

LEARNING GRAPH PROCESSES WITH MULTIPLE DYNAMICAL MODELS

Qin Lu[†], Vassilis N. Ioannidis[†], Georgios B. Giannakis[†] and Mario Coutino[‡]

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

[‡]Dept. of EE, Mathematics and CS, Delft University of Technology, The Netherlands

ABSTRACT

Network-science related applications frequently deal with inference of spatio-temporal processes. Such inference tasks can be aided by a graph whose topology contributes to the underlying spatio-temporal dependencies. Contemporary approaches extrapolate dynamic processes relying on a fixed dynamical model, that is not adaptive to changes in the dynamics. Alleviating this limitation, the present work adopts a candidate set of graph-adaptive dynamical models with one active at any given time. Given partially observed nodal samples, a scalable Bayesian tracker is leveraged to infer the graph processes and learn the active dynamical model simultaneously in a data-driven fashion. The resulting algorithm is termed graph-adaptive interacting multiple dynamical models (Grad-IMDM). Numerical tests with synthetic and real data corroborate that the proposed Grad-IMDM is capable of tracking the graph processes and adapting to the dynamical model that best fits the data.

Index Terms— Spatiotemporal process, multiple dynamical models, Bayesian tracker

1. INTRODUCTION

Given limited data at a subset of nodes, various applications deal with inference of processes across all network nodes, which can be tackled thanks to the underlying graph topology that captures nodal inter-dependencies [6, 8]. While attempts have been made towards this semi-supervised learning (SSL) task, existing works either assume no dynamics or fixed dynamic pattern in the nodal process, rendering them incapable of handling scenarios when dynamics change over time; see, e.g., [4, 7, 14]. For example, stock prices, related via correlation graphs, undergo different evolution patterns in periods of economic depression and prosperity, and thus a universal dynamical model is not appropriate for tracking financial data.

Past works. Existing approaches to reconstructing *time-invariant* graph processes often rely on the smoothness [9, 16], ‘graph bandlimitedness’ [15], sparsity and overcomplete dictionaries [5], most of which can be unified under the framework of learning using graph kernels; see e.g., [14]. Towards

SSL of *time-varying* processes over graphs, the property of graph bandlimitness allows for extrapolation of nodal processes in a dynamical *model-free* fashion; see, [4, 17]. Graph kernel-based methods that leverage a *single* structured dynamical model have also been developed in [7, 11]. In spite of the aforementioned works, inference of nodal processes that undergo changes in the dynamics has not been addressed so far. Outside graph-based applications, changing dynamical patterns also exist in the motions of maneuvering objects, such as aircrafts and drones. Tracking these objects has been well pursued relying on multiple dynamical models in the field of target tracking; see, e.g., [2, 3].

Contributions. This paper considers a candidate set of dynamical models to accommodate the changing dynamics of graph processes, when the active one per slot is *unknown*. Each dynamical model is characterized by a first-order Markovian process that encodes in the graph topology both temporal variations across slots as well as spatial correlations per slot. Leveraging partially observed nodal samples, this paper aims at jointly tracking the graph processes and selecting the active model via the so-termed graph-adaptive interacting multiple dynamical models (Grad-IMDM). Numerical tests on synthetic and real data show the capability of our Grad-IMDM in reconstructing the graph processes while adapting to the best model on-the-fly.

2. PROBLEM FORMULATION

Consider a graph with N nodes collected in the vertex set $\mathcal{V} := \{1, \dots, N\}$. The connectivity pattern between nodes at slot t is captured by the $N \times N$ adjacency matrix $\mathbf{A}_t$, whose (n, n') th entry is the nonnegative weight of the edge connecting node n with n' . The graph is considered undirected with no self-loops, that is, $\mathbf{A}_t^\top = \mathbf{A}_t$ and $\mathbf{A}_t(n, n) = 0$, where $^\top$ stands for transposition. The Laplacian matrix is then $\mathbf{L}_t := \text{diag}(\mathbf{A}_t \mathbf{1}_N) - \mathbf{A}_t$, in which $\mathbf{1}_N$ is the $N \times 1$ all-one vector. With $\mathcal{T} := \{1, 2, \dots\}$ denoting the set of slot indices, a dynamic graph process, which represents the time-varying nodal feature, is the mapping $x : \mathcal{V} \times \mathcal{T} \mapsto \mathbb{R}$. Specifically, $x_t(n)$ may represent the price of stock n at day t in the aforementioned stock network. The values over all nodes at time t are collected in $\mathbf{x}_t := [x_t(1) \dots x_t(N)]^\top$.

