ROBUST NETWORK TOPOLOGY INFERENCE

Santiago Segarra[†]*, Antonio G. Marques*[*]*, Gonzalo Mateos*[‡] *and Alejandro Ribeiro*[§]

[†]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA
[*]Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain
[‡]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
[§]Dept. of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

We address the problem of identifying a graph structure from the observation of signals defined on its nodes. Fundamentally, the unknown graph encodes direct relationships between signal elements, which we aim to recover from observable indirect relationships generated by a diffusion process on the graph. We put forth a novel network topology inference approach whereby we: i) identify the eigenvectors of a matrix representation of the graph from realizations of the diffused signal; and ii) rely on these (possibly imperfect) *spectral templates* to estimate the eigenvalues by imposing desirable properties on the graph to be recovered. Robust algorithms with quantifiable performance are developed for the pragmatic settings where the eigenvectors are estimated with errors, or, when the eigenbasis is only partially known. Numerical tests showcase the effectiveness of the proposed algorithm in recovering amino-acid networks.

***Index Terms*—** Network topology inference, graph signal processing, network deconvolution.

1. INTRODUCTION

Consider a network whose abstraction is a weighted and undirected graph $\mathcal{G}$, consisting of a node set $\mathcal{N}$ of known cardinality N, an edge set $\mathcal{E}$ of unordered pairs of elements in $\mathcal{N}$, and edge weights $A_{ij} \in \mathbb{R}$ such that $A_{ij} = A_{ji} \neq 0$ for all $(i,j) \in \mathcal{E}$. The edge weights A_{ij} are collected as entries of the symmetric adjacency matrix $\mathbf{A}$ and the node degrees in the diagonal matrix $\mathbf{D} := \text{diag}(\mathbf{A1})$. These are used to form the combinatorial Laplacian matrix $\mathbf{L}_c := \mathbf{D} - \mathbf{A}$ and the normalized Laplacian $\mathbf{L} := \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. More broadly, one can define a generic *graph-shift operator* (GSO) $\mathbf{S} \in \mathbb{R}^{N\times N}$ as any matrix having the same sparsity pattern of $\mathcal{G}$ [1]. Although the choice of $\mathbf{S}$ can be adapted to the problem at hand, most existing works set it to either $\mathbf{A}$, $\mathbf{L}_c$, or $\mathbf{L}$ [2].

Our focus in this paper is on identifying graphs that explain the structure of a random signal. Formally, let $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$ be a graph signal in which the ith element x_i denotes the signal value at node i of an *unknown graph* $\mathcal{G}$ with shift operator $\mathbf{S}$. Further suppose that we are given a zero-mean white signal $\mathbf{w}$ with covariance matrix $\mathbf{C}_w = \mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right] = \mathbf{I}$. We say that the graph $\mathbf{S}$ represents the structure of the signal $\mathbf{x}$ if there exists a diffusion process in the GSO $\mathbf{S}$ that produces the signal $\mathbf{x}$ from the white signal $\mathbf{w}$, that is

$$\mathbf{x} \;=\; \alpha_0 \textstyle\prod_{l=1}^{\infty}(\mathbf{I} - \alpha_l \mathbf{S})\mathbf{w} \;=\; \sum_{l=0}^{\infty} \beta_l \mathbf{S}^l \mathbf{w}. \qquad (1)$$

This is equivalent to saying that the *graph* process $\mathbf{x}$ is stationary in $\mathbf{S}$; see [3] and Sec. 2 for further details. While $\mathbf{S}$ encodes only one-hop interactions, each successive application of the shift in (1) percolates (correlates) $\mathbf{w}$ over $\mathcal{G}$; see e.g. [4]. The product and sum representations in (1) are common (and equivalent) models for the generation of random signals. Indeed, any process that can be understood as the linear propagation of a white input through a static graph can be written in the form in (1). The justification to say that $\mathbf{S}$ represents the structure of $\mathbf{x}$ is that we can think of the edges of $\mathbf{S}$ as direct (one-hop) relations between the elements of the signal. The diffusion described by (1) generates indirect relations. Our goal is to recover the fundamental relations described by $\mathbf{S}$ from a set $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^P$ of P independent samples of the random signal $\mathbf{x}$.

