Network Topology Inference via Elastic Net Structural Equation Models

Panagiotis A. Traganitis, Yanning Shen and Georgios B. Giannakis

Dept. of ECE & Digital Technology Center, Univ. of Minnesota, USA.

Emails: {traga003,shenx513,georgios}@umn.edu

Abstract—Linear structural equation models (SEMs) have been very successful in identifying the topology of complex graphs, such as those representing social and brain networks. In many cases however, the presence of highly correlated nodes hinders performance of the available SEM estimators that rely on the least-absolute shrinkage and selection operator (LASSO). To this end, an elastic net based SEM is put forth, to infer causal relations between nodes belonging to networks, in the presence of highly correlated data. An efficient algorithm based on the alternating direction method of multipliers (ADMM) is developed, and preliminary tests on synthetic as well as real data demonstrate the effectiveness of the proposed approach.

Index Terms—Networks, Topology inference, Structural Equation Models, Elastic Net

I. INTRODUCTION

Networks have ubiquitous presence in a plethora of disciplines such as sociology, communications and machine learning among others, where their ability to model a multitude of complex systems [18] has rendered them indispensable. These complex systems may include naturally emerging networks, such as social and communication or power networks, or model-induced ones, employed to simplify the representation of a system, such as brain networks [22]. Given a graph representation of a network, various tools from graph theory and network science [18] can be employed to draw inferences from nodal variable dependencies. Examples of such inferences include behavioral prediction of complex systems [13], and detection of communities over social or brain graphs [4], among others. In addition, many machine learning [2] and signal processing [23] tasks can be performed over a graph. All these tasks, however, presume knowledge of the network graph representation. While this information may be naturally available in some networks, such as power or communication networks, in many cases, such as brain networks, it has to be inferred.

Network *topology inference* aims to discover the (typically sparse) connectivity between nodes, given only nodal measurements, and thus has practical implications in a multitude of settings. Examples of such applications include discovery of causal links between brain regions, or identifying how contagions spread [1].

Prior works. Several approaches have been proposed for inferring the topology of networks. Probabilistic models rely on maximum likelihood estimation to obtain edge weights [8], [17], [19]. *Structural equation models* (SEMs) are simple yet capable of capturing causal relationships [11]. The basic premise of a SEM is that a node measurement depends linearly on those of its neighbors, plus possibly an additive exogenous input. Linear SEMs have wide applicability in fields as diverse as sociology [7], psychometrics [16] and genetics [3], and have recently been employed to track dynamic topologies of social networks [1], by leveraging the typically sparse connectivity of a network. In addition, nonlinear SEMs have been advocated to model nonlinear phenomena [9], [12], [14], and also for capturing nonlinear connectivity between pairs of nodes [20], [21]. All aforementioned approaches, however, employ LASSO [10] type solvers, which tend to ignore multiple edges that arise when data are highly correlated.

The aim of the present work is to introduce a novel method that enables topology inference of networks by employing an elastic net [24] solver, that performs well even in scenarios where some of the data are highly correlated. In addition, even when data are not highly correlated, the proposed elastic net SEM performs at least as well as the regular LASSO solver.

Notation. Boldface uppercase (lowercase) letters indicate matrices (column vectors). The vector containing the diagonal elements of a matrix is denoted by $\mathrm{diag}(\cdot)$, while $\mathbf{0}$ and $\mathbf{1}$ denote the all-zeros and all-ones vectors, respectively. Calligraphic uppercase letters denote sets, and $|\mathcal{A}|$ represents the cardinality of $\mathcal{A}$. Operators $\|\cdot\|_2$ and $\|\cdot\|_1$ stand for the L_2- and L_1-norms of a vector, respectively, $(\cdot)^\top$ denotes vector and matrix transposition, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. The matrix operator $\|\cdot\|_0$ denotes the number of nonzero entries of its argument, while $\mathrm{Bdiag}\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ denotes a block-diagonal matrix, with the matrices $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ in its diagonal.

