

Network Topology Inference from Spectral Templates

Santiago Segarra, *Member, IEEE*, Antonio G. Marques, *Senior Member, IEEE*, Gonzalo Mateos, *Senior Member, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

Abstract—We address the problem of identifying the structure of an *undirected* graph 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. The fresh look advocated here leverages concepts from convex optimization and stationarity of graph signals, in order to identify the graph shift operator (a matrix representation of the graph) given only its *eigenvectors*. These *spectral templates* can be obtained, e.g., from the sample covariance of independent graph signals diffused on the sought network. The novel idea is to find a graph shift that, while being consistent with the provided spectral information, endows the network with certain desired properties such as sparsity. To that end, we develop efficient inference algorithms stemming from provably tight convex relaxations of natural nonconvex criteria, particularizing the results for two shifts: the adjacency matrix and the normalized Laplacian. Algorithms and theoretical recovery conditions are developed not only when the templates are perfectly known, but also when the eigenvectors are noisy or when only a subset of them are given. Numerical tests showcase the effectiveness of the proposed algorithms in recovering synthetic and real-world networks.

Index Terms—Graph signal processing, graph sparsification, network deconvolution, network topology inference.

I. INTRODUCTION

ADVANCING a holistic theory of networks necessitates fundamental breakthroughs in modeling, identification, and controllability of distributed network processes – often conceptualized as *signals defined on the vertices of a graph* [3], [4]. Under the assumption that the signal properties are related to the topology of the graph where they are supported, the goal

Manuscript received July 17, 2017; accepted July 19, 2017. Date of publication July 24, 2017; date of current version August 14, 2017. This work was supported in part by Spanish MINECO under Grant TEC2013-41604-R and in part by USA NSF CCF-1217963. Part of the results in this paper appeared in [1] and [2]. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega. (*Corresponding author: Santiago Segarra.*)

S. Segarra is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: segarra@mit.edu).

A. G. Marques is with the Department of Signal Theory and Communications, King Juan Carlos University, Madrid 28933, Spain (e-mail: antonio.garcia.marques@urjc.es).

G. Mateos is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: gmateos@ece.rochester.edu).

A. Ribeiro is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: aribeiro@seas.upenn.edu).

Digital Object Identifier 10.1109/TSIPN.2017.2731051

of graph signal processing (GSP) is to develop algorithms that fruitfully leverage this relational structure [5], [6]. Instrumental to that end is the so-termed graph-shift operator (GSO) [6], a matrix capturing the graph's local topology and whose eigenbasis is central to defining graph Fourier transforms [7]. Most GSP works assume that the GSO (hence the graph) is known, and then analyze how the algebraic and spectral characteristics of the GSO impact the properties of the signals and filters defined on such a graph. Here instead we take the reverse path and investigate how to use information available from graph signals to infer the underlying graph topology; see also [1], [8]–[12].

Our focus in this paper is on identifying undirected 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 which encodes direct relationships between the elements of the signal from observable indirect relationships generated by a diffusion process. Such a problem is shown to be underdetermined and related to the concept of stationarity of graph signals [13]–[15]. More precisely, it is established that the sought GSO must have the same eigenvectors as the stationary signal's covariance matrix. This motivates a 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 (possibly imperfect and incomplete) *spectral templates* to recover the GSO by estimating its eigenvalues.

Network topology inference from a set of (graph-signal) observations is a prominent problem in Network Science [4], [16]. 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 [4, Ch. 7.3.1]. This approach is not without merit and widely used in practice, but it exhibits several drawbacks, the main one being that links are formed taking into account only pairwise interactions, ignoring that the observed correlations can be due to latent network effects. Acknowledging these limitations, alternative methods rely on partial correlations [4], [17], Gaussian graphical models [18]–[21], structural equation models [22], [23], Granger causality [16], [24], or their nonlinear (kernelized) variants [25], [26]. Differently, recent 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 [9] 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 [8] to infer graph Laplacians, seeking that input graph signals are smooth over the learned topologies; see also [11], [26] and a recent variant where signals are assumed to be spanned by few atoms from a graph dictionary consisting of heat diffusion kernels [27]. Different from [8], [9], [12], [27] 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 and addressed the problem of identifying a GSO based on its eigenvectors. One is [1], which assumes perfect knowledge of the spectral templates. The other is [10], which only focuses on a sparse Laplacian GSO.