In several privacy-concerned or sampling-cost-limited ap-

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

plications, only a subset of the nodal samples is observed, yielding the following observation model

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{e}_t \quad (1)$$

where $\mathbf{H} \in \{0, 1\}^{M \times N}$ is the $M \times N$ ($M < N$) sampling matrix, whose rows sum to 1, and $\mathbf{e}_t$ is zero-mean, temporally independent, Gaussian noise with covariance matrix $\mathbf{R}$.

Given the graph topologies $\{\mathbf{A}_t\}_{t=1}^T$ and the observations $\mathbf{Z}_T := [\mathbf{z}_1, \dots, \mathbf{z}_T]$ over T slots, our objective is to extrapolate all the nodal features contained in $\mathbf{X}_T := [\mathbf{x}_1, \dots, \mathbf{x}_T]$. It may be impossible to handle this SSL task when $\mathbf{x}_t$ is nonstationary without building structured dynamical models that account for the spatiotemporal nature of the process. Attempts that rely on a *single* dynamical model per graph have been made towards this end; see, e.g., [7, 11]. However, these contemporary approaches fall short when the dynamics change over time, such as the evolution patterns of stock prices in periods of economic depression and prosperity.

To address this limitation, we account for possible spatiotemporal dynamics by constructing multiple topology-dependent dynamical models. Specifically, the evolution from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ is modeled as the first-order vector process

$$\mathbf{x}_t = f^{l_t}(\mathbf{A}_t) \mathbf{x}_{t-1} + \boldsymbol{\eta}_t^{k_t} \quad (2)$$

where the transition function f^{l_t} , taking as input the graph topology $\mathbf{A}_t$, is chosen from a given set $\{f^1, \dots, f^L\}$ that captures candidate temporal dynamics; and the spatial correlation is accounted for through the zero-mean Gaussian noise $\boldsymbol{\eta}_t^{k_t}$ via its covariance $(r^{k_t}(\mathbf{L}_t))^\dagger$ († for pseudo-inverse), which is a Laplacian kernel [9]. Here the so-termed energy mapping r^{k_t} is also selected from a known set $\{r^1, \dots, r^K\}$, which can promote desirable properties, such as diffusion, smoothness, or bandlimitedness.

Considering all possible combinations for f^{l_t} and r^{k_t} yields $S = LK$ dynamical models per slot, with the *unknown* active one indicated by (l_t, k_t) , which can be mapped into $\sigma_t \in \{1, \dots, S\}$. Hence, the active dynamical model at slot t is given by the pair $(\mathbf{F}_t^{\sigma_t}, \mathbf{K}_t^{\sigma_t})$, where $\mathbf{F}_t^{\sigma_t} = f^{l_t}(\mathbf{A}_t)$ and $\mathbf{K}_t^{\sigma_t} = (r^{k_t}(\mathbf{L}_t))^\dagger$. Given $\mathbf{Z}_T$ and candidate dynamical models $\{\{\mathbf{F}_t^s, \mathbf{H}_t^s\}_{s=1}^S\}_{t=1}^T$, our goal is adapted to estimate $\mathbf{X}_T$ and the active model indices (or modes) $\{\sigma_t\}_{t=1}^T$ jointly. Cast in the Bayesian framework, the objective function is chosen to be the batch posterior probability density function (pdf) $p(\mathbf{X}_T|\mathbf{Z}_T) \propto p(\mathbf{Z}_T|\mathbf{X}_T)p(\mathbf{X}_T)$. Temporal independence in (1) and (2) allows for the factorizations of the batch likelihood and prior as

$$p(\mathbf{Z}_T|\mathbf{X}_T) = \prod_{t=1}^T p(\mathbf{z}_t|\mathbf{x}_t) = \prod_{t=1}^T \mathcal{N}(\mathbf{z}_t; \mathbf{H}\mathbf{x}_t, \mathbf{R})$$

$$p(\mathbf{X}_T) = \prod_{t=1}^T p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \prod_{t=1}^T \left(\sum_{s=1}^S w_t^s \mathcal{N}(\mathbf{x}_t; \mathbf{F}_t^s \mathbf{x}_{t-1}, \mathbf{K}_t^s) \right)$$

where $w_t^s \in \{0, 1\}$, $w_t^s = 1$ if $\sigma_t = s$ and $\sum_{s=1}^S w_t^s = 1$.