II. NETWORK MODEL AND PROBLEM STATEMENT

Consider a network consisting of N nodes be modeled as a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices/nodes, with $|\mathcal{V}| = N$, and $\mathcal{E}$ is the set of edges between nodes. This graph can be further described using a binary $N \times N$ adjacency matrix $\mathbf{A}$ whose (i, j)-th entry is given by

$$\alpha_{ij} \begin{cases} \neq 0 & \text{if } (i,j) \in \mathcal{E} \\ = 0 & \text{otherwise;} \end{cases} \tag{1}$$

hence, α_{ij} is nonzero if there exists a *directed* edge between nodes i and j. Accordingly, weights assigned to edges can be captured by an $N \times N$ matrix $\mathbf{W}$.
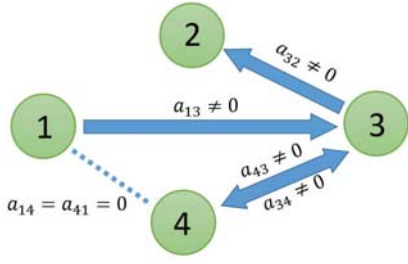
Fig. 1. Example of possible network topology and its connection with structural equation models [20].

Consider now a process observed over the entire network, with y_{im} denoting the m-th observation at node i. A linear SEM [1], [3] obeys the relationship

$$y_{im} = \sum_{j \neq i} \alpha_{ij} y_{jm} + b_{ii} x_{im} + \epsilon_{im} \quad (2)$$

where the m-th observation at node i depends linearly on the corresponding endogenous observations of the neighboring nodes of i, with the addition of a possible exogenous variable x_{im}. Here, ϵ_{im} captures unmodeled dynamics, such as noise. An example of such a model is shown in Fig. 1.

Upon defining the $M \times 1$ vectors $\boldsymbol{y}_i := [y_{i1}, \ldots, y_{iM}]^\top$, $\boldsymbol{x}_i := [x_{i1}, \ldots, x_{iM}]^\top$, the $N \times 1$ vector $\boldsymbol{b} = [b_{11}, b_{22}, \ldots, b_{NN}]^\top$, and the $M \times N$ matrices $\mathbf{Y} := [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]$, $\mathbf{X} := [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$, (2) can be cast into matrix form

$$\mathbf{Y} = \mathbf{YA} + \mathbf{XB} + \mathbf{E} \quad (3)$$

where $\mathbf{B}$ is an $N \times N$ diagonal matrix with $\boldsymbol{b}$ as its diagonal, and $\mathbf{E}$ collects all the noise variables.

Given nodal measurements across the entire network, $\mathbf{Y}$, and exogenous inputs $\mathbf{X}$ the task of topology inference is to find the unknown adjacency matrix $\mathbf{A}$ of the underlying network. Note that typical connectivity of real-world networks is sparse, as nodes usually connect to few other nodes, thus the adjacency matrix $\mathbf{A}$ is expected to be sparse.

III. TOPOLOGY INFERENCE ALGORITHM

Having established the network model, we next consider estimating the wanted adjacency matrices in the noisy SEM of (3). In order to estimate the unknowns in (3), or (2), the following sparsity promoting optimization problem is proposed:

$$\min_{\mathbf{A},\mathbf{B}} \frac{1}{2}\|\mathbf{Y} - \mathbf{YA} - \mathbf{XB}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \frac{\lambda_2}{2}\|\mathbf{A}\|_F^2$$
$$\text{subject to } \text{diag}(\mathbf{A}) = \mathbf{0} \quad (4)$$

where the constraint $\text{diag}(\mathbf{A}) = \mathbf{0}$ ensures that there are no self-loops, λ_1 and λ_2 are regularization scalars for the L_1 and Frobenius norms, and $\|\cdot\|_1$ denotes the L_1-norm of the vectorized matrix. Also, note that the objective function in (4) is convex. The following proposition justifies the use of the elastic net penalty, the weighted sum of L_1 and L_2 norms,

instead of just using the sparsity promoting L_1 norm, typically employed in LASSO.