After surveying the required GSP background, in Section II we formulate the problem of identifying a GSO that explains the fundamental structure of a random signal diffused on a graph. The novel idea is to search among all feasible networks for the one that endows the resulting graph-signal transforms with prescribed spectral properties (those guaranteeing graph stationarity [13]–[15]), while the inferred graph also exhibits desirable structural characteristics such as sparsity or minimum-energy edge weights. It is argued that the required spectral templates can be pragmatically obtained, e.g., via principal component analysis (PCA) of a set of graph signals resulting from network diffusion dynamics. Additional sources for the spectral templates are provided in Section II-C. Using the templates as input, a fairly general optimization problem is formulated in Section II-A to identify the undirected network structure. For concreteness, emphasis is laid on the recovery of two particular GSOs; namely the adjacency matrix and the normalized graph Laplacian, but our methodology can be applied to other matrix representations of graphs. In Section III-A we derive conditions under which the feasible set of the optimization problem reduces to a singleton, a situation in which pursuit of additional network structure is rendered vacuous. When multiple solutions exist, provably-tight convex relaxations – leading to computationally-efficient algorithms – are developed to identify the sparsest GSO consistent with the given eigenspace (Section III-B). While arguably less useful in practice, the idealized setting in Section III has important conceptual value towards formulating and understanding the pragmatic case where knowledge of the spectral templates is imperfect (Section IV). In Section IV-A we establish that the proposed algorithm can identify the underlying network topology robustly. Last but not least, in Section IV-B we investigate the case where only a subset of the GSO’s eigenvectors are known. Such incomplete spectral templates arise, for example, when the observed graph signals are bandlimited. Comprehensive numerical tests corroborate our theoretical findings and confirm that the novel approach compares favorably with respect to: (i) established methods based on (partial) correlations; and (ii) recent graph signal processing-based topology inference algorithms (Section V). Test cases include the recovery of social and structural brain networks from synthetically-generated signals, as well as the identification of

the structural properties of proteins [28] and the most relevant collaborations in two co-authorship networks [29].

Notation: The entries of a matrix \mathbf{X} and a (column) vector \mathbf{x} are denoted by X_{ij} and x_i , respectively. Sets are represented by calligraphic capital letters and $\mathbf{X}_{\mathcal{I}}$ denotes a submatrix of \mathbf{X} formed by selecting the rows of \mathbf{X} indexed by \mathcal{I} . The notation T and \dagger stands for transpose and pseudo-inverse, respectively; $\mathbf{0}$ and $\mathbf{1}$ refer to the all-zero and all-one vectors. For a vector \mathbf{x} , $\text{diag}(\mathbf{x})$ is a diagonal matrix whose i th diagonal entry is x_i ; when applied to a matrix, $\text{diag}(\mathbf{X})$ is a vector with the diagonal elements of \mathbf{X} . The operators \circ , \otimes , and \odot stand for the Hadamard (elementwise), Kronecker, and Khatri-Rao (column-wise Kronecker) matrix products. $\|\mathbf{X}\|_p$ denotes the ℓ_p norm of the *vectorized* form of \mathbf{X} , whereas $\|\mathbf{X}\|_{M(p)}$ is the matrix norm induced by the vector ℓ_p norm. $\ker(\mathbf{X})$ and $\text{Im}(\mathbf{X})$ refer to the null space and the span of the columns of \mathbf{X} , respectively.

II. PROBLEM STATEMENT

A weighted and undirected graph \mathcal{G} consists of a node set \mathcal{N} of 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 $\mathbf{D} := \text{diag}(\mathbf{A}\mathbf{1})$. 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 GSO $\mathbf{S} \in \mathbb{R}^{N \times N}$ as any matrix whose off-diagonal sparsity pattern is equal to that of the adjacency matrix of \mathcal{G} [6]. 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} .

The main focus in this paper is on identifying graphs that explain the structure of a random signal. Formally, let $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathbb{R}^N$ be a graph signal in which the i th element x_i denotes the signal value at node i of an unknown graph \mathcal{G} with symmetric shift operator \mathbf{S} . Further suppose that we are given a zero-mean white signal \mathbf{w} with covariance matrix $\mathbb{E}[\mathbf{w}\mathbf{w}^T] = \mathbf{I}$. We say that \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

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

While \mathbf{S} encodes only one-hop interactions, each successive application of the shift percolates (correlates) the original information across an iteratively increasing neighborhood; see e.g. [30]. 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, undirected graph can be written in the form in (1). These include processes generated by graph filters with time-varying coefficients or those generated by the so-called *diffusion Laplacian kernels* [31], to name a few.

The justification to say that \mathbf{S} is the structure of \mathbf{x} is that we can think of the edges of \mathbf{S} as direct (one-hop) relationships between the elements of the signal. The diffusion described by (1) generates indirect relationships. One of our goals is to

recover the fundamental relationships described by \mathbf{S} from a set $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^R$ of R independent samples of the random signal \mathbf{x} .

We show next that this is an underdetermined problem. Since we focus on the inference of undirected graphs, the shift operator \mathbf{S} is symmetric and diagonalizable. Hence, upon defining the orthogonal eigenvector matrix $\mathbf{V} := [\mathbf{v}_1, \dots, \mathbf{v}_N]$ and the eigenvalue matrix $\mathbf{\Lambda} := \text{diag}(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda} := [\lambda_1, \dots, \lambda_N]^T$, it holds that

$$\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^T. \quad (2)$$

Further observe that while the diffusion expressions in (1) are polynomials on the GSO of possibly infinite degree, the Cayley-Hamilton theorem implies that they are equivalent to polynomials of degree smaller than N . Upon defining the vector of coefficients $\mathbf{h} := [h_0, \dots, h_{L-1}]^T$ and the graph filter $\mathbf{H} \in \mathbb{R}^{N \times N}$ as $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l$, the generative model in (1) can be rewritten as

$$\mathbf{x} = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \right) \mathbf{w} = \mathbf{H}\mathbf{w} \quad (3)$$

for some particular \mathbf{h} and L . Since a graph filter \mathbf{H} is a polynomial on \mathbf{S} [6], 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 (3) 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, substitute (3) into the covariance matrix definition and use the fact that $\mathbb{E}[\mathbf{w}\mathbf{w}^T] = \mathbf{I}$ to write

$$\mathbf{C}_x = \mathbb{E} \left[\mathbf{H}\mathbf{w}(\mathbf{H}\mathbf{w})^T \right] = \mathbf{H}\mathbb{E}[\mathbf{w}\mathbf{w}^T]\mathbf{H}^T = \mathbf{H}\mathbf{H}^T. \quad (4)$$

