SEMI-BLIND INFERENCE OF TOPOLOGIES AND SIGNALS OVER GRAPHS

Vassilis N. Ioannidis, Yanning Shen, and Georgios B. Giannakis

ECE Dept. and Digital Tech. Center, Univ. of Minnesota, Mpls, MN 55455, USA
E-mails: {ioann006, shenx513, georgios}@umn.edu

ABSTRACT

Network science provides valuable insights across numerous disciplines including sociology, biology, neuroscience and engineering. A task of major practical importance in these application domains is inferring the network topology from noisy observations over a limited subset of nodes. This work presents a novel approach for *joint* inference of the network topology and estimation of graph signals from partial nodal observations based on structural equation models (SEMs). SEMs have well-documented merits in identifying the directed topology of complex graphs by capturing causal relationships among nodes. The resultant algorithm iterates between inferring a directed graph that "best" fits the data, and estimating the graph signals over the learned graph. Numerical tests with synthetic as well as real data corroborate the effectiveness of the joint inference approach.

Index Terms— Graph signal reconstruction, topology inference, directed graphs, structural equation models.

1. INTRODUCTION

Modeling vertex attributes (or features) as a signal (or a function) that takes values over a graph, provides valuable prior information for data processing tasks, such as filtering, denoising, inference, and compressing, to leverage information captured by the network topology that is presumed available [16, 26]. However, if the network topology is unavailable or inaccurate, performance of the associated data processing task can degrade severely.

Topology identification is possible when observations at all nodes can be collected, by unveiling the hidden (possibly causal) connectivities across nodes. However, in many real settings one can afford to collect nodal observations only from a subset of nodes due to application-specific restrictions, e.g. sampling massive graphs may be prohibitive; and in social networks individuals may be reluctant to share personal information due to privacy concerns. In this context, the present paper addresses the challenging task of *jointly* inferring the network topology and estimating graph signals, given noisy observations at just a subset of nodes. The novel approach will rely on linear structural equation models (SEMs) [14].

SEMs provide a statistical modeling framework for inference of causal relationships among nodes [14]. Linear SEMs have been widely adopted in fields as diverse as sociometrics [9], psychometrics [19], and genetics [4]. Recently, dynamic and nonlinear SEMs have been also developed for tracking dynamic topologies and modeling nonlinear interactions [2, 24, 25]. Other approaches identify undirected topologies under the assumption that the graph signals are smooth [6] or that the observations are generated by a diffusion process [23, 28]. In all these contemporary approaches, it is assumed that samples of the graph process are available for all nodes. However, acquiring network-wide observations may incur prohibitive sampling costs, especially when dealing with massive networks.

Methods for inference (or reconstruction) of graph signals, typically assume that the network topology is *known* and *undirected*. Parametric approaches also adopt the graph-bandlimited model, which postulates that the signal of interest lies in a graph-related B-dimensional subspace [1, 18, 20]. Nonparametric techniques employ kernels on graphs to estimate the graph signals [7, 21, 27], while multi-kernel learning bestows data-driven kernel selection [11, 13]. On the other hand, semi-parametric methods incorporate known signal structure to kernel-based learning without sacrificing flexibility of the overall model [12]. As expected, performance of the aforementioned techniques deteriorates when the assumed topology is not available or inaccurate.

This paper develops a novel approach for joint inference of even directed network topologies and functions over the underlying graph. The algorithm relies on SEMs, and it is *semi-blind* because it only requires observations on just a subset of nodes without requiring knowledge of the topology.

2. MODELING AND PROBLEM FORMULATION

Consider a network with N nodes whose topology is modeled by the graph $\mathcal{G} := (\mathcal{V}, \mathbf{A})$, where $\mathcal{V} := \{v_1, \ldots, v_N\}$ is the set of vertices and $\mathbf{A}$ denotes the $N \times N$ adjacency matrix, whose (n, n')-th entry $A_{n,n'}$ represents the weight of the directed edge from $v_{n'}$ to v_n. A real-valued function (or signal) on $\mathcal{G}$ is a map $f_l : \mathcal{V} \to \mathbb{R}$. In social networks (e.g.,

The work in this paper was supported by NSF grant 1500713, and NIH 1R01GM104975-01.

