the structures in (1).

3. EXPECTATION MAXIMIZATION ALGORITHM

Our plan is to tackle Problem 1 via a MAP estimation framework. Observe that while (1) may be regarded as a special case of the classical Gaussian mixture model, it entails additional structures which shall be leveraged to improve the estimation performance.

To obtain an effective prior for the MAP framework, we study the mixture model (1) and the graph filters under H1. Inspired by [10], we observe an *intrinsic decomposition* for the matrix product:

$$\mathcal{H}(\boldsymbol{A}^{(c)})\boldsymbol{B} = (\mathcal{H}(\boldsymbol{A}^{(c)}) - \rho\boldsymbol{I})\boldsymbol{B} + \rho\boldsymbol{B} \equiv \boldsymbol{L}_c + \boldsymbol{B}_\rho, \quad (3)$$

for any $\rho \geq 0$. Note that $\mathcal{H}_\rho(\boldsymbol{A}^{(c)}) := \mathcal{H}(\boldsymbol{A}^{(c)}) - \rho\boldsymbol{I}$ is yet another graph filter. The following result [9, Corollary 1] holds:

Corollary 1 *Under H1, there exists $\rho > 0$ such that $\mathcal{H}_\rho(\cdot)$ satisfies H1 with low pass ratio given by $\tilde{\eta} < \eta$. Furthermore, let $\widetilde{\boldsymbol{v}}_1^{(c)}$ be the top left singular vector of $\boldsymbol{L}_c$. It holds*

$$\|\widetilde{\boldsymbol{v}}_1^{(c)} - \boldsymbol{c}_{\mathsf{eig}}^{(c)}\| = \mathcal{O}(\tilde{\eta}). \quad (4)$$

In addition, as shown in [10], there exists $\rho > 0$ such that $\mathcal{H}_\rho(\cdot)$ satisfies H1 with low pass ratio given by $\tilde{\eta} \ll 1$. Under this choice of ρ, $\boldsymbol{L}_c$ is approximately rank-one and $\widetilde{\boldsymbol{v}}_1^{(c)}$ gives a good estimate of the eigen-centrality vector, i.e., tackling goal (B). For a number of applications (e.g., stock data, opinion dynamics), the matrix $\boldsymbol{B}$ is sparse as the latter models the pattern of influences from external sources to the graph. In conclusion, $\mathcal{H}(\boldsymbol{A}^{(c)})\boldsymbol{B}$ admits a 'low-rank + sparse' structure and the priors for $\{\boldsymbol{L}_c\}_{c=1}^C$, $\boldsymbol{B}_\rho$ is given by:

$$p(\{\boldsymbol{L}_c\}_{c=1}^C, \boldsymbol{B}_\rho) \propto \exp\big(-\lambda_S\|\boldsymbol{B}_\rho\|_1 - \lambda_L\textstyle\sum_{c=1}^C\|\boldsymbol{L}_c\|_*\big), \quad (5)$$

where $\lambda_S, \lambda_L \geq 0$ are regularization parameters and we have taken the approximation by assuming $\{\boldsymbol{L}_c\}_{c=1}^C$, $\boldsymbol{B}_\rho$ are independent.

We denote $\Theta = \big\{\{\boldsymbol{L}_c\}_{c=1}^C, \boldsymbol{B}_\rho, \{P_c\}_{c=1}^C\big\}$ as the set of unknown parameters under the mixture model (1), (3). The MAP estimation problem is given by:

$$\max_{\Theta} \frac{1}{m}\sum_{\ell=1}^m \log p(\boldsymbol{y}_\ell|\Theta, \boldsymbol{z}_\ell) + \log p(\Theta), \quad (6)$$

The MAP problem (6) is challenging as the log-likelihood function $\log p(\boldsymbol{y}_\ell|\Theta, \boldsymbol{z}_\ell)$ is intractable in general. Our strategy is to apply a customized EM algorithm [20] developed below.