Thus, taking the logarithm of $p(\mathbf{X}_T|\mathbf{Z}_T)$ yields the following optimization problem

$$\arg \min_{\substack{\{\mathbf{x}_t\}_{t=1}^T \\ \{w_t^s\}_{s=1}^S}} \frac{1}{2} \sum_{t=1}^T [\|\mathbf{z}_t - \mathbf{H}\mathbf{x}_t\|_{\mathbf{R}}^2 + \sum_{s=1}^S w_t^s \|\mathbf{x}_t - \mathbf{F}_t^s \mathbf{x}_{t-1}\|_{\mathbf{K}_t^s}^2]$$

$$\text{s.to } w_t^s \in \{0, 1\}, \quad \sum_{s=1}^S w_t^s = 1 \quad (4)$$

which is identified as a mixed-integer program with prohibitive computational complexity. Next, we will develop a scalable solver that is also able to handle streaming observations.

3. GRAPH-ADAPTIVE BAYESIAN TRACKER

Aiming at a computationally efficient online solver of (4), we will adapt the interacting multi-model algorithm [3], which is a Bayesian tracker for maneuvering targets with applications in target tracking [12] and air traffic control [10], but without graph-related information. Relying on topology-dependent dynamical models (2), our resultant approach is naturally termed graph-adaptive interacting multiple dynamical model (Grad-IMDM). Grad-IMDM starts by replacing the hard constraint $w_t^s \in \{0, 1\}$ with the soft one $w_t^s \in [0, 1]$, which allows one to interpret it as the posterior probability mass function (pmf) of mode s being active at slot t given $\mathbf{Z}_t$, namely $w_t^s = \Pr(\sigma_t = s|\mathbf{Z}_t)$. Furthermore, the evolving mode σ_t is modeled with a first-order Markov chain parameterized by the $S \times S$ mode transition matrix $\boldsymbol{\Pi}$, whose (s, s') th entry $\pi_{ss'} = \Pr(\sigma_t = s|\sigma_{t-1} = s')$ denotes the transition probability from mode s' at slot $t-1$ to mode s at slot t .

Grad-IMDM then leverages the current observation $\mathbf{z}_t$ to propagate the posterior marginal state pdf $p(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1})$ to $p(\mathbf{x}_t|\mathbf{Z}_t)$, which, by invoking Bayes' rule and the total probability theorem (TPT), is given by

$$p(\mathbf{x}_t|\mathbf{Z}_t) = \sum_{s=1}^S w_t^s p(\mathbf{x}_t|\sigma_t = s, \mathbf{Z}_t). \quad (5)$$

To effect tractability, Grad-IMDM makes the Gaussian approximation $p(\mathbf{x}_t|\sigma_t = s, \mathbf{Z}_t) \approx \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}^s, \boldsymbol{\Sigma}_{t|t}^s)$, where $\hat{\mathbf{x}}_{t|t}^s$ and $\boldsymbol{\Sigma}_{t|t}^s$ are the mode-conditioned posterior mean and covariance. Hence, $p(\mathbf{x}_t|\mathbf{Z}_t)$ in (5) is approximated by a Gaussian mixture (GM), that is characterized by the parameter set $\mathcal{P}_t := \{w_t^s, \hat{\mathbf{x}}_{t|t}^s, \boldsymbol{\Sigma}_{t|t}^s, s = 1, \dots, S\}$. Consequently, the propagation of posterior pdf (5) from slot $t-1$ to t amounts to the update of $\mathcal{P}_{t-1}$ to $\mathcal{P}_t$, which are implemented via the prediction and correction steps described next.

Prediction. Given $\mathcal{P}_{t-1}$, Grad-IMDM leverages the mode and state evolution models to make predictions about the

mode-conditioned state pdf and the mode pmf, respectively.