Twitter) over which information diffuses, f_{ln} could represent the timestamp when subscriber n tweeted about a viral story l. The linear SEM [9] postulates that f_{ln} depends linearly on $\{f_{ln'}\}_{n' \neq n}$, that amounts to $f_{ln} = \sum_{n \neq n'} A_{n,n'} f_{ln'} + \eta_{ln}$, where the unknown $A_{n,n'}$ captures the causal influence of node $v_{n'}$ upon node v_n, and η_{ln} accounts for unmodeled dynamics. The SEM suggests that f_{ln} is influenced directly by its neighboring nodes in the set $\mathcal{N}_n := \{v_{n'} : A_{n,n'} \neq 0\}$. Since real-world networks often exhibit edge sparsity, $\mathbf{A}$ has only a few nonzero entries. With the $N \times 1$ vectors $\mathbf{f}_l := [f_{l1}, \ldots, f_{lN}]^\top$, and $\boldsymbol{\eta}_l := [\eta_{l1}, \ldots, \eta_{lN}]^\top$, the linear SEM can be written in matrix-vector form as

$$\mathbf{f}_l = \mathbf{A}\mathbf{f}_l + \boldsymbol{\eta}_l. \tag{1}$$

Suppose that M_l noisy observations $y_{lm} = f_{ln_m} + \epsilon_{lm}$, $m = 1, \ldots, M_l$ are available, where $\mathcal{M}_l := \{n_1, \ldots, n_{M_l}\}$ contains the indices $1 \leq n_1 \leq \ldots \leq n_{M_l} \leq N$ of the sampled vertices, and ϵ_{lm} models the observation error. With $\mathbf{y}_l := [y_{l1}, \ldots, y_{lM_l}]^\top$, and $\boldsymbol{\epsilon}_l := [\epsilon_{l1}, \ldots, \epsilon_{lM_l}]^\top$, the observation model is

$$\mathbf{y}_l = \mathbf{M}_l \mathbf{f}_l + \boldsymbol{\epsilon}_l \tag{2}$$

where $\mathbf{M}_l$ is an $M_l \times N$ matrix with entries $\{(m, n_m)\}_{m=1}^{M_l}$, set to one, and the rest set to zero.

Problem statement. The goal of this paper is to *jointly* infer the nodal outputs $\{\mathbf{f}_l\}_{l=1}^L$, and the unknown directed adjacency matrix $\mathbf{A}$ given the observations $\{\mathbf{y}_l\}_{l=1}^L$ collected in accordance to the sampling matrices $\{\mathbf{M}_l\}_{l=1}^L$. As estimating $\mathbf{A}$ and $\{\mathbf{f}_l\}_{l=1}^L$ relies on partial observations this is a semi-blind inference task.

Remark 1. Recognizing the limitations of linear SEMs for modeling *nonlinear* interactions, nonlinear SEMs have emerged recently; see e.g., [10]. Inspired by these works, nonlinear variants of the present approach can also be developed to broaden the scope of our models, but are omitted due to space limitations.

Remark 2. Despite the wide applicability of a single-layer static network that is considered in the present work, *multi-layer* networks are also of great interest since they allow multiple types of relationships among nodes, called layers [15]. Moreover, real-world networks may vary over time, which can not be captured by static SEMs. This motivates our future work to accommodate multi-layer as well as dynamic SEMs.

3. JOINTLY INFERRING TOPOLOGY AND SIGNALS

Given $\{\mathbf{y}_l\}_{l=1}^L$ in (2), this section develops a novel approach to infer $\mathbf{A}$, and $\{\mathbf{f}_l\}_{l=1}^L$. To this end, the following regularized least-squares (LS) problem is considered

$$\min_{\mathbf{A} \in \mathcal{A}, \{\mathbf{f}_l\}_{l=1}^L} \mathcal{F}(\mathbf{A}, \{\mathbf{f}_l\}_{l=1}^L) := \sum_{l=1}^L \|\mathbf{f}_l - \mathbf{A}\mathbf{f}_l\|_2^2 + \sum_{l=1}^L \frac{\mu}{M_l} \|\mathbf{y}_l - \mathbf{M}_l \mathbf{f}_l\|_2^2 + \lambda_1 \|\mathbf{A}\|_1 + \lambda_2 \|\mathbf{A}\|_F^2 \tag{3}$$

