SEMI-SUPERVISED LEARNING OF PROCESSES OVER MULTI-RELATIONAL GRAPHS

Qin Lu *Vassilis N. Ioannidis* *Georgios B. Giannakis*

Dept. of ECE and Digital Technology Center, University of Minnesota, USA

ABSTRACT

Semi-supervised learning (SSL) of dynamic processes over graphs is encountered in several applications of network science. Most of the existing approaches are unable to handle graphs with multiple relations, which arise in various real-world networks. This work deals with SSL of dynamic processes over multi-relational graphs (MRGs). Towards this end, a structured dynamical model is introduced to capture the spatio-temporal nature of dynamic graph processes, and incorporate contributions from multiple relations of the graph in a probabilistic fashion. Given nodal samples over a subset of nodes and the MRG, the expectation-maximization (EM) algorithm is adapted to extrapolate nodal features over unobserved nodes, and infer the contributions from the multiple relations in the MRG simultaneously. Experiments with real data showcase the merits of the proposed approach.

***Index Terms*—** Dynamic graph processes, multi-relational graphs, EM, semi-supervised learning.

1. INTRODUCTION

A number of network science related applications rely on features over all nodes in the graph, which cannot be accommodated with limited nodal observations, due to privacy concerns or sampling costs. The semi-supervised learning (SSL) task of reconstructing nodal features over unobserved nodes can be addressed with the aid of the underlying graph topology that captures nodal inter-dependencies [6, 10]. In Facebook for instance, where nodes and edges represent users and their friendships, one can infer the income of a specific user from her/his friends' income.

Temporal dynamics in the nodal processes further challenge the SSL task, which has been tackled in several works. Relying on the so-termed graph bandlimited model, inference of slow-varying processes over graphs has been pursued in [5, 15]. On the other hand, graph kernel based estimators have been leveraged to reconstruct general dynamic processes [8, 13]. Most recently, semi-supervised tracking of dynamic processes over switching graphs has been handled by an online Bayesian algorithm [11]. However, all these approaches deal with *single-relational* graphs per slot.

This work was supported in part by NSF grants 1508993, 1711471, and 1901134.

In many contemporary applications, multiple types of relations over the same set of nodes coexist, giving rise to a *multi-relational* graph (MRG) [9, 12]. In the particular case of social networks, each relation of the MRG accounts for a specific form of social interaction, such as friendship, family bonds, or coworker ties [16]. Generalizing traditional SSL approaches for single-relational graphs to MRGs is challenging. While SSL of static attributes over MRGs has been studied recently [7], SSL of dynamic nodal processes over MRGs has not been addressed.

Contributions. Alleviating the aforementioned challenges, the present work deals with SSL of dynamic processes over MRGs. Specifically, a dynamical model is constructed to capture the spatio-temporal nature of the dynamic processes per relation. Using this model, the contributions from multiple relations of the MRG to the dynamics in the nodal processes are fused in a probabilistic fashion. A novel EM solver with convergence guarantees is developed to predict the process values over the unobserved nodes.

2. MODELING AND PROBLEM FORMULATION

Consider a network with N nodes comprising the vertex set $\mathcal{V} := \{1, \ldots, N\}$. In several applications, nodes in the network may be connected via multiple relations. For example, Facebook users can be also connected through the Linkedin and Twitter social networks. To account for such multiple relations of nodes in $\mathcal{V}$, an *undirected* MRG is defined as $\mathcal{G} := \{\mathcal{V}, \mathcal{A}\}$, where $\mathcal{A}$ is the adjacency set $\mathcal{A} := \{\mathbf{A}^s, s = 1, \ldots, S\}$, and $\mathbf{A}^s$ is the $N \times N$ *symmetric* adjacency matrix associated with relation s. The Laplacian matrix associated with $\mathbf{A}^s$ is $\mathbf{L}^s := \mathbf{D}^s - \mathbf{A}^s$, where $\mathbf{D}^s = \mathrm{diag}\{\mathbf{A}^s \mathbf{1}_N\}$ with $\mathbf{1}_N$ denoting the $N \times 1$ all-one column.