If we further use the spectral decomposition of the shift in (2) to express the filter as $\mathbf{H} = \sum_{l=0}^{L-1} h_l (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l \boldsymbol{\Lambda}^l) \mathbf{V}^T$, we can write the covariance matrix as

$$\mathbf{C}_x = \mathbf{V} \left(\sum_{l=0}^{L-1} h_l \boldsymbol{\Lambda}^l \right)^2 \mathbf{V}^T. \quad (5)$$

A consequence of (5) 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 in (5), which includes indirect relationships between components, and \mathbf{S} in (2), which includes 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 (5) 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.¹ 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

¹To simplify exposition, the general description of the recovery problem in this section assumes that neither \mathbf{S} nor \mathbf{C}_x have repeated eigenvalues. Technical modifications in the formulation to accommodate setups where the eigenvalues are not all distinct are discussed in Section IV-B.

ambiguity, which amounts to selecting the eigenvalues of the shift, we assume that the GSO of interest is optimal in some sense. This is the subject of Section II-A, but before a remark is in order.

Remark 1 (Graph stationarity): Recently, a group of works has generalized the definition of stationarity to graph processes [13]–[15]. Such a generalization is not trivial because graph signals need not have a time-varying interpretation, and also due to the irregularity of the underlying graph domain. The aforementioned works consider that a graph signal is stationary in a particular GSO \mathbf{S} if either the signal can be expressed as the output of a graph filter with white inputs [13, Def. 2], or if its covariance matrix is simultaneously diagonalizable with \mathbf{S} [13, Def. 3]. These are precisely the conditions in (3) and (5), respectively. Hence, our problem of identifying a GSO that explains the fundamental structure of \mathbf{x} is equivalent to the problem of identifying a shift \mathbf{S} in which the signal \mathbf{x} is stationary.

A. Optimal GSO

Many large-scale, real-world networks are sparse [4], so it is often meaningful to infer a GSO where most of the entries in \mathbf{S} are zero. Let \mathcal{S} be a convex set that specifies the type of shift operator we want to identify (details on \mathcal{S} are provided in Section II-B) and let $\|\mathbf{S}\|_0$ count the number of nonzero entries in the GSO. We then want to identify $\mathbf{S}_0^* \in \mathcal{S}$ with the smallest number of nonzero entries (e.g., those corresponding to direct relationships among signal elements), namely

$$\begin{aligned} \mathbf{S}_0^* &:= \underset{\{\mathbf{S}, \boldsymbol{\lambda}\}}{\text{argmin}} \quad \|\mathbf{S}\|_0, \\ \text{s.t.} \quad \mathbf{S} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \quad \mathbf{S} \in \mathcal{S}. \end{aligned} \quad (6)$$

To simplify notation we have purposely ignored the optimal eigenvalues λ_0^* that belong to the argument of the minimum. 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 (6) are linear on the unknown eigenvalues λ_k .

The definition in (6) provides a formal description of a sparse GSO \mathbf{S} that is considered to be the best possible description of the structure of the signal \mathbf{x} . Our goal is to find estimators of this operator as described in the following two formal problem statements.

Problem 1: Given the eigenvectors \mathbf{V} , identify the optimal graph-shift operator $\mathbf{S}_0^* \in \mathcal{S}$ defined in (6).

Problem 2: Given a set $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^R$ of R 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 $\mathbf{S}_0^* \in \mathcal{S}$ defined in (6).

Finding the sparsest graph efficiently requires suitable relaxations for the non-convex ℓ_0 objective. To find the structure of \mathbf{x} when the covariance matrix \mathbf{C}_x is known, we solve Problem 1 by using the eigenvectors of \mathbf{C}_x in (6) and show that the estimation of the eigenvalues yields consistent estimators of sparse network structures. Problem 1 is addressed in Section III

and while it could be perceived as idealized and rather limited in scope, its conceptual value is key towards formulating and analyzing the more challenging and pragmatic Problem 2. Nevertheless, Section II-C illustrates additional settings in which the estimation of a GSO with (possibly approximated) prescribed eigenvalues is of practical interest. To solve Problem 2 we first use independent samples of the random signal to *estimate* the covariance eigenvectors. Then we estimate the eigenvalues using reformulations of (6) which are robust to errors stemming from the aforementioned eigenvector estimation step; see Section IV for a detailed treatment of Problem 2.