where $\mu \geq 0$ tunes the relative importance of the fitting term; $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ control the effect of the ℓ_1-norm and the Frobenius-norm, respectively, and $\mathcal{A} := \{\mathbf{A} : \mathbf{A} \in \mathbb{R}^{N \times N}, \{A_{n,n} = 0\}_{n=1}^N\}$. SEMs provide the flexibility to model positive as well as negative influences, nonetheless prior knowledge as positive edges can be easily included in (3). The weighted sum of $\|\cdot\|_1$ and $\|\cdot\|_F$ is the so-termed elastic net penalty and promotes connections between highly correlated nodal measurements. The elastic net targets the "sweet spot" between the ℓ_1 regularizer that effects sparsity, and the $\|\cdot\|_F$ regularizer, which advocates fully connected networks.

The bilinear product $\mathbf{A}\mathbf{f}_l$ renders (3) nonconvex. The next subsection develops a computationally efficient block coordinate descent (BCD) solver of (3).

3.1. Joint BCD algorithm

Even though (3) is nonconvex in both $\mathbf{A}$ and $\mathbf{f}_l$, it is convex with respect to (w.r.t.) each variable separately. This motivates an iterative BCD algorithm that alternates between estimating $\{\mathbf{f}_l\}_{l=1}^L$ and $\mathbf{A}$. Given $\hat{\mathbf{A}}(i)$ at iteration i, the estimates $\{\hat{\mathbf{f}}_l(i)\}_{l=1}^L$ are found by solving the quadratic problem

$$\min_{\{\mathbf{f}_l\}_{l=1}^L} \sum_{l=1}^L \|\mathbf{f}_l - \hat{\mathbf{A}}(i)\mathbf{f}_l\|_2^2 + \sum_{l=1}^L \frac{\mu}{M_l} \|\mathbf{y}_l - \mathbf{M}_l \mathbf{f}_l\|_2^2 \tag{4}$$

where the regularization terms in (3) do not appear. Clearly, (4) conveniently decouples across l, and thus reduces to

$$\min_{\mathbf{f}_l} \; g_i(\mathbf{f}_l) := \tfrac{M_l}{\mu} \|\big[\mathbf{I}_N - \hat{\mathbf{A}}(i)\big]\mathbf{f}_l\|_2^2 + \|\mathbf{y}_l - \mathbf{M}_l \mathbf{f}_l\|_2^2. \tag{5}$$

The first quadratic in (5) can be written as $\|\big[\mathbf{I}_N - \hat{\mathbf{A}}(i)\big]\mathbf{f}_l\|_2^2 = \sum_{n=1}^N (f_{ln} - \sum_{n' \in \mathcal{N}_n} \hat{A}_{n,n'} f_{ln'})^2$, and it can be viewed as a regularizer for $\mathbf{f}_l$, promoting graph functions with similar values at neighboring nodes. Notice that (5) may not be strongly convex, since $\mathbf{I}_N - \hat{\mathbf{A}}(i)$ could be rank deficient. Nonetheless, since $g_i(\cdot)$ is smooth (5) can be readily solved via gradient descent (GD) iterations

$$\mathbf{f}_l{}^{(\tau)} = \mathbf{f}_l{}^{(\tau-1)} - \theta \nabla g_i(\mathbf{f}_l{}^{(\tau)}) \tag{6}$$

where $\nabla g_i(\mathbf{f}_l) := \frac{M_l}{\mu}\big([\mathbf{I}_N - \hat{\mathbf{A}}(i)]^\top [\mathbf{I}_N - \hat{\mathbf{A}}(i)] + \mathbf{M}_l{}^\top \mathbf{M}_l\big)\mathbf{f}_l - \mathbf{M}_l \mathbf{y}_l$, and $\theta > 0$ is the stepsize chosen e.g. by the Armijo rule [3]. The computational cost of (6) is dominated by the matrix-vector multiplication of $\mathbf{I}_N - \hat{\mathbf{A}}(i)$ with $\mathbf{f}_l$, which is