Relation to prior work. Network topology inference from a set of nodal observations is a prominent problem in Network Science [5, 6]. Since networks encode similarities between nodes, several approaches infer the so-termed *association networks* by constructing graphs whose edge weights correspond to correlations or coherence measures indicating a nontrivial level of association between signal profiles at incident nodes [5, Ch. 7.3.1]. Association approaches form links taking into account only pairwise interactions. To account for the fact that the observed correlations can be due to latent network effects, alternative methods rely on partial correlations [5, 7], Gaussian graphical models [8–11], structural equation models [12, 13], Granger causality [6, 14], or their nonlinear (kernelized) variants [15, 16]. Differently, recent graph signal processing (GSP)-based network inference frameworks postulate that the network exists as a latent underlying structure, and that observations are generated as a result of a network process defined in such graph. For instance, network structure is estimated in [17] to unveil unknown relations among nodal time series adhering to an autoregressive model involving graph-filter dynamics. A factor analysis-based approach is put forth in [18] to infer graph Laplacians, seeking that input graph signals are smooth over the learned topologies; see also [19]. Different from [17–19] that operate on the graph domain, the goal here is to identify graphs that endow the given observations with desired spectral (frequency-domain) characteristics. Two works have recently explored this approach by identifying a GSO given its eigenvectors. One is [20], which assumes perfect knowledge of the eigenvectors. The other is [21], which only focuses on a Laplacian GSO.

Paper outline. In Sec. 2 we formulate the problem of identifying a GSO that explains the fundamental structure of a random signal diffused on a graph. Such a problem is shown to be underdetermined and related to the concept of stationarity of graph signals [3, 22]. More precisely, it is established that the sought GSO must have the same eigenvectors as the signal's covariance matrix. This motivates a novel two-step network topology inference approach whereby we: i) leverage results from GSP theory to identify the GSO's eigenbasis from realizations of the diffused signal; and ii) rely on these *spectral templates* to recover the GSO by estimating its eigenvalues. Different from [20], we introduce an inference method for the pragmatic case where knowledge of the spectral templates is imperfect (Sec.

Work in this paper is supported by USA NSF CCF-1217963 and Spanish MINECO grants TEC2013-41604-R and TEC2016-75361-R.

3.1), and establish that the proposed algorithm can identify the underlying network topology robustly. Last but not least, in Sec. 3.2 we investigate the case where only a subset of the GSO's eigenvectors are known. Such incomplete spectral templates arise e.g., when the observed graph signals are bandlimited. Numerical tests in Sec. 4 showcase the effectiveness of the proposed algorithm in identifying the structural properties of proteins from a mutual information graph of the co-variation between the constitutional amino-acids [23]. Concluding remarks are given in Sec. 5.

2. PROBLEM STATEMENT

This paper aims at identifying graphs that explain the structure of a random signal, meaning that there exists a diffusion process in the GSO that can generate the observed signal. Alternatively, we can say that the goal is to recover the GSO that encodes direct relationships between the elements of the signal from observable indirect relationships generated by a diffusion process. We show next that this is an underdetermined problem closely related to the notion of stationary signals on graphs [3, 22]. Begin by assuming that the shift operator $\mathbf{S}$ is symmetric. Define then the eigenvector matrix $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ and the eigenvalue matrix $\mathbf{\Lambda} := \text{diag}(\lambda_1, \ldots, \lambda_N)$ to write $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. Further note that while the diffusion expressions in (1) are polynomials on the GSO of possibly infinite degree, the Cayley-Hamilton theorem implies they are equivalent to polynomials of degree smaller than N. Upon defining the vector of coefficients $\mathbf{h} := [h_0, \ldots, h_{L-1}]^T$ and the graph filter $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l$, the signal model in (1) can be rewritten as

$$\mathbf{x} = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l\right) \mathbf{w} = \mathbf{H}\mathbf{w}, \tag{2}$$

for some particular $\mathbf{h}$ and $L \leq N$. Since a graph filter $\mathbf{H}$ is a polynomial on $\mathbf{S}$ [1], graph filters are linear graph-signal operators that have the *same eigenvectors* as the shift (i.e., the operators $\mathbf{H}$ and $\mathbf{S}$ commute). More important for the present paper, the filter representation in (2) can be used to show that *the eigenvectors of* $\mathbf{S}$ *are also eigenvectors of the covariance matrix* $\mathbf{C}_x := \mathbb{E}[\mathbf{x}\mathbf{x}^T]$. To that end, use (2) and the fact that $\mathbf{C}_w = \mathbf{I}$ to write $\mathbf{C}_x = \mathbb{E}[\mathbf{H}\mathbf{w}\big(\mathbf{H}\mathbf{w}\big)^T] = \mathbf{H}\mathbf{H}^T$. If we further use the spectral decomposition of the shift to express the filter as $\mathbf{H} = \sum_{l=0}^{L-1} h_l(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l)\mathbf{V}^T$, we can write the covariance matrix as

$$\mathbf{C}_x \;=\; \mathbf{V}\left|\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\right|^2 \mathbf{V}^T \;:=\; \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^T, \tag{3}$$

where the matrix squared-modulus operator $|\cdot|^2$ should be understood entrywise, and we defined the power spectral density vector $\mathbf{p} := \text{diag}(|\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l|^2)$ in the second equality.