Remark 2 (Sparse precision matrices): The precision matrix \mathbf{C}_x^{-1} (alternatively, the pseudo-inverse \mathbf{C}_x^\dagger if the covariance is degenerate) is a possible solution to the problem of finding a GSO that explains the structure of \mathbf{x} . This establishes a clear connection between (6) and the problem of finding sparse estimates of precision matrices [4, Ch. 7]. For the specific cases in which the precision matrix \mathbf{C}_x^{-1} is the *sparsest* matrix that explains the structure of \mathbf{x} , this matrix is also the solution to (6) and we have that $\mathbf{S}_0^* = \mathbf{C}_x^{-1}$. In general, however, \mathbf{C}_x^{-1} may not be sparse and, even if it is, there may be *sparser* graphs that explain \mathbf{x} . A case of particular relevance is that of structural equation models where the observations are modeled as $\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{w}$ [26]. If the so-called exogenous input \mathbf{w} is white, it follows readily that $\mathbf{C}_x = (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{I} - \mathbf{A})^{-T}$. With this model, classical precision-based methods will identify $(\mathbf{I} - \mathbf{A}^T)(\mathbf{I} - \mathbf{A})$ as the underlying topology. Differently, (6) will generate the more parsimonious \mathbf{A} . We can then think of (6) as a general formulation that reduces to the problem of finding a sparse precision matrix for the cases when the random signal \mathbf{x} is indeed associated with a sparse \mathbf{C}_x^{-1} .

Remark 3 (Beyond sparse GSOs): Different from the minimum zero-norm formulation in (6), we can introduce alternative criteria in the form of generic convex functions $f(\mathbf{S}, \boldsymbol{\lambda})$ and define the shift operator that is optimal with respect to these criteria. Beyond $f(\mathbf{S}, \boldsymbol{\lambda}) = \|\mathbf{S}\|_0$ as in (6), other possible convex choices for the criterion are to: (i) Adopt $f(\mathbf{S}, \boldsymbol{\lambda}) = f(\mathbf{S}) = \|\mathbf{S}\|_F$ which finds a GSO that minimizes the total energy stored in the weights of the edges. (ii) Set $f(\mathbf{S}, \boldsymbol{\lambda}) = f(\mathbf{S}) = \|\mathbf{S}\|_\infty$ which yields shifts \mathbf{S} associated with graphs of uniformly low edge weights. This can be meaningful, e.g., when identifying graphs subject to capacity constraints. (iii) Minimize $f(\mathbf{S}, \boldsymbol{\lambda}) = f(\boldsymbol{\lambda}) = -\lambda_2$, where λ_2 is the second smallest eigenvalue of \mathbf{S} . If the GSO is further assumed to be a Laplacian matrix, then $\mathbf{I} - \mathbf{S}$ corresponds to a shift operator with fast mixing times [32].

B. A Priori Knowledge About the GSO

The constraint $\mathbf{S} \in \mathcal{S}$ in (6) 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

$$\mathcal{S}_A := \left\{ \mathbf{S} \mid S_{ij} \geq 0, \mathbf{S} \in \mathcal{M}^N, S_{ii} = 0, \sum_j S_{j1} = 1 \right\}. \quad (7)$$

The first condition in \mathcal{S}_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 choice of the first node is (almost) arbitrary; any node with at least one neighbor in the sought graph suffices. Although not considered here, additional sources of information such as knowing the existence (or not) of particular edges can be incorporated into \mathcal{S} as well.

Alternatively, when $\mathbf{S} = \mathbf{L}$ represents a normalized Laplacian [5], the associated \mathcal{S}_L is

$$\mathcal{S}_L := \left\{ \mathbf{S} \mid S_{ij} \in [-1, 0] \text{ for } i \neq j, \mathbf{S} \in \mathcal{M}_+^N, S_{ii} = 1 \text{ for all } i, \lambda_1 = 0, \lambda_i \leq 2 \right\}. \quad (8)$$

In \mathcal{S}_L we impose that \mathbf{S} is symmetric and positive semi-definite, its diagonal entries are 1, its off-diagonal entries are non-positive, and its eigenvalues are no larger than 2. Moreover, since \mathbf{S} is a normalized Laplacian we know that the vector $\sqrt{\mathbf{d}}$ containing as entries the square roots of the node degrees is an eigenvector whose associated eigenvalue is zero, and this is incorporated into the last constraint. Notice that for this last constraint to be implementable we should be able to identify $\sqrt{\mathbf{d}}$ among all the spectral templates in \mathbf{V} . This can be done since $\sqrt{\mathbf{d}}$ is the only eigenvector whose entries have all the same sign [33]. In the same way that fixing a scale discards the solution $\mathbf{S} = \mathbf{0}$ for adjacency matrices, the constraint $\lambda_1 = 0$ rules out the uninformative solution $\mathbf{S} = \mathbf{I}$ from the feasible set \mathcal{S}_L . For the cases where the underlying graph has D connected components, \mathcal{S}_L can be modified accordingly by forcing $\lambda_i = 0$ for $i = 1, \dots, D$. The corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_D$ can be identified from the eigenbasis of \mathbf{C}_x as those corresponding to a repeated eigenvalue and spanning $\sqrt{\mathbf{d}}$.