Algorithm 1: Joint Infer. of Signals and Graphs (JISG)

Input: Observations $\{\mathbf{y}_l\}_{l=1}^L$; sampling matrices $\{\mathbf{M}_l\}_{l=1}^L$; and regularization parameters $\{\mu, \lambda_1, \lambda_2\}$

Intialize: $\hat{\mathbf{f}}_l(0) = \mathbf{M}_l{}^\top \mathbf{y}_l,\ l = 1, \ldots, L$
while iterates not converge **do**
  Estimate $\hat{\mathbf{A}}(i)$ from (7) using ADMM.
  Update $\{\hat{\mathbf{f}}_l(i)\}_{l=1}^L$ using (5) and (6).
  $i = i + 1$
end while

Output: $\{\hat{\mathbf{f}}_l{}^*\}_{l=1}^L, \hat{\mathbf{A}}^*$.

proportional to $\mathcal{O}(k_{\text{nnz}})$, where k_{nnz} denotes the number of non-zero entries of $\hat{\mathbf{A}}(i)$. Moreover, the learned $\hat{\mathbf{A}}(i)$ is expected to be sparse due to the ℓ_1 regularizer in (3), which renders first-order iterations (6) computationally attractive, especially when graphs are large. The GD iterations (6) are run in parallel across l until convergence to a minimizer of (5).

On the other hand, with $\{\hat{\mathbf{f}}_l(i)\}_{l=1}^L$ available, $\hat{\mathbf{A}}(i)$ is found after solving

$$\min_{\mathbf{A}\in\mathcal{A}} \sum_{l=1}^{L} \|\hat{\mathbf{f}}_l(i) - \mathbf{A}\hat{\mathbf{f}}_l(i)\|_2^2 + \lambda_1\|\mathbf{A}\|_1 + \lambda_2\|\mathbf{A}\|_F^2 \quad (7)$$

where the fitting term has been omitted from (3). Note that (7) is a strongly convex problem with linear constraints, and as such it admits a unique minimizer. To find it, we adopt the alternating methods of multipliers (ADMM), which guarantees convergence to the global minimum in a finite number of iterations; see e.g. [8].

The BCD solver for joint inference of topology and signals evolving over a graph (JISG) is summarized as Algorithm 1. JISG converges at least to a stationary point of (3), as asserted by the ensuing lemma.

Lemma 1. *The sequence of iterates* $\left\{\{\hat{\mathbf{f}}_l(i)\}_{l=1}^L, \hat{\mathbf{A}}(i)\right\}_i$*, resulting from solving* (4) *and* (7)*, is bounded and converges monotonically to a stationary point of* (3)*.*

Proof. The basic convergence results of BCD iterations have been established in [29]. First, notice that all the terms in (3) are Gâteaux-differentiable over their open domain except the non-differentiable ℓ_1 norm, which is however separable. These observations establish, based on [29, Lemma 3.1], that $\mathcal{F}(\mathbf{A}, \{\mathbf{f}_l\}_{l=1}^L)$ is regular at each coordinatewise minimum point $\hat{\mathbf{A}}^*, \{\hat{\mathbf{f}}_l{}^*\}_{l=1}^L$, and therefore every such a point is a stationary point of (3). Moreover, $\mathcal{F}(\mathbf{A}, \{\mathbf{f}_l\}_{l=1}^L)$ is continuous and convex per variable. Hence, by appealing to [29, Theorem 5.1], the sequence of iterates generated by JISG converges monotonically to a coordinatewise minimum point of $\mathcal{F}$ and consequently to a stationary point of (3). □

A few additional remarks are now in order.

Remark 3. A popular alternative to the elastic net regularizer is the nuclear norm $\rho(\mathbf{A}) = \|\mathbf{A}\|_*$ that promotes low rank of the learned adjacency matrix - a well-motivated attribute when the graph is expected to exhibit clustered structure [5].

Remark 4. Oftentimes, prior information about $\mathcal{G}$ may be available, e.g. the support of $\mathbf{A}$; the edge weight $A_{n,n'}$ for some n, n'. Such prior information can be easily incorporated in (3) by adjusting $\mathcal{A}$, and the ADMM solver accordingly.

Remark 5. The SEM-based graph function estimator in (4) can be employed on its own, when the topology is known. As a byproduct of using SEMs, (4) estimates functions over *directed* graphs as well, while kernel-based approaches [27] and estimators that rely on the graph Fourier transform [26] are usually confined to undirected graphs.

4. NUMERICAL TESTS

The tests in this section evaluate the performance of the proposed joint inference approach in comparison with state of the art graph signal inference and topology identification techniques using synthetic and real data. The network topology perfomance is measured by the edge identification error rate (EIER), defined as EIER $:= \frac{\|\mathbf{S}-\hat{\mathbf{S}}\|_0}{N(N-1)} \times 100\%$, with the operator $\|\cdot\|_0$ denoting the number of nonzero entries of its argument, and $\mathbf{S}$ ($\hat{\mathbf{S}}$) give the support of $\mathbf{A}$ ($\hat{\mathbf{A}}$). For the estimated adjacency an edge is declared present if $|\hat{A}_{n,n'}|$ exceeds a threshold chosen to yield the smallest EIER. The signal inference performance is assessed by comparing the normalized mean-square error NMSE $:= \sum_{l=1}^L \|\hat{\mathbf{f}}_l - \mathbf{f}_l\|^2/\|\mathbf{f}_l\|^2$ Parameters μ, λ_1 and λ_2 are selected via cross validation. All results represent averages over 10 independent Monte Carlo runs. Unless otherwise stated, $\mathcal{M}_l$ is chosen uniformly at random without replacement over $\mathcal{V}$ for each l with constant size over time; that is, $M_l = M,\ \forall l$.

First, a synthetic network of size $N = 81$ was generated using the *Kronecker product* model, that effectively captures properties of real graphs [17]. It relies on the "seed matrix"

$$\mathbf{D}_0 := \begin{bmatrix} 0.6 & 0.1 & 0.7 \\ 0.3 & 0.1 & 0.5 \\ 0 & 1 & 0.1 \end{bmatrix}$$

that produces the $N \times N$ matrix as $\mathbf{D} := \mathbf{D}_0 \otimes \mathbf{D}_0 \otimes \mathbf{D}_0 \otimes \mathbf{D}_0$, where $\otimes$ denotes Kronecker product. The entries of $\mathbf{A}$ were selected as $A_{n,n'} \sim \text{Bernoulli}(D_{n,n'})\ \forall n, n'$, and the resulting matrix was rendered symmetric by adding its transpose. The graph signals were generated using the graph-bandlimited model $\mathbf{f}_l = \sum_{i=1}^{10} \gamma_l{}^{(i)} \mathbf{u}^{(i)},\ l = 1, \ldots, L$, where $L = 100$, $\gamma_l{}^{(i)} \sim \mathcal{N}(0, 1)$, and $\{\mathbf{u}^{(i)}\}_{i=1}^{10}$ are the eigenvectors associated with the 10 smallest eigenvalues of the Laplacian matrix $\mathbf{L} := \text{diag}\{\mathbf{A}\mathbf{1}\} - \mathbf{A}$.

The compared estimators for graph signal inference include the bandlimited estimator (BL) [1, 20, 30] with bandwidth B; and the multi-kernel learning (MKL) estimator that

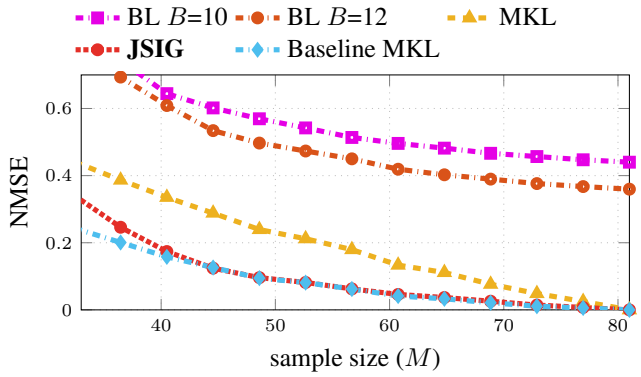


Fig. 1: Graph signal inference performance based on NMSE ($\mu = 10^4, \lambda_1 = 0.5, \lambda_2 = 0.1$).