Proposition 1 ([24]). Suppose that the nodal measurements $\{\boldsymbol{y}_i\}$ have unit norm. Let $\boldsymbol{a}^* := [a_1^*, \ldots, a_N^*]^\top$ be the optimal solution to the following optimization problem

$$f(\boldsymbol{a}) = \frac{1}{2}\|\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}\|_2^2 + \lambda_1\|\boldsymbol{a}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{a}\|_2^2 \quad (5)$$

and suppose $a_i^* a_j^* > 0$. With $\psi_{ij} := \boldsymbol{y}_i^\top \boldsymbol{y}_j$ it then holds that

$$|a_i^* - a_j^*| \leq \frac{\sqrt{2(1-\psi_{ij})}}{\lambda_2}\|\boldsymbol{z}\|_2 \quad (6)$$

Proof: Since $\boldsymbol{a}^*$ is the minimizer of (5) it holds that

$$f(\boldsymbol{a}^*) \leq f(\mathbf{0}) \Rightarrow \|\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*\|_2 \leq \|\boldsymbol{z}\|_2. \quad (7)$$

In addition the gradient of f at $\boldsymbol{a}^*$ will vanish, that is

$$-\mathbf{Y}^\top(\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*) + \lambda_1\partial\|\boldsymbol{a}^*\|_1 + \lambda_2\boldsymbol{a}^* = \mathbf{0}. \quad (8)$$

Now consider the i-th and j-th rows of (8)

$$-\boldsymbol{y}_i^\top(\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*) + \lambda_1\text{sign}(a_i^*) + \lambda_2 a_i^* = 0 \quad (9)$$

$$-\boldsymbol{y}_j^\top(\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*) + \lambda_1\text{sign}(a_j^*) + \lambda_2 a_j^* = 0. \quad (10)$$

Subtracting (10) from (9) yields

$$\lambda_2(a_i^* - a_j^*) = (\boldsymbol{y}_i - \boldsymbol{y}_j)^\top(\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*) \quad (11)$$

Taking the norm of both sides and invoking the Cauchy-Schwarz inequality

$$|a_i^* - a_j^*| \leq \frac{\sqrt{2(1-\psi_{ij})}}{\lambda_2}\|\boldsymbol{z} - \mathbf{Y}\boldsymbol{a}^*\|_2 \leq \frac{\sqrt{2(1-\psi_{ij})}}{\lambda_2}\|\boldsymbol{z}\|_2 \quad (12)$$

where the last inequality follows from (7). ∎

Proposition 1 suggests that when two nodal measurements are highly correlated, the elastic net solver will likely have these two nodes connected to the same set of other nodes. This is in contrast to LASSO [10] solvers for (3), where in the presence of highly correlated data, the L_1 norm regularization promotes only one connection.

Note that if $\lambda_2 = 0$ then (4) reduces to the SEM obtained via LASSO [1]. In order to solve (4), the alternating direction method of multipliers (ADMM [6]) will be employed. Consider the auxiliary variables $\mathbf{C}$ and $\mathbf{D}$, and re-write (4) as

$$\min_{\mathbf{A},\mathbf{C},\mathbf{B},\mathbf{D}} \frac{1}{2}\|\mathbf{Y} - \mathbf{YA} - \mathbf{XB}\|_F^2 + \lambda_1\|\mathbf{C}\|_1 + \frac{\lambda_2}{2}\|\mathbf{A}\|_F^2$$
$$\text{subject to } \text{diag}(\mathbf{C}) = \mathbf{0}, \quad \mathbf{A} = \mathbf{C}, \quad \mathbf{B} = \mathbf{D} \quad (13)$$