The expression in (3) is precisely the requirement for a graph signal to be stationary [3, Def. 3]; hence, the problem of identifying a GSO that explains the fundamental structure of $\mathbf{x}$ is equivalent to identifying a shift on which the signal $\mathbf{x}$ is stationary. A consequence of this fact, which is apparent from (3), is that the *eigenvectors* of the shift $\mathbf{S}$ and the covariance $\mathbf{C}_x$ are the same. Alternatively, one can say that the difference between $\mathbf{C}_x$, which includes indirect relationships between components, and $\mathbf{S}$, which contains exclusively direct relationships, is only on their *eigenvalues*. While the diffusion in (1) obscures the eigenvalues of $\mathbf{S}$, the eigenvectors $\mathbf{V}$ remain present in $\mathbf{C}_x$ as templates of the original spectrum.

Identity (3) also shows that the problem of finding a GSO that generates $\mathbf{x}$ from a white input $\mathbf{w}$ with unknown coefficients [cf. (1)] is *underdetermined*. As long as the matrices $\mathbf{S}$ and $\mathbf{C}_x$ have the same eigenvectors, filter coefficients that generate $\mathbf{x}$ through a diffusion process on $\mathbf{S}$ exist. In fact, the covariance matrix $\mathbf{C}_x$ itself is a GSO that can generate $\mathbf{x}$ through a diffusion process and so is the precision matrix $\mathbf{C}_x^{-1}$. To sort out this ambiguity, which amounts to selecting the eigenvalues of $\mathbf{S}$, we assume that the GSO of interest is optimal in some sense. To be more precise, we can introduce criteria in the form of generic (often convex) functions $f(\mathbf{S}, \mathbf{\lambda})$ and define the shift operator that is optimal with respect to these criteria

$$\mathbf{S}^* := \underset{\{\mathbf{S},\mathbf{\lambda}\}}{\text{argmin}}\; f(\mathbf{S}, \mathbf{\lambda}),\; \text{s. to}\; \mathbf{S} = \textstyle\sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T,\; \mathbf{S} \in \mathcal{S}, \tag{4}$$

where $\mathbf{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$, and $\mathcal{S}$ is a convex set that specifies the type of GSO we want to identify. Also, we have written $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T$ to emphasize that if the eigenvectors $\mathbf{v}_k$ are known, the constraints in (4) are linear on the unknown eigenvalues λ_k.

Criteria. Possible choices for the criteria in (4) are to: (i) Adopt $f(\mathbf{S}, \mathbf{\lambda}) = \|\mathbf{S}\|_{\text{F}}$, which finds a GSO minimizing the total energy stored in the weights of the edges. (ii) Make $f(\mathbf{S}, \mathbf{\lambda}) = \|\mathbf{S}\|_\infty$, which yields GSOs for graphs with uniformly low edge weights. This can be meaningful, e.g., when identifying graphs subject to capacity constraints. (iii) Minimize the ℓ_0 norm $f(\mathbf{S}, \mathbf{\lambda}) = \|\mathbf{S}\|_0$ which is non-convex, but of widespread interest in identifying sparse graphs (e.g., of direct relations among signal elements). For later use, let $\mathbf{S}_0^*$ be the (sparsest) solution of (4) for the choice $f(\mathbf{S}, \mathbf{\lambda}) = \|\mathbf{S}\|_0$.

Constraints. The constraint $\mathbf{S} \in \mathcal{S}$ in (4) incorporates a priori knowledge about $\mathbf{S}$. If we let $\mathbf{S} = \mathbf{A}$ represent the adjacency matrix of an undirected graph with non-negative weights and no self-loops, we can explicitly write $\mathcal{S}$ as follows

$$\mathcal{S}_{\text{A}} := \{\mathbf{S} \,|\, S_{ij} \geq 0,\; \mathbf{S} \in \mathcal{M}^N,\; S_{ii} = 0,\; \textstyle\sum_j S_{j1} = 1\}. \tag{5}$$

The first condition in $\mathcal{S}_{\text{A}}$ encodes the non-negativity of the weights whereas the second condition incorporates the fact that the unknown graph is undirected, hence, $\mathbf{S}$ must belong to the set $\mathcal{M}^N$ of real and symmetric $N \times N$ matrices. The third condition encodes the absence of self-loops, thus, each diagonal entry of $\mathbf{S}$ must be null. Finally, the last condition fixes the scale of the admissible graphs by setting the weighted degree of the first node to 1, and also rules out the trivial solution $\mathbf{S} = \mathbf{0}$. Naturally, the identification of other GSOs can be of interest as well, including for instance $\mathbf{L}$, $\mathbf{L}_c$, and the random walk Laplacian [24]. These can be accommodated in our proposed framework via minor modifications to the set $\mathcal{S}$; see [25].

Independently of the criteria, the definition in (4) provides a formal description of a GSO that is considered to be the best possible description of the structure of the signal $\mathbf{x}$. Our goal is to find estimators of these operators as per the following problem statement.