EM Algorithm for (6). We take the graph identifier $w \in \{1, \ldots, C\}$ as a latent r.v. following the model distribution $q(\cdot|\boldsymbol{y}, \Theta, \boldsymbol{z})$ selected as the conditional probability mass function. Fixing $\widetilde{\Theta} = \big\{\{\widetilde{\boldsymbol{L}}_c\}_{c=1}^C, \widetilde{\boldsymbol{B}_\rho}, \{\widetilde{P}_c\}_{c=1}^C\big\}$, we apply the Jensen's inequality:

$$\begin{aligned}\log p(\boldsymbol{y}|\Theta, \boldsymbol{z}) &= \log\textstyle\sum_{c=1}^C p(\boldsymbol{y}, w=c|\Theta, \boldsymbol{z}) \\ &\geq \mathbb{E}_{w\sim q(\cdot|\boldsymbol{y},\widetilde{\Theta},\boldsymbol{z})}\big[\log p(\boldsymbol{y}, w|\Theta, \boldsymbol{z})\big] + \text{constant}.\end{aligned} \quad (7)$$

Furthermore, the non-constant term in the right hand side of (7) can be written as:

$$\mathbb{E}_{w\sim q(\cdot|\boldsymbol{y},\widetilde{\Theta},\boldsymbol{z})}\big[\textstyle\sum_{c=1}^C \mathbb{1}(w=c)\{\log(P_c) + \log p(\boldsymbol{y}|\boldsymbol{B}_\rho, \boldsymbol{L}_c, \boldsymbol{z})\}\big]$$

Recall from (1) that $p(\boldsymbol{y}|\Theta, \boldsymbol{z})$ follows a Gaussian distribution:

$$\log p(\boldsymbol{y}|\boldsymbol{B}_\rho, \boldsymbol{L}_c, \boldsymbol{z}) = \frac{1}{\sigma^2}\Big\{\langle(\boldsymbol{L}_c + \boldsymbol{B}_\rho)\boldsymbol{z}|\boldsymbol{y}\rangle - \frac{\|(\boldsymbol{L}_c + \boldsymbol{B}_\rho)\boldsymbol{z}\|^2}{2}\Big\} + \text{constant}. \quad (8)$$

Let us define the following sufficient statistics:

$$\overline{P}_c^{\widetilde{\Theta}} = \tfrac{1}{m}\textstyle\sum_{\ell=1}^m \mathbb{E}_{w\sim q(\cdot|\boldsymbol{y}_\ell,\widetilde{\Theta},\boldsymbol{z}_\ell)}\big[\mathbb{1}(w=c)\big], \quad (9)$$

Algorithm 1 EM Algorithm for (6)

1: **Input**: graph signals $\boldsymbol{Y}$, excitation parameters $\boldsymbol{Z}$, # graphs C.
2: Initialize the conditional expectations

$$w_c^0(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell),\ \ell = 1, ..., m, \quad (11)$$

and evaluate the sufficient statistics in (9) accordingly. Denote the latter as $\{\overline{P}_c^0, \overline{\boldsymbol{YZ}}_c^0, \overline{\boldsymbol{ZZ}}_c^0\}_{c=1}^C$.
3: **for** $t = 1, 2, \ldots, T_{\max}$ **do**
4: M-step: maximize the surrogate function through solving

$$\Theta^t \in \arg\max_{\Theta}\ \mathcal{L}(\Theta\,|\,\{\overline{P}_c^{t-1}, \overline{\boldsymbol{YZ}}_c^{t-1}, \overline{\boldsymbol{ZZ}}_c^{t-1}\}_{c=1}^C), \quad (12)$$

see (13), (14) for details. Denote the above solution as $\Theta^t = \big\{\{\boldsymbol{L}_c^t\}_{c=1}^C, \boldsymbol{B}_\rho^t, \{P_c^t\}_{c=1}^C\big\}$.
5: E-step: update the conditional expectations as

$$w_c^t(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell) = \frac{P_c^t \exp(\frac{-1}{2\sigma^2}||\boldsymbol{y}_\ell - (\boldsymbol{B}_\rho^t + \boldsymbol{L}_c^t)\boldsymbol{z}_\ell||^2)}{\sum_{c'=1}^C P_{c'}^t \exp(\frac{-1}{2\sigma^2}||\boldsymbol{y}_\ell - (\boldsymbol{B}_\rho^t + \boldsymbol{L}_{c'}^t)\boldsymbol{z}_\ell||^2)},$$

for all $\ell = 1, \ldots, m$. Then evaluate sufficient statistics in (9) accordingly. Denote the latter as $\{\overline{P}_c^t, \overline{\boldsymbol{YZ}}_c^t, \overline{\boldsymbol{ZZ}}_c^t\}_{c=1}^C$.
6: **end for**
7: **Output**: the converged parameters $\Theta^{T_{\max}}$ and the conditional expectations $w_c^{T_{\max}}(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell)$, $\ell = 1, \ldots, m$.

$$\overline{\boldsymbol{YZ}}_c^{\widetilde{\Theta}} = \tfrac{1}{m}\textstyle\sum_{\ell=1}^m \mathbb{E}_{w\sim q(\cdot|\boldsymbol{y}_\ell,\widetilde{\Theta},\boldsymbol{z}_\ell)}\big[\mathbb{1}(w=c)\boldsymbol{y}_\ell\boldsymbol{z}_\ell^\top\big],$$
$$\overline{\boldsymbol{ZZ}}_c^{\widetilde{\Theta}} = \tfrac{1}{m}\textstyle\sum_{\ell=1}^m \mathbb{E}_{w\sim q(\cdot|\boldsymbol{y}_\ell,\widetilde{\Theta},\boldsymbol{z}_\ell)}\big[\mathbb{1}(w=c)\boldsymbol{z}_\ell\boldsymbol{z}_\ell^\top\big].$$

Notice that the above quantities can be computed from the conditional expectations $\mathbb{E}_{w\sim q(\cdot|\boldsymbol{y}_\ell,\widetilde{\Theta},\boldsymbol{z}_\ell)}\big[\mathbb{1}(w=c)\big]$, $\ell = 1, \ldots, m$. Lastly, we observe that the MAP objective function (6) is lower bounded by the following surrogate function:

$$\begin{aligned}&\mathcal{L}(\Theta\,|\,\{\overline{P}_c^{\widetilde{\Theta}}, \overline{\boldsymbol{YZ}}_c^{\widetilde{\Theta}}, \overline{\boldsymbol{ZZ}}_c^{\widetilde{\Theta}}\}_{c=1}^C) \\ &= -\lambda_S||\boldsymbol{B}_\rho||_1 + \textstyle\sum_{c=1}^C\Big\{-\lambda_L||\boldsymbol{L}_c||_* + \overline{P}_c^{\widetilde{\Theta}}\log(P_c) \\ &+ \frac{2\langle\boldsymbol{L}_c + \boldsymbol{B}_\rho|\overline{\boldsymbol{YZ}}_c^{\widetilde{\Theta}}\rangle - \mathrm{Tr}\big((\boldsymbol{L}_c + \boldsymbol{B}_\rho)\overline{\boldsymbol{ZZ}}_c^{\widetilde{\Theta}}(\boldsymbol{L}_c + \boldsymbol{B}_\rho)^\top\big)}{2\sigma^2}\Big\}.\end{aligned} \quad (10)$$

Importantly, we observe the surrogate function is *concave* in Θ when given the sufficient statistics $\{\overline{P}_c^{\widetilde{\Theta}}, \overline{\boldsymbol{YZ}}_c^{\widetilde{\Theta}}, \overline{\boldsymbol{ZZ}}_c^{\widetilde{\Theta}}\}_{c=1}^C$. We summarize the EM algorithm in Algorithm 1.

As observed in the pseudo code, the EM algorithm alternates between a M-step which maximizes $\mathcal{L}(\Theta|\Theta^{t-1})$ and an E-step which aims at computing the conditional expectations in (9). In particular for the M-step, $\{\boldsymbol{L}_c^t\}_{c=1}^C$, $\boldsymbol{B}_\rho^t$ can be obtained by solving:

$$\begin{aligned}\min_{\{\boldsymbol{L}_c\}_{c=1}^C, \boldsymbol{B}_\rho}\ &\tfrac{1}{2\sigma^2}\textstyle\sum_{c=1}^C||(\boldsymbol{L}_c + \boldsymbol{B}_\rho)\overline{\boldsymbol{Z}}_c^{t-1} - \overline{\boldsymbol{YZ}}_c^{t-1}(\overline{\boldsymbol{Z}}_c^{t-1})^{-1}||_F^2 \\ &+ \lambda_L\textstyle\sum_{c=1}^C||\boldsymbol{L}_c||_* + \lambda_S||\boldsymbol{B}_\rho||_1,\end{aligned} \quad (13)$$

where $\overline{\boldsymbol{Z}}_c^{t-1}$ is defined as the matrix square root of $\overline{\boldsymbol{ZZ}}_c^{t-1}$. Note that the above problem is akin to the classical robust PCA problem [21]. This is a consequence of the 'low-rank + sparse' prior imposed by (5). Meanwhile, $\{P_c^t\}_{c=1}^C$ is obtained by

$$P_c^t = \big(\textstyle\sum_{c'=1}^C \overline{P}_{c'}^{t-1}\big)^{-1}\overline{P}_c^{t-1},\ \ c = 1, \ldots, C. \quad (14)$$

It is known that the EM algorithm converges to a stationary point of (6) [28]. Notice that Algorithm 1 has a per-iteration complexity of $\mathcal{O}(n^3)$ when (13) is solved using the algorithm in [29].

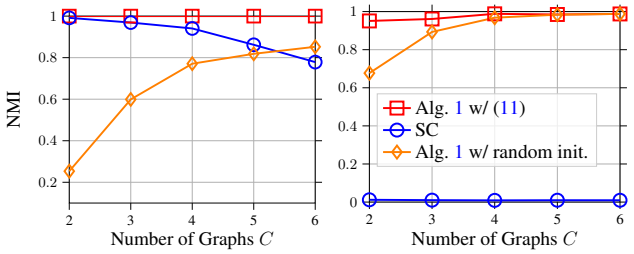

Fig. 3. **Goal** (A). Normalized mutual information against the number of graphs C with (Left) $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}^{(w_\ell)})$ and (Right) $\mathcal{H}_{\mathsf{w}}(\boldsymbol{A}^{(w_\ell)})$.

To treat Problem 1, we observe that the output $w_c^T(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell)$ is the conditional probability for the event $w_\ell = c$ given Θ^T. Goal (A) can be tackled by $\hat{w}_\ell = \arg\max_{c=1,\ldots,C} w_c^T(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell)$. Goal (B) is tackled by applying SVD on $\{\boldsymbol{L}_c\}_{c=1}^C$ as we recall Corollary 1.

Practical Implementation. Note that the performance with EM algorithm is known to be sensitive to the initialization in (11). We propose to adopt the output from spectral clustering for the initialization. Note that although the SC method alone does not provide accurate graph identification, the information extracted from it is found to be sufficient to initialize the EM algorithm. In particular, we take

$$w_c^0(\boldsymbol{y}_\ell, \boldsymbol{z}_\ell) = \frac{\exp(-||\boldsymbol{v}_{1:C}^{y_\ell} - \bar{\boldsymbol{v}}_c||^2)}{\sum_{c'=1}^C \exp(-||\boldsymbol{v}_{1:C}^{y_\ell} - \bar{\boldsymbol{v}}_{c'}||^2)}, \tag{15}$$

where $\{\boldsymbol{v}_1^y, \ldots, \boldsymbol{v}_C^y\}$ are the top-C eigenvectors of the data correlation matrix $\boldsymbol{C}_y$ and $\bar{\boldsymbol{v}}_c$ is the centroid vector of the cth cluster.

4. EXPERIMENT

Our goal is to compare the performance in solving Problem 1: goal (A) graph identifiers detection and goal (B) centrality estimation for data generated on a mixture graph model. For every experiment, we repeat 100 Monte-carlo trials and generate $m = 400C$ data samples with noise variance $\sigma^2 = (0.1)^2$ on C different core-periphery (CP) graphs. For any $c = 1, \ldots, C$, the c-th CP graph is constructed by a stochastic block model with 2 blocks and $n = 100$ nodes. The node set $\mathcal{V} = \{1, ..., n\}$ is partitioned into $\mathcal{V}_{\mathsf{o}}^{(c)}$ and $\mathcal{V}_{\mathsf{p}}^{(c)}$. For any $i, j \in \mathcal{V}$, an edge is always assigned if $i, j \in \mathcal{V}_{\mathsf{o}}^{(c)}$; with probability 0.2 if $i \in \mathcal{V}_{\mathsf{o}}^{(c)}, j \in \mathcal{V}_{\mathsf{p}}^{(c)}$; and with probability 0.05 if $i, j \in \mathcal{V}_{\mathsf{p}}^{(c)}$. We uniformly select $|\mathcal{V}_{\mathsf{o}}^{(c)}| = 10$ nodes at random to construct the core node set. For the matrix $\boldsymbol{B} \in \Re^{n \times k}$, we randomly select 10% of the entries in $\boldsymbol{B}$ are non-zero which means $B_{ij} = M_{ij}\widetilde{B}_{ij}$, where $M_{ij}, \widetilde{B}_{ij}$ are independent r.v.s, $M_{ij} \in \{0, 1\}$ is Bernouli with $\mathbb{E}[M_{ij}] = 0.1$, and $\widetilde{B}_{ij} \sim \mathcal{U}([0.1, 1])$. For the excitation parameter matrix $\boldsymbol{Z} \in \Re^{k \times m}$, we have 60% of the entries in $\boldsymbol{Z}$ are non-zero so $Z_{ij} = N_{ij}\widetilde{Z}_{ij}$, where $N_{ij}, \widetilde{Z}_{ij}$ are independent r.v.s, $N_{ij} \in \{0, 1\}$ is Bernouli with $\mathbb{E}[N_{ij}] = 0.6$, and $\widetilde{Z}_{ij} \sim \mathcal{U}([0.1, 1])$. We set the excitation rank $k = 40$ for each experiment.

For each generated data $\boldsymbol{y}_\ell$, we consider its graph filter as $\mathcal{H}(\boldsymbol{A}^{(w_\ell)}) = (\boldsymbol{I} - \alpha\boldsymbol{A}^{(w_\ell)})^{-1}$ where the graph identifier w_ℓ is drawn from $\mathcal{U}\{1, \ldots, C\}$. To evaluate the clustering accuracy [cf. Goal (A)], we compute the normalized mutual information (NMI) [30] between the detected graph identifiers $\{\hat{w}_\ell\}_{\ell=1}^m$ and the ground truth graph identifiers $\{w_\ell\}_{\ell=1}^m$, NMI ≈ 1 indicates a high clustering accuracy. To evaluate the centrality estimation performance [cf. Goal (B)], we compare the detected central nodes set with the ground truth set $\mathcal{V}_{\mathsf{o}}^{(c)}$. A node is detected as central node if its estimated centrality is in the top-10. Let $\widehat{\mathcal{V}}_{\mathsf{o}}^{(c)}$ be the set of central nodes detected by the algorithms. We define the average error rate:

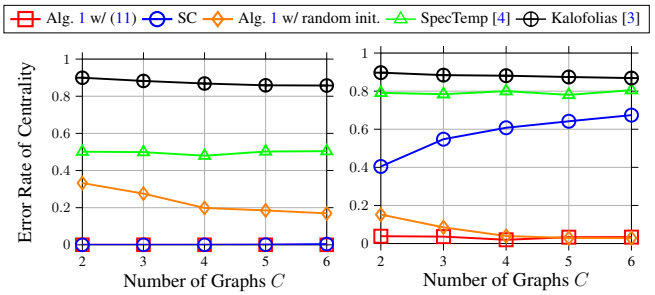

Fig. 4. **Goal** (B). Error rate against the number of graphs C with (Left) $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}^{(w_\ell)})$ and (Right) $\mathcal{H}_{\mathsf{w}}(\boldsymbol{A}^{(w_\ell)})$.

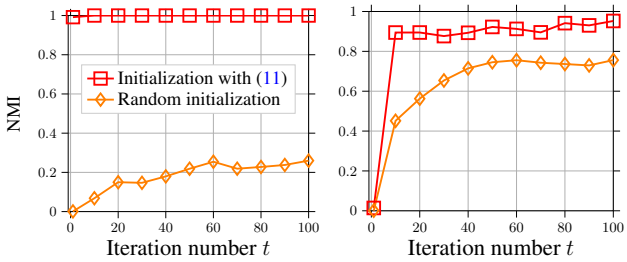

Fig. 5. **Goal** (A). Normalized mutual information against inner iteration number T_{max} by Algorithm 1 with (Left) $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}^{(w_\ell)})$ and (Right) $\mathcal{H}_{\mathsf{w}}(\boldsymbol{A}^{(w_\ell)})$ on $C = 2$.

$$\text{Error rate} = (1/C)\textstyle\sum_{c=1}^C \mathbb{E}\big[\frac{1}{10}|\mathcal{V}_{\mathsf{o}}^{(c)} \cap \widehat{\mathcal{V}}_{\mathsf{o}}^{(c)}|\big]. \tag{16}$$

In Fig. 3 and Fig. 4, we compare the performance in tackling Problem 1 when the graph signals are generated from two types of graph filters with different low pass strengths. One is $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}^{(w_\ell)}) = (\boldsymbol{I} - \frac{1}{40}\boldsymbol{A}^{(w_\ell)})^{-1}$ and the other is $\mathcal{H}_{\mathsf{w}}(\boldsymbol{A}^{(w_\ell)}) = (\boldsymbol{I} - \frac{1}{80}\boldsymbol{A}^{(w_\ell)})^{-1}$. Note that $\mathcal{H}_{\mathsf{w}}(\cdot)$ is a 'weaker' low pass graph filter as it has a low pass ratio $\eta \approx 1$. We set $T_{\mathsf{max}} = 100$ for Algorithm 1 with initialization (15). For the benchmark with SC, we modify Algorithm 1 by replacing its E step with spectral clustering and then complete M step in one step to estimate $\{\boldsymbol{L}_c\}_{c=1}^C$. Furthermore, we consider two graph topology learning methods: SpecTemp [4] and the method by Kalofolias [3], as benchmark. We first separate data into C groups based on the graph identifiers detected from Algorithm 1, then respectively apply these methods on separated data groups to estimate C graph topology and their corresponding centrality vectors. In both tests, we find that Algorithm 1 outperforms the existing algorithms. It is robust to different types of low pass graph filters and the number of graphs. Furthermore, the centrality estimation performance improves with the clustering accuracy. We also found that the initialization scheme (15) can improve the performances.

Lastly, in Fig. 5, we examine the evolution of NMI (clustering accuracy) against the EM iteration number under the same settings in the above with $C = 2$ graphs. We compare how the initialization (11) help with the convergence of the EM algorithm. From the figure, we found that with (11), the EM algorithm converges quickly within the first 20 iterations, while with a random initialization, the EM algorithm may get stuck at local minima where the solution accuracy does not improve significantly after 100 iterations, especially with the low pass graph filter given by $\mathcal{H}_{\mathsf{s}}(\boldsymbol{A}^{(w_\ell)})$.

Conclusions. We proposed a mixture model of graph signals and developed an EM algorithm to tackle the joint centrality estimation and graph identification problem [cf. Problem 1]. The performance of the proposed algorithm is demonstrated via numerical experiments, where it successfully achieves both goals of the problem.

5. REFERENCES

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

[2] 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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[3] V. Kalofolias, "How to learn a graph from smooth signals," in *Artificial Intelligence and Statistics*, 2016, pp. 920–929.

[4] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

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

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

[7] Y. He and H.-T. Wai, "Estimating centrality blindly from low-pass filtered graph signals," in *ICASSP*, 2020, pp. 5330–5334.

[8] T. M. Roddenberry and S. Segarra, "Blind inference of eigenvector centrality rankings," *IEEE Transactions on Signal Processing*, 2021.

[9] Y. He and H.-T. Wai, "Detecting central nodes from low-rank excited graph signals via structured factor analysis," *arXiv preprint arXiv:2109.13573*, 2021.

[10] H.-T. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind community detection from low-rank excitations of a graph filter," *IEEE Transactions on Signal Processing*, vol. 68, pp. 436–451, 2019.

[11] Y. Yuan, D. W. Soh, H. H. Yang, and T. Q. S. Quek, "Learning overlapping community-based networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 4, pp. 684–697, 2019.

[12] M. T. Schaub, S. Segarra, and J. N. Tsitsiklis, "Blind identification of stochastic block models from dynamical observations," *SIAM Journal on Mathematics of Data Science*, vol. 2, no. 2, pp. 335–367, Jan 2020.

[13] T. Hoffmann, L. Peel, R. Lambiotte, and N. S. Jones, "Community detection in networks without observing edges," *Science Advances*, vol. 6, no. 4, p. eaav1478, Jan 2020.

[14] T. M. Roddenberry, M. T. Schaub, H.-T. Wai, and S. Segarra, "Exact blind community detection from signals on multiple graphs," *IEEE Transaction on Signal Processing*, 2020.

[15] D. Hallac, Y. Park, S. Boyd, and J. Leskovec, "Network inference via the time-varying graphical lasso," in *KDD*, 2017, pp. 205–213.

[16] B. Baingana and G. B. Giannakis, "Tracking switched dynamic network topologies from information cascades," *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 985–997, 2016.

[17] K. Yamada, Y. Tanaka, and A. Ortega, "Time-varying graph learning with constraints on graph temporal variation," *arXiv preprint arXiv:2001.03346*, 2020.

[18] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[19] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[21] A. Agarwal, S. Negahban, M. J. Wainwright *et al.*, "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions," *The Annals of Statistics*, vol. 40, no. 2, pp. 1171–1197, 2012.

[22] C. Kaushik, T. M. Roddenberry, and S. Segarra, "Network topology change-point detection from graph signals with prior spectral signatures," in *ICASSP*, 2021, pp. 5395–5399.

[23] S. P. Chepuri and G. Leus, "Subgraph detection using graph signals," in *Asiolmar*, 2016, pp. 532–534.

[24] S. Shaked and T. Routtenberg, "Identification of edge disconnections in networks based on graph filter outputs," *arXiv preprint arXiv:2102.06428*, 2021.

[25] G. J. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, 2004.

[26] H. P. Maretic and P. Frossard, "Graph laplacian mixture model," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 261–270, 2020.

[27] R. Ramakrishna, H.-T. Wai, and A. Scaglione, "A user guide to low-pass graph signal processing and its applications," *IEEE Signal Processing Magazine*, 2020.

[28] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[29] A. Aravkin, S. Becker, V. Cevher, and P. Olsen, "A variational approach to stable principal component pursuit," in *30th Conference on Uncertainty in Artificial Intelligence (UAI) 2014*, no. CONF, 2014.

[30] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *The Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, 2010.