A dynamic graph process is defined as the mapping $x : \mathcal{V} \times \mathcal{T} \mapsto \mathbb{R}$, where $\mathcal{T} := \{1, 2, ...\}$ denotes the set of time indices representing time-varying nodal features. For instance, $x_t(n)$ may represent the income of person n at year t. The values over all the nodes at slot t are collected in the vector $\mathbf{x}_t := [x_t(1), \ldots, x_t(N)]^\top$, where $^\top$ stands for transposition.

Accounting for partially observed nodal samples, the observation model per slot t is given by

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{e}_t \tag{1}$$

where the time-varying observation matrix $\mathbf{H}_t \in \{0,1\}^{M_t \times N}$ samples M_t ($M_t < N$) nodes at slot t, and $\mathbf{e}_t$ is the observation noise that accounts for uncertainties and is modeled as temporally uncorrelated, and Gaussian distributed with zero mean and covariance matrix $\mathbf{R}_t$. Equivalently, (1) can be rewritten in the form of the data likelihood as $p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t; \mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t)$.

Given the MRG $\mathcal{G}$ and observations over T slots collected in $\mathbf{Y} := [\mathbf{y}_1 \dots \mathbf{y}_T]$, our goal is to estimate $\mathbf{X} := [\mathbf{x}_1 \dots \mathbf{x}_T]$. This SSL task faces two challenges. The first is to account jointly for spatio-temporal dependencies. The second is to judiciously combine the contributions from the S relations of an MRG. Fortunately, the aforementioned two challenges can be addressed by building structured dynamical models for $\mathbf{x}_t$.

2.1. Modeling dynamics of processes over an MRG

To capture the spatio-temporal nature of the dynamic graph processes for relation s, the evolution from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ is modeled by a first-order Markov process with state transition probability density function (pdf) given by

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, s) = \mathcal{N}(\mathbf{x}_t; \mathbf{F}^s\mathbf{x}_{t-1}, \mathbf{K}^s) \tag{2}$$

where the state transition matrix $\mathbf{F}^s := f(\mathbf{A}^s)$ is a known function f of the adjacency matrix of relation s, and the covariance matrix $\mathbf{K}^s := r^\dagger(\mathbf{L}^s)$ is modeled as a graph Laplacian kernel; see also [14]. While the temporal variation is accounted for via the transition term $\mathbf{F}^s\mathbf{x}_{t-1}$, the spatial correlation per slot is modeled by the Laplacian kernel, and thus (2) takes into account the *spatio-temporal* nature of the dynamic graph processes per relation.

Further, to address the second challenge, the dynamical models (2) for the S relations of the MRG are fused in a probabilistic fashion, yielding the overall state transition pdf

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}; \boldsymbol{\pi}_t) = \sum_{s=1}^{S} \pi_{t,s} p(\mathbf{x}_t|\mathbf{x}_{t-1}, s) \tag{3a}$$

$$= \sum_{s=1}^{S} \pi_{t,s} \mathcal{N}(\mathbf{x}_t; \mathbf{F}^s\mathbf{x}_{t-1}, \mathbf{K}^s) \tag{3b}$$

where $\boldsymbol{\pi}_t := [\pi_{t,1} \dots \pi_{t,S}]^\top$ holds the *unknown* probabilities $\pi_{t,s} \in [0,1]$ that weighs the contribution from the sth relation of the MRG at slot t, while satisfying $\sum_{s=1}^{S} \pi_{t,s} = 1$.

3. GRAPH-AWARE BATCH SOLVER

Building on (3) and (1), the maximum a posteriori (MAP) estimate of $\mathbf{X}$ and the maximum likelihood (ML) estimate of $\boldsymbol{\Pi} := [\boldsymbol{\pi}_1 \dots \boldsymbol{\pi}_T]$ given observations $\mathbf{Y}$, are obtained as

$$\left\{\hat{\mathbf{X}}, \hat{\boldsymbol{\Pi}}\right\} = \arg\max_{\mathbf{X},\boldsymbol{\Pi}} \log p(\mathbf{X}|\mathbf{Y}; \boldsymbol{\Pi})$$