Naturally, the identification of other GSOs can be of interest as well, including for instance the combinatorial Laplacian \mathbf{L}_c and the random walk Laplacian [34]. These can be accommodated in our proposed framework via minor modifications to the set \mathcal{S} ; see e.g. Section V-C for an experiment on recovering \mathbf{L}_c . For concreteness, we henceforth focus exclusively on adjacency and normalized Laplacian matrices, and the theoretical guarantees presented refer to the specific recovery of these two GSOs.

C. Additional Sources for the Spectral Templates

The central focus of this paper is to solve (6) when eigenvectors \mathbf{v}_k are estimated from a sample set \mathcal{X} (cf. Problem 2). Notwithstanding, the network topology inference problem in (6) is applicable as long as eigenvectors or eigenvector estimates are available. Four settings where solving Problem 1 is of independent interest are outlined next.

GSO associated with orthogonal transformations: Expressing signals \mathbf{x} in an alternative domain $\tilde{\mathbf{x}}$ by using an orthonormal transform $\tilde{\mathbf{x}} := \mathbf{K}^T \mathbf{x}$, such as Fourier, wavelets, Karhunen-Loève, or discrete-cosine, is a cornerstone operation in signal processing. If we set $\mathbf{V} = \mathbf{K}$ in (6) we formulate the problem of identifying a graph shift $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = \mathbf{K}\boldsymbol{\Lambda}\mathbf{K}^T$ whose graph

Fourier transform [7] $\tilde{\mathbf{x}} := \mathbf{V}^T \mathbf{x} = \mathbf{K}^T \mathbf{x}$ is the given orthonormal transform of interest. This is important because it reveals the proximity structure between signal components that is implicitly assumed and exploited by the transform \mathbf{K} .

Design of graph filters: In addition to describing linear diffusion dynamics [cf. (3)], graph filters represent linear transformations that can be implemented in a distributed manner [35]–[37]. In the context of distributed algorithms, consider the approximation of a prescribed linear transformation represented by the square matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$ using a graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ [30]. A necessary condition for this approximation to be exact is that the eigenvectors of the shift $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$ and those of the transformation $\mathbf{B} = \mathbf{V}_B \mathbf{\Lambda}_B \mathbf{V}_B^T$ must coincide; [30, Prop. 1]. Since \mathbf{V}_B can be obtained from the prescribed \mathbf{B} , the optimization in (6) can be solved using $\mathbf{V} = \mathbf{V}_B$ as input. Solving Problem 1 using \mathbf{V}_B enables us to design the sparsest \mathbf{S} which facilitates implementation of a given transformation \mathbf{B} via distributed graph filtering.

Graph sparsification: Given a GSO \mathbf{T} , we can use our framework to obtain a sparse shift \mathbf{S} with the same eigenvectors as \mathbf{T} , and desirable properties encoded in \mathcal{S} . This *graph sparsification* problem can be addressed by solving (6) using as inputs the eigenvectors of \mathbf{T} . Note that, different from the setup where the goal is to explain the structure of \mathbf{x} (cf. Problem 2), the matrix \mathbf{T} is not necessarily a covariance matrix. Moreover, in general the input \mathbf{T} need not belong to \mathcal{S} . Thus, (6) is effectively promoting sparsity while ensuring that the output GSO adheres to the specifications in \mathcal{S} .

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 setting $\mathbf{S} = \mathbf{T}(\mathbf{I} + \mathbf{T})^{-1}$ [38]. This solution assumes a diffusion as in (1) that results in 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 1, with \mathbf{V} given by the eigenvectors of \mathbf{T} .

III. TOPOLOGY INFERENCE FROM SPECTRAL TEMPLATES

The focus of this section is on Problem 1, that is to find the sparsest graph shift \mathbf{S} that is diagonalized by given spectral templates $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$. The structure of the feasible set in (6) plays a critical role in solving this network inference problem. In fact, it can be shown that in a number of setups the feasible set reduces to a singleton, or otherwise to a low-dimensional subspace. This is important because even if the objective is non-convex as in (6), searching over a small space is not necessarily hard.

We first investigate the size of the feasible set and provide conditions under which it reduces to a singleton, thus rendering the objective function inconsequential to the optimization. For the cases where there are multiple feasible solutions, the optimization (6) is non-convex and in fact NP-hard [39]. We then

propose convex relaxations that can be solved in polynomial time and are provably tight under some technical conditions (see Theorems 1 and 2 for details).

