ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448337

LEARNING MULTIPLEX GRAPH WITH INTER-LAYER COUPLING

Chenyue Zhang, Hoi-To Wai

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

ABSTRACT

In many real-life systems, the interactions among entities are complex and varied. This necessitates the use of a multiplex graph model with heterogeneous layers of graphs to effectively describe these interactions. The current paper focuses on incorporating high-order relations, specifically inter-layer couplings or connections, in multiplex graph learning. Through developing a high-order smoothness criterion, we propose an algorithm that integrates inter-layer connections to perform inference from multi-attribute graph signals. We show that it is essential to consider high-order interactions in the inference process. We validate our claims through numerical experiments, demonstrating their efficacy in capturing the intricate relationships within multiplex networks.

Index Terms— graph signal processing, multiplex graph learning, multi-attribute graph signals, smooth graph signals

1. INTRODUCTION

In recent years, there has been a rising interest in machine learning and signal processing focusing on graph-based modeling. Methods for extracting information such as graph topology, centrality, etc., from network data have found applications in science and business applications. This has led to the emergence of studies on graph signal processing (GSP) which extend traditional signal processing on time-domain to irregular (graph) data domains. A key concept of GSP on data science involves the abstraction of network data as graph signals generated from graph filters excited by external influences. Equipped with such model, a notable trend is the development of graph learning methods for uncovering the latent graph topology from network data, as outlined in [1,2] and related references. These methods have applications spanning the analysis of social dynamics, financial systems, power systems, and biological networks [3].

Prior research on graph topology learning using GSP models have primarily focused on single layer networks [1, 2], where signals are represented as graph nodes connected by a single type of edges. However, real-world network data often exhibit multifaceted or multi-way interactions [4]. Their corresponding networked systems are marked by the heterogeneity of their connections, which can be represented as *multiplex* graphs. As multiplex graph comprises of inter-connected layers of graphs with distinct structures, such configuration encompasses coupling interactions both *within* individual layers and *across* multiple network layers.

Inter-layer coupling across network layers often plays a crucial role in real-world network systems. For example, in cyber-physical systems, a failure in the physical-resource network can impact the computational-resource network [5]. Similarly, in supra-diffusion dynamics, inter-layer coupling substantially influences diffusion rates [4]. Additionally, it plays a crucial role in understanding synchronization patterns and stability in multiplex oscillators [6]. Other examples, such as multi-dimensional opinion dynamics on correlated topics [7], protein-protein interactions [8], animal networks [9], and image pixel relations [10], underscore the significance of coupling interactions within individual layers and across multiple network layers in various dynamic phenomena. Attempting to capture or learn these systems by treating them with a simple graph model, without taking into account their inter-layer coupling structure, may miss out on these essential properties.

This paper addresses the challenge of learning the inter-coupling topology in multiplex networks. Our contributions span from the fundamental aspect of graph signals modeling to the algorithm aspect of developing tractable inference algorithms, as summarized:

- We propose a multiplex graph filter model for modeling multi-attribute graph signals. This model incorporates distinct nonlinear intra-layer and inter-layer coupling dynamics in the form of a multinomial function. We demonstrate that it covers a wide range of common dynamics observed in multiplex networks.
- We develop a high-order smoothness total variation (TV) criterion through approximating the multiplex filter by a composition of low pass graph filters on high-order graph shifts. The new TV criterion allows for fitting the graph signals stemmed from the unique dynamics on multiplex networks characterized by coupling both across and within layers.
- We derive an alternating optimization (AO) procedure for learning the multiplex graph model through the proposed smoothness metric. We show that the subproblems in the AO procedure can be solved efficiently as thru reduction to lower dimensional graph learning problems.

Finally, we present numerical experiments to support our claims. Our results underscore the importance of treating the coupling effects of graph signals in multiplex graph learning.

Related Works. The studies of graph signal processing with multilayer graph are relatively rare. The closest work to ours is [11] which proposed a tensor GSP model for multilayer graph. However, [11] focused on graph filters with a single shift operator that do not consider distinct inter-layer coupling dynamics, nor do they study multiplex graph learning method. On the other hand, there has been growing interests in GSP for *product graphs*, which is a special case of multiplex graphs but with homogeneous layer-graphs. This stemmed from the series of works [12–15] for modeling time-vertex data. Recent works consider the product graph learning problem via exploring the smoothness criterion [16, 17], spectral template [18], and spectral methods with emphasis on the inter-layer coupling effects [19]. In comparison, our work tackles the multiplex graph learning problem and suggests insights on how to model the corresponding graph data with respect to a multiplex graph model.

Other than the signal processing community, numerous papers in network science [20, 21], have delved into uncovering hidden structures within multiplex networks. Aspects such as centrality and com-

This work is supported in part by HKRGC project #24203520. Emails: {czhang,htwai}@se.cuhk.edu.hk

munity detection that are catered for multiplex graph have been considered, also see [22]. Many of these papers assume prior knowledge of the graph topology, while our work focuses on the inverse problem of learning the multiplex graph topology.

2. SIGNAL MODEL

Multiplex Graph. We consider a multiplex graph with L layers that is denoted by the triplet $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{G}^C \rangle$. The set $\mathcal{V}$ with $|\mathcal{V}| = N$ represents the set of nodes. The collection $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_L)$ describes L sets of edges between the nodes in $\mathcal{V}$. For each ℓ, the edge set $\mathcal{E}_\ell = \{(v_i, v_j) : v_i \neq v_j, v_i, v_j \in \mathcal{V}\}$ induces the ℓth layer-graph denoted by $\mathcal{G}_\ell = \langle \mathcal{V}, \mathcal{E}_\ell \rangle$. Meanwhile, $\mathcal{G}^C = \langle [L], \mathcal{E}^C \rangle$ with the layer set $[L] = \{1, \ldots, L\}$ describes the *coupling graph* consisting of the coupling edges between a node and its clone in the layer-graphs. We let $\boldsymbol{A}_\ell \in \mathbb{R}^{N \times N}$, $\boldsymbol{C} \in \mathbb{R}^{L \times L}$ be the (weighted) adjacency matrix of $\mathcal{G}_\ell$, $\mathcal{G}^C$, respectively. For simplicity, these adjacency matrices are assumed to be symmetric.

It is convenient to describe $\mathcal{G}$ using a simple (single-layer) graph. We define the set of *supranodes* at the ℓth layer-graph $V_\ell := \{(v, \ell) : v \in \mathcal{V}\}$ and the corresponding edge set $E_\ell = \{((v_i, \ell), (v_i, \ell)) : (v_i, v_j) \in \mathcal{E}_\ell\}$. Note that $V_\ell, \ell \in [L]$ consists of 'clones' of the nodes in $\mathcal{V}$ at different layers. Subsequently, the *intra-layer graph* that consists of NL supranodes is taken as the union graph $G^L = \langle V, E^L \rangle$ with $V = \cup_{\ell=1}^L V_\ell$, $E^L = \cup_{\ell=1}^L E_\ell$. The coupling graph can be similarly extended as $G^C = \langle V, E^C \rangle$ with $E^C = \{((v, \ell), (v, \ell')) : v \in \mathcal{V}, (\ell, \ell') \in \mathcal{E}^C\}$. Observe that both G^L, G^C are simple graphs defined on NL supranodes. We define their adjacency matrices respectively as

$$\boldsymbol{\mathcal{A}}_L := \text{blkdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_L), \quad \boldsymbol{\mathcal{A}}_C := \boldsymbol{C} \otimes \boldsymbol{I}_N, \tag{1}$$

where the block diagonal elements of $\boldsymbol{\mathcal{A}}_L$ are in order $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_L$ and $\otimes$ is the Kronecker product. Moreover, a common approach [4] to encode multiplex graphs is via the supra-adjacency matrix:

$$\boldsymbol{\mathcal{A}} := \boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_C. \tag{2}$$

The simple graph induced by the supra-adjacency matrix $\boldsymbol{\mathcal{A}}$ effectively treats every (intra-layer and coupling) edges in $\mathcal{G}$ as equal.

Multi-attribute Graph Signals. Our next endeavor is to study the multi-attribute graph signal on $\mathcal{G}$, $y : \mathcal{V} \to \mathbb{R}^L$, where each node is associated with L attributes, each associated with one layer-graph in $\mathcal{G}$. A natural way to represent the graph signal is to consider y as a graph signal defined on the supranode set V, given by the vector $\boldsymbol{y} = (y_1(\mathcal{V})^\top \cdots y_L(\mathcal{V})^\top)^\top \in \mathbb{R}^{NL}$.

The observed graph signal $\boldsymbol{y}$ is modeled as the output of a *graph filter* that captures the interplay between the intra-layer graph and coupling graph. The mth signal observation is modeled by

$$\boldsymbol{y}^{(m)} = \mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\boldsymbol{x}^{(m)} + \boldsymbol{w}^{(m)}, \tag{3}$$

where $\boldsymbol{x}^{(m)}$ is an unknown excitation and $\boldsymbol{w}^{(m)}$ models the observation noise. For simplicity, we assume that $\boldsymbol{x}^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{w}^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and the random variables (r.v.s) are independently distributed. We model the multiplex graph filter $\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$ after the general product graph filter in [12–15, 23]:

$$\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) = \sum_{t=0}^{T-1} \sum_{j=0}^{2^t - 1} h_{t,j} \prod_{i=1}^{t} \boldsymbol{\mathcal{A}}_L^{b_{t,j}^{(i)}} \boldsymbol{\mathcal{A}}_C^{1 - b_{t,j}^{(i)}}, \tag{4}$$

where $b_{t,j}^{(i)} \in \{0, 1\}$, $i = 1, \ldots, t$ is a binary sequence that encodes the integer $j \in \{0, \ldots, 2^t - 1\}$, $h_{t,j} \in \mathbb{R}$ is a filter coefficient, and $T \in \mathbb{N}$ is the filter's order (can be infinite). We have defined $\boldsymbol{\mathcal{A}}^0 = \boldsymbol{I}_{NL}$ for notation convenience.

To justify (4), we notice that the intra-layer and coupling graphs may yield different dynamics models as their respective edges play different roles in the multiplex graph. Eq. (4) is thus designed to capturing such distinct dynamics. As a special case of (4), the bivariate filter $\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) = \sum_{i=0}^{T_1} \sum_{j=0}^{T_2} h_{ij} \boldsymbol{\mathcal{A}}_L^i \boldsymbol{\mathcal{A}}_C^j$ gives a joint filter where the intra-layer, coupling dynamics depends on $\{h_{ij}\}_{j=1}^{T_2}$, $\{h_{ij}\}_{i=1}^{T_1}$, respectively. We further instantiate (3), (4) by:

Example 1. *Consider an extension of multi-dimensional opinion dynamics with agent-independent logical matrices [7]. At time $t \geq 0$ of the mth discussion, the evolution of multi-dimensional opinions for ℓth agent is described by:*

$$\boldsymbol{x}_\ell(t+1) = \boldsymbol{A}_\ell \textstyle\sum_{\ell'=1}^{L} C_{\ell,\ell'} \boldsymbol{x}_{\ell'}(t) + \boldsymbol{x}_\ell^{(m)}, \tag{5}$$

where $\boldsymbol{C} \in \mathbb{R}^{L \times L}$ encodes the mutual trusts between L agents and $\boldsymbol{A}_\ell \in \mathbb{R}^{N \times N}$ is the logical matrix between N topics, $\boldsymbol{x}^{(m)}$ denotes the initial beliefs. Stacking up $\boldsymbol{x}_\ell(t)$ to be $\boldsymbol{x}(t)$ such that its ℓth block represents the opinions held by the ℓth agent. These matrices are properly scaled to ensure that the vector $(\boldsymbol{I}_{NL} - \boldsymbol{\mathcal{A}}_L \boldsymbol{\mathcal{A}}_C)\boldsymbol{1}$ is positive. Our observation is given by the steady-state opinions:

$$\boldsymbol{y}^{(m)} = \lim_{t \to \infty} \boldsymbol{x}(t) = (\boldsymbol{I}_{NL} - \boldsymbol{\mathcal{A}}_L \boldsymbol{\mathcal{A}}_C)^{-1} \boldsymbol{x}^{(m)}. \tag{6}$$

Example 2. *The supra-diffusion process [4] captures the dynamics such as social interactions, epidemics, and transportation on a graph composed of multiple interconnected layers. Specifically, driven by the excitation $\boldsymbol{x}^{(m)} = (\boldsymbol{x}_1^{(m)}, \ldots, \boldsymbol{x}_L^{(m)})$, at layer ℓ, the node states at time t evolve as:*

$$\frac{d\boldsymbol{x}_\ell(t)}{dt} = -\boldsymbol{x}_\ell(t) + \boldsymbol{A}_\ell \boldsymbol{x}_\ell(t) + \sum_{\ell'=1}^{L} C_{\ell,\ell'} \boldsymbol{x}_{\ell'}(t) + \boldsymbol{x}_\ell^{(m)}. \tag{7}$$

This is a diffusion process defined on the supra-adjaceny matrix $\boldsymbol{\mathcal{A}}$. With properly scaled $\boldsymbol{\mathcal{A}}_L$, $\boldsymbol{\mathcal{A}}_C$, (7) admits a unique equilibrium:

$$\boldsymbol{y}^{(m)} = (\boldsymbol{I}_{NL} - (\boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_C))^{-1} \boldsymbol{x}^{(m)}, \tag{8}$$

Although (4) describes a flexible model that allows for different dynamics between the intra-layer and coupling graphs, the model lacks a spectral domain interpretation which is essential to graph frequency analysis. One of the key issues lies with the fact that (4) is non-symmetric, even when its component matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_L, \boldsymbol{C}$ are. This limits our ability to analyze the multi-attribute graph signals in (3), and subsequently, develop an inference scheme.

3. MULTIPLEX GRAPH LEARNING

This section presents a method for learning the multiplex graph $\mathcal{G}$ from samples of graph signals observed in (3), i.e., to estimate $\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C$ from M samples $\{\boldsymbol{y}^{(m)}\}_{m=1}^M$. Our idea is to leverage the smoothness of graph signals in the graph learning process. Inspired by [24], we adopt a quadratic total variation (TV) criterion (a.k.a. Dirichlet energy). However, we depart from the original interpretation in [24] and re-consider the TV criterion as an approximate matched filter objective for fitting *low pass* graph signals. Using this interpretation, we derive a *high-order smoothness* criterion for fitting the multi-attribute graph signals generated by strong coupling dynamics across layers.

Our endeavor begins with the following TV criterion:

$$\text{TV}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) := \textstyle\sum_{i,j=1}^{NL} \left[\hat{h}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\right]_{ij} S_{ij}, \tag{9}$$

where $[\boldsymbol{\mathcal{A}}]_{ij}$ denotes the (i,j)th entry of a matrix $\boldsymbol{\mathcal{A}}$,

$$S_{ij} := (1/M)\sum_{m=1}^{M} |y_i^{(m)} - y_j^{(m)}|^2, \; i,j = 1,\ldots,NL. \quad (10)$$

and $\hat{h} : \mathbb{R}^{NL\times NL} \times \mathbb{R}^{NL\times NL} \to \mathbb{R}^{NL\times NL}$ fuses the intra-layer and coupling graphs whose form will be discussed later. Observe that $\mathrm{TV}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$ is the inner product $\langle \hat{h}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) \,|\, \boldsymbol{S}\rangle$.

We note that TV measures the 'smoothness' of the observed graph signals with respect to (w.r.t.) a fused simple graph with the weighted adjacency matrix given by $\hat{h}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$. To gain more insights, we shall take a closer look at the signal model. The distance matrix $\boldsymbol{S}$ can be written as

$$\mathbf{1}\mathbb{E}_m[\boldsymbol{y}^{(m)} \odot \boldsymbol{y}^{(m)}]^\top + \mathbb{E}_m[\boldsymbol{y}^{(m)} \odot \boldsymbol{y}^{(m)}]\mathbf{1}^\top - 2\mathbb{E}_m[\boldsymbol{y}^{(m)}(\boldsymbol{y}^{(m)})^\top]$$

where $\odot$ denotes element-wise product and the expectation $\mathbb{E}_m[\cdot]$ is taken as the empirical mean over M samples. Recall that $\boldsymbol{x}^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and assume that $\boldsymbol{w}^{(m)} \approx \mathbf{0}$, $M \gg 1$ yields

$$\begin{aligned} &\mathbb{E}_m[\boldsymbol{y}^{(m)} \odot \boldsymbol{y}^{(m)}] \approx \big(\cdots \|\boldsymbol{e}_i^\top \mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\|^2 \cdots\big)^\top, \\ &\mathbb{E}_m[\boldsymbol{y}^{(m)}(\boldsymbol{y}^{(m)})^\top] \approx \mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)^\top. \end{aligned} \quad (11)$$

As $\|\boldsymbol{e}_i^\top \mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\|^2$ is approximately constant, the TV criterion $\mathrm{TV}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$ can be *minimized* if the inner product $\langle \hat{h}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) \,|\, \mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)^\top\rangle$ is *large*. This suggests that $\hat{h}(\cdot,\cdot)$ should be a *matched* multiplex graph filter.

To determine $\hat{h}(\cdot,\cdot)$, we consider the following (possibly non-unique) decomposition for the multiplex graph filter (4):

$$\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)^\top = \quad (12)$$

$$\overline{\mathcal{H}}_1(\boldsymbol{\mathcal{A}}_L) + \overline{\mathcal{H}}_2(\boldsymbol{\mathcal{A}}_C) + \overline{\mathcal{H}}_3(\boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C + \boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L) + \widetilde{\mathcal{H}}_{\sf res}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$$

where $\widetilde{\mathcal{H}}_{\sf res}(\cdot,\cdot)$ is the residual term. The graph filters $\overline{\mathcal{H}}_1(\boldsymbol{\mathcal{A}}_L)$, $\overline{\mathcal{H}}_2(\boldsymbol{\mathcal{A}}_C)$ model the dynamics within and between the layer-graphs, respectively. Meanwhile, the graph filter $\overline{\mathcal{H}}_3(\boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C + \boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L)$ captures dynamics between two-hops neighbors in the multiplex graph that *couple* the layer-graphs together, e.g., as seen in (6). The modeling of such two-hops neighbors marks a significant difference between the multiplex graph and a simple graph induced by the supra-adjacency matrix (2). They represent cross-layer interactions that emphasize the inter-layer coupling of multilayer graph. Fig. 1 illustrates the edges modeled by the corresponding GSOs.

Notice that the matrices $\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C, \boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C$ are symmetric. Taking them as the GSOs of the respective graph filters, their eigen decompositions admit a natural graph spectral interpretation. To this end, we assume that

H1. *For each $i = 1,2,3$, the graph filter $\overline{\mathcal{H}}_i(\boldsymbol{\mathcal{A}}) := \sum_{t=0}^{T-1} h_t^{(i)} \boldsymbol{\mathcal{A}}^t$ is low pass at all graph frequencies [25]. That is, the scalar polynomial function $\overline{h}_i(\lambda) := \sum_{t=0}^{T-1} h_t^{(i)} \lambda^t$ is non-decreasing in λ.*

It can be verified that H1 holds for the respective decomposition of the form (12) with the dynamics in Examples 1, 2.

Under H1 and suppose that the residual term $\widetilde{\mathcal{H}}_{\sf res}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$ is negligible, setting $\hat{h}(\cdot,\cdot)$ to be a linear mixture of the GSOs yields an approximate matched graph filter to (12). We propose

$$\hat{h}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) := \boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_C + \lambda(\boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C), \quad (13)$$

where $\lambda \geq 0$ is a tunable parameter. Notice that other forms of $\hat{h}(\cdot,\cdot)$ to match with $\mathcal{H}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C)$ are available. However, they may not yield tractable objective function for the TV in (9), especially when the exact forms of $\overline{\mathcal{H}}_i(\cdot)$ are unknown. Finally, substituting (13) into (9) yields a *high-order smoothness* TV objective that captures the two-hops inter-layer coupling in multiplex graph.

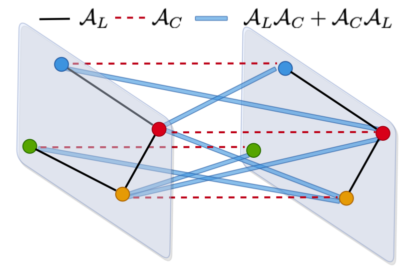

Fig. 1. Illustrating the intra-layer graph $\boldsymbol{\mathcal{A}}_L$, coupling graph $\boldsymbol{\mathcal{A}}_C$ and the two-hops cross-layer coupling graph $\boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C + \boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L$.

3.1. Alternating Optimization Algorithm

Equipped with the high-order smoothness TV criterion (9), (13), we are ready to discuss the solution strategy for learning $\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C$. Our focus is the following non-convex optimization problem:

$$\begin{aligned} \min_{\substack{\boldsymbol{C}\in\mathcal{W}^L,\\ \boldsymbol{A}_\ell\in\mathcal{W}^N, \ell\in[L]}} \quad & \mathrm{TV}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) + \rho_L \textstyle\sum_{\ell=1}^L \|\boldsymbol{A}_\ell\|_F^2 + \rho_C\|\boldsymbol{C}\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{\mathcal{A}}_L = \mathrm{blkdiag}(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_L),\; \boldsymbol{\mathcal{A}}_C = \boldsymbol{C}\otimes\boldsymbol{I}_N, \end{aligned} \quad (14)$$

where the structures of $\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C$ are taken into account (cf. (1)), $\rho_L, \rho_C > 0$ are regularization parameters, and the feasible set is:

$$\mathcal{W}^N := \{\boldsymbol{A} \in \mathbb{R}_+^{N\times N} : \mathbf{1}^\top\boldsymbol{A}\mathbf{1} = N, \boldsymbol{A} = \boldsymbol{A}^\top, \mathrm{diag}(\boldsymbol{A}) = \mathbf{0}\}. \quad (15)$$

The TV criterion can be written as

$$\begin{aligned} \mathrm{TV}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) = & \langle \boldsymbol{\mathcal{A}}_L \,|\, \boldsymbol{S}\rangle + \langle \boldsymbol{\mathcal{A}}_C \,|\, \boldsymbol{S}\rangle \\ & + \lambda\langle \boldsymbol{\mathcal{A}}_L\boldsymbol{\mathcal{A}}_C + \boldsymbol{\mathcal{A}}_C\boldsymbol{\mathcal{A}}_L \,|\, \boldsymbol{S}\rangle \end{aligned} \quad (16)$$

where the first two terms are

$$\langle \boldsymbol{\mathcal{A}}_L \,|\, \boldsymbol{S}\rangle = \textstyle\sum_{\ell=1}^L \langle \boldsymbol{A}_\ell \,|\, \mathrm{c}_\ell(\boldsymbol{S})\rangle, \; \langle \boldsymbol{\mathcal{A}}_C \,|\, \boldsymbol{S}\rangle = \sum_{i=1}^N \langle \boldsymbol{C} \,|\, \mathrm{l}_i(\boldsymbol{S})\rangle$$

such that $\mathrm{c}_\ell(\boldsymbol{S})$ takes the ℓth principal submatrix of $\boldsymbol{S}$, and $\mathrm{l}_i(\boldsymbol{S})$ collects the ith diagonal elements in each of the $N\times N$ block matrices in $\boldsymbol{S}$, i.e., $[\mathrm{l}_i(\boldsymbol{S})]_{k,\ell} := [\boldsymbol{S}]_{(k-1)N+i,(\ell-1)N+i}$, $k,\ell \in [L]$.

When $\lambda = 0$, i.e., without the last bilinear term in (16), problem (14) can be solved independently for each of $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_L,\boldsymbol{C}$. In this case, the TV criterion can be seen as an extension of the 'stacked-up' formulation in [16] which takes only the layer-wise and node-wise distances between graph signals. For instance, $\langle \boldsymbol{A}_\ell \,|\, \mathrm{c}_\ell(\boldsymbol{S})\rangle$ fits the graph signal distances *within* the ℓ-th layer-graph $\boldsymbol{A}_\ell$, which does not account for influences from the node's clones in other layers. This may be insufficient to expose the diverse coupling between layer-graphs for multiplex networks.

In general, problem (14) is non-convex due to the bilinear term in (16). As a remedy, we adopt a standard alternating optimization (AO) approach. Observe that when we fix $\boldsymbol{C} = \overline{\boldsymbol{C}}$, the AO subproblem is separable such that for any $\ell \in [L]$, optimizing (14) w.r.t. $\boldsymbol{A}_\ell$ yields the convex problem:

$$\min_{\boldsymbol{A}_\ell\in\mathcal{W}^N} \big\langle \boldsymbol{A}_\ell \,|\, \mathrm{c}_\ell\big(\boldsymbol{S} + \lambda(\boldsymbol{\mathcal{A}}_{\overline{C}}\boldsymbol{S} + \boldsymbol{S}\boldsymbol{\mathcal{A}}_{\overline{C}})\big)\big\rangle + \rho_L\|\boldsymbol{A}_\ell\|_F^2,$$

with $\boldsymbol{\mathcal{A}}_{\overline{C}} := \overline{\boldsymbol{C}} \otimes \boldsymbol{I}_N$ and we recall that $\mathrm{c}_\ell(\cdot)$ is a linear transformation. Similarly, as we fix $\boldsymbol{A}_\ell = \overline{\boldsymbol{A}}_\ell$ for $\ell \in [L]$, optimizing (14) w.r.t. $\boldsymbol{C}$ yields the convex problem:

$$\min_{\boldsymbol{C}\in\mathcal{W}^L} \sum_{i=1}^N \big\langle \boldsymbol{C} \,|\, \mathrm{l}_i\big(\boldsymbol{S} + \lambda(\boldsymbol{\mathcal{A}}_{\overline{L}}\boldsymbol{S} + \boldsymbol{S}\boldsymbol{\mathcal{A}}_{\overline{L}})\big)\big\rangle + \rho_C\|\boldsymbol{C}\|_F^2$$

Algorithm 1 Alternating Optimization (AO) procedure for (14)

Input: Graph signals $\{\boldsymbol{y}^{(m)}\}_{m=1}^{M}$, maximum iterations T_0, parameters λ, ρ_L, ρ_C, initialization for the coupling graph $\boldsymbol{C}^{(0)}$.

Compute the pairwise distance matrix $\boldsymbol{S}$ by (10).

for $t = 0$ to $T_0 - 1$ **do**

Set $\boldsymbol{P}^{(t)} = \boldsymbol{S} + \lambda((\boldsymbol{C}^{(t)} \otimes \boldsymbol{I})\boldsymbol{S} + \boldsymbol{S}(\boldsymbol{C}^{(t)} \otimes \boldsymbol{I}))$ and solve the subproblems for $\ell \in [L]$:

$$\boldsymbol{A}_\ell^{(t+1)} \in \underset{\boldsymbol{A}_\ell \in \mathcal{W}^N}{\arg\min} \langle \boldsymbol{A}_l \,|\, \mathrm{c}_\ell(\boldsymbol{P}^{(t)}) \rangle + \rho_L \|\boldsymbol{A}_\ell\|_F^2.$$

Set $\boldsymbol{Q}^{(t+1)} = \boldsymbol{S} + \lambda(\bigoplus_{\ell=1}^{L} \boldsymbol{A}_\ell^{(t+1)} \boldsymbol{S} + \boldsymbol{S} \bigoplus_{\ell=1}^{L} \boldsymbol{A}_\ell^{(t+1)})$ and solve the subproblem:

$$\boldsymbol{C}^{(t+1)} \in \underset{\boldsymbol{C} \in \mathcal{W}^L}{\arg\min} \sum_{i=1}^{N} \langle \boldsymbol{C} \,|\, \mathrm{l}_i(\boldsymbol{Q}^{(t+1)}) \rangle + \rho_C \|\boldsymbol{C}\|_F^2.$$

Note that $\bigoplus_\ell$ denotes the direct sum of matrices.

end for

Output: layer graphs $\boldsymbol{A}_1^{(T)}, \ldots, \boldsymbol{A}_L^{(T)}$, coupling graph $\boldsymbol{C}^{(T)}$.

with $\boldsymbol{\mathcal{A}}_{\overline{L}} = \mathrm{blkdiag}(\overline{\boldsymbol{A}}_1, \ldots, \overline{\boldsymbol{A}}_L)$. Notice that the above subproblems are tractable as they only involve the optimization of $N \times N$ and $L \times L$ symmetric matrices.

The AO procedure is summarized in Algorithm 1. It is known that the procedure finds a stationary point of (14) [26] as $T_0 \to \infty$.

4. NUMERICAL EXPERIMENT

This section presents preliminary experiments on the proposed multiplex graph learning method using synthetic data. We aim to verify if modeling the inter-layer coupling of dynamics between two-hops neighbors is necessary for reliably learning the multiplex graph.

Throughout this section, we set the coupling graph $\mathcal{G}^C$ to be an $L = 4$-nodes graph with a star graph topology. The adjacency matrix is given by $\boldsymbol{C} = \sum_{\ell=2}^{4} (\boldsymbol{e}_{\ell-1}(\boldsymbol{e}_\ell)^\top + \boldsymbol{e}_\ell(\boldsymbol{e}_{\ell-1})^\top)$, where $\boldsymbol{e}_\ell \in \mathbb{R}^L$ is the ℓth canonical basis vector. Meanwhile, for each $\ell \in [L]$, the intra-layer graph $\mathcal{G}_\ell$ is generated as Erdos-Renyi graph with connection probability p_ℓ. For different layers, we have $p_1 = 0.2$, $p_2 = 0.3, p_3 = 0.4$, and $p_4 = 0.5$, Correspondingly, for $\ell \in [L]$, $\boldsymbol{A}_\ell$ is the unweighted adjacency matrix of the above intra-layer graph.

For the proposed algorithm, we set the regularization parameters as $\rho_L = 1$, $\rho_C = 15$. Additionally, the maximum iteration for Algorithm 1 is $T_0 = 100$, and the procedure is terminated when the changes in iterate satisfy $\max\{\|\boldsymbol{\mathcal{A}}_C^{(t+1)} - \boldsymbol{\mathcal{A}}_C^{(t)}\|_F, \|\boldsymbol{\mathcal{A}}_L^{(t+1)} - \boldsymbol{\mathcal{A}}_L^{(t)}\|_F\} \leq 10^{-3}$. The convex sub-problems are solved using the `cvx` package on MATLAB. As a benchmark, we consider solving a special case of (14) with $\lambda = 0$. This formulation with $\lambda = 0$ is a direct extension over [16]. We recall that it ignores the inter-layer coupling term in (13) that allows to capture the inter-layer coupling between the 'two-hops' neighbors in $\mathcal{G}$.

The distance matrix $\boldsymbol{S}$ in (10) is computed from M synthetic graph signals generated from (3). For each graph signal, we generate $\boldsymbol{x}^{(m)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{w}^{(m)} \sim \mathcal{N}(\boldsymbol{0}, 0.01\boldsymbol{I})$, and simulate the graph filters that are modeled after Examples 1, 2:

- 'Weak coupling': $\mathcal{H}_{\mathsf{wk}}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) = (\boldsymbol{I} - \tau_{\mathsf{wk}}(\boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_C))^{-1}$.
- 'Strong coupling': $\mathcal{H}_{\mathsf{str}}(\boldsymbol{\mathcal{A}}_L, \boldsymbol{\mathcal{A}}_C) = (\boldsymbol{I} - \tau_{\mathsf{str}} \boldsymbol{\mathcal{A}}_L \boldsymbol{\mathcal{A}}_C)^{-1}$.

We also set $\tau_{\mathsf{wk}} = 1/\max_i d_i^{\mathsf{wk}}, \tau_{\mathsf{str}} = 1/\max_i d_i^{\mathsf{str}}$ with $d_i^{\mathsf{wk}} = \sum_{j=1}^{NL} [\boldsymbol{\mathcal{A}}_L + \boldsymbol{\mathcal{A}}_C]_{ij}$ and $d_i^{\mathsf{str}} = \sum_{j=1}^{NL} [\boldsymbol{\mathcal{A}}_L \boldsymbol{\mathcal{A}}_C]_{ij}$. Note that the

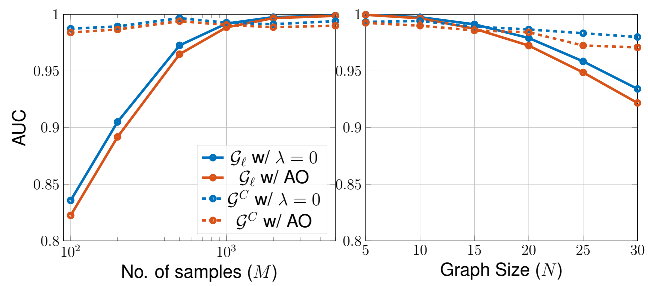

Fig. 2. **Topology Reconstruction under weak coupling** $\mathcal{H}_{\mathsf{wk}}(\cdot)$ AUC performance against (Left) sample size M with $N = 15$; (Right) graph size N with $M = 1000$.

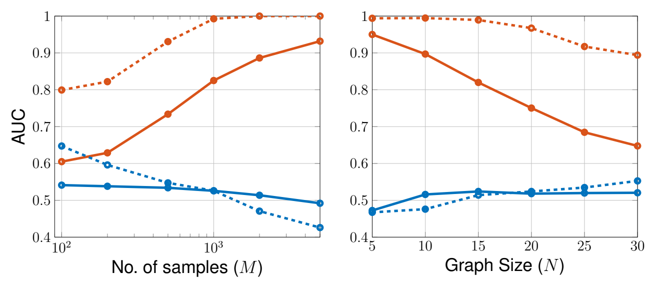

Fig. 3. **Topology Reconstruction under strong coupling** $\mathcal{H}_{\mathsf{str}}(\cdot)$ AUC performance against (Left) sample size M with $N = 15$; (Right) graph size N with $M = 1000$.

case with $\mathcal{H}_{\mathsf{wk}}(\cdot)$, modeled after the supra-diffusion process, refers to a situation when the inter-layer coupling is weak; while $\mathcal{H}_{\mathsf{str}}(\cdot)$, modeled after the multi-dimensional opinion dynamics, refers to the case with strong inter-layer coupling. We measure the estimation quality of the intra-layer graphs $\boldsymbol{\mathcal{A}}_L$ and coupling graph $\boldsymbol{\mathcal{A}}_C$ separately. Lastly, we measure the expected area under ROC (AUC) from 100 Monte-Carlo trials in detecting if an edge exists in the respective graph topologies.

In the first experiment shown in Fig. 2, we consider the case of 'weak coupling' and compare the AUC performance against the sample size (M), the graph size (N). We set $\lambda = 0.1$ for the proposed algorithm. For the comparison against M on the left panel, we fix $N = 15$ and observe that the AUC performance improves in general as M increases. Meanwhile, for the comparison against N on the right panel, we fix $M = 1000$ and observe that the AUC performance deteriorates as N increases. Observe that the proposed AO algorithm attains similar performance as the benchmark with $\lambda = 0$.

In the second experiment shown in Fig. 3, we consider the case of 'strong coupling' and perform a similar comparison for the AUC performance as in Fig. 2. We set $\lambda = 5$ for the proposed algorithm. While similar trends are observed as N, M increases, we notice that the benchmark with $\lambda = 0$ struggles to estimate the intra-layer and inter-layer graph topologies in this scenario. On the other hand, the AO approach continues to deliver reliable estimation over the range of system configurations tested.

Conclusions. This paper focuses on learning multiplex network from multi-attribute graph signals. We introduced a general multiplex graph filter model and devised a high-order smoothness metric for learning the graph topologies with an emphasis on the inter-layer coupling structure. An efficient AO procedure is developed subsequently. Our result shows that in learning multiplex graphs, the distinct dynamics of inter and intra-layer interactions can have an important role, which should be handled properly.

5. REFERENCES

[1] 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.

[2] 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.

[3] M. Newman, *Networks*. Oxford university press, 2018.

[4] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, "Diffusion dynamics on multiplex networks," *Physical review letters*, vol. 110, no. 2, p. 028701, 2013.

[5] Z. Huang, C. Wang, M. Stojmenovic, and A. Nayak, "Characterization of cascading failures in interdependent cyber-physical systems," *IEEE Transactions on Computers*, vol. 64, no. 8, pp. 2158–2168, 2014.

[6] L. Tang, X. Wu, J. Lü, J.-a. Lu, and R. M. D'Souza, "Master stability functions for complete, intralayer, and interlayer synchronization in multiplex networks of coupled rössler oscillators," *Physical Review E*, vol. 99, no. 1, p. 012304, 2019.

[7] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2270–2285, 2016.

[8] N. Pržulj, "Protein-protein interactions: Making sense of networks via graph-theoretic modeling," *Bioessays*, vol. 33, no. 2, pp. 115–123, 2011.

[9] K. R. Finn, M. J. Silk, M. A. Porter, and N. Pinter-Wollman, "The use of multilayer network analysis in animal behaviour," *Animal behaviour*, vol. 149, pp. 7–22, 2019.

[10] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[11] S. Zhang, Q. Deng, and Z. Ding, "Signal processing over multilayer graphs: Theoretical foundations and practical applications," *IEEE Internet of Things Journal*, 2023.

[12] A. Natali, E. Isufi, and G. Leus, "Forecasting multi-dimensional processes over graphs," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5575–5579.

[13] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, "A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 817–829, 2017.

[14] N. Perraudin, A. Loukas, F. Grassi, and P. Vandergheynst, "Towards stationary time-vertex signal processing," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2017, pp. 3914–3918.

[15] J. Jiang, H. Feng, D. B. Tay, and S. Xu, "Theory and design of joint time-vertex nonsubsampled filter banks," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1968–1982, 2021.

[16] S. K. Kadambari and S. P. Chepuri, "Product graph learning from multi-domain data with sparsity and rank constraints," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5665–5680, 2021.

[17] A. Kalaitzis, J. Lafferty, N. D. Lawrence, and S. Zhou, "The bigraphical lasso," in *International Conference on Machine Learning*. PMLR, 2013, pp. 1229–1237.

[18] A. Einizade and S. H. Sardouie, "Learning product graphs from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, 2023.

[19] C. Zhang, Y. He, and H.-T. Wai, "Product graph learning from multi-attribute graph signals with inter-layer coupling," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[20] F. Tudisco, F. Arrigo, and A. Gautier, "Node and layer eigenvector centralities for multiplex networks," *SIAM Journal on Applied Mathematics*, vol. 78, no. 2, pp. 853–876, 2018.

[21] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, "Eigenvector-based centrality measures for temporal networks," *Multiscale Modeling & Simulation*, vol. 15, no. 1, pp. 537–574, 2017.

[22] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov, "Clustering on multi-layer graphs via subspace analysis on grassmann manifolds," *IEEE Transactions on signal processing*, vol. 62, no. 4, pp. 905–918, 2013.

[23] L. Butler, A. Parada-Mayorga, and A. Ribeiro, "Convolutional learning on multigraphs," *IEEE Transactions on Signal Processing*, vol. 71, pp. 933–946, 2023.

[24] 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.

[25] R. Ramakrishna, H.-T. Wai, and A. Scaglione, "A user guide to low-pass graph signal processing and its applications: Tools and applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 74–85, 2020.

[26] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear gauss–seidel method under convex constraints," *Operations research letters*, vol. 26, no. 3, pp. 127–136, 2000.