Note that, here $\mathbf{D}$ is a diagonal matrix. The augmented Lagrangian of (13) is then

$$\mathcal{L} = \frac{1}{2}\|\mathbf{Y} - \mathbf{YA} - \mathbf{XB}\|_F^2 + \lambda_1\|\mathbf{C}\|_1 + \frac{\lambda_2}{2}\|\mathbf{A}\|_F^2 +$$
$$\text{tr}\left(\mathbf{U}_1^\top(\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C}))\right) + \frac{\rho}{2}\|\mathbf{A} - \mathbf{C} + \text{diag}(\mathbf{C})\|_F^2 +$$
$$\text{tr}(\mathbf{U}_2^\top(\mathbf{B} - \mathbf{D})) + \frac{\rho}{2}\|\mathbf{B} - \mathbf{D}\|_F^2 \quad (14)$$

Algorithm 1 Elastic Net SEM Topology Inference

Input: Nodal measurements $\mathbf{Y}$; Exogenous inputs $\mathbf{X}$; parameters $\{\lambda_1, \lambda_2, \rho\}$

Output: Estimate of network adjacency matrix $\mathbf{A}$

1: Initialize all variables to 0.
2: **while** Not converged **do**
3: Update $\mathbf{A}$ using (15).
4: Update $\mathbf{B}$ using (16).
5: Update auxiliary variable $\mathbf{C}$ using (17).
6: Update auxiliary variable $\mathbf{D}$ using (19).
7: Update Lagrange multipliers using (20)
8: **end while**

---

where $\mathbf{U}_1$ and $\mathbf{U}_2$ denote Lagrange multipliers, while ρ is a positive scalar. Henceforth, matrix superscripts denote ADMM iteration indices. The update for $\mathbf{A}$ can be obtained by taking the derivative of $\mathcal{L}$ with respect to (w.r.t.) $\mathbf{A}$ and equating it to zero

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}} = \mathbf{0} \Rightarrow \tag{15}$$
$$(\mathbf{K}_Y + (\lambda_2 + \rho)\mathbf{I}) \mathbf{A}^i = \mathbf{K}_Y - \mathbf{K}_{Y,X}\mathbf{B} - \mathbf{U}_1^{i-1} + \rho\mathbf{C}^{i-1}.$$

Here $\mathbf{I}$ is the identity matrix of appropriate dimension, and $\mathbf{K}_{Y,X}$ denotes the inner product matrix between the columns of $\mathbf{Y}$ and $\mathbf{X}$, that is $\mathbf{K}_{Y,X} := \mathbf{Y}^\top\mathbf{X}$. Also, let by definition $\mathbf{K}_Y := \mathbf{K}_{Y,Y}$. The update for $\mathbf{B}$ can be obtained in a similar manner

$$\frac{\partial \mathcal{L}}{\partial \mathbf{B}} = \mathbf{0} \Rightarrow \tag{16}$$
$$(\mathbf{K}_X + \rho\mathbf{I}) \mathbf{B}^i = \mathbf{K}_{X,Y}(\mathbf{I} - \mathbf{A}^i) - \mathbf{U}_2^{i-1} + \rho\mathbf{D}^{i-1}.$$

Accordingly, the update for the $\mathbf{C}$ is given by

$$\mathbf{J} = \mathcal{T}_{\lambda_1/\rho}\left(\mathbf{A}^i + \frac{1}{\rho}\mathbf{U}_1^{i-1}\right) \Rightarrow$$
$$\mathbf{C}^i = \mathbf{J} - \mathrm{diag}(\mathbf{J}) \tag{17}$$

where $\mathcal{T}_\kappa(\cdot)$ denotes the elementwise soft-thresholding operator defined as

$$\mathcal{T}_\kappa(x) := \begin{cases} x - \kappa & , x > \kappa \\ 0 & , |x| \le \kappa \\ x + \kappa & , x < -\kappa. \end{cases} \tag{18}$$

The diagonal entries of $\mathbf{D}$ are updated as follows

$$d = \frac{1}{\rho}\mathrm{diag}(\mathbf{U}_2^{i-1}) + \mathrm{diag}(\mathbf{B}^i) \Rightarrow$$
$$\mathbf{D}^i = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_N \end{bmatrix} \tag{19}$$

where d_k is the k-th entry of d. Finally, the Lagrange multipliers are updated as

$$\mathbf{U}_1^i = \mathbf{U}_1^{i-1} + \rho\left(\mathbf{A}^i - \mathbf{C}^i\right)$$
$$\mathbf{U}_2^i = \mathbf{U}_2^{i-1} + \rho\left(\mathbf{B}^i - \mathbf{D}^i\right). \tag{20}$$