1) *Predictive mode-conditioned state pdf.* Grad-IMDM predicts the state pdf conditioning on σ_t as

$$p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_{t-1}) = \sum_{s=1}^S \Pr(\sigma_{t-1} = s | \sigma_t = s', \mathbf{Z}_{t-1}) p(\mathbf{x}_t | \sigma_t = s', \sigma_{t-1} = s, \mathbf{Z}_{t-1}) \quad (6)$$

where the first factor $\Pr(\sigma_{t-1} = s | \sigma_t = s', \mathbf{Z}_{t-1}) := w_{t-1|t}^{s|s'}$ can be viewed as a backward mode transition probability, while the second factor is the predicted state pdf conditioned on mode s' at slot t and mode s at slot $t-1$. Upon appealing to Bayes' rule and the TPT, the first factor boils down to

$$\begin{aligned} w_{t-1|t}^{s|s'} &= \frac{\Pr(\sigma_{t-1} = s | \mathbf{Z}_{t-1}) \Pr(\sigma_t = s' | \sigma_{t-1} = s, \mathbf{Z}_{t-1})}{\sum_{s=1}^S \Pr(\sigma_{t-1} = s | \mathbf{Z}_{t-1}) \Pr(\sigma_t = s' | \sigma_{t-1} = s, \mathbf{Z}_{t-1})} \\ &= \frac{w_{t-1}^s \pi_{s's}}{\sum_{s=1}^S w_{t-1}^s \pi_{s's}}. \end{aligned} \quad (7)$$

As for the second factor in (6), state equation (2) implies that

$$p(\mathbf{x}_t | \sigma_t = s', \sigma_{t-1} = s, \mathbf{Z}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}^{s',s}, \Sigma_{t|t-1}^{s',s})$$

where the predictive mean and the covariance are respectively

$$\hat{\mathbf{x}}_{t|t-1}^{s',s} = \mathbf{F}_t^{s'} \hat{\mathbf{x}}_{t-1|t-1}^s \quad (8a)$$

$$\Sigma_{t|t-1}^{s',s} = \mathbf{F}_t^{s'} \Sigma_{t-1|t-1}^s (\mathbf{F}_t^{s'})^\top + \mathbf{K}_t^{s'}. \quad (8b)$$

Although (7) and (8) yield the predicted GM pdf $p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_{t-1})$, evolving it to its posterior in (6) is challenging, simply because a GM pdf is a non-Gaussian pdf. Towards obtaining a tractable mode-conditioned Gaussian posterior, we will approximate (6) by the following Gaussian pdf

$$p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_{t-1}) \approx \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}^{s'}, \Sigma_{t|t-1}^{s'}) \quad (9)$$

where $\hat{\mathbf{x}}_{t|t-1}^{s'}$ and $\Sigma_{t|t-1}^{s'}$ are chosen to match the first two moments of (6) as

$$\hat{\mathbf{x}}_{t|t-1}^{s'} = \sum_{s=1}^S w_{t-1|t}^{s|s'} \hat{\mathbf{x}}_{t|t-1}^{s',s} \quad (10a)$$

$$\begin{aligned} \Sigma_{t|t-1}^{s'} &= \sum_{s=1}^S w_{t-1|t}^{s|s'} \left(\Sigma_{t|t-1}^{s',s} + (\hat{\mathbf{x}}_{t|t-1}^{s',s} - \hat{\mathbf{x}}_{t|t-1}^{s'}) (\hat{\mathbf{x}}_{t|t-1}^{s',s} - \hat{\mathbf{x}}_{t|t-1}^{s'})^\top \right). \end{aligned} \quad (10b)$$

2) *Predictive mode pmf.* Given the Markov transition model, Bayes' rule and the TPT yield the predictive mode pmf

$$w_{t|t-1}^{s'} := \Pr(\sigma_t = s' | \mathbf{Z}_{t-1}) = \sum_{s=1}^S \Pr(\sigma_t = s', \sigma_{t-1} = s | \mathbf{Z}_{t-1})$$

$$\begin{aligned} &= \sum_{s=1}^S \Pr(\sigma_t = s' | \sigma_{t-1} = s, \mathbf{Z}_{t-1}) \Pr(\sigma_{t-1} = s | \mathbf{Z}_{t-1}) \\ &= \sum_{s=1}^S \pi_{s's} w_{t-1}^s. \end{aligned} \quad (11)$$

Correction. Grad-IMDM propagates the predictive pdf (9) and pmf (11) to their posterior counterparts upon receiving $\mathbf{z}_t$.