A. Size of the Feasibility Set

The feasible set of problem (6) for both \mathcal{S}_L and \mathcal{S}_A is in general small. To be more precise, some notation must be introduced. Define $\mathbf{W} := \mathbf{V} \odot \mathbf{V} \in \mathbb{R}^{N^2 \times N^2}$, where \odot denotes the Khatri-Rao product. Notice that from the definition of \mathbf{S} we can write $\mathbf{s} := \text{vec}(\mathbf{S})$ as $\mathbf{s} = \mathbf{W} \boldsymbol{\lambda}$. Hence, each row of \mathbf{W} represents the N weighting coefficients that map $\boldsymbol{\lambda}$ to the corresponding entry of \mathbf{S} . Further, define the set \mathcal{D} containing the indices of \mathbf{s} corresponding to the diagonal entries of \mathbf{S} and select the corresponding rows of \mathbf{W} to form $\mathbf{W}_{\mathcal{D}} \in \mathbb{R}^{N \times N}$. Also, define the matrix $\mathbf{U} := \mathbf{V}^1 \circ \mathbf{V}^1 \in \mathbb{R}^{N \times N}$, where \circ denotes the elementwise product and $\mathbf{V}^1 := [\mathbf{1}, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_N]$. Using these conventions, the following result holds.

Proposition 1: Assume that (6) is feasible, then it holds that:

- a) If $\mathcal{S} = \mathcal{S}_A$, then $\text{rank}(\mathbf{W}_{\mathcal{D}}) \leq N - 1$. Similarly, if $\mathcal{S} = \mathcal{S}_L$, then $\text{rank}(\mathbf{U}) \leq N - 1$.
- b) If $\text{rank}(\mathbf{W}_{\mathcal{D}}) = N - 1$ when $\mathcal{S} = \mathcal{S}_A$ or $\text{rank}(\mathbf{U}) = N - 1$ when $\mathcal{S} = \mathcal{S}_L$, then the feasible set of (6) is a singleton.

Proof: We show statements a) and b) for the case $\mathcal{S} = \mathcal{S}_A$. The proofs for $\mathcal{S} = \mathcal{S}_L$ are analogous and thus omitted.

Note that we may write $\mathbf{W}_{\mathcal{D}} \boldsymbol{\lambda} = \text{diag}(\mathbf{S}) = \mathbf{0}$ for all feasible $\boldsymbol{\lambda}$. Hence, feasibility implies that $\mathbf{W}_{\mathcal{D}}$ is rank-deficient as stated in a). To show b), assume that $\text{rank}(\mathbf{W}_{\mathcal{D}}) = N - 1$ so that $\boldsymbol{\lambda}$ in $\ker(\mathbf{W}_{\mathcal{D}})$ is unique up to a scaling factor. However, since one of the conditions in \mathcal{S}_A forces the first row of \mathbf{S} to sum up to 1 [cf. (7)], this scaling ambiguity is resolved and the unique feasible $\boldsymbol{\lambda}$ (and hence \mathbf{S}) is obtained. ■

Proposition 1 offers sufficient conditions under which (6) reduces to a feasibility problem. More specifically, when condition b) is met, the objective in (6) is inconsequential since there exists only one feasible \mathbf{S} . In practice, we have observed that for random graphs with real-valued weights, condition b) is satisfied except for graphs with a very simple structure such as trees. Interestingly, this implies that most *random weighted* graphs can be uniquely determined by their eigenvectors. For these cases, the analysis of the solution to the convex relaxation of (6) carried out in the remaining of this section serves as a theoretical underpinning for the more practical findings in Section IV. On the other hand, when the random graphs are unweighted or the edges take values in a small finite set, condition b) is often-times not satisfied. For these cases, however, among all feasible GSOs the sparsest one can be (in theory) recovered by solving (6). The level of rank deficiency of $\mathbf{W}_{\mathcal{D}}$ and \mathbf{U} for the adjacency and normalized Laplacian cases, respectively, offers a practical indicator of the difficulty in recovering \mathbf{S} via convex relaxations of (6); see also Section III-B and Fig. 1(c). Consequently, the recovery of highly structured graphs such as grids, cycles, and other regular graphs – which in practice lead to low-rank $\mathbf{W}_{\mathcal{D}}$ and \mathbf{U} – tends to be more challenging than for random graphs.

Proposition 1 can also help in selecting the type of GSO (adjacency or Laplacian) that is more likely to explain the structure

of the signal \mathbf{x} . This is relevant for settings where the type of shift is unknown a priori, and we may exploit the structure of \mathbf{V} to make an informed decision on the GSO. For instance, at first one can look at the columns of \mathbf{V} and check whether one of them is a constant vector, which is a requirement for the shift to be a combinatorial Laplacian. One could also examine $\mathbf{W}_{\mathcal{D}}$ or \mathbf{U} . For example if $\mathbf{W}_{\mathcal{D}}$ has full rank, it follows from Proposition 1 that \mathbf{S} cannot be an adjacency matrix with zeros in the diagonal. Analogously, in the pragmatic setting of Section IV-A where the eigenvectors are estimated imperfectly, one can check how close the noisy versions of $\mathbf{W}_{\mathcal{D}}$ and \mathbf{U} are to being rank deficient and choose the GSO type accordingly.