The steps of our topology inference algorithm are listed in Alg. 1. Since (4) is convex, this ADMM procedure will converge in a finite number of iterations. The update complexity of $\mathbf{A}$ and $\mathbf{B}$ is $\mathcal{O}(N^3)$, while the update complexity of the auxiliary variables $\mathbf{C}$ and $\mathbf{D}$ is $\mathcal{O}(N^2)$ and $\mathcal{O}(N)$ respectively. This brings the overall complexity of the algorithm to $\mathcal{O}(I(N^3 + N^2 + N))$, where I is the number of required ADMM iterations until convergence.

Remark 1. The proposed ADMM solver for the Elastic Net SEM, can also solve LASSO SEM's by setting $\lambda_2 = 0$.

Remark 2. All the variable updates are separable per node, i.e. each column of $\mathbf{A}$ can be updated separately, which lends it self naturally to a *distributed* implementation of the algorithm; see also [6].

Remark 3. The present topology identification approach can be extended to cope with dynamically changing networks, by employing an exponentially weighted least-squares cost in (4), along the lines of [1].

IV. NUMERICAL TESTS

The proposed scheme is validated in this section using synthetic and real data. In all tests Elastic Net SEM is compared to LASSO SEM [1], both implemented using the ADMM algorithm outlined in Section III. An edge is declared present if $\hat{\alpha}_{ij} \ge 10^{-1}$. Given the support $\mathbf{S}$ of the ground truth adjacency matrix $\mathbf{A}$ with entries

$$[\mathbf{S}]_{ij} = \begin{cases} 1, & \text{if } a_{ij} \ne 0 \\ 0, & \text{otherwise} \end{cases}$$

and the support $\hat{\mathbf{S}}$ of estimated adjacency matrix $\hat{\mathbf{A}}$ with entries,

$$[\hat{\mathbf{S}}]_{ij} = \begin{cases} 1, & \text{if } \hat{a}_{ij} \ge 10^{-1} \\ 0, & \text{otherwise} \end{cases}$$

the metric evaluated is the *edge identification error rate (EIER)* given by

$$EIER = \frac{\|\mathbf{S} - \hat{\mathbf{S}}\|_0}{N(N-1)} \times 100\%.$$

The software employed to conduct all experiments is MATLAB [15]. All results represent averages over 10 independent Monte Carlo runs. In all experiments the ADMM parameter ρ is set to 10.

A. Synthetic data

A synthetic network with $L = 4$ non-overlapping communities and $N = \sum_{\ell=1}^{L} N_\ell$ is generated. Here, N_ℓ denotes the number of nodes in the ℓ-th community, which were set as $N_\ell = \{4, 8, 16, 32\}$. The connectivity pattern in each community is generated based on the following seed matrix

$$\mathbf{S}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{21}$$

The adjacency matrix in each community is generated as $[\mathbf{A}_\ell]_{ij} \sim \text{Bernoulli}(0.9[\mathbf{S}_\ell]_{ij})$, where $\mathbf{S}_\ell = \mathbf{S}_0 \otimes \mathbf{1}_{\ell \times 1}$, and $\otimes$ denotes the Kronecker product. The overall adjacency matrix of the network is then defined as $\mathbf{A} = \text{Bdiag}\{\mathbf{A}_1, \ldots, \mathbf{A}_L\}$, while the matrix of exogenous effects is set as $\mathbf{B} = \mathbf{I}$. The number of observations per node is $M = 54$. For the nodes in each community, the $N_\ell \times M$ exogenous variable matrix $\mathbf{X}_\ell$ was generated as $\mathbf{X}_\ell = \bar{\mathbf{X}}_\ell \otimes \mathbf{1}_{\ell \times 1}$, with each entry of $\bar{\mathbf{X}}_\ell$ drawn from a standardized normal distribution. The exogenous variable matrix for the entire network is formed as $\mathbf{X} = [\mathbf{X}_1^\top, \ldots, \mathbf{X}_L^\top]^\top$. Setting $\sigma_\epsilon = 0.01$, noise terms were sampled independently as $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Finally, the measurement matrices were generated based on the linear SEM of (3) as $\mathbf{Y} = (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{BX} + \mathbf{E})$.

The Elastic Net SEM parameters are $\lambda_1 = 0.005, \lambda_2 = 0.1$, while for LASSO SEM $\lambda = 0.005$. Fig. 2a shows the heatmap of the ground truth adjacency matrix $\mathbf{A}$ for one instance of this network. For the same instance, Figs. 2b and 2c depict the estimated adjacency matrices for Elastic Net SEM and LASSO SEM, respectively. Results for this network are listed in Tab. I. Clearly, Elastic Net SEM is able to identify more edges in this scenario, while LASSO SEM performs worse. As the data are generated to be highly correlated, this experiment showcases the shortcomings of the LASSO solver for SEM [1] compared to the Elastic Net. Indeed, LASSO SEM tends to ignore many edges that correspond to highly correlated data.

Algorithm	Average EIER
Elastic Net SEM	**5.3056**
SEM	9.0833

TABLE I
AVERAGE NUMBER OF MISIDENTIFIED EDGES FOR ELASTIC NET SEM AND SEM FOR A SYNTHETIC NETWORK.

B. Real data

Further tests were conducted based on real gene regulatory network data [3]. Nodes in this networks represent 39 immune-related genes, while the measurements consist of gene expression data from 69 unrelated Nigerian individuals [5]. The gene expression levels were treated as endogenous inputs, while genotypes of the genes involved were considered as the exogenous inputs. Note that, in this scenario, there is no ground-truth adjacency matrix, thus only Elastic Net SEM and LASSO SEM are compared. Fig. 3 shows the results for this dataset. The parameters for this experiment were $\lambda_1 = 600, \lambda_2 = 600$ for Elastic Net SEM, and $\lambda = 600$ for LASSO SEM. While both algorithms provide similar adjacency matrices, note that Elastic Net SEM is able to identify two more edges than SEM. This could possibly facilitate the discovery of novel causal patterns, that may not be captured by LASSO-based SEM.

V. CONCLUSIONS AND FUTURE WORK

This paper introduced a novel approach for network topology inference, termed Elastic Net SEM, which is based on linear structural equation models. The proposed method exploits
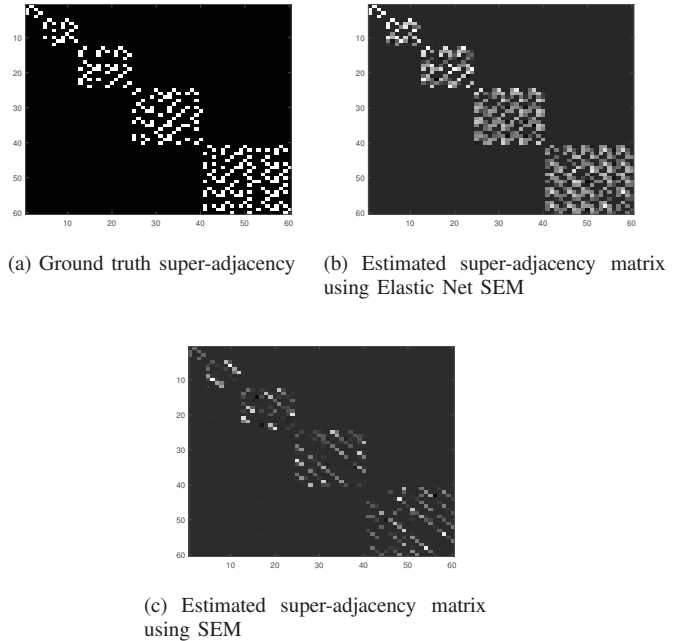

(a) Ground truth super-adjacency

(b) Estimated super-adjacency matrix using Elastic Net SEM

(c) Estimated super-adjacency matrix using SEM

Fig. 2. Heatmaps of adjacency matrices for a 4-community synthetic network. White (black) squares indicate the presence (absence) of an edge.