Algorithm	Duration
JSIG	**120**
MKL baseline	1,600

Table 1: Execution times in milliseconds, averaged over different M values, for the algorithms in Fig. 1. Tests in Sec. 5.1 are run using Matlab on a PC with a Intel Core (TM) i7-4790, clocked at 3.6GHz with 32GB RAM memory.

employs a dictionary comprising 100 diffusion kernels with parameter σ^2 uniformly spaced between 0.01 and 2 and selects the kernel that "fits" best the observed data [22]. These reconstruction algorithms assume the topology is known and symmetric, which may not always be the case. To capture potential model mismatch, BL and MKL use $\mathbf{A} + \mathbf{E}$ with $E_{n,n'} \sim \mathcal{N}(0, 0.05)$ instead of $\mathbf{A}$. Fig. 1 shows the NMSE of various approaches with increasing M, where $M_l = M \; \forall l$, and the baseline is the MKL that considers the true topology $\mathbf{A}$. The reconstruction performance of JSIG, that simultaneously learns the topology, is superior compared to that of BL and MKL, while it matches the baseline performance. Moreover, JSIG exhibits faster computational time relative to the MKL baseline; see Table 1. No noise was added to the observations; that is, $\boldsymbol{\epsilon}_l = \mathbf{0}, \; \forall l$.

Moreover, for the same simulation setting, the topology inference performance was evaluated, by comparing with the elastic net (EN) SEM that learns the network topology from observations across all nodes, meaning $\{\mathbf{y}_l = \mathbf{f}_l\}_{l=1}^L$. Fig. 2 plots the EIER with increasing M for JSIG while EN-SEM uses $M = N$. The semi-blind novel approach achieves similar performance with the baseline, which can not cope with missing nodal measurements.

Further tests were conducted using real gene exrpession data [4]. Nodes in this network represent $N = 39$ immune-related genes, while the measurements consist of gene expression data from $L = 69$ unrelated Nigerian individuals. The graph function f_{ln} measures the expression level of gene n for individual l. This experiment evaluates the topology in-

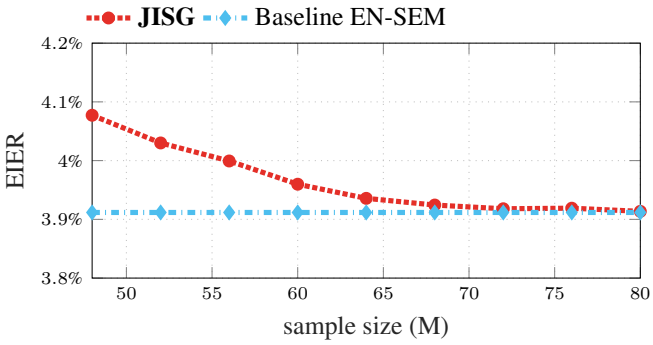
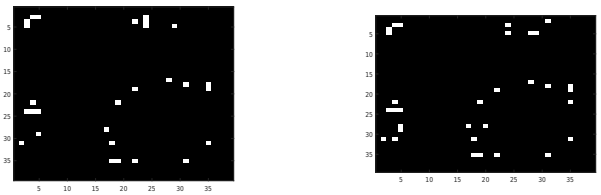


Fig. 2: Network topology inference performance based on EIER. EN-SEM uses $M = N$ and the same parameters λ_1, λ_2 as JSIG. ($\mu = 10^8, \lambda_1 = 100, \lambda_2 = 1$).



(a) EN-SEM. (b) JSIG.

Fig. 3: Heatmaps of estimated adjacency matrices for the gene regulatory network. White (black) indicate presence (absence) of an edge. ($\mu = 10^8, \lambda_1 = 10^{-2}, \lambda_2 = 10^{-4}$).

ference performance of JSIG with $M_l = 31$ genes for all individuals sampled at random. Since no ground-truth topology is available here, the estimated adjacency of EN-SEM, that relies on all the observations, was used for $\mathbf{A}$. Fig. 3 depicts heatmaps of the estimated adjacencies, where an edge is identified if $|\hat{A}_{n,n'}| > 0.068$. As observed, JSIG learns a similar topology with EN-SEM having EEIR$= 0.5\%$ and $\|\mathbf{A} - \hat{\mathbf{A}}\|_F^2 / \|\mathbf{A}\|_F^2 = 0.02$, and imputes the missing values with NMSE $= 0.017$. Therefore, our joint approach may discover causal patterns even when gene expression data contain missing values.