1) *Posterior mode-conditioned state pdf.* First, Bayes' rule updates the predictive state pdf (10) as

$$\begin{aligned} p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_t) &= p(\mathbf{x}_t | \sigma_t = s', \mathbf{z}_t, \mathbf{Z}_{t-1}) \\ &= \frac{p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_{t-1}) p(\mathbf{z}_t | \mathbf{x}_t, \sigma_t = s', \mathbf{Z}_{t-1})}{p(\mathbf{z}_t | \sigma_t = s', \mathbf{Z}_{t-1})} \end{aligned} \quad (12)$$

where $p(\mathbf{z}_t | \mathbf{x}_t, \sigma_t = s', \mathbf{Z}_{t-1}) = p(\mathbf{z}_t | \mathbf{x}_t)$ by independence. Since $p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_{t-1})$ and $p(\mathbf{z}_t | \mathbf{x}_t)$ are Gaussian, $p(\mathbf{x}_t | \sigma_t = s', \mathbf{Z}_t)$ will also be Gaussian with the first two moments in (13d) and (13e) given by Kalman updates [2]

$$\hat{\mathbf{z}}_{t|t-1}^{s'} = \mathbf{H} \hat{\mathbf{x}}_{t|t-1}^{s'} \quad (13a)$$

$$\Phi_t^{s'} = \mathbf{H} \Sigma_{t|t-1}^{s'} \mathbf{H}^\top + \mathbf{R} \quad (13b)$$

$$\mathbf{G}_t^{s'} = \Sigma_{t|t-1}^{s'} \mathbf{H}^\top (\Phi_t^{s'})^{-1} \quad (13c)$$

$$\hat{\mathbf{x}}_{t|t}^{s'} = \hat{\mathbf{x}}_{t|t-1}^{s'} + \mathbf{G}_t^{s'} (\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}^{s'}) \quad (13d)$$

$$\Sigma_{t|t}^{s'} = \Sigma_{t|t-1}^{s'} - \mathbf{G}_t^{s'} \Phi_t^{s'} (\mathbf{G}_t^{s'})^\top. \quad (13e)$$

2) *Posterior mode pmf.* Then, the mode probabilities are updated as

$$\begin{aligned} w_t^{s'} &= \Pr(\sigma_t = s' | \mathbf{z}_t, \mathbf{Z}_{t-1}) \\ &= \frac{p(\mathbf{z}_t | \sigma_t = s', \mathbf{Z}_{t-1}) w_{t|t-1}^{s'}}{\sum_{s=1}^S p(\mathbf{z}_t | \sigma_t = s, \mathbf{Z}_{t-1}) w_{t|t-1}^s} \end{aligned} \quad (14)$$

where $w_{t|t-1}^{s'}$ is given by (11) and $p(\mathbf{z}_t | \sigma_t = s', \mathbf{Z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_{t|t-1}^{s'}, \Phi_t^{s'})$ from (13a) and (13b), thus rendering (14) computable.

Finally, the marginal posterior state pdf is

$$p(\mathbf{x}_t | \mathbf{Z}_t) = \sum_{s'=1}^S w_t^{s'} \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}^{s'}, \Sigma_{t|t}^{s'}) \approx \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \Sigma_{t|t})$$

where the single Gaussian approximant of GM has moments

$$\hat{\mathbf{x}}_{t|t} = \sum_{s'=1}^S w_t^{s'} \hat{\mathbf{x}}_{t|t}^{s'} \quad (15a)$$

$$\Sigma_{t|t} = \sum_{s'=1}^S w_t^{s'} \left(\Sigma_{t|t}^{s'} + (\hat{\mathbf{x}}_{t|t}^{s'} - \hat{\mathbf{x}}_{t|t}) (\hat{\mathbf{x}}_{t|t}^{s'} - \hat{\mathbf{x}}_{t|t})^\top \right). \quad (15b)$$

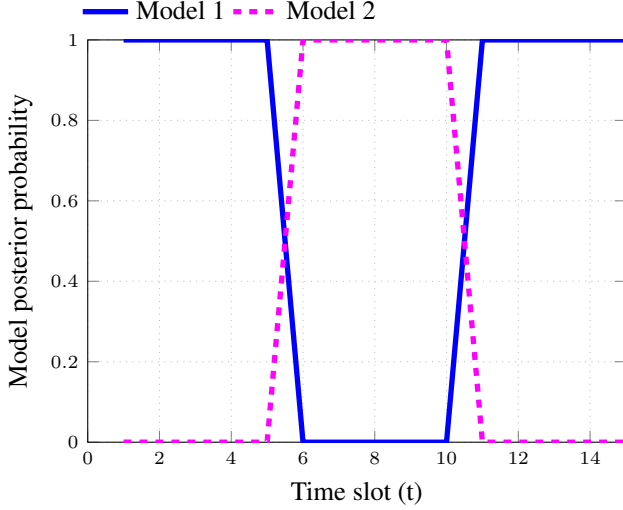


Fig. 1: Posterior model pmfs of Grad-IMDM for synthetic data.