Problem 1 *Given a set* $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^P$ *of* P *independent samples of the random signal* $\mathbf{x}$*, estimate the optimal description of the structure of* $\mathbf{x}$ *in the form of the graph-shift operator* $\mathbf{S}^*$ *defined in* (4)*.*

Problem 1 is a simple convex optimization problem if the objective in (4) is convex, but necessitates relaxations for the minimum zero-norm formulation. To solve Problem 1 we first use independent samples of the random signal to estimate the covariance eigenvectors. Then we estimate the eigenvalues using reformulations of (4) robust to errors stemming from the eigenvector estimation step.

3. ROBUST RECOVERY FROM IMPERFECT TEMPLATES

Whenever the number of observed graph signals in $\mathcal{X}$ is limited or the observations are corrupted by noise, assuming perfect knowledge of the spectral templates $\mathbf{V}$ in (4) may be unrealistic. To overcome this practical hurdle, this section deals with robust network inference problems from imperfect (noisy or incomplete) spectral templates.

3.1. Noisy spectral templates

We first address the case where knowledge of an approximate version of the spectral templates $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N]$ is available, e.g., from the eigenvectors of a *sample* covariance matrix $\hat{\mathbf{C}}_x$. The question here is how to update the general formulation in (4) to account for the discrepancies between the estimated spectral templates $\hat{\mathbf{V}}$ and the actual eigenvectors of $\mathbf{S}$. A tractable alternative is to form $\mathbf{S}' := \sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T$ and search for a shift $\mathbf{S}$ that possesses the desired properties while being close to $\mathbf{S}'$. Formally, one can solve

$$\hat{\mathbf{S}}^* := \underset{\{\mathbf{S},\boldsymbol{\lambda},\mathbf{S}'\}}{\operatorname{argmin}} \; f(\mathbf{S},\boldsymbol{\lambda}) \qquad (6)$$
$$\text{s. to } \; \mathbf{S}' = \textstyle\sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T, \;\; \mathbf{S} \in \mathcal{S}, \;\; d(\mathbf{S},\mathbf{S}') \leq \epsilon,$$

where $d(\cdot,\cdot)$ is a *convex* matrix distance. For instance, if $\|\mathbf{S}-\mathbf{S}'\|_{\mathrm{F}}$ is chosen, the focus is on the similarities across the entries of the shifts, while $\|\mathbf{S}-\mathbf{S}'\|_2$ focuses on their spectrum. The value of ϵ must be chosen based on a priori information on the imperfections, such as the number of signals used to form $\hat{\mathbf{C}}_x$, or the statistics of the noise. Note that (6) is also relevant to setups where the templates $\hat{\mathbf{V}}$ are not necessarily noisy but the goal is to enlarge the set of feasible GSOs. This can be of interest if, e.g., finding an $\mathbf{S}$ that is both sparse and with the exact templates collected in $\hat{\mathbf{V}}$ is impossible.

The difficulty in solving (6) is determined by f and $\mathcal{S}$. For the particular case of sparse shifts, the ℓ_1 norm relaxation of (6) yields

$$\hat{\mathbf{S}}_1^* := \underset{\{\mathbf{S},\boldsymbol{\lambda},\mathbf{S}'\}}{\operatorname{argmin}} \; \|\mathbf{S}\|_1 \qquad (7)$$
$$\text{s. to } \; \mathbf{S}' = \textstyle\sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T, \;\; \mathbf{S} \in \mathcal{S}, \;\; d(\mathbf{S},\mathbf{S}') \leq \epsilon,$$

where iteratively re-weighted schemes are also possible [25]. Note that $\|\mathbf{S}\|_1$ in (7) refers to the ℓ_1 norm of the vectorized version of $\mathbf{S}$. Moreover, further uncertainties can be introduced in the definition of the feasible set $\mathcal{S}$, e.g. in the scale of the admissible graphs for the case of $\mathcal{S} = \mathcal{S}_{\mathrm{A}}$ (cf. Proposition 1 and [25] for additional details).

To assess the effect of the noise in recovering the sparsest $\mathbf{S}$, some notation must be introduced. Define $\hat{\mathbf{W}} := \hat{\mathbf{V}} \odot \hat{\mathbf{V}} \in \mathbb{R}^{N^2 \times N}$, where $\odot$ denotes the Khatri-Rao product. Let $\mathbf{s}_0^* := \operatorname{vec}(\mathbf{S}_0^*)$, denote by $\mathcal{D}$ the diagonal indices such that $(\mathbf{s}_0^*)_{\mathcal{D}} = \operatorname{diag}(\mathbf{S}_0^*)$ and partition its complement $\mathcal{D}^c$ into $\mathcal{K}$ and $\mathcal{K}^c$, with the former indicating the positions of the nonzero entries of $\mathbf{s}_{0\mathcal{D}^c}^* := (\mathbf{s}_0^*)_{\mathcal{D}^c}$, where matrix *calligraphic subscripts* select rows. Denoting by † the matrix pseudo-inverse, we define $\hat{\mathbf{M}} := (\mathbf{I} - \hat{\mathbf{W}}\hat{\mathbf{W}}^\dagger)_{\mathcal{D}^c} \in \mathbb{R}^{N^2-N \times N^2}$, i.e., the orthogonal projector onto the kernel of $\hat{\mathbf{W}}^T$ constrained to the off-diagonal elements in $\mathcal{D}^c$. With $\mathbf{e}_1$ denoting the first canonical basis vector, we construct

$$\hat{\mathbf{R}} := [\hat{\mathbf{M}}, \, \mathbf{e}_1 \otimes \mathbf{1}_{N-1}] \in \mathbb{R}^{N^2-N \times N^2+1}, \qquad (8)$$

by horizontally concatenating $\hat{\mathbf{M}}$ and a column vector of size $|\mathcal{D}^c|$ with ones in the first $N-1$ positions and zeros elsewhere. Further, we drop the non-negativity constraint in $\mathcal{S}_{\mathrm{A}}$ – to obtain $\tilde{\mathcal{S}}_{\mathrm{A}}$ – and incorporate the scale ambiguity by augmenting $d(\mathbf{S},\mathbf{S}')$ as $\tilde{d}(\mathbf{S},\mathbf{S}') = (d(\mathbf{S},\mathbf{S}')^2 + (\sum_j S_{j1} - 1)^2)^{1/2}$. With this notation, the following result on robust recovery of network topologies holds[1].

Proposition 1 *When $d(\mathbf{S},\mathbf{S}') = \|\mathbf{S}-\mathbf{S}'\|_{\mathrm{F}}$, and assuming that there exists at least one $\mathbf{S}'$ such that $\tilde{d}(\mathbf{S}_0^*,\mathbf{S}') \leq \epsilon$, the solution $\hat{\mathbf{s}}_1^* := \operatorname{vec}(\hat{\mathbf{S}}_1^*)$ to (7) for $\mathcal{S} = \tilde{\mathcal{S}}_{\mathrm{A}}$ with scale ambiguity satisfies*

$$\|\hat{\mathbf{s}}_1^* - \mathbf{s}_0^*\|_1 \leq C\epsilon, \quad \textit{with } \; C = 2C_1 + 2C_2C_3, \qquad (9)$$

[1] Due to space constraints, the proofs can be found in an online appendix at http://www.seas.upenn.edu/~ssegarra/wiki/uploads/Research/RobTopID.pdf

if the two following conditions are satisfied:
1) $\operatorname{rank}(\hat{\mathbf{R}}_{\mathcal{K}}) = |\mathcal{K}|$*; and*
2) There exists a constant $\delta > 0$ such that

$$\psi := \|\mathbf{I}_{\mathcal{K}^c}(\delta^{-2}\hat{\mathbf{R}}\hat{\mathbf{R}}^T + \mathbf{I}_{\mathcal{K}^c}^T\mathbf{I}_{\mathcal{K}^c})^{-1}\mathbf{I}_{\mathcal{K}}^T\|_\infty < 1. \qquad (10)$$

Constants C_1, C_2, and C_3 are given by

$$C_1 = \frac{\sqrt{|\mathcal{K}|}}{\sigma_{\min}(\hat{\mathbf{R}}_{\mathcal{K}}^T)}, \;\; C_2 = \frac{1 + \|\hat{\mathbf{R}}^T\|_2 C_1}{1-\psi}, \;\; C_3 = \|\hat{\mathbf{R}}^\dagger\|_2 N, \qquad (11)$$

where $\sigma_{\min}(\cdot)$ denotes the minimum singular value.

When given noisy versions $\hat{\mathbf{V}}$ of the spectral templates of our target GSO, Proposition 1 quantifies the effect that the noise has on the recovery. More precisely, the recovered shift is guaranteed to be at a maximum distance from the desired shift bounded by the tolerance ϵ times a constant, which depends on $\hat{\mathbf{R}}$ and the support $\mathcal{K}$. This also implies that as the number of observed signals increases we recover the true GSO. In particular, as the number of observed signals increases, the sample covariance $\hat{\mathbf{C}}_x$ tends to the covariance $\mathbf{C}_x$ and, for the cases where the latter has no repeated eigenvalues, the noisy eigenvectors $\hat{\mathbf{V}}$ tend to the eigenvectors $\mathbf{V}$ of the desired shift; see, e.g., [26, Th. 3.3.7]. In particular, with better estimates $\hat{\mathbf{V}}$ the tolerance ϵ in (7) needed to guarantee feasibility can be made smaller, entailing a smaller discrepancy between the recovered $\mathbf{S}_1^*$ and the sparsest shift $\mathbf{S}_0^*$. In the limit when $\hat{\mathbf{V}} = \mathbf{V}$ and under no additional uncertainties, the tolerance ϵ can be made zero and (9) guarantees perfect recovery under conditions *1)* and *2)*.

3.2. Incomplete spectral templates

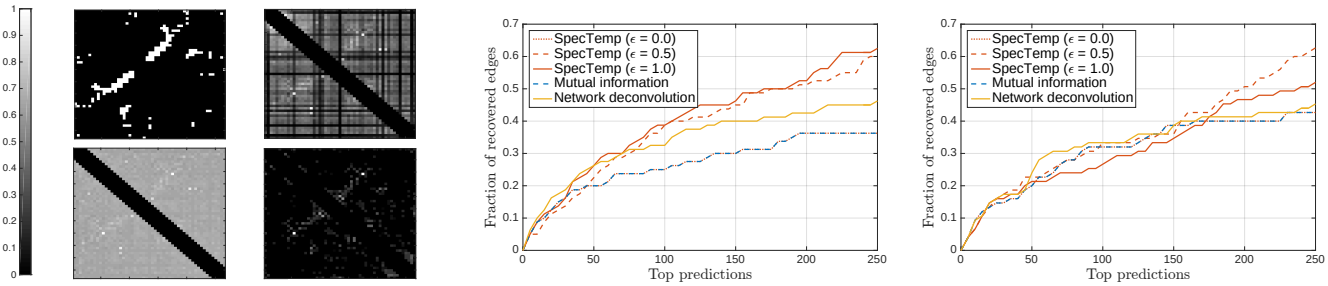
Thus far we have assumed that (an estimate of) the entire set of eigenvectors $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ is known. However, there are scenarios where only some of the eigenvectors (say K out of N) are available. This would be the case when e.g., $\mathbf{V}$ is found as the eigenbasis of $\mathbf{C}_x$ and the given signal ensemble is bandlimited. More generally, if $\mathbf{C}_x$ contains repeated eigenvalues there is a rotation ambiguity in the definition of the associated eigenvectors. Hence, in this case, we keep the eigenvectors that can be unambiguously characterized and, for the remaining ones, we include the rotation ambiguity as an additional constraint in our optimization problem.

Formally, assume that the K first eigenvectors $\mathbf{V}_K = [\mathbf{v}_1, ..., \mathbf{v}_K]$ are those which are known. Then, the network topology inference problem with incomplete spectral templates can be formulated as

$$\bar{\mathbf{S}}_1^* := \underset{\{\mathbf{S},\mathbf{S}_{\bar{K}},\boldsymbol{\lambda}\}}{\operatorname{argmin}} \; \|\mathbf{S}\|_1 \qquad (12)$$
$$\text{s. to } \; \mathbf{S} = \mathbf{S}_{\bar{K}} + \textstyle\sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{S}_{\bar{K}}\mathbf{V}_K = \mathbf{0},$$

where we already particularized the objective to the ℓ_1 convex relaxation. The formulation in (12) enforces $\mathbf{S}$ to be partially diagonalized by the known spectral templates $\mathbf{V}_K$, while its remaining component $\mathbf{S}_{\bar{K}}$ is forced to belong to the orthogonal complement of range$(\mathbf{V}_K)$. Notice that, as a consequence, the rank of $\mathbf{S}_{\bar{K}}$ is at most $N-K$. An advantage of using only partial information of the eigenbasis as opposed to the whole $\mathbf{V}$ is that the set of feasible solutions in (12) is larger than that in (4) when $f(\mathbf{S},\boldsymbol{\lambda}) = \|\mathbf{S}\|_1$. This is particularly important when the templates do not come from a prescribed GSO but, rather, one has the freedom to choose $\mathbf{S}$ provided it satisfies certain spectral properties (see [4] for examples in the context of distributed estimation).

GSO recovery guarantees can be derived for (12). To formally state these, define $\mathbf{W}_K := \mathbf{V}_K \odot \mathbf{V}_K$ and $\boldsymbol{\Upsilon} := [\mathbf{I}_{N^2}, \mathbf{0}_{N^2 \times N^2}]$. Also, define matrices $\mathbf{B}^{(i,j)} \in \mathbb{R}^{N \times N}$ for $i < j$ such that $B_{ij}^{(i,j)} =$



(a) (b) (c)

Fig. 1: (a) Real and inferred contact networks between amino-acid residues for protein BPT1 BOVIN. Ground truth contact network (top left), mutual information of the co-variation of amino-acid residues (top right), contact network inferred by network deconvolution (bottom left), contact network inferred by our method based on spectral templates (bottom right). (b) Fraction of the real contact edges between amino-acids recovered for each method as a function of the number of edges considered. (c) Counterpart of (b) for protein YES HUMAN.

1, $B_{ji}^{(i,j)} = -1$, and all other entries are zero. Based on this, we denote by $\mathbf{B} \in \mathbb{R}^{\binom{N}{2} \times N^2}$ a matrix whose rows are the vectorized forms of $\mathbf{B}^{(i,j)}$ for all $i, j \in \{1, 2, \ldots, N\}$ where $i < j$. In this way, $\mathbf{Bs} = \mathbf{0}$ when $\mathbf{s}$ is the vectorized form of a symmetric matrix. Further, we define the following matrices

$$\mathbf{P}_1 := \begin{bmatrix} \mathbf{I} - \mathbf{W}_K \mathbf{W}_K^\dagger \\ \mathbf{I}_{\mathcal{D}} \\ \mathbf{B} \\ \mathbf{0}_{NK \times N^2} \\ (\mathbf{e}_1 \otimes \mathbf{1}_N)^T \end{bmatrix}^T, \quad \mathbf{P}_2 := \begin{bmatrix} \mathbf{W}_K \mathbf{W}_K^\dagger - \mathbf{I} \\ \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{\binom{N}{2} \times N^2} \\ \mathbf{I} \otimes V_K^T \\ \mathbf{0}_{1 \times N^2} \end{bmatrix}^T, \quad (13)$$

and $\mathbf{P} := [\mathbf{P}_1^T, \mathbf{P}_2^T]^T$. With this notation and denoting by $\mathcal{J}$ the indices of the support of $\mathbf{s}_0^* = \text{vec}(\mathbf{S}_0^*)$, the following result holds.

Theorem 1 *Whenever $\mathcal{S} = \mathcal{S}_{\text{A}}$ and assuming problem* (12) *is feasible, $\bar{\mathbf{S}}_1^* = \mathbf{S}_0^*$ if the two following conditions are satisfied:*
1) $\text{rank}([\mathbf{P}_{1\mathcal{J}}^T, \mathbf{P}_2^T]) = |\mathcal{J}| + N^2$*; and*
2) There exists a constant $\delta > 0$ such that

$$\eta := \|\mathbf{\Upsilon}_{\mathcal{J}^c}(\delta^{-2}\mathbf{P}\mathbf{P}^T + \mathbf{\Upsilon}_{\mathcal{J}^c}^T \mathbf{\Upsilon}_{\mathcal{J}^c})^{-1} \mathbf{\Upsilon}_{\mathcal{J}}^T\|_\infty < 1. \quad (14)$$

The theorem provides sufficient conditions for the relaxed problem in (12) to recover the sparsest graph, even when not all the eigenvectors are known. In practice it is observed that for smaller number K of known spectral templates the value of η in (14) tends to be larger, indicating a less favorable setting for recovery.

Notice that scenarios that combine the settings in Secs. 3.1 and 3.2, i.e., where the knowledge of the K templates is imperfect, can be handled by combining the formulations in (7) and (12). This can be achieved upon implementing the following modifications to (12): considering the shift $\mathbf{S}'$ as a new optimization variable, replacing the first constraint in (12) with $\mathbf{S}' = \mathbf{S}_{\bar{K}} + \sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^T$, and adding $d(\mathbf{S}, \mathbf{S}') \leq \epsilon$ as a new constraint [cf. (7)].

Laplacian shifts. Counterparts to the optimizations in (7) and (12) as well as for the recovery guarantees in Proposition 1 and Theorem 1 can be derived for the case of (normalized) Laplacian operators. This requires changing the definition of $\mathcal{S}$ and accounting for the fact that the Laplacian has a zero eigenvalue; see [25] for details.

4. NETWORK DECONVOLUTION

The network deconvolution problem is the identification of an adjacency matrix $\mathbf{S}$ that encodes direct dependencies when given an adjacency $\mathbf{T}$ that includes indirect relationships. The problem is a generalization of channel deconvolution and can be solved by making $\mathbf{T} = \mathbf{S}(\mathbf{I} - \mathbf{S})^{-1}$ [27]. This solution assumes a diffusion as in (1) but for the particular case of a single-pole-single-zero graph filter. A more general approach is to assume that $\mathbf{T}$ can be written as a polynomial of $\mathbf{S}$ but be agnostic to the form of the filter. This leads to problem formulation (4) with $\mathbf{V}$ given by the eigenvectors of $\mathbf{T}$.

In this context, our goal is to identify the structural properties of proteins from a mutual information graph of the co-variation between the constitutional amino-acids [23]; see [27] for details. For example, for a particular protein, we want to recover the structural graph in the top left of Fig. 1(a) when given the graph of mutual information in the top right corner. Notice that the structural contacts along the first four sub-diagonals of the graphs were intentionally removed to assess the capability of the methods in detecting the contacts between distant amino-acids. The graph recovered by network deconvolution [27] is illustrated in the bottom left corner of Fig. 1(a) whereas the one recovered using SpecTemp [the proposed approach in (7)] is depicted in the bottom right corner. Comparing both recovered graphs, SpecTemp leads to a sparser graph that follows more closely the desired structure. To quantify this, in Fig. 1(b) we plot the fraction of the real contact edges recovered for each method as a function of the number of edges considered, as done in [27]. E.g., if for a given method the 100 edges with largest weight in the recovered graph contain 40% of the edges in the ground truth graph we say that the 100 top edge predictions achieve a fraction of recovered edges of 0.4. As claimed in [27], network deconvolution improves the estimation when compared to raw mutual information data. Nevertheless, from Fig. 1(b) it follows that SpecTemp outperforms network deconvolution. Notice that when $\epsilon = 0$ [cf. (7)] we are forcing the eigenvectors of $\mathbf{S}$ to coincide exactly with those of the matrix of mutual information $\mathbf{S}'$. However, since $\mathbf{S}'$ is already a valid adjacency matrix, we recover $\mathbf{S} = \mathbf{S}'$. By contrast, for larger values of ϵ the additional flexibility in the choice of the eigenvectors allows us to recover shifts $\mathbf{S}$ that more closely resemble the ground truth. For example, when considering the top 200 edges, the mutual information and the network deconvolution methods recover 36% and 43% of the desired edges, respectively, while our method for $\epsilon = 1$ achieves a recovery of 53%. In Fig. 1(c) we present this same analysis for a different protein and similar results can be appreciated.

5. CONCLUSIONS

The problem of identifying a graph-shift operator $\mathbf{S}$ – encoding the topology of a graph $\mathcal{G}$ of interest – given its eigenbasis $\mathbf{V}$ was studied. We first formulated the problem when perfect knowledge of $\mathbf{V}$ is available. We then investigated *robust* recovery for the cases where we have access to: 1) a noisy version of $\mathbf{V}$; and 2) subset of the columns of $\mathbf{V}$. We formulated optimization problems to recover $\mathbf{S}$, presented convex relaxations, and derived theoretical results characterizing the recovery. Finally, we illustrated the performance of the proposed approach via the identification of amino-acid graphs.

6. REFERENCES

[1] A. Sandryhaila and J.M.F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[2] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, Mar. 2013.

[3] A.G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *arXiv preprint arXiv:1603.04667*, 2016.

[4] S. Segarra, A.G. Marques, and A. Ribeiro, "Distributed linear network operators using graph filters," *arXiv preprint arXiv:1510.03947*, 2015.

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

[6] O. Sporns, *Discovering the Human Connectome*, MIT Press, Boston, MA, 2012.

[7] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[8] B.M. Lake and J.B. Tenenbaum, "Discovering structure by learning sparse graph," in *Annual Cognitive Sc. Conf.*, 2010, pp. 778 – 783.

[9] M. Slawski and M. Hein, "Estimation of positive definite M-matrices and structure learning for attractive gaussian markov random fields," *Linear Algebra and its Applications*, vol. 473, pp. 145–179, 2015.

[10] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Stat.*, vol. 34, pp. 1436–1462, 2006.

[11] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Shanghai, China, Mar. 20-25, 2016.

[12] X. Cai, J.A. Bazerque, and G.B. Giannakis, "Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations," *PLoS, Computational Biology*, June 2013.

[13] B. Baingana, G. Mateos, and G.B. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 563–575, Aug. 2014.

[14] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S.L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality," *PNAS*, vol. 101, pp. 98499854, 2004.

[15] G.V. Karanikolas, G.B. Giannakis, K. Slavakis, and R.M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Shanghai, China, Mar. 20-25, 2016.

[16] Y. Shen, B. Baingana, and G.B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *arXiv preprint arXiv:1605.03122*, 2016.

[17] J. Mei and J.M.F. Moura, "Signal processing on graphs: Estimating the structure of a graph," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, pp. 5495-5499, 2015.

[18] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *arXiv preprint arXiv:1406.7842v2*, 2015.

[19] V. Kalofolias, "How to learn a graph from smooth signals," in *Intl. Conf. Artif. Intel. Stat. (AISTATS)*. J. Mach. Learn. Res., pp. 920-929, 2016.

[20] S. Segarra, A.G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *IEEE Wrkshp. Statistical Signal Process. (SSP)*, Palma de Mallorca, Spain, Jun. 26-29, 2016.

[21] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and Michael G. Rabbat, "Characterization and inference of weighted graph topologies from observations of diffused signals," *arXiv preprint arXiv:1605.02569*, 2016.

[22] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *arXiv preprint arXiv:1601.02522*, 2016.

[23] D.S. Marks, L.J. Colwell, R. Sheridan, T.A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, "Protein 3D structure computed from evolutionary sequence variation," *PLoS ONE*, vol. 6, no. 12, pp. e28766, 2011.

[24] F. Chung, *Spectral Graph Theory*, vol. 92, American Mathematical Soc., 1997.

[25] S. Segarra, A.G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *arXiv preprint arXiv:1608.03008*, 2016.

[26] J.M. Ortega, *Numerical Analysis: A Second Course*, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1990.

[27] S. Feizi, D. Marbach, M. Medard, and M. Kellis, "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nat Biotech*, vol. 31, no. 8, pp. 726–733, 2013.