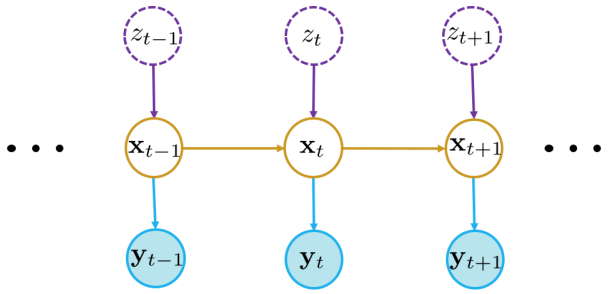

Fig. 1: Probabilistic graphical model for dynamic processes over MRGs.

$$= \arg\max_{\mathbf{X},\boldsymbol{\Pi}} \log p(\mathbf{Y}|\mathbf{X}) + \log p(\mathbf{X}; \boldsymbol{\Pi}) \tag{4}$$

where the conditional independence of the variables in (1) and (3) across slots allows factorization of the batch data likelihood and prior as

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{y}_t; \mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t) \tag{5}$$

$$p(\mathbf{X}; \boldsymbol{\Pi}) = \prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1}; \boldsymbol{\pi}_t)\ . \tag{6}$$

Although the non-convex objective in (4) is solvable using the gradient-ascent based algorithm, its convergence to local optima entails judicious selection of step size per iteration. Also, first-order approaches do not fully leverage the potential benefits arising from a data generative model. These considerations prompt one to employ the celebrated expectation-maximization (EM) algorithm [4], which is an iterative optimization technique to find the ML [4] or MAP [17] solutions for probabilistic models with latent variables [3]. Next, an EM-based solver is developed to *jointly* seek the MAP and ML estimates in (4) with provable convergence at least to a local maximum.

3.1. EM for joint MAP and ML estimation

Aiming to adapt the EM as a solver of (4), we start by rewriting the Gaussian mixture prior (3) as

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}; \boldsymbol{\pi}_t) = \sum_{s=1}^{S} \Pr(z_t = s) p(\mathbf{x}_t|\mathbf{x}_{t-1}, z_t = s) \tag{7}$$

where the latent variable $z_t \in \{1, \dots, S\}$ that is assumed independent across slots, has probability mass function (pmf) $\boldsymbol{\pi}_t$. The probabilistic graphical model for random variables $\{\mathbf{x}_t, \mathbf{y}_t, z_t\}$ based on (1) and (7), is shown in Fig. 1.

After introducing the hidden variables $\mathbf{z} := [z_1 \dots z_T]$, it can be verified for any pdf $q(\mathbf{z})$ that

$$\log p(\mathbf{X}|\mathbf{Y};\boldsymbol{\Pi}) = \mathrm{KL}(q(\mathbf{z}) \| p(\mathbf{z}|\mathbf{X},\mathbf{Y};\boldsymbol{\Pi})) + \mathcal{L}(q,\mathbf{X},\boldsymbol{\Pi}) \tag{8}$$

where

$$\mathcal{L}(q,\mathbf{X},\mathbf{\Pi}) := \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{X},\mathbf{Y},\mathbf{z};\mathbf{\Pi})] - \log p(\mathbf{Y}) + H(q(\mathbf{z})). \tag{9}$$

In (8) and (9), $\mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z})) := \int q(\mathbf{z}) \log \frac{q(\mathbf{z})}{p(\mathbf{z})} d\mathbf{z} \geq 0$ is the Kullback-Leibler divergence of the pdfs $p(\mathbf{z})$ and $q(\mathbf{z})$, and $H(q(\mathbf{z})) := -\int q(\mathbf{z}) \log q(\mathbf{z}) d\mathbf{z}$ is the entropy of $q(\mathbf{z})$. It is evident that (9), which is a functional of $q(\mathbf{z})$ and a function of $\mathbf{X}$ and $\mathbf{\Pi}$, amounts to a lower bound for $\log p(\mathbf{X}|\mathbf{Y};\mathbf{\Pi})$. As we shall see later, the EM algorithm maximizes the lower bound (9) iteratively by alternating between E- and M-steps per iteration.