The desired nodal feature estimate is given by (15a), whose covariance matrix (15b) provides additional uncertainty quantification.

4. EXPERIMENTAL RESULTS

The performance of the proposed Grad-IMDM approach is assessed on synthetic and real data. The performance metric is the normalized mean-square error (NMSE) over unobserved nodes as $\text{NMSE}(t) := \|\mathbf{H}^c(\hat{\mathbf{x}}_{t|t} - \mathbf{x}_t)\|_2^2 / \|\mathbf{H}^c \mathbf{x}_t\|_2^2$, where $\mathbf{H}^c$ is the sampling matrix for the unobserved nodes.

1) *Synthetic Data.* A dynamic process is generated over a graph having $N = 60$ nodes and adjacency matrix $\mathbf{A}$ given by an undirected Erdős-Rényi random graph with edge existence probability 0.2. Process $\mathbf{x}_t$ is generated according to (2) with $f^{l_t}(\mathbf{A}) = c^{l_t}(\mathbf{A} + \mathbf{I}_N)$, where $c^{l_t} = 1$ for $t \in [1, 5] \cup [11, 15]$ and $c^{l_t} = 1.5$ for $t \in [6, 10]$. Process noise covariance $(\mathbf{r}^{k_t}(\mathbf{L}_t))^{\dagger}$ is a diffusion kernel (see in [7, Table I]) with parameter 2. The observations adhere to (1) with $M = 30$, $\mathbf{H} = [\mathbf{I}_M, \mathbf{0}_{M, N-M}]$ and $\mathbf{R} = 4\mathbf{I}_M$. To evaluate the average performance, 100 Monte-Carlo runs are conducted with independent realizations of the process and observation noises. As shown in Fig. 1 which reports the posterior model pmfs, the proposed Grad-IMDM selects the right dynamical model adaptively on-the-fly. This observation is further validated by Fig. 2, where Grad-IMDM demonstrates comparable tracking performance to the model-clairvoyant oracle approach — the kernel Kalman filter (KKF) [13].

2) *Real Data.* This dataset records measurements of delays in $N = 70$ paths, each of which connects two of 9 end-nodes via a subset of 26 directed links on the Internet2backbone [1]. A symmetric graph adjacency matrix is constructed as in [7] with the weight reflecting the similarity of links be-

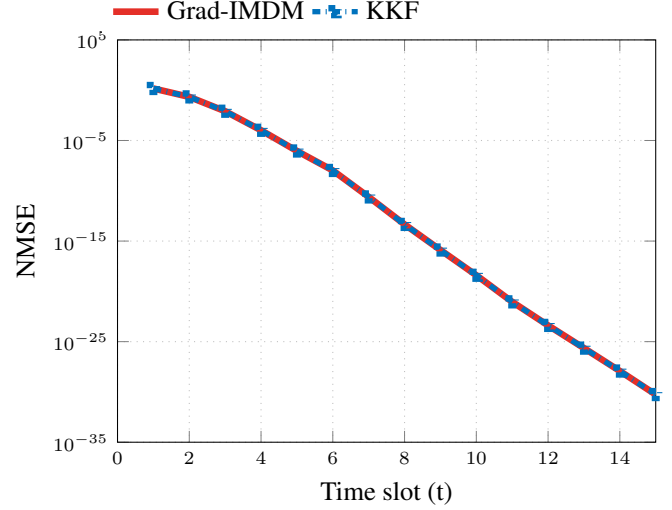


Fig. 2: NMSE for synthetic data.