5. CONCLUSIONS AND FUTURE WORK

This paper introduced a novel approach to jointly infer sparse directed network topologies and graph signals based on SEMs. The resultant optimization task is solved by an efficient algorithm with provable convergence that alternates between reconstructing the graph signal and inferring the topology using ADMM. Numerical tests on synthetic and real data-sets demonstrate the competitive performance of JISG in both inferring the graph signals and the network topology. Future research will focus on learning structured network topologies, extensions to time evolving and multi-layer networks, and corresponding identifiability analyses.

6. REFERENCES

[1] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Sig. Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.

[2] B. Baingana, G. Mateos, and G. B. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE J. Sel. Topics Sig. Proc.*, vol. 8, no. 4, pp. 563–575, Aug. 2014.

[3] D. Bertsekas, *Nonlinear Programming*. Athena Scientific Belmont, 1999.

[4] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations," *PLoS Comp. Biol.*, vol. 9, no. 5, p. e1003068, May 2013.

[5] Y. Chen, A. Jalali, S. Sanghavi, and H. Xu, "Clustering partially observed graphs via convex optimization," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2213–2238, 2014.

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

[7] P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *IEEE Trans. Sig. Process.*, vol. 62, no. 13, pp. 3305–3320, Jul. 2014.

[8] G. B. Giannakis, Q. Ling, G. Mateos, I. D. Schizas, and H. Zhu, "Decentralized learning for wireless communications and networking," in *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, 2016, pp. 461–497.

[9] A. S. Goldberger, "Structural equation methods in the social sciences," *Econometrica*, vol. 40, no. 6, pp. 979–1001, Nov. 1972.

[10] J. R. Harring, B. A. Weiss, and J.-C. Hsu, "A comparison of methods for estimating quadratic effects in nonlinear structural equation models." *Psychological Methods*, vol. 17, no. 2, pp. 193–214, Jun. 2012.

[11] V. N. Ioannidis, M. Ma, A. Nikolakopoulos, G. B. Giannakis, and D. Romero, "Kernel-based inference of functions on graphs," in *Adaptive Learning Methods for Nonlinear System Modeling*, D. Comminiello and J. Principe, Eds. Elsevier, 2018.

[12] V. N. Ioannidis, A. N. Nikolakopoulos, and G. B. Giannakis, "Semi-parametric graph kernel-based reconstruction," in *Global Conf. Sig. Inf. Process.*, Montreal, Canada, Nov. 2017.

[13] V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Learning dynamic processes over dynamic graphs via a multi-kernel kriged Kalman filter," *IEEE Trans. Signal Process., to appear 2018.*

[14] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions*. Sage, 2009.

[15] 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.

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

[17] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," *J. Mach. Learn. Res.*, vol. 11, no. 2, pp. 985–1042, Feb. 2010.

[18] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Sig. Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.

[19] B. Muthén, "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators," *Psychometrika*, vol. 49, no. 1, pp. 115–132, Mar. 1984.

[20] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Global Conf. Sig. Inf. Process.*, Austin, Texas, 2013, pp. 491–494.

[21] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE J. Sel. Topics Sig. Process.*, vol. 11, no. 6, pp. 1–14, Sep. 2017.

[22] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Sig. Process.*, vol. 65, no. 3, pp. 764–778, Feb. 2017.

[23] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. on Sig. and Info. Process. over Net.*, vol. 3, no. 3, pp. 467–483, Sep. 2017.

[24] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Trans. Sig. Process.*, vol. 65, no. 14, pp. 4004–4018, July 2017.

[25] ——, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Trans. Sig. Proc.*, vol. 65, no. 10, pp. 2503–2516, May 2017.

[26] 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 Sig. Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[27] A. J. Smola and R. I. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*. Springer, 2003, pp. 144–158.

[28] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, 2017.

[29] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. of Opt.n Theo. and Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[30] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *Trans. on Sig. Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.