Let $\hat{\mathbf{X}}^{(n)}$ and $\hat{\mathbf{\Pi}}^{(n)}$ denote the estimates of $\mathbf{X}$ and $\mathbf{\Pi}$ at the end of iteration n. The E-Step at iteration $n+1$ maximizes (9) wrt $q(\mathbf{z})$ with $\hat{\mathbf{X}}^{(n)}$ and $\hat{\mathbf{\Pi}}^{(n)}$ fixed, yielding $q^*(\mathbf{z}) = p(\mathbf{z}|\hat{\mathbf{X}}^{(n)},\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)})$ and $\mathcal{L}(q^*,\hat{\mathbf{X}}^{(n)},\hat{\mathbf{\Pi}}^{(n)}) = \log p(\hat{\mathbf{X}}^{(n)}|\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)})$. Subsequently, with $q^*(\mathbf{z})$ fixed, the M-step maximizes $\mathcal{L}(q^*,\mathbf{X},\mathbf{\Pi})$ wrt $\mathbf{X}$ and $\mathbf{\Pi}$, which boils down to maximizing the following auxiliary function

$$Q(\mathbf{X},\mathbf{\Pi};\hat{\mathbf{X}}^{(n)},\hat{\mathbf{\Pi}}^{(n)}) := \mathbb{E}_{q^*(\mathbf{z})}[\log p(\mathbf{X},\mathbf{z},\mathbf{Y};\mathbf{\Pi})] \ . \tag{10}$$

Thus, the updates of $\mathbf{X}$ and $\mathbf{\Pi}$ at iteration $n+1$ are given by

$$\left\{\hat{\mathbf{X}}^{(n+1)},\hat{\mathbf{\Pi}}^{(n+1)}\right\} = \arg\max_{\mathbf{X},\mathbf{\Pi}} Q(\mathbf{X},\mathbf{\Pi};\hat{\mathbf{X}}^{(n)},\hat{\mathbf{\Pi}}^{(n)}) \ . \tag{11}$$

Since $\mathcal{L}(q^*,\hat{\mathbf{X}}^{(n+1)},\hat{\mathbf{\Pi}}^{(n+1)}) \geq \mathcal{L}(q^*,\hat{\mathbf{X}}^{(n)},\hat{\mathbf{\Pi}}^{(n)})$ and $\mathrm{KL}(q^*(\mathbf{z})\|p(\mathbf{z}|\hat{\mathbf{X}}^{(n+1)},\mathbf{Y};\hat{\mathbf{\Pi}}^{(n+1)})) \geq 0$, it holds that

$$\log p(\hat{\mathbf{X}}^{(n+1)}|\mathbf{Y};\hat{\mathbf{\Pi}}^{(n+1)}) \geq \log p(\hat{\mathbf{X}}^{(n)}|\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)}) \tag{12}$$

thereby proving convergence of the EM iterates at least to one of the local maxima of $\log p(\mathbf{X}|\mathbf{Y};\mathbf{\Pi})$.

Next, we will specialize the solution of (4) using the EM algorithm outlined so far.

3.2. EM for SSL of dynamic processes over MRGs

Our graph-aware batch solver of (4) entails alternating between the following E- and M-steps per iteration.

E-Step at iteration $n+1$ calls for evaluating $q^*(\mathbf{z}) = p(\mathbf{z}|\hat{\mathbf{X}}^{(n)},\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)})$, and further the Q-function (10) with $\hat{\mathbf{X}}^{(n)}$ and $\hat{\mathbf{\Pi}}^{(n)}$ available from iteration n. Since z_t is independent across slots (cf. Fig. 1), we have $q^*(\mathbf{z}) = \prod_{t=1}^T q^*(z_t)$, where $q^*(z_t) = p(z_t|\hat{\mathbf{X}}^{(n)},\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)})$. Invoking Bayes' rule, it follows that

$$\begin{aligned} w_{t,s}^{(n)} &:= \Pr(z_t = s|\hat{\mathbf{X}}^{(n)},\mathbf{Y};\hat{\mathbf{\Pi}}^{(n)}) \\ &= \frac{\Pr(z_t = s;\hat{\mathbf{\Pi}}^{(n)}) p(\hat{\mathbf{X}}^{(n)}|z_t = s) p(\mathbf{Y}|\hat{\mathbf{X}}^{(n)})}{p(\hat{\mathbf{X}}^{(n)}) p(\mathbf{Y}|\hat{\mathbf{X}}^{(n)})} \\ &= \frac{\hat{\pi}_{t,s}^{(n)} \mathcal{N}(\hat{\mathbf{x}}_t^{(n)};\mathbf{F}^s\hat{\mathbf{x}}_{t-1}^{(n)},\mathbf{K}^s)}{\sum_{s'=1}^S \hat{\pi}_{t,s'}^{(n)} \mathcal{N}(\hat{\mathbf{x}}_t^{(n)};\mathbf{F}^{s'}\hat{\mathbf{x}}_{t-1}^{(n)},\mathbf{K}^{s'})} \ . \end{aligned} \tag{13}$$