B. Convex Relaxation

One usual approach to handle the non-convex ℓ_0 (pseudo) norm objective in (6) is to relax it to an iteratively re-weighted ℓ_1 norm. Specifically, with t denoting an iteration index, we aim to solve a sequence $t = 1, \dots, T$ of weighted ℓ_1 -norm minimization problems

$$\mathbf{S}_{\omega}^* := \underset{\{\mathbf{S}, \lambda\}}{\operatorname{argmin}} \sum_{i,j} \omega_{ij}(t) |S_{ij}| \quad \text{s.t. } \mathbf{S} = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \mathbf{S} \in \mathcal{S}, \quad (9)$$

with weights $\omega_{ij}(t) := \tau / (|S_{ij}(t-1)| + \delta_{\omega})$, for appropriately chosen positive constants τ and δ_{ω} . Intuitively, the goal of the re-weighted scheme in (9) is that if $|S_{ij}(t-1)|$ is small, in the next iteration the penalization $\omega_{ij}(t)$ is large, promoting further shrinkage of S_{ij} towards zero [40].

Naturally, under condition b) in Proposition 1 the solutions \mathbf{S}_0^* of (6) and \mathbf{S}_{ω}^* of (9) are guaranteed to coincide given that the feasible set is reduced to a singleton. Moreover, even when condition b) is not satisfied, there exist weights ω_{ij} that guarantee the equivalence of both solutions. To state this formally, define the set \mathcal{J} containing the indices identifying the support of \mathbf{S}_0^* and denote by \mathcal{J}^c its complement. Whenever \mathbf{S}_0^* is the unique solution to (6), it is not hard to establish that by setting weights in (9) as $\omega_{ij} = 1$ for $(i, j) \in \mathcal{J}^c$ and $\omega_{ij} = 0$ otherwise, then \mathbf{S}_{ω}^* is unique and equal to \mathbf{S}_0^* .

The upshot of this simple observation is that there exist optimal weights so that the sparsest solution \mathbf{S}_0^* can be recovered by solving a convex optimization problem. Being convex, this approach incurs substantially lower computational complexity when compared to the original NP-hard optimization. This result confers validity to the re-weighted formulation in (9), nonetheless, we can neither choose these weights without knowing \mathbf{S}_0^* a priori nor there is a guarantee that the succession of weights $\omega_{ij}(t)$ converges to these optimal weights. Hence, we now focus on the derivation of theoretical guarantees for a particular set of weights that can be set a priori, namely, we consider the formulation in which each entry of the GSO is equally weighted. This boils down to solving the convex optimization

$$\mathbf{S}_1^* := \underset{\{\mathbf{S}, \lambda\}}{\operatorname{argmin}} \|\mathbf{S}\|_1 \quad \text{s.t. } \mathbf{S} = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \mathbf{S} \in \mathcal{S}. \quad (10)$$

Interestingly, under certain conditions we can ensure that the solution \mathbf{S}_1^* to the relaxed problem (10) coincides with \mathbf{S}_0^* . To be more specific, define $\mathbf{s}_0^* := \operatorname{vec}(\mathbf{S}_0^*)$, denote by \mathcal{D}^c the complement of \mathcal{D} and partition \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 we recall that matrix *calligraphic subscripts* select rows. Recalling that \dagger denotes the matrix pseudo-inverse, we define

$$\mathbf{M} := (\mathbf{I} - \mathbf{W}\mathbf{W}^{\dagger})_{\mathcal{D}^c} \in \mathbb{R}^{N^2 - N \times N^2}, \quad (11)$$

i.e., the orthogonal projector onto the kernel of \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 the matrix

$$\mathbf{R} := [\mathbf{M}, \mathbf{e}_1 \otimes \mathbf{1}_{N-1}] \in \mathbb{R}^{N^2 - N \times N^2 + 1}, \quad (12)$$

by horizontally concatenating \mathbf{M} and a column vector of size $|\mathcal{D}^c|$ with ones in the first $N - 1$ positions and zeros elsewhere. With this notation in place, the following recovery result holds.

Theorem 1: If $\mathcal{S} = \mathcal{S}_A$ and (10) is feasible, then $\mathbf{S}_1^* = \mathbf{S}_0^*$ provided that the two following conditions are satisfied:

A-1) $\operatorname{rank}(\mathbf{R}_{\mathcal{K}}) = |\mathcal{K}|$; and

A-2) There exists a constant $\delta > 0$ such that

$$\psi_{\mathbf{R}} := \|\mathbf{I}_{\mathcal{K}^c} (\delta^{-2} \mathbf{R}\mathbf{R}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T\|_{M(\infty)} < 1. \quad (13)$$

Proof: See Appendix A. \blacksquare

Theorem 1 offers *sufficient* conditions under which the relaxation (10) is tight, namely that solving (10) incurs no loss of optimality and $\mathbf{S}_1^* = \mathbf{S}_0^*$. In other words, under A1)-A2) one can still recover the sparsest adjacency matrix defined in (6), by solving a convex optimization problem in polynomial time. Simulations in Section V reveal that the bound (13) imposed on $\psi_{\mathbf{R}}$ is tight by providing examples where $\psi_{\mathbf{R}}$ is equal to 1 and for which recovery fails. In Theorem 1, condition A-1) ensures that the solution to (10) is unique, a necessary requirement to guarantee sparse recovery. Condition A-2) is derived from the construction of a dual certificate specially designed to ensure that the unique solution to (10) also has minimum ℓ_0 norm [41].

