Joint graph learning from Gaussian observations in the presence of hidden nodes

Samuel Rey*, Madeline Navarro†, Andrei Buciulea*, Santiago Segarra†, and Antonio G. Marques*

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10051977

***Abstract*—Graph learning problems are typically approached by focusing on learning the topology of a single graph when signals from all nodes are available. However, many contemporary setups involve *multiple* related networks and, moreover, it is often the case that only a subset of nodes is observed while the rest remain *hidden*. Motivated by this, we propose a *joint graph* learning method that takes into account the presence of *hidden (latent) variables*. Intuitively, the presence of the hidden nodes renders the inference task ill-posed and challenging to solve, so we overcome this detrimental influence by harnessing the *similarity* of the estimated graphs. To that end, we assume that the observed signals are drawn from a Gaussian Markov random field with latent variables and we carefully model the graph similarity among hidden (latent) nodes. Then, we exploit the structure resulting from the previous considerations to propose a convex optimization problem that solves the joint graph learning task by providing a regularized maximum likelihood estimator. Finally, we compare the proposed algorithm with different baselines and evaluate its performance over synthetic and real-world graphs.**

***Index Terms*—Graph learning, network topology inference, Gaussian graphical models, latent variables, multi-layer graphs**

I. Introduction

The rising popularity of graph-based methods to deal with data defined over irregular domains has propelled the development of graph learning algorithms [1]–[3]. Indeed, the problem of graph learning, which seeks to learn the topology of a graph from a set of nodal observations, is among the most active research areas of graph signal processing (GSP), a field devoted to the development of tools for processing graph signals [4]–[7]. Fundamental to learning the topology of the graph is assuming that the (statistical) properties of the observed signals depend on the unknown topology, with different assumptions leading to different graph learning methods. Noteworthy examples of these methods include correlation networks [8], sparse structural equation models [9], [10], graph stationarity models [11]–[13], smooth (local total variation) models [14]–[16], and models based on more sophisticated graph-topology priors [17], [18], just to name a few. Of particular interest for the paper at hand are Gaussian graphical models, a popular and flexible family of graph learning algorithms that builds upon the assumption that the observed signals are drawn from a Gaussian Markov random field (GMRF) [19]–[24].

Work supported by the Spanish Grants SPGRAPH (PID2019-105032GB-I00/AEI/10.13039/501100011033), FPU17/04520, EST21/00420, and CAM PEJ-2020-AI/TIC-18964, and the USA NSF award CCF-2008555. *Dept. of Signal Theory and Comms., King Juan Carlos University, Madrid, Spain. †Dept. of ECE, Rice University, Houston, USA. Contact author: antonio.garcia.marques(AT)urjc.es.

Despite the growing interest aroused by the graph learning problem, the focus is typically placed on learning an *individual* (single) graph under the assumption that observations from *all* nodes are available. Nonetheless, many contemporary scenarios such as social networks or brain analytics involve multiple related networks, each of them with a subset of available signals. When several related networks are available, we can harness the graph similarity to boost the performance of the graph learning algorithms [13], [25]–[28]. Moreover, in many relevant scenarios, we only have access to observations from a subset of nodes, while the rest remain unobserved or hidden. Intuitively, ignoring the presence of hidden variables will hinder the performance of the graph learning algorithms but, at the same time, accounting for the influence of the hidden nodes renders the graph learning task an ill-posed problem. This challenging setting has been studied in the context of learning a single graph [29]–[32]. However, modeling the influence of hidden nodes in the context of joint graph learning becomes even more critical since it is unclear how to measure the graph similarity between nodes that remain unobserved.

Motivated by the previous discussion, we propose a convex optimization framework to solve the joint graph learning problem in the presence of hidden variables under the assumption that the observed signals are drawn from a GMRF. The main motivation is to exploit the known benefits of joint network topology inference methods to improve the performance of the challenging setup resulting from the presence of hidden variables. Nonetheless, to put forward such an approach we need to carefully model the influence of the hidden nodes in (i) the assumption of GMRF signals that relates the observations and the graph; and (ii) the similarity of the graphs being estimated. This is achieved by carefully leveraging the structure inherent to the presence of hidden nodes into a regularized maximum likelihood estimator.

Even though the motivation is similar, the resulting algorithm differs substantially from other related works such as [28]. The main difference lies in the assumption relating the observed signals and the graph topology, which results in modeling the influence of the hidden nodes in a different way, and hence, the structure of the problem is completely different. Furthermore, the different structure of the problem at hand also requires a new approach for exploiting the graph similarity assumption when hidden nodes are involved.

The remainder of the paper is as follows. Section II introduces fundamental concepts about GSP and Gaussian graphical models. Section III states the joint graph learning problem with hidden nodes and presents the proposed solution. Finally,

Authorized licensed use limited to: Fondren Library Rice University. Downloaded on July 02,2023 at 21:10:18 UTC from IEEE Xplore. Restrictions apply.

Section IV offers a numerical evaluation of the proposed algorithm and Section V provides some concluding remarks.

II. Graph learning from Gaussian observations

A. Fundamentals of GSP

Graphs. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph described by the set of N nodes $\mathcal{V}$, and the set of links $\mathcal{E}$. The elements in $\mathcal{E}$ are unordered pairs (i, j) with $i, j \in \mathcal{V}$ and such that $(i, j) \in \mathcal{E}$ implies that nodes i and j are connected. For any graph $\mathcal{G}$, the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a sparse matrix encoding the connectivity of the graph with $A_{ij} = 0$ if and only if $(i, j) \notin \mathcal{E}$. Then, the value A_{ij} of the non-zero entries captures the strength of the link between the nodes i and j.

Graph signals and graph-shift operator. In addition to graphs, we are also concerned with graphs signals, a family of signals defined on the set of nodes $\mathcal{V}$. Formally, a graph signal can be denoted as a vector $\mathbf{x} = [x_1, ..., x_N]^\top \in \mathbb{R}^N$, where x_i denotes the signal value observed at node i. Because the signal is defined on top of the graph, the key assumption in GSP is that the properties of $\mathbf{x}$ are related to the topology of $\mathcal{G}$. In this sense, if the graph encodes similarity between nodes, then the signal values at two neighboring nodes are expected to be close. Of particular interest when processing graph signals is the graph-shift operator (GSO), a linear operator applied to graph signals that captures the topology of the graph [4]. The GSO is represented by the matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ whose entries satisfy that $S_{ij} \neq 0$ only if $i = j$ or $(i, j) \in \mathcal{E}$. Typical choices for the GSO are the adjacency matrix $\mathbf{A}$, the combinatorial graph Laplacian $\mathbf{L} := \text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$, or their normalized counterparts [4], [5].

B. Gaussian graphical models

Arguably, Gaussian graphical models are one of the most commonly used graph learning methods [19], [21], [22]. The main assumption adopted by these models is that the observed signals are drawn from a multivariate Gaussian distribution where the zeros in the inverse of the covariance correspond to edges not present in $\mathcal{E}$. Such a distribution is commonly known as a Gaussian Markov random field (GMRF), and the inverse of the covariance is referred to as the precision matrix. Upon setting the positive definite precision matrix to be the GSO $\mathbf{S}$, the GMRF assumption implies that $\mathbf{x}$ is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{S}^{-1})$. In other words, if we denote the covariance of $\mathbf{x}$ as $\mathbf{C} := \mathbf{E}[\mathbf{x}\mathbf{x}^\top]$, it follows that the mapping between $\mathbf{C}$ and $\mathbf{S}$ is given by $\mathbf{C} = \mathbf{S}^{-1}$. Then, Gaussian graphical models propose to estimate the graph topology encoded in $\mathcal{G}$ via the regularized maximum likelihood estimator

$$\begin{aligned} \max_{\mathbf{S}} \quad & \log\det(\mathbf{S}) - \text{tr}(\hat{\mathbf{C}}\mathbf{S}) - \lambda r(\mathbf{S}) \\ \text{s. t.} \quad & \mathbf{S} \succeq \mathbf{0}, \end{aligned} \tag{1}$$

where $\hat{\mathbf{C}}$ denotes the sample covariance matrix, and the term $r(\mathbf{S})$ promotes desired properties on $\mathbf{S}$. Indeed, for the particular case when $r(\mathbf{S})$ is set to $r(\mathbf{S}) = \|\mathbf{S}\|_1$, the estimator in (1) yields the celebrated graphical Lasso algorithm [21], [26].

III. Joint graph learning with hidden nodes

To formally state the joint graph learning problem in the presence of hidden nodes some definitions are in order. Let us consider that there are K unknown graphs $\{\mathcal{G}^{(k)}\}_{k=1}^K$, all of them defined over a common set of nodes $\mathcal{V}$ of cardinality N. For the k-th graph, we have access to M_k graph signals collected in the matrix $\mathbf{X}^{(k)} := [\mathbf{x}_1, ..., \mathbf{x}_{M_k}]^\top \in \mathbb{R}^{N \times M_k}$.

The presence of hidden nodes implies that only a subset of nodes $\mathcal{O} \subset \mathcal{V}$ of cardinality O is observed. Meanwhile, the $H = N - O$ remaining nodes collected in the set $\mathcal{H} = \mathcal{V} \setminus \mathcal{O}$ stay unobserved. Furthermore, consider that the sets $\mathcal{O}$ and $\mathcal{H}$ are the same for the K graphs. Without loss of generality, let us assume that the nodes in $\mathcal{O}$ correspond to the first nodes in the graph, so the signal values associated with the observed nodes correspond to the first O rows of $\mathbf{X}^{(k)}$, which we denote as $\mathbf{X}_{\mathcal{O}}^{(k)}$. Then, let us define the sample covariance matrix of the k-th graph as $\hat{\mathbf{C}}^{(k)} := \frac{1}{M_k}\mathbf{X}^{(k)}(\mathbf{X}^{(k)})^\top$. The distinction between observed and hidden nodes endows the GSO and the sampled covariance matrix with the following block structure

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{\mathcal{O}} & \mathbf{S}_{\mathcal{O}\mathcal{H}} \\ \mathbf{S}_{\mathcal{H}\mathcal{O}} & \mathbf{S}_{\mathcal{H}} \end{bmatrix}, \ \hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{\mathcal{O}} & \hat{\mathbf{C}}_{\mathcal{O}\mathcal{H}} \\ \hat{\mathbf{C}}_{\mathcal{H}\mathcal{O}} & \hat{\mathbf{C}}_{\mathcal{H}} \end{bmatrix}. \tag{2}$$

The $O \times O$ submatrices $\mathbf{S}_{\mathcal{O}}^{(k)}$ and $\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}$ respectively denote the block of the GSO capturing the connection between observed nodes and the block of the sample covariance of the signal values at the observed nodes (i.e., the sample covariance of $\mathbf{X}_{\mathcal{O}}$). On the other hand, the remaining blocks of $\mathbf{S}$ and $\hat{\mathbf{C}}$ involve edges and covariances between hidden nodes.

Based on the previous definitions, the problem of joint graph learning in the presence of hidden nodes is formally stated next.

Problem 1. *Find the matrices $\{\mathbf{S}_{\mathcal{O}}^{(k)}\}_{k=1}^K$ encoding the connectivity between the observed nodes $\mathcal{O}$ for all of the K graphs, given the $O \times M_k$ matrices $\{\mathbf{X}_{\mathcal{O}}^{(k)}\}_{k=1}^K$ collecting the signal values at the observed nodes under the assumptions that:*
***(AS1)** The number of observed nodes is considerably larger than the number of hidden nodes, i.e., $O \gg H$.*
***(AS2)** The columns of $\mathbf{X}^{(k)}$ are independent realizations of a zero mean multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, (\mathbf{S}^{(k)})^{-1})$.*
***(AS3)** The distance between the K graphs is small according to some metric $d(\mathbf{S}^{(k)}, \mathbf{S}^{(k')})$.*

Learning the topology encoded in the GSOs $\{\mathbf{S}_{\mathcal{O}}^{(k)}\}_{k=1}^K$ while accounting for the influence of hidden nodes is an ill-conditioned problem since there are no observations from the nodes in $\mathcal{H}$. Therefore, **(AS1)** ensures the tractability of the problem by assuming that most of the nodes are observed. Second, **(AS2)** establishes a connection between the unknown graph $\mathcal{G}^{(k)}$ and the signals collected in $\mathbf{X}^{(k)}$ via a GMRF model. However, although this assumption involves the whole matrices $\mathbf{X}^{(k)}$, only the submatrices $\mathbf{X}_{\mathcal{O}}^{(k)}$ are observed. A similar problem arises with **(AS3)**. The last assumption guarantees

Authorized licensed use limited to: Fondren Library Rice University. Downloaded on July 02,2023 at 21:10:18 UTC from IEEE Xplore. Restrictions apply.

that the K graphs are closely related so we can harness this relation in a joint graph learning algorithm. Nonetheless, how to exploit the similarity between subgraphs whose nodes are not being observed is not a trivial question.

To address these issues, in the following, we leverage the block structure introduced in (2) to model how the presence of the hidden nodes carries over to the assumptions **(AS1)** and **(AS2)**, and then we leverage the resulting relations to solve the joint graph learning problem with hidden variables via an optimization problem.

A. Joint graph learning as an optimization problem

Let us start by modeling how the presence of hidden variables impacts the GMRF assumption **(AS2)**, which is fundamental for the graph learning problem at hand. Recall that the GMRF assumption implies that the mapping between the covariance matrix and the GSO is given by $\mathbf{C}^{(k)} = (\mathbf{S}^{(k)})^{-1}$. Then, due to the presence of hidden nodes, we only have access to the observed sampled covariance matrix $\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}$, and hence, we are interested in establishing a relation between the matrices $\mathbf{S}_{\mathcal{O}}^{(k)}$ and $\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}$. To that end, similar to [29], [32], we leverage the block structure of $\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}$ and $\mathbf{S}_{\mathcal{O}}^{(k)}$, and employ the Schur complement to obtain the following expression

$$(\hat{\mathbf{C}}_{\mathcal{O}}^{(k)})^{-1} = \mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{S}_{\mathcal{O}\mathcal{H}}^{(k)}(\mathbf{S}_{\mathcal{H}}^{(k)})^{-1}\mathbf{S}_{\mathcal{H}\mathcal{O}}^{(k)} = \mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)}. \quad (3)$$

Since all the blocks of the GSO associated with hidden nodes are unknown, we *lift* the problem by defining the matrices $\mathbf{P}^{(k)} := \mathbf{S}_{\mathcal{O}\mathcal{H}}^{(k)}(\mathbf{S}_{\mathcal{H}}^{(k)})^{-1}\mathbf{S}_{\mathcal{H}\mathcal{O}}^{(k)} \in \mathbb{R}^{O\times O}$. These matrices capture the influence of the hidden nodes and, since they involve the product of the $H \times H$ matrices $\mathbf{S}_{\mathcal{H}}^{(k)}$, from **(AS1)** it follows that the matrices $\mathbf{P}^{(k)}$ are low-rank matrices such that $\text{rank}(\mathbf{P}^{(k)}) \leq H$. Furthermore, we can leverage these matrices to harness the graph similarity **(AS3)** between hidden nodes, as we will explain in more detail later.

Then, a reasonable approach to tackle Problem 1 is to modify the regularized maximum likelihood estimator from (1) based on the aforementioned considerations. This result in the non-convex optimization problem

$$\begin{aligned}\min_{\{\mathbf{S}_{\mathcal{O}}^{(k)},\mathbf{P}^{(k)}\}_{k=1}^{K}} \quad & \sum_{k=1}^{K} \text{tr}\left((\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)})\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}\right) - \log\det(\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)}) \\ & + \rho_k \|\mathbf{S}_{\mathcal{O}}^{(k)}\|_0 \\ & + \sum_{k<k'} \rho_{kk'} d_S(\mathbf{S}_{\mathcal{O}}^{(k)}, \mathbf{S}_{\mathcal{O}}^{(k')}) + \beta_{kk'} d_P(\mathbf{P}^{(k)}, \mathbf{P}^{(k')}) \\ \text{s. t.} \quad & \mathbf{P}^{(k)} \succcurlyeq 0; \;\; \mathbf{S}^{(k)} - \mathbf{P}^{(k)} \succ 0; \;\; \mathbf{S}_{\mathcal{O}}^{(k)} \in \mathcal{S}, \\ & \text{rank}(\mathbf{P}^{(k)}) \leq H. \end{aligned} \quad (4)$$

First, note that the maximum likelihood terms involve the expression $\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)}$ to account for the influence of hidden nodes in assumption **(AS2)**. Second, the rank constraint $\text{rank}(\mathbf{P}^{(k)}) \leq H$ captures the fact that the matrices $\mathbf{P}^{(k)}$ involve the multiplication by $H \times H$ matrices. Then, the ℓ_0 promotes sparsity on the estimated GSOs and the set $\mathcal{S}$ represents the set of admissible GSOs; see [11].

Regarding **(AS3)**, to harness the similarity of the K graphs on the whole GSOs (instead of only on the observed blocks), the objective function in (4) includes two distances. Inspired by other standard joint inference approaches [26], we capture the similarity between the observed GSOs via the function $d_S(\cdot,\cdot)$. Then, the function $d_P(\cdot,\cdot)$ captures the graph similarity between the unobserved blocks of the GSOs by minimizing the distance between the matrices $\mathbf{P}^{(k)}$. This important distinction allows us to incorporate additional structure reducing the degrees of freedom and rendering the problem more manageable.

While there are several ways of measuring graph similarity, in this work we focus on the commonly used assumption that the K graphs present a similar support [13], [26]. The influence on the observed nodes is straightforward since such an assumption implies that the sparsity pattern of the matrices $\mathbf{S}_{\mathcal{O}}^{(k)}$ is similar, so we can set

$$d_S(\mathbf{S}_{\mathcal{O}}^{(k)}, \mathbf{S}_{\mathcal{O}}^{(k')}) = \|\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{S}_{\mathcal{O}}^{(k')}\|_0. \quad (5)$$

On the other hand, to select an appropriate function $d_P(\cdot,\cdot)$ we recall that $\mathbf{P}^{(k)} = \mathbf{S}_{\mathcal{O}\mathcal{H}}^{(k)}(\mathbf{S}_{\mathcal{H}}^{(k)})^{-1}\mathbf{S}_{\mathcal{H}\mathcal{O}}^{(k)}$. From this expression, it follows that the submatrices $\mathbf{S}_{\mathcal{O}\mathcal{H}}^{(k)}$ and $\mathbf{S}_{\mathcal{H}\mathcal{O}}^{(k)}$ determine the sparsity pattern of the matrices $\mathbf{P}^{(k)}$, and hence, the graph similarity assumption also implies that the sparsity pattern of the matrices $\mathbf{P}^{(k)}$ has to be similar across the different graphs. Therefore, we select the function

$$d_P(\mathbf{P}^{(k)}, \mathbf{P}^{(k')}) = \|\mathbf{P}^{(k)} - \mathbf{P}^{(k')}\|_0. \quad (6)$$

Although the resulting distance is simple, it builds upon the known structure of the matrices $\{\mathbf{P}^{(k)}\}_{k=1}^{K}$ to harness the similarity assumption over hidden nodes.

Once the graph similarity distance is set, the final ingredient to approach the joint graph learning problem in the presence of hidden nodes is to deal with the non-convexity of (4). The non-convexity of the optimization problem arises from the presence of the ℓ_0 norm and the rank constraint. Following the classical approach in the literature, we replace the ℓ_0 norm by its convex surrogate the ℓ_1 norm, and instead of setting a maximum value for the rank of each $\mathbf{P}^{(k)}$, we minimize its nuclear norm. This result in the following convex optimization problem

$$\begin{aligned}\min_{\{\mathbf{S}_{\mathcal{O}}^{(k)},\mathbf{P}^{(k)}\}_{k=1}^{K}} \quad & \sum_{k=1}^{K} \text{tr}\left((\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)})\hat{\mathbf{C}}_{\mathcal{O}}^{(k)}\right) - \log\det(\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{P}^{(k)}) \\ & + \rho_k \|\mathbf{S}_{\mathcal{O}}^{(k)}\|_1 + \beta_k \|\mathbf{P}^{(k)}\|_* \\ & + \sum_{k<k'} \rho_{kk'} \|\mathbf{S}_{\mathcal{O}}^{(k)} - \mathbf{S}_{\mathcal{O}}^{(k')}\|_1 + \beta_{kk'} \|\mathbf{P}^{(k)} - \mathbf{P}^{(k')}\|_1 \\ \text{s. t.} \quad & \mathbf{P}^{(k)} \succcurlyeq 0; \;\; \mathbf{S}^{(k)} - \mathbf{P}^{(k)} \succ 0; \;\; \mathbf{S}_{\mathcal{O}}^{(k)} \in \mathcal{S}, \end{aligned} \quad (7)$$

where $\|\mathbf{S}_{\mathcal{O}}^{(k)}\|_1$ denotes the ℓ_1 norm of the vectorized form of $\mathbf{S}_{\mathcal{O}}^{(k)}$, and $\|\mathbf{P}^{(k)}\|_*$ denotes the nuclear norm of $\mathbf{P}^{(k)}$.

IV. Numerical results

In this section, we evaluate the performance of the proposed algorithms over both synthetic and real-world graphs. The code related to the implementation of our graph learning

Authorized licensed use limited to: Fondren Library Rice University. Downloaded on July 02,2023 at 21:10:18 UTC from IEEE Xplore. Restrictions apply.

algorithm and the code for all the simulations presented in this paper is available in GitHub[1].

The results of the experiments conducted are shown in Fig. 1, where we compare the estimated graphs obtained with our algorithm in (7), denoted as "Joint-Hid" in the legend, with other popular graphical models. The algorithms considered as baselines are: (i) the graphical Lasso algorithm, denoted as "GL" [21]; (ii) group graphical Lasso, a variant of the graphical Lasso algorithm modified to jointly estimate several related graphs that is denoted as "GGL" [26]; and (iii) latent-variable graphical Lasso, a variant of graphical Lasso that accounts for the presence of hidden nodes that is denoted as "LVGL" [29]. Then, to prevent additional errors associated with the scale ambiguity that often appears in graph learning algorithms, we consider the following mean error metric

$$\frac{1}{K}\sum_{k=1}^{K}\left\|\frac{\hat{\mathbf{S}}_{\mathcal{O}}^{(k)}}{\|\hat{\mathbf{S}}_{\mathcal{O}}^{(k)}\|_F}-\frac{\mathbf{S}_{\mathcal{O}}^{(k)}}{\|\mathbf{S}_{\mathcal{O}}^{(k)}\|_F}\right\|_F^2, \tag{8}$$

where $\hat{\mathbf{S}}_{\mathcal{O}}^{(k)}$ and $\mathbf{S}_{\mathcal{O}}^{(k)}$ respectively denote the estimated and the true observed GSO. The error reported in the experiments is the average of 100 independent realizations. The different experiments and the obtained results are detailed next.

Test case 1. We start by assessing the benefits of implementing a joint graph learning approach. The results are depicted in Fig. 1a, where the x-axis shows the number of related graphs that need to be estimated and the y-axis the error computed as in (8). The first graph $\mathcal{G}^{(1)}$ is generated as an Erdős-Rényi (ER) with $N = 20$ nodes and a link probability of $p = 0.2$ and the remaining $K - 1$ graphs are generated by randomly rewiring a fix number of edges. Among the 20 nodes, $H = 2$ nodes are selected as hidden nodes according to a uniform distribution so only $O = 18$ nodes are observed. Then, for each graph we create $M_k = 200$ graph signals drawn from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, (\mathbf{S}^{(k)})^{-1})$. From the results in Fig. 1a, we observe that the error of "Joint-Hid" and "GGL" drops as the number of related graphs increases, showcasing the benefit of the joint graph learning approach. In addition, we also observe the importance of accounting for the presence of hidden variables since "LVGL" outperforms both "GL" and "GGL" when the number of graphs being estimated is small. Finally, we remark that the "Joint-Hid" algorithm outperforms the other alternatives, which was expected since it considers both the influence of hidden nodes and the joint estimation of the graphs.

Test case 2. The second experiment compares the performance of different algorithms as the number of observed signals increases. The results are depicted in Fig. 1b where the x-axis shows the number of samples M_k, which is the same for the $K = 4$ graphs. In this case, the graphs considered are small world (SW) graphs [33] with $N = 20$ nodes, each node is connected to 3 neighbors, and the rewiring probability is $p = 0.15$. From the 20 nodes, $O = 18$ are observed nodes while the remaining $H = 2$ stay hidden. The results in Fig. 1b show that the performance of all the algorithms is similar when the number of signals is small but, as more signals are available, the proposed "Joint-Hid" method swiftly outperforms the alternatives. On the opposite side, the "GL" method requires a larger amount of signals to obtain an error below 0.1. As a result, this experiment highlights that following a joint approach that takes into account the hidden variables allows more efficient usage of the available signals, rendering a smaller error with fewer observed signals.

Test case 3. Finally, we close the numerical evaluation by studying the influence of the number of hidden nodes. In this case, we employ $K = 4$ real-world graphs extracted from the network of students of the University of Ljubljana dataset[2]. All the graphs are composed of the same set of 32 nodes, which represent students, and the different networks capture different types of interactions among the students. The graph signals have been generated according to a GMRF as in the previous experiments. The error as the number of observed nodes increases (and hence, the number of hidden nodes becomes smaller), is illustrated in Fig. 1c. Looking at the figure, it is clear that the proposed "Joint-Hid" algorithm outperforms the other alternatives. Moreover, we note that, as the number of hidden nodes gets closer to 0, the performances of "GL" and "LVGL" seems to converge. The same happens with "GGL" and "Joint-Hid". Interestingly, this behavior suggests that no evident prejudice follows from accounting for hidden nodes when all the nodes are being observed.

V. Conclusions

In this paper, we approached the challenging task of joint graph learning of several closely related graphs in the presence of hidden nodes. To establish a relation between the observed signals and the unknown graphs, we assumed that the signals were drawn from a GMRF model, i.e., $\mathbf{X}^{(k)} \sim \mathcal{N}(\mathbf{0}, (\mathbf{S}^{(k)})^{-1})$. Because the presence of hidden nodes renders the problem ill-conditioned, we further assumed that the number of observed nodes is much larger than the number of hidden nodes ($O \gg H$) to ensure the tractability of the problem. Then, we modified the GMRF assumption to account for the hidden nodes by exploiting the block structure of $\mathbf{S}^{(k)}$ and $\hat{\mathbf{C}}^{(k)}$, resulting in the definition of the low-rank matrices $\mathbf{P}^{(k)}$. Moreover, after studying the definition of these matrices, we arrived at the conclusion that we could harness the graph similarity between blocks of $\mathbf{S}^{(k)}$ associated with hidden nodes by estimating matrices $\mathbf{P}^{(k)}$ with a similar sparsity pattern. This observation allowed us to leverage the graph similarity on the *whole graphs*. Finally, we presented a convex algorithm employing the ℓ_1 norm and the nuclear norm as convex surrogates for the ℓ_0 norm and the matrix rank. The performance of the proposed algorithm was compared with other popular Gaussian graphical models over synthetic and real-world graphs.

[1]https://github.com/reysam93/hidden_joint_gaussian_inf

[2]The original data can be found at http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:pajek:students

Authorized licensed use limited to: Fondren Library Rice University. Downloaded on July 02,2023 at 21:10:18 UTC from IEEE Xplore. Restrictions apply.

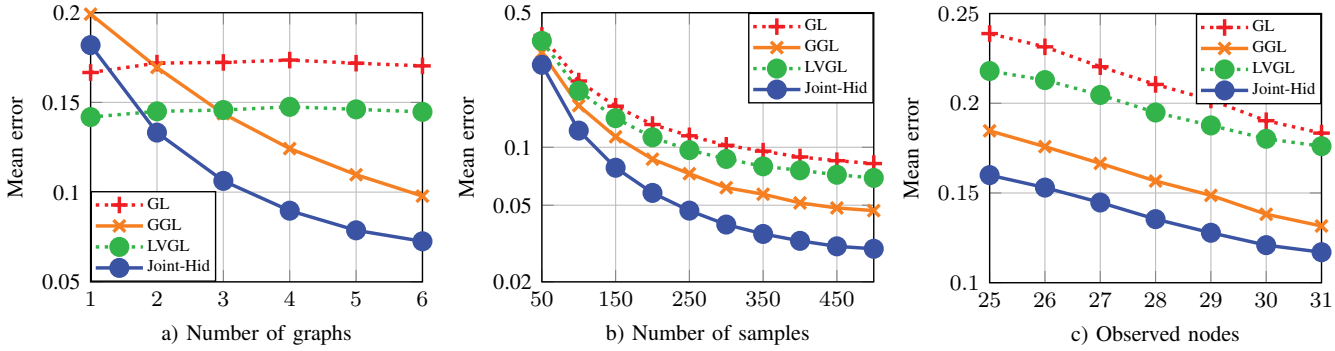

Fig. 1: Mean error of the estimated matrices $\hat{\mathbf{S}}^{(k)}$ through several algorithms and in different types of graphs. The error is computed according to (8) and we report the mean of 100 independent realizations. a) Graphs being estimated are ER graphs with $N = 20$ nodes and $p = 0.15$. b) Graphs are drawn from an SW random graph model with $N = 20$ nodes, 3 neighbors per node, and a rewiring probability of $p = 0.15$. c) Graphs are obtained from the students of Ljubljana dataset.

REFERENCES

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

[2] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "Graph topology inference based on sparsifying transform learning," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1712–1727, 2019.

[3] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu, "Graph learning: A survey," *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 109–127, 2021.

[4] 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.

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

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

[7] P. Djuric and C. Richard, *Cooperative and Graph Signal Processing: Principles and Applications*, Academic Press, 2018.

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

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

[10] B. Baingana, G. Mateos, and G. B. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 563–575, Aug. 2014.

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

[12] R. Shafipour and G. Mateos, "Online topology inference from streaming stationary graph signals with partial connectivity information," *Algorithms*, vol. 13, no. 9, pp. 228, 2020.

[13] M. Navarro, Y. Wang, A. G. Marques, C. Uhler, and S. Segarra, "Joint inference of multiple graphs from matrix polynomials," *J. Mach. Learn. Res.*, vol. 23, no. 76, pp. 1–35, 2022.

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

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

[16] S. S. Saboksayr and G. Mateos, "Accelerated graph learning from smooth signals," *IEEE Signal Process. Lett.*, vol. 28, pp. 2192–2196, 2021.

[17] S. Rey, T. M. Roddenberry, S. Segarra, and A. G. Marques, "Enhanced graph-learning schemes driven by similar distributions of motifs," *arXiv preprint arXiv:2207.04747*, 2022.

[18] M. Navarro and S. Segarra, "Joint network topology inference via a shared graphon model," *IEEE Trans. Signal Process.*, 2022.

[19] S. L. Lauritzen, *Graphical Models*, vol. 17, Clarendon Press, 1996.

[20] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Statist.*, vol. 34, pp. 1436–1462, 2006.

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

[22] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu, "High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence," *Electron. J. Statistics*, vol. 5, pp. 935–980, 2011.

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

[24] S. Kumar, J. Ying, J. Cardoso, and D. Palomar, "Structured graph learning via Laplacian spectral constraints," *Advances Neural Inf. Process. Syst.*, vol. 32, 2019.

[25] Y. Murase, J. Török, H. H. Jo, K. Kaski, and J. Kertész, "Multilayer weighted social network model," *Physical Review E*, vol. 90, no. 5, pp. 052810, 2014.

[26] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *J. Roy. Statistical Soc.: Ser. B (Statistical Methodology)*, vol. 76, no. 2, pp. 373–397, 2014.

[27] J. Arroyo, A. Athreya, J. Cape, G. Chen, C. E. Priebe, and J. T. Vogelstein, "Inference for multiple heterogeneous networks with a common invariant subspace," *J. Mach. Learn. Res.*, vol. 22, no. 142, pp. 1–49, 2021.

[28] S. Rey, A. Buciulea, M. Navarro, S. Segarra, and A. G. Marques, "Joint inference of multiple graphs with hidden variables from stationary graph signals," in *IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*. IEEE, 2022, pp. 5817–5821.

[29] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *Ann. Statist.*, vol. 40, no. 4, pp. 1935–1967, 2012.

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

[31] A. Chang, T. Yao, and G. I. Allen, "Graphical models and dynamic latent factors for modeling functional brain connectivity," *2019 IEEE Data Science Wrksp. (DSW)*, pp. 57–63, 2019.

[32] A. Buciulea, S. Rey, and A. G. Marques, "Learning graphs from smooth and graph-stationary signals with hidden variables," *IEEE Trans. Signal Inf. Process. Netw.*, 2022.

[33] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

Authorized licensed use limited to: Fondren Library Rice University. Downloaded on July 02,2023 at 21:10:18 UTC from IEEE Xplore. Restrictions apply.