Algorithm 1 Find $\tilde{\mathbf{F}}_t$ and $\tilde{\mathbf{K}}_t$ based on (18)

Input: $\mathbf{F}^s$, $\mathbf{K}^s$, $w_{t,s}, t = 1,\ldots,T, s = 1,\ldots,S$

Step 1: Obtain $\mathbf{C}_t$, $\mathbf{D}_t$, $t = 1,\ldots,T$
$\mathbf{D}_T = \sum_{s=1}^S w_{T,s}(\mathbf{K}^s)^{-1}$
$\mathbf{C}_T = -\sum_{s=1}^S w_{T,s}(\mathbf{K}^s)^{-1}\mathbf{F}^s$
for $t = 1$ to $T-1$ **do**
  $\mathbf{D}_t = \sum_{s=1}^S w_{t,s}(\mathbf{K}^s)^{-1} + w_{t+1,s}\mathbf{F}^{s\top}(\mathbf{K}^s)^{-1}\mathbf{F}^s$
  $\mathbf{C}_t = -\sum_{s=1}^S w_{t,s}(\mathbf{K}^s)^{-1}\mathbf{F}^s$
end for

Step 2: Obtain $\tilde{\mathbf{F}}_t, \tilde{\mathbf{K}}_t, t = 1,\ldots,T$
$\tilde{\mathbf{K}}_T^{-1} = \mathbf{D}_T$
for $t = T$ to 2 **do**
  $\tilde{\mathbf{F}}_t = -\tilde{\mathbf{K}}_t\mathbf{C}_t$
  $\tilde{\mathbf{K}}_{t-1}^{-1} = \mathbf{D}_{t-1} - \tilde{\mathbf{F}}_t^\top\tilde{\mathbf{K}}_t^{-1}\tilde{\mathbf{F}}_t$
end for
$\tilde{\mathbf{F}}_1 = -\tilde{\mathbf{K}}_1\mathbf{C}_1$

Output: $\tilde{\mathbf{F}}_t, \tilde{\mathbf{K}}_t, t = 1,\ldots,T$

Next, based on the graphical model in Fig. 1, the joint pdf $p(\mathbf{X},\mathbf{Y},\mathbf{z};\mathbf{\Pi})$ in the Q-function (10) is given by

$$\begin{aligned} p(\mathbf{X},\mathbf{Y},\mathbf{z};\mathbf{\Pi}) &= p(\mathbf{X},\mathbf{z};\mathbf{\Pi})p(\mathbf{Y}|\mathbf{X}) \\ &= \prod_{t=1}^T p(z_t)p(\mathbf{x}_t|\mathbf{x}_{t-1},z_t)p(\mathbf{y}_t|\mathbf{x}_t) \\ &= \prod_{t=1}^T \left(\prod_{s=1}^S \pi_{t,s}^{I(z_t=s)} \mathcal{N}(\mathbf{x}_t;\mathbf{F}^s\mathbf{x}_{t-1},\mathbf{K}^s)^{I(z_t=s)}\right) \\ &\quad \times \mathcal{N}(\mathbf{y}_t;\mathbf{H}_t\mathbf{x}_t,\mathbf{R}_t) \end{aligned} \tag{14}$$

where the indicator function $I(z_t = s)$ is

$$I(z_t = s) = \begin{cases} 1 & z_t = s \\ 0 & z_t \neq s \end{cases} \ . \tag{15}$$

Thus, the Q-function (10) can be expressed as

$$\begin{aligned} Q\left(\mathbf{X},\mathbf{\Pi};\hat{\mathbf{X}}^{(n)},\hat{\mathbf{\Pi}}^{(n)}\right) &= \mathbb{E}_{q^*(\mathbf{z})}\left[\log\left(p\left(\mathbf{X},\mathbf{z},\mathbf{Y};\mathbf{\Pi}\right)\right)\right] \\ &= \sum_{t=1}^T \left(\sum_{s=1}^S w_{t,s}^{(n)}\left(\log\pi_{t,s} + \frac{\|\mathbf{x}_t - \mathbf{F}^s\mathbf{x}_{t-1}\|_{\mathbf{K}^s}^2}{2}\right)\right. \\ &\quad \left. + \frac{\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|_{\mathbf{R}_t}^2}{2}\right) + C \end{aligned} \tag{16}$$

where the operator $\|\mathbf{x}\|_{\mathbf{A}}^2 := \mathbf{x}^\top\mathbf{A}^{-1}\mathbf{x}$ with $^{-1}$ representing matrix inversion, $w_{t,s}^{(n)} = \mathbb{E}_{q^*(\mathbf{z})}[I(z_t = s)]$, and C is a constant unrelated with $\mathbf{X}$ and $\mathbf{\Pi}$.

M-Step yields $\hat{\mathbf{\Pi}}^{(n+1)}$ and $\hat{\mathbf{X}}^{(n+1)}$ updates by maximizing the Q-function (16) wrt $\mathbf{\Pi}$ and $\mathbf{X}$ as in (11). Taking the

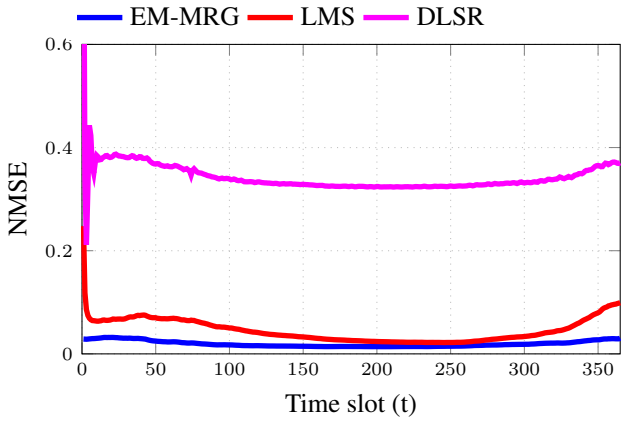

Fig. 2: NMSEs over unobserved nodes .

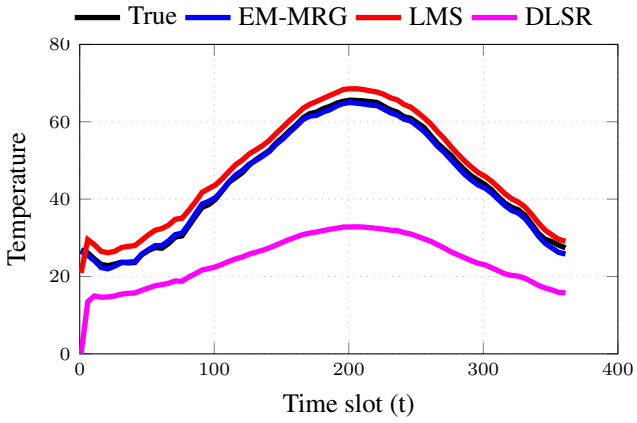

Fig. 3: True and estimated temperature values over an unobserved location.

derivative of (16) wrt $\pi_{t,s}$, $t = 1, \ldots, T$, $s = 1, \ldots, S$, yields

$$\hat{\pi}_{t,s}^{(n+1)} = w_{t,s}^{(n)} \ . \tag{17}$$

As for the update of $\mathbf{X}$, upon reorganizing

$$\sum_{t=1}^{T}\sum_{s=1}^{S} w_{t,s}\|\mathbf{x}_t - \mathbf{F}^s\mathbf{x}_{t-1}\|_{\mathbf{K}^s}^2 = \sum_{t=1}^{T}\|\mathbf{x}_t - \tilde{\mathbf{F}}_t\mathbf{x}_{t-1}\|_{\tilde{\mathbf{K}}_t}^2 \tag{18}$$

where the steps to obtain $\tilde{\mathbf{F}}_t$ and $\tilde{\mathbf{K}}_t$ are summarized in Alg. 1. The cost function related to $\mathbf{X}$ is then given by

$$Q_{\mathbf{X}} = \sum_{t=1}^{T}\left(\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|_{\mathbf{R}_t}^2 + \|\mathbf{x}_t - \tilde{\mathbf{F}}_t\mathbf{x}_{t-1}\|_{\tilde{\mathbf{K}}_t}^2\right) \ . \tag{19}$$

Clearly, the optimal $\mathbf{X}$ that maximizes (19) can be obtained by the well-known Kalman smoother, whose implementation details can be found in [2] for instance.