Recall that the ℓ_{∞} norm in (13) is the maximum ℓ_1 norm across the rows of the argument matrix, which has $|\mathcal{K}^c|$ rows each containing $|\mathcal{K}|$ elements. It is thus expected that sparser graphs (small $|\mathcal{K}|$) might have smaller values of $\psi_{\mathbf{R}}$. Furthermore, to have an intuitive understanding of $\psi_{\mathbf{R}}$ it is helpful to see that condition A-2) is always satisfied whenever $\mathbf{R}\mathbf{R}^T$ is non-singular. More specifically, for small values of δ we have that $\psi_{\mathbf{R}} \approx \delta^2 \|\mathbf{I}_{\mathcal{K}^c} (\mathbf{R}\mathbf{R}^T)^{-1} \mathbf{I}_{\mathcal{K}}^T\|_{M(\infty)}$, which can be made arbitrarily small and, in particular, strictly smaller than 1. Matrix $\mathbf{R}\mathbf{R}^T$ can be shown to be invertible whenever $\operatorname{rank}(\mathbf{W}_{\mathcal{D}}) = N - 1$ (cf. Proposition 1). Thus, in the extreme case where the feasible set is a singleton, Theorem 1 guarantees recovery, as expected. A more general characterization of the classes of random graphs that tend to satisfy (13) with high probability is of interest, but left as future research.

The recovery result of Theorem 1 can be replicated for the case where the shift of interest is a normalized Laplacian, i.e., when $\mathcal{S} = \mathcal{S}_L$. To state this formally, if we define $\mathbf{Q} := (\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^{\dagger})_{\mathcal{D}^c}$, where $\tilde{\mathbf{U}} := \tilde{\mathbf{V}} \odot \tilde{\mathbf{V}}$ for $\tilde{\mathbf{V}} := [\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_N]$ the following result holds. .

Theorem 2: If $\mathcal{S} = \mathcal{S}_L$ and (10) is feasible, then $\mathbf{S}_1^* = \mathbf{S}_0^*$ provided that the two following conditions are satisfied:

L-1) $\text{rank}(\mathbf{Q}_{\mathcal{K}}) = |\mathcal{K}|$; and

L-2) There exists a constant $\delta > 0$ such that

$$\psi_{\mathbf{Q}} := \|\mathbf{I}_{\mathcal{K}^c} (\delta^{-2} \mathbf{Q} \mathbf{Q}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T\|_{M(\infty)} < 1. \quad (14)$$

Proof: The proof follows the same steps as those in Theorem 1 and, thus, is omitted. ■

IV. IMPERFECT SPECTRAL TEMPLATES

Whenever the number of observed graph signals is limited or the observations are noisy as in Problem 2, assuming perfect knowledge of the spectral templates \mathbf{V} may be unrealistic. This section broadens the scope of the network inference problems dealt with so far, to accommodate imperfect spectral templates that can either be noisy or incomplete. Specifically, we investigate pragmatic scenarios where: i) only an approximate version of \mathbf{V} can be obtained (e.g., from the eigenvectors of a *sample* covariance matrix); and ii) where only a subset of \mathbf{V} is available (e.g., when the observed signals are bandlimited and one can only estimate the non-zero frequencies that are present).

A. Noisy Spectral Templates

We first address the case where knowledge of an approximate version of the spectral templates $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_N]$ is available. The question here is how to update the general formulation in (6) to account for the discrepancies between the estimated spectral templates $\hat{\mathbf{V}}$ and the actual eigenvectors of \mathbf{S} . An instructive reformulation is to include $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ as decision variables and formulate the following problem

$$\begin{aligned} & \min_{\{\mathbf{S}, \lambda, \mathbf{V}\}} \|\mathbf{S}\|_0 \\ & \text{s.t. } \mathbf{S} = \sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \quad \mathbf{S} \in \mathcal{S}, \quad d(\mathbf{v}_k, \hat{\mathbf{v}}_k) \leq \epsilon_k \text{ for all } k, \end{aligned} \quad (15)$$

where $d(\cdot, \cdot)$ is a *convex* vector distance function, such as the ℓ_p norm of the vector difference for $p \geq 1$. The idea in (15) is to find a sparse \mathbf{S} that satisfies the desired properties in \mathcal{S} , while its eigenvectors \mathbf{v}_k are each of them close to the observed ones $\hat{\mathbf{v}}_k$. The value of ϵ_k must be chosen based on a priori information on the imperfections, such as the number of signals used to estimate the sample covariance, the magnitude of the eigenvalues (eigenvectors associated with small eigenvalues are harder to estimate [42, Ch. 3]), or the statistics of the observation noise. While conceptually simple, (15) is a formidable problem since both λ_k and \mathbf{v}_k are optimization variables, rendering the first constraint non-convex.

To bypass this issue, our approach 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

$$\begin{aligned} & \hat{\mathbf{S}}^* := \underset{\{\mathbf{S}, \lambda, \mathbf{S}'\}}{\text{argmin}} \|\mathbf{S}\|_0 \\ & \text{s.t. } \mathbf{S}' = \sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T, \quad \mathbf{S} \in \mathcal{S}, \quad d(\mathbf{S}, \mathbf{S}') \leq \epsilon, \end{aligned} \quