

LEARNING PRODUCT GRAPHS FROM MULTIDOMAIN SIGNALS

Sai Kiran Kadambari and Sundeep Prabhakar Chepuri

Indian Institute of Science, Bangalore, India

ABSTRACT

In this paper, we focus on learning the underlying product graph structure from multidomain training data. We assume that the product graph is formed from a Cartesian graph product of two smaller factor graphs. We then pose the product graph learning problem as the factor graph Laplacian matrix estimation problem. To estimate the factor graph Laplacian matrices, we assume that the data is smooth with respect to the underlying product graph. When the training data is noise free or complete, learning factor graphs can be formulated as a convex optimization problem, which has an explicit solution based on the water-filling algorithm. The developed framework is illustrated using numerical experiments on synthetic data as well as real data related to air quality monitoring in India.

Index Terms— Graph learning, graph signal processing, Laplacian matrix estimation, product graphs, graph learning, topology inference.

1. INTRODUCTION

Leveraging the underlying structure in data is central to many machine learning and signal processing tasks [1–4]. In many cases, data resides on irregular (non-Euclidean) domains. Some examples include datasets from meteorological stations, traffic networks, social and biological networks, to name a few.

Graphs offer a natural way to describe the structure and explain complex relationships in such network datasets. More specifically, data (signal) is indexed by the nodes of a graph and the edges encode the relationship between the function values at the nodes. When the number of nodes in the graph is very large, signal processing or machine learning operations defined on graphs require more memory and computational resources, e.g., computing the graph Fourier transform [2] using an eigendecomposition requires $O(N^3)$ operations for a graph with N nodes [5]. Whenever the underlying graph can be factorized into two or more factor graphs with fewer nodes, the computational costs of these operations can be reduced significantly [1]. Product graphs can efficiently represent multidomain data. For instance, in brain imaging, the data is multidomain as each spatially distributed sensor gathers temporal data. Similarly, in movie recommender systems (such as Netflix) the rating data matrix has a user dimension as well as a movie dimension. The graph underlying such multidomain data can be factorized so that each graph factor corresponds to one of the domains.

Having a good quality graph is essential for most of the machine learning or signal processing problems over graphs. However, in some applications, the underlying graph may not be readily available, and it has to be estimated from the available training data. Given the training data, the problem of estimating the graph Laplacian or the weighted adjacency matrix, assuming that the graph signals are smooth on the underlying topology has been considered

in [6–11]. Even though the available training data might be multidomain, existing graph learning methods [6–11] ignore the product structure in the graph. For example, when dealing with data collected from spatially distributed sensors over a period of time, existing methods learn a graph that best explains the spatial domain while ignoring the temporal structure. In [9], time-varying graphs are estimated from smooth signals, where the second domain is assumed to be regular.

In this paper, instead of ignoring the structure in any of the domains or treating one of the domains as regular, we propose to learn the underlying graphs related to each domain. This corresponds to learning the factors of the product graph. Specifically, the contributions of this paper are as follows. We propose a framework for estimating the graph Laplacian matrices of the factors of the product graph from the training data. The product graph learning problem can be solved optimally using a *water-filling* approach. Numerical experiments based on synthetic and real data related to air quality monitoring are provided to demonstrate the developed theory. Since the real dataset has many missing entries, we present an alternating minimization algorithm for *joint matrix completion and product graph learning*.

Throughout this paper, we will use upper and lower case bold-face letters to denote matrices and column vectors, respectively. We will denote sets using calligraphic letters. $\mathbf{1}$ ($\mathbf{0}$) denotes the vector/matrix of all ones (zeros) of appropriate dimension. \mathbf{I}_P denotes the identity matrix of dimension P . $\text{diag}[\cdot]$ is a diagonal matrix with its argument along the main diagonal. $\text{vec}(\cdot)$ denotes the matrix vectorization operation. X_{ij} and x_i denote the (i, j) th element and i th element of \mathbf{X} and \mathbf{x} , respectively. \oplus represents the Kronecker sum and \otimes represents the Kronecker product.

2. PRODUCT GRAPH SIGNALS

Consider a graph $\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N)$ with N vertices (or nodes), where \mathcal{V}_N denotes the set of vertices and \mathcal{E}_N denotes the edge set. The structure of the graph with N nodes is captured by the weighted adjacency matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ whose (i, j) th entry denotes the weight of the edge between node i and node j . When there is no edge between node i and node j , the (i, j) th entry of \mathbf{W} is zero. We assume that the graph is undirected with positive edge weights. The corresponding graph Laplacian matrix is a symmetric matrix of size N , given by $\mathbf{L}_N = \text{diag}[\mathbf{d}] - \mathbf{W}$, where $\mathbf{d} \in \mathbb{R}^N$ is a degree vector given by $\mathbf{d} = \mathbf{W}\mathbf{1}$.

Consider two graphs $\mathcal{G}_P = (\mathcal{V}_P, \mathcal{E}_P)$ and $\mathcal{G}_Q = (\mathcal{V}_Q, \mathcal{E}_Q)$ with P and Q nodes, respectively. Let the corresponding graph Laplacian matrices be $\mathbf{L}_P \in \mathbb{R}^{P \times P}$ and $\mathbf{L}_Q \in \mathbb{R}^{Q \times Q}$. Let the Cartesian product of the two graphs \mathcal{G}_P and \mathcal{G}_Q be denoted by \mathcal{G}_N with $|\mathcal{V}_N| = |\mathcal{V}_P||\mathcal{V}_Q|$ nodes, i.e., $PQ = N$. In other words, the graphs \mathcal{G}_P and \mathcal{G}_Q are the factors of \mathcal{G}_N . Then the graph Laplacian matrix \mathbf{L}_N can be expressed in terms of \mathbf{L}_P and \mathbf{L}_Q as

$$\mathbf{L}_N = \mathbf{L}_P \oplus \mathbf{L}_Q = \mathbf{I}_Q \otimes \mathbf{L}_P + \mathbf{L}_Q \otimes \mathbf{I}_P. \quad (1)$$

This work was supported by the Robert Bosch Centre for Cyber Physical Systems, IISc.

Let us collect the set of graph signals $\{\mathbf{x}_i\}_{i=1}^T$ with $\mathbf{x}_i \in \mathbb{R}^N$ defined on the product graph \mathcal{G}_N in an $N \times T$ matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$. As each node in the product graph \mathcal{G}_N is related to a pair of vertices in its graph factors, we can reshape any product graph signal \mathbf{x}_i as $\mathbf{X}_i \in \mathbb{R}^{P \times Q}$, $i = 1, 2, \dots, T$, such that $\mathbf{x}_i = \text{vec}(\mathbf{X}_i)$. This means that each product graph signal represents a multidomain graph signal where the columns and rows of \mathbf{X}_i are graph signals associated to the graph factor \mathcal{G}_P and \mathcal{G}_Q , respectively.

In this work, we will assume that \mathbf{X} is smooth with respect to (w.r.t.) the graph \mathcal{G}_N . The amount of smoothness is quantified by the Laplacian quadratic form $\text{tr}(\mathbf{X}^T \mathbf{L}_N \mathbf{X})$, where small values of $\text{tr}(\mathbf{X}^T \mathbf{L}_N \mathbf{X})$ imply that the data \mathbf{X} is smooth on the graph \mathcal{G}_N .

3. TASK-COGNIZANT PRODUCT GRAPH LEARNING

Suppose we are given the training data $\mathbf{X} \in \mathbb{R}^{N \times T}$ defined on the product graph \mathcal{G}_N , where each column of \mathbf{X} , i.e., $\mathbf{x}_i \in \mathbb{R}^N$, represents a multidomain graph data $\mathbf{X}_i \in \mathbb{R}^{P \times Q}$. Assuming that the given data is smooth on the graph \mathcal{G}_N and the product graph \mathcal{G}_N can be factorized as the Cartesian product of two graphs \mathcal{G}_P and \mathcal{G}_Q as in (1), we are interested in estimating the graph Laplacian matrices \mathbf{L}_P and \mathbf{L}_Q . To do so, we assume that P and Q are known.

Typically, we might not have access to the original graph data \mathbf{X} , but we might observe data related to \mathbf{X} . Let us call this data $\mathbf{Y} \in \mathbb{R}^{N \times T}$. For example, \mathbf{Y} could be a noisy or an incomplete version of \mathbf{X} . Given \mathbf{Y} , the joint estimation of \mathbf{X} , \mathbf{L}_P and \mathbf{L}_Q , may be mathematically formulated as the following optimization problem

$$\begin{aligned} \underset{\mathbf{L}_P \in \mathcal{L}_P, \mathbf{L}_Q \in \mathcal{L}_Q, \mathbf{X}}{\text{minimize}} \quad & f(\mathbf{X}, \mathbf{Y}) + \alpha \text{tr}(\mathbf{X}^T (\mathbf{L}_P \oplus \mathbf{L}_Q) \mathbf{X}) \\ & + \beta_1 \|\mathbf{L}_P\|_F^2 + \beta_2 \|\mathbf{L}_Q\|_F^2, \end{aligned} \quad (2)$$

where $\mathcal{L}_N := \{\mathbf{L} \in \mathbb{R}^{N \times N} | \mathbf{L}\mathbf{1} = \mathbf{0}, \text{tr}(\mathbf{L}) = N, L_{ij} = L_{ji} \leq 0, i \neq j\}$ is the space of all the valid Laplacian matrices of size N . We use the trace equality constraint in this set to avoid a trivial solution. The loss function $f(\mathbf{X}, \mathbf{Y})$ is appropriately chosen depending on the nature of the observed data. More importantly, we learn the product graph suitable for the task of minimizing $f(\mathbf{X}, \mathbf{Y})$. For instance, if the observed graphs signals are noisy as $\mathbf{y}_i = \mathbf{x}_i + \mathbf{n}_i$ for $1 \leq i \leq T$ with the noise vector \mathbf{n}_i , then $f(\mathbf{X}, \mathbf{Y})$ is chosen as $\|\mathbf{X} - \mathbf{Y}\|_F^2$. A smoothness promoting quadratic term is added to the objective function with a positive regularizer α . The squared Frobenius norm of the Laplacian matrices with tuning parameters β_1 and β_2 in the objective function controls the distribution of edge weights. By varying α , β_1 and β_2 , we can control the sparsity (i.e., the number of zeros) of \mathbf{L}_P and \mathbf{L}_Q , while setting $\beta_1 = \beta_2 = 0$ gives the sparsest graph.

4. SOLVER BY ADAPTING EXISTING WORKS

Ignoring the structure of \mathcal{G}_N that it can be factorized as \mathcal{G}_P and \mathcal{G}_Q , and given \mathbf{X} (or its noisy version), one can learn the graph Laplacian matrix \mathbf{L}_N using any one of the methods in [6–9]. In this section, we will develop a solver for the optimization problem (2) based on the existing method in [6].

There exists several numerical approaches for factorizing product graphs, i.e., to find the graph factors \mathbf{L}_P and \mathbf{L}_Q from \mathbf{L}_N [12, 13]. One of the simplest and efficient ways to obtain the graph factors is by solving the convex optimization problem

$$\underset{\mathbf{L}_P \in \mathcal{L}_P, \mathbf{L}_Q \in \mathcal{L}_Q}{\text{minimize}} \quad \|\mathbf{L}_N - \mathbf{L}_P \oplus \mathbf{L}_Q\|_F^2. \quad (3)$$

The above problem can be solved using any one of the off-the-shelf convex optimization toolboxes. However, notice that this is a two-step approach, which requires computing a size- N Laplacian matrix in the first step using [6], for instance. The first step is computationally expensive as the number of nodes in the product graph \mathcal{G}_N increases exponentially with the increase in the number of nodes in either of the factor graphs \mathcal{G}_P and \mathcal{G}_Q . In other words, finding the product graph Laplacian matrix and then factorizing it is computationally much expensive and sometimes infeasible because of the huge memory requirement for storing the data. Therefore, in the next section, we propose a computationally efficient solution.

5. PROPOSED SOLVER

In this section, instead of a two-step approach, as discussed in Section 4, we provide a one-step solution for estimating the Laplacian matrices of the factors of the product graph by leveraging the properties of the Cartesian graph product and the structure of the data.

The quadratic smoothness promoting term $\text{tr}(\mathbf{X}^T (\mathbf{L}_P \oplus \mathbf{L}_Q) \mathbf{X})$ can be written as

$$\text{tr}(\mathbf{X}^T (\mathbf{L}_P \oplus \mathbf{L}_Q) \mathbf{X}) = \sum_{i=1}^T \mathbf{x}_i^T (\mathbf{L}_P \oplus \mathbf{L}_Q) \mathbf{x}_i. \quad (4)$$

Using the property of the Cartesian product that for any given matrices \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{X} of appropriate dimensions, the equation $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}$ is equivalent to $(\mathbf{A} \oplus \mathbf{B}^T) \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{C})$. Using the property that $\text{tr}(\mathbf{A}^T \mathbf{X}) = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{X})$, (4) simplifies to

$$\begin{aligned} \sum_{i=1}^T \mathbf{x}_i^T (\mathbf{L}_P \oplus \mathbf{L}_Q) \mathbf{x}_i &= \sum_{i=1}^T \text{vec}(\mathbf{X}_i)^T \text{vec}(\mathbf{L}_P \mathbf{X}_i + \mathbf{X}_i \mathbf{L}_Q) \\ &= \sum_{i=1}^T \text{tr}(\mathbf{X}_i^T \mathbf{L}_P \mathbf{X}_i) + \text{tr}(\mathbf{X}_i \mathbf{L}_Q \mathbf{X}_i^T). \end{aligned}$$

This means that the amount of smoothness of the multidomain signal \mathbf{x}_i w.r.t. \mathcal{G}_N is equal to the sum of the amount of smoothness of the signals collected in the rows and columns of \mathbf{X}_i w.r.t. \mathcal{G}_P and \mathcal{G}_Q , respectively.

Assuming that the training data is noise free and complete, that is, $\mathbf{Y} = \mathbf{X}$, and \mathbf{X} available, (2) simplifies to

$$\begin{aligned} \underset{\mathbf{L}_P \in \mathcal{L}_P, \mathbf{L}_Q \in \mathcal{L}_Q}{\text{minimize}} \quad & \alpha \sum_{i=1}^T [\text{tr}(\mathbf{X}_i^T \mathbf{L}_P \mathbf{X}_i) + \text{tr}(\mathbf{X}_i \mathbf{L}_Q \mathbf{X}_i^T)] \\ & + \beta_1 \|\mathbf{L}_P\|_F^2 + \beta_2 \|\mathbf{L}_Q\|_F^2 \end{aligned} \quad (5)$$

with variables \mathbf{L}_P and \mathbf{L}_Q . The optimization problem (5) has a unique minimizer as it is convex in \mathbf{L}_P and \mathbf{L}_Q . In fact, it can be expressed as a quadratic program (QP) with fewer variables as compared to (5) by exploiting the fact that \mathbf{L}_P and \mathbf{L}_Q are symmetric matrices. In essence, we need to solve for only the lower or upper triangular elements of \mathbf{L}_P and \mathbf{L}_Q .

Let the vectorized form of the lower triangular elements of \mathbf{L}_P and \mathbf{L}_Q be denoted by $\text{vecl}(\mathbf{L}_P) \in \mathbb{R}^{P(P+1)/2}$ and $\text{vecl}(\mathbf{L}_Q) \in \mathbb{R}^{Q(Q+1)/2}$, respectively. Furthermore, $\text{vecl}(\mathbf{L}_P)$ and $\text{vecl}(\mathbf{L}_Q)$, may be, respectively, related to $\text{vec}(\mathbf{L}_P)$ and $\text{vec}(\mathbf{L}_Q)$ as $\text{vec}(\mathbf{L}_P) = \mathbf{A} \text{vecl}(\mathbf{L}_P)$, $\text{vec}(\mathbf{L}_Q) = \mathbf{B} \text{vecl}(\mathbf{L}_Q)$ using matrices \mathbf{A} and \mathbf{B} of appropriate dimensions. Now, using the fact that $\text{tr}(\mathbf{X}^T \mathbf{L}_P \mathbf{X}) = \text{vec}(\mathbf{X} \mathbf{X}^T)^T \text{vec}(\mathbf{L}_P)$ and $\text{tr}(\mathbf{X} \mathbf{L}_Q \mathbf{X}^T) = \text{vec}(\mathbf{X}^T \mathbf{X})^T \text{vec}(\mathbf{L}_Q)$ and the properties of Frobenius norm, we may rewrite (5) as

$$\underset{\mathbf{z} \in \mathbb{R}^K}{\text{minimize}} \quad \frac{1}{2} \mathbf{z}^T \mathbf{P} \mathbf{z} + \mathbf{q}^T \mathbf{z} \quad \text{subject to} \quad \mathbf{C} \mathbf{z} = \mathbf{d}, \mathbf{z} \geq \mathbf{0}, \quad (6)$$

where $\mathbf{z} = [\text{vec}^T(\mathbf{L}_P), \text{vec}^T(\mathbf{L}_Q)]^T$ is the optimization variable of length $K = 0.5(P^2 + Q^2 + P + Q)$. Here, $\mathbf{P} = \text{diag}[2\beta_1 \mathbf{A}^T \mathbf{A}, 2\beta_2 \mathbf{B}^T \mathbf{B}]$ is the diagonal matrix of size K and $\mathbf{q} = \alpha \sum_{i=1}^T \mathbf{q}_i$ with $\mathbf{q}_i = [\text{vec}(\mathbf{X}_i \mathbf{X}_i^T)^T \mathbf{A}, \text{vec}(\mathbf{X}_i^T \mathbf{X}_i)^T \mathbf{B}]^T \in \mathbb{R}^K$. The matrix $\mathbf{C} \in \mathbb{R}^{(P+Q+2) \times K}$ and the vector $\mathbf{d} \in \mathbb{R}^{P+Q+2}$ are defined appropriately to represent the trace equality constraints (in the constraint sets) in (5). The problem (6) is a QP in its standard form and can be solved optimally using any one of the off-the-shelf solvers such as CVX [14]. To obtain the graph factors \mathbf{L}_P and \mathbf{L}_Q using the solver based on the existing methods as described in Section 4 requires solving a QP with $N(N+1)/2$ variables for \mathbf{L}_N [6], and subsequently, factorizing \mathbf{L}_N as \mathbf{L}_P and \mathbf{L}_Q as in (3), which requires solving one more QP with K variables. In contrast, the proposed method requires solving only one QP with K variables. Thus, the proposed method is computationally efficient as compared to the solver based on the existing methods.

6. WATER-FILLING SOLUTION

In this section, we derive an explicit solution for the optimization problem (6) based on the well-known *water-filling* approach. By writing the KKT conditions and solving for z_i leads to an explicit solution.

The Lagrangian function for (6) is given by

$$\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \mathbf{z}^T \mathbf{P} \mathbf{z} + \mathbf{q}^T \mathbf{z} + \boldsymbol{\mu}^T (\mathbf{d} - \mathbf{C} \mathbf{z}) - \boldsymbol{\lambda}^T \mathbf{z},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^K$ and $\boldsymbol{\mu} \in \mathbb{R}^{P+Q+2}$ are the Lagrange multipliers corresponding to the inequality and equality constraints and recall that \mathbf{P} is a diagonal matrix. Then the KKT conditions are given by

$$\begin{aligned} P_{i,i} z_i^* + q_i - \mathbf{c}_i^T \boldsymbol{\mu}^* - \lambda_i^* &= 0, \\ \sum_{i=1}^N \mathbf{c}_i z_i^* &= \mathbf{d}, \quad z_i^* \geq 0, \quad \lambda_i^* \geq 0, \quad \lambda_i^* z_i^* = 0, \end{aligned}$$

where \mathbf{c}_i is the i th column of \mathbf{C} . We can now solve the KKT conditions to find z_i^* , $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$. To do so, we eliminate the slack variable λ_i^* and solve for z_i^* . This results in

$$z_i^* = \max \left\{ 0, P_{i,i}^{-1} (\mathbf{c}_i^T \boldsymbol{\mu}^* - q_i) \right\}.$$

To find $\boldsymbol{\mu}^*$, we may use z_i^* in the second KKT condition to obtain $\sum_{i=1}^N \mathbf{c}_i \max \{ 0, P_{i,i}^{-1} (\mathbf{c}_i^T \boldsymbol{\mu}^* - q_i) \} = \mathbf{d}$. Since this is a linear function in $\boldsymbol{\mu}_i^*$, we can compute $\boldsymbol{\mu}_i^*$ using a simple iterative method. This approach is similar to the water-filling algorithm with multiple water levels [15].

7. NUMERICAL RESULTS

In this section, we will evaluate the performance of the proposed water-filling based solver (henceforth, referred to as Solver 1) and compare it with the solver based on the existing methods described in Section 4 (henceforth, referred to as Solver 2) on synthetic and real datasets.

7.1. Results on synthetic data

To evaluate the performance of the proposed method, we generate synthetic data on a known graph, which can be factorized. We will use these graph factors as a reference (i.e., ground truth) and compare the estimated graph factors with the ground truth in terms of

Method	\mathbf{L}_P	\mathbf{L}_Q	\mathbf{L}_N
Solver 1	0.9615	0.9841	0.9755
Solver 2	0.7556	0.7842	0.7612

Table 1: F-measure of the proposed solver (Solver 1) and the solver based on the existing methods (Solver 2).

F-measure, which is a measure of the percentage of correctly recovered edges [6]. Specifically, we generate a graph with $N = 150$ nodes, which is formed by the Cartesian product of two community graphs with $P = 10$ and $Q = 15$ nodes, respectively.

We generate $T = 50$ signals $\{\mathbf{x}_i\}_{i=1}^{50}$ that are smooth w.r.t. \mathcal{G}_N by using the factor analysis model described in [6]. Given \mathbf{X} , we learn the graph factors using Solver 1 (using the water-filling method) and Solver 2 (wherein we learn the complete graph Laplacian matrix of $N = 150$ nodes as in [6] and then factorize it by solving (3)).

In Table 1, we show the best (in terms of the tuning parameters) F-measure scores obtained by Solver 1 and Solver 2. The F-measure of Solver 1 is close to 0.98, which means that the proposed method learns the graph factors close to the ground truth. In Fig. 1, we plot the Laplacian matrices of the ground truth, learned Laplacian matrices from Solver 1 and Solver 2. We see that the graph factors estimated from Solver 1 are more consistent with the ground truth than the factors estimated using Solver 2. Moreover, we stress the fact that the proposed water-filling method is computationally efficient as compared to Solver 2.

7.2. Joint matrix completion and learning graph factors

We now test the performance on real data. For this purpose, we use the PM_{2.5} data collected over 40 air quality monitoring stations in different locations in India for each day of the year 2018 [16]. However, there are many missing entries in this dataset. Given this multidomain data that has spatial and temporal dimensions, the aim is to learn the graph factors that best explain the data. More precisely, we aim to learn the graph factors that capture the relationships between spatially distributed sensors and a graph that captures the seasonal variations of the PM_{2.5} data. Since the dataset has missing entries, we impute the missing entries using a graph Laplacian regularized nuclear norm minimization [17]. That is, we use $f(\mathbf{X}, \mathbf{Y}) := \sum_{i=1}^T \|\mathcal{A}(\mathbf{X}_i - \mathbf{Y}_i)\|_F^2 + \gamma \|\mathbf{X}_i\|_*$, where \mathcal{A} denotes the known observation mask that selects the available entries, $\|\cdot\|_*$ denotes the nuclear norm, and $\gamma \geq 0$ is the tuning parameter. As the problem (2) is not convex in $\{\mathbf{X}, \mathbf{L}_P, \mathbf{L}_Q\}$ due to the coupling between the variables in the smoothness promoting terms, we use alternating minimization method. Specifically, in an alternating manner, we solve for $\{\mathbf{L}_P, \mathbf{L}_Q\}$ by fixing \mathbf{X} using the solver described in Section 6, and then solve for \mathbf{X} by fixing $\{\mathbf{L}_P, \mathbf{L}_Q\}$ as

$$\underset{\{\mathbf{X}_i\}_{i=1}^T}{\text{minimize}} \quad f(\mathbf{X}, \mathbf{Y}) + \alpha \sum_{i=1}^T [\text{tr}(\mathbf{X}_i^T \mathbf{L}_P \mathbf{X}_i) + \text{tr}(\mathbf{X}_i \mathbf{L}_Q \mathbf{X}_i^T)],$$

where recall that $\text{vec}(\mathbf{X}_i)$ forms the i th column of \mathbf{X} .

The learned graph factors that represent the multidomain graph signals are shown in Fig. 2. Fig. 2a shows the graph factor that encodes the relationships between the spatially distributed air quality monitoring stations. Each colored dot indicates the concentration of PM_{2.5} on a random day across different locations. We can see from Fig. 2a that the graph connections are not necessarily related to the geographical proximity of the cities, but based on the similarity of the PM_{2.5} values.

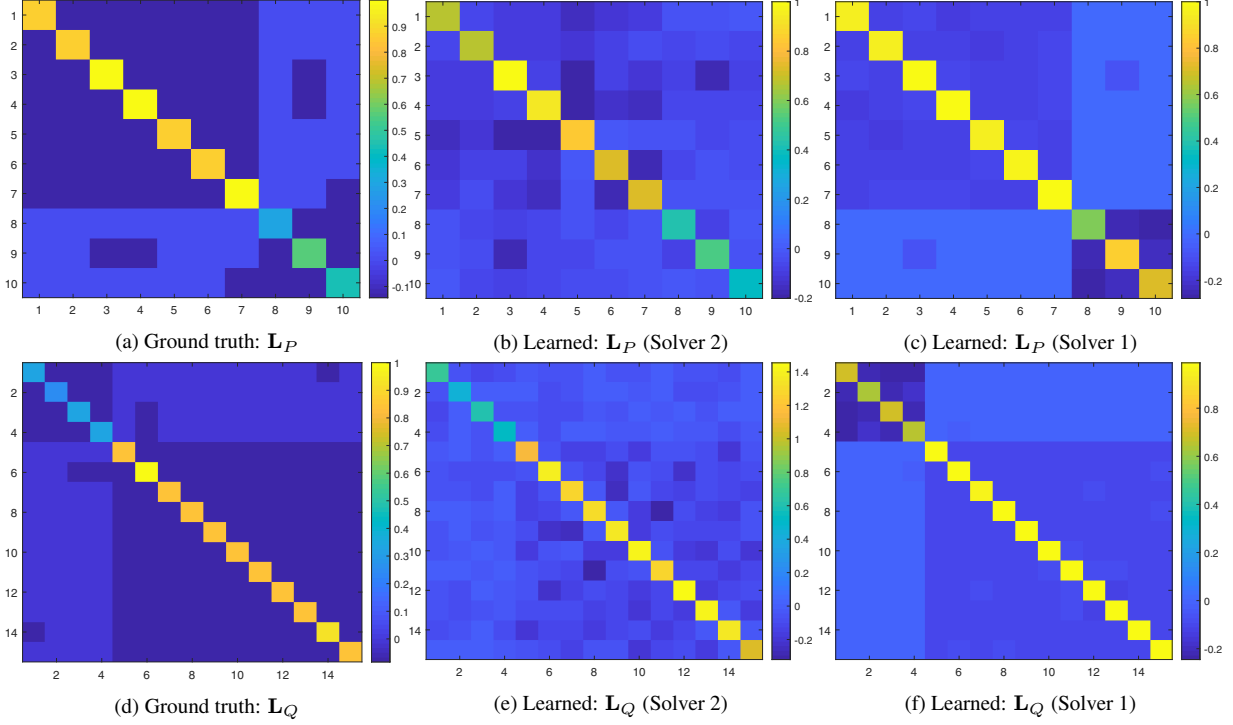


Fig. 1: Product graph learning on synthetic data

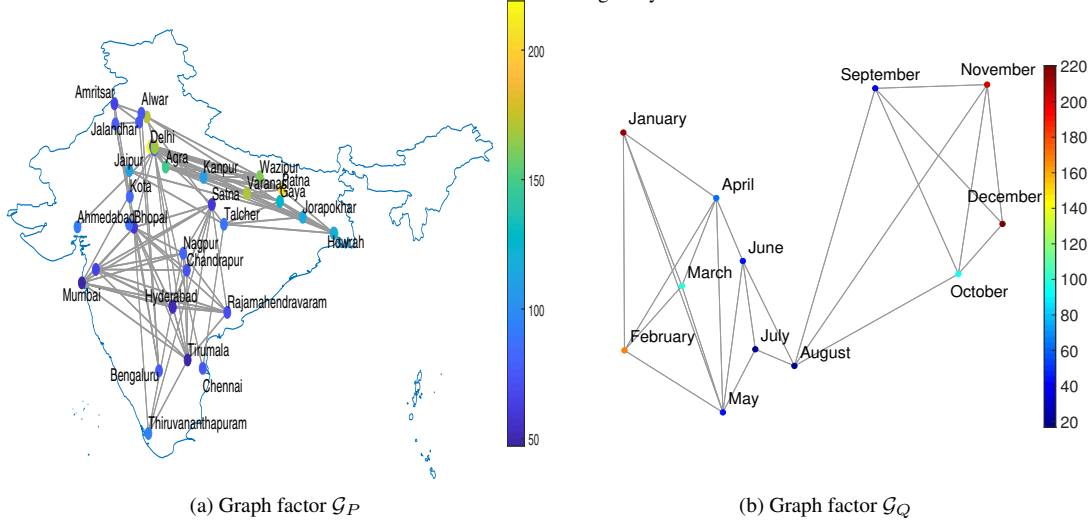


Fig. 2: Factor graph learning on air pollution data: The colored dots indicates the $PM_{2.5}$ values. (a) Graph factor \mathcal{G}_P represents the spatially distributed sensor graph. (b) Graph factor \mathcal{G}_Q is a graph capturing the temporal/seasonal variation of the $PM_{2.5}$ data across months.

The concentration of $PM_{2.5}$ is relatively lesser during the summer/monsoon months (i.e., June, July, and August) as compared to the other months. The graph capturing these temporal variations in the $PM_{2.5}$ data is shown in Fig. 2b. Each color dot indicates the concentration of $PM_{2.5}$ averaged over each month at a random location. We can see in Fig. 2b that months with similar average concentration of $PM_{2.5}$ are connected. Moreover, the months having low $PM_{2.5}$ values (during the monsoon months) are clustered together and the remaining months are clustered into two groups based on their occurrence before and after the monsoon. This shows that the obtained time graph has clusters that capture the seasonal variations hence confirms the quality of the learned graph.

8. CONCLUSIONS

We developed a framework for learning graphs that can be factorized as the Cartesian product of two smaller graphs. The estimation of the Laplacian matrices of the factor graphs is posed as a convex optimization problem. The proposed solver is computationally efficient and has an explicit solution based on the water-filling method. The performance of the proposed method is significantly better than other intuitive ways to solve the problem. We present numerical experiments based on air pollution data collected across different locations in India. Since this dataset has missing entries, we developed a task-cognizant learning method to solve the joint matrix completion and product graph learning problem.

9. REFERENCES

- [1] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.
- [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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [3] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning theory and kernel machines*. Springer, 2003, pp. 144–158.
- [4] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2010.
- [5] M. Puschel and J. M. F. Moura, "Algebraic signal processing theory: Cooley–tukey type algorithms for dct and dst," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1502–1521, April 2008.
- [6] 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, Dec 2016.
- [7] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. of Artificial Intelligence and Statistics (AISTATS)*, 2016, pp. 920–929.
- [8] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero III, "Learning sparse graphs under smoothness prior," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017.
- [9] V. Kalofolias, A. Loukas, D. Thanou, and P. Frossard, "Learning time varying graphs," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, pp. 2826–2830.
- [10] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.
- [11] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.
- [12] R. Hammack, W. Imrich, and S. Klavzar, *Handbook of Product Graphs, Second Edition*, 2nd ed. Boca Raton, FL, USA: CRC Press, Inc., 2011.
- [13] W. Imrich, S. Klavzar, and D. F. Rall, *Topics in Graph Theory: Graphs and Their Cartesian Product*. AK Peters Ltd, 2008.
- [14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [15] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, Feb 2005.
- [16] "Central control room for air quality management, india," <https://app.cpcbcr.com/>.
- [17] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Matrix completion on graphs," *arXiv preprint arXiv:1408.1717*, 2014.