The E- and M-steps proceed repeatedly until at least one of the convergence conditions is met: either the maximum number of iterations is reached, or, the difference of the loss between successive iterations drops below a preselected threshold.

4. EXPERIMENTAL RESULTS

In this section, the performance of our novel approach that we henceforth abbreviate as "EM-MRG," is tested on the temperature dataset [1]. The latter records hourly temperature measurements at $N = 109$ measuring stations across the continental United States in 2010. Relying on the geographical distances of the measuring stations, we constructed an MRG with two relations following the approach in [13], where 7- and 11-nearest neighbors are considered respectively. In the dynamical model (2) associated with relation s, $\mathbf{F}^s = 0.05(\mathbf{I}_N + \mathbf{A}^s)$, and $\mathbf{K}^s$ was set to be a diffusion kernel [14]. The observation matrix was constant across slots and the number of observed nodes was 44. The performance metric is the normalized mean-square error (NMSE) over unobserved nodes, which is given by $\text{NMSE}(t) := \|\mathbf{H}_t^{\text{c}}\left(\hat{\mathbf{x}}_{t|t} - \mathbf{x}_t\right)\|_2^2 / \|\mathbf{H}_t^{\text{c}}\mathbf{x}_t\|_2^2$, where $\mathbf{H}_t^{\text{c}}$ is the sampling matrix for the unobserved nodes. To allevaite the effect of sampling, the NMSE was averaged over 100 random sampling realizations.

EM-MRG is compared with two existing algorithms, namely the adaptive least mean-square (LMS) algorithm [5], and the distributed least-squares reconstruction (DLSR) scheme in [15]. Since the competing alternatives are unable to handle MRGs in a principled manner, their performance were assessed using the average of the nodal feature estimates associated with the two relations in the MRG.

The average NMSEs over time slots for the competing approaches are depicted in Fig. 2. It is evident that EM-MRG outperforms the competing alternatives in terms of extrapolating unobserved nodal processes, which is also corroborated by Fig. 3, where the reconstructed temperature values over an unobserved node is shown. Thus, the proposed EM-MRG approach can make the best of the relational information for an MRG to reconstruct missing nodal features.

5. CONCLUSIONS

This contribution dealt with SSL of dynamic processes over MRGs. It put forth a structured dynamical model that first accounts for the spatio-temporal dynamics of processes per relation, and further combines the contributions from multiple relations of an MRG. Given partially observed nodal samples and relational information from the MRG, the EM algorithm was leveraged to reconstruct graphs signals and also output the weight that measures the contribution of each relation. Numerical tests validated the performance of the novel approach.

6. REFERENCES

[1] "1981-2010 U.S. climate normals," https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data, [Online; accessed 29-July-2019].

[2] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. John Wiley & Sons, 2004.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

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

[5] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean-squares estimation of graph signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 555–568, Dec. 2016.

[6] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 787–807, May 2018.

[7] V. N. Ioannidis, A. G. Marques, and G. B. Giannakis, "A recurrent graph neural network for multi-relational data," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, 2019, pp. 8157–8161.

[8] V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Inference of spatio-temporal functions over graphs via multikernel kriged Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 66, no. 12, pp. 3228–3239, Jun. 2018.

[9] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.

[10] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.

[11] Q. Lu, V. Ioannidis, and G. B. Giannakis, "Semi-supervised tracking of dynamic processes over switching graphs," in *Proc. of IEEE Data Science Workshop*, Minneapolis, MN, Jun. 2019.

[12] M. Nickel, V. Tresp, and H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in *Proc. Intl. Conf. Mach. Learn.*, vol. 11, 2011, pp. 809–816.

[13] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 856–869, Sep. 2017.

[14] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, 2017.

[15] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 728–740, Jun. 2015.

[16] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.

[17] P. Willett, Y. Ruan, and R. Streit, "PMHT: Problems and some solutions," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 738–754, 2002.