(a) Estimated adjacency matrix using Elastic Net SEM

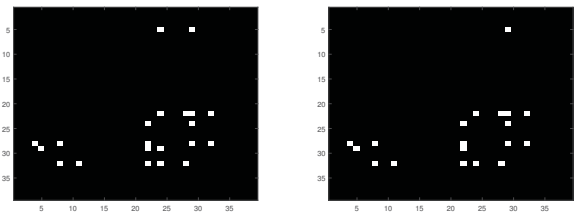(b) Estimated adjacency matrix using SEM

Fig. 3. Heatmaps of adjacency matrices for the gene regulatory network, with $N = 39$ nodes in total. White (black) squares indicate the presence (absence) of an edge.

the sparse connectivity of the network, through the elastic net, to identify possible directed edges, even in the presence of highly correlated data, a scenario where LASSO typically fails. Elastic Net SEM was efficiently implemented using an ADMM algorithm and preliminary tests on synthetic and real data showcase promising results compared to the LASSO-based SEM. Future research will focus on extensive numerical tests with real datasets, extensions to multi-layer networks, as well as distributed implementations and corresponding identifiability analysis.

REFERENCES

[1] 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.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[3] 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, May 2013.

[4] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[5] K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal *et al.*, "A second generation human haplotype map of over 3.1 million snps," *Nature*, vol. 449, no. 7164, pp. 851–861, 2007.

[6] 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 and Imaging, Science and Engineering*, R. Glowinski, S. Osher, and W. Yin, Eds. Springer, 2016.

[7] A. S. Goldberger, "Structural equation methods in the social sciences," *Econometrica*, pp. 979–1001, Nov. 1972.

[8] M. Gomez Rodriguez, J. Leskovec, and B. Schölkopf, "Structure and dynamics of information pathways in online media," in *Proceedings of the sixth ACM Intl. Conf. on Web Search and Data Mining*. ACM, 2013, pp. 23–32.

[9] 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, p. 193, Jun. 2012.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2001.

[11] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions*. Sage, 2009, vol. 2nd ed.

[12] A. Kelava, B. Nagengast, and H. Brandt, "A nonlinear structural equation mixture modeling approach for nonnormally distributed latent predictor variables," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 21, no. 3, pp. 468–481, Jun. 2014.

[13] E. D. Kolaczyk and G. Csárdi, *Statistical analysis of network data with R*. Springer, 2014, vol. 65.

[14] S.-Y. Lee and X.-Y. Song, "Model comparison of nonlinear structural equation models with fixed covariates," *Psychometrika*, vol. 68, no. 1, pp. 27–47, Mar. 2003.

[15] MATLAB, *version 8.6.0 (R2015b)*. Natick, Massachusetts: The Math-Works Inc., 2015.

[16] 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.

[17] S. Myers and J. Leskovec, "On the convexity of latent social network inference," in *Advances in Neural Information Processing Systems*, 2010, pp. 1741–1749.

[18] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, pp. 167–256, June 2003.

[19] M. G. Rodriguez, D. Balduzzi, and B. Schölkopf, "Uncovering the temporal dynamics of diffusion networks," in *28th Int. Conf. Mach. Learn.*, Bellevue, WA, July 2011.

[20] Y. Shen, B. Baingana, and G. B. Giannakis, "Nonlinear structural equation models for network topology inference," in *Proc. of Annual Conference on Information Science and Systems*, Princeton NJ, USA, March 2016, pp. 163–168.

[21] ——, "Kernel-based structural equation models for topology identification of directed networks," 2016. [Online]. Available: http://arxiv.org/abs/1605.03122

[22] ——, "Nonlinear structural vector autoregressive models for inferring effective brain network connectivity," *arXiv preprint arXiv:1610.06551*, 2016.

[23] 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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[24] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.