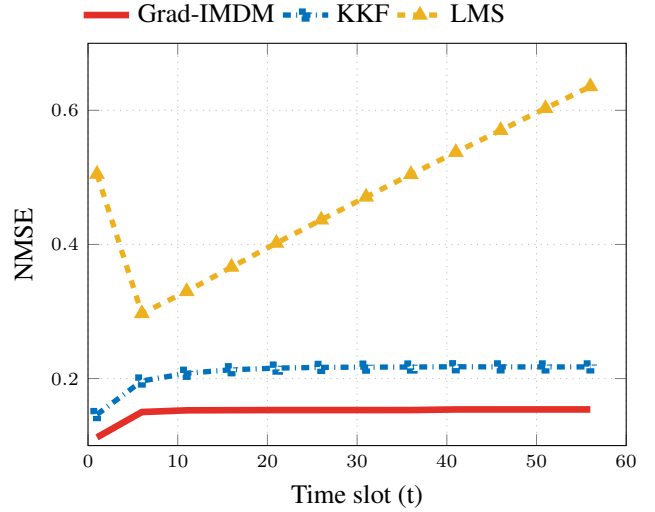


Fig. 3: NMSE for network delay data.

tween two paths. The state transition matrix is selected to be $\mathbf{F}_t^{l_t} = c^{l_t}(\mathbf{A}^{\sigma_t} + \mathbf{I}_N)$, where the unknown scalar c^{l_t} takes value from 0.15 to 0.2 uniformly spaced by 0.01. Process noise covariance $\mathbf{K}_t^{k_t}$ is chosen from a set of diffusion kernels with parameter σ^{k_t} taking value from 1 to 2 with uniform spacing equal to 0.2. The observation matrix is constant per t with $M = 20$ sampled nodes. The tracking performance is averaged over 100 realizations of the sampling schemes. The competing approaches include the single-model KKF with $c^{l_t} = 0.2$ and $\sigma^{k_t} = 1.8$, and the adaptive least mean-square (LMS) algorithm [4], which tracks slow-varying B -bandlimited graph signals. Figs. 3 and 4 showcase the superior tracking performance of our novel Grad-IMDM compared with the rest two competing alternatives.

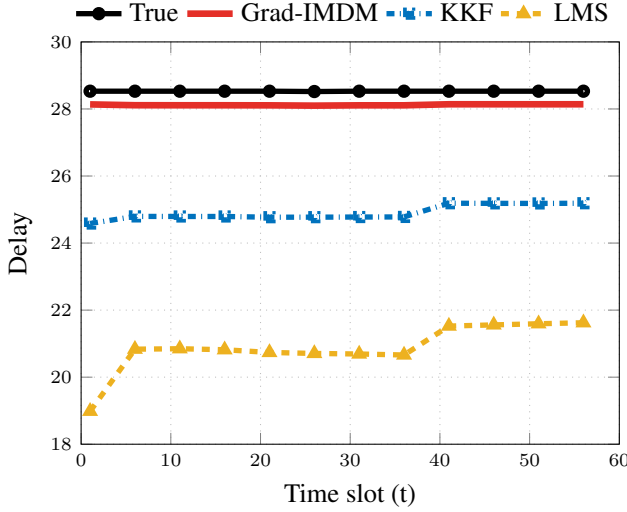


Fig. 4: True and estimated network delays over an unobserved path.

5. CONCLUSIONS

This contribution dealt with SSL of dynamic graph processes using multiple topology-dependent dynamical models, each of which accounts for a certain spatiotemporal property. With sequential arrival of observations over a subset of nodes, a graph-adaptive Bayesian tracker, termed Grad-IMDM, was developed to reconstruct the unobserved nodal features and select the fitted dynamical model jointly. Numerical tests on synthetic and real data corroborated the outstanding performance of the novel Grad-IMDM approach.

6. REFERENCES

- [1] "One-way ping internet2," <http://software.internet2.edu/owamp/>, [Online; accessed 29-April-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] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [4] 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, 2016.
- [5] P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *IEEE Transactions on Signal Processing*, vol. 62, no. 13, pp. 3305–3320, 2014.
- [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, 2018.
- [7] 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, 2018.
- [8] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.
- [9] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. Intl. Conf. on Machine Learning*, 2002, pp. 315–322.
- [10] X.-R. Li and Y. Bar-Shalom, "Design of an interacting multiple model algorithm for air traffic control tracking," *IEEE Transactions on Control Systems Technology*, vol. 1, no. 3, pp. 186–194, 1993.
- [11] Q. Lu, V. Ioannidis, and G. B. Giannakis, "Semi-supervised tracking of dynamic processes over switching graphs," Minneapolis, MN, 2019.
- [12] E. Mazar, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, 1998.
- [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, 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] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4363–4378, 2016.
- [16] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*. Springer, 2003, pp. 144–158.
- [17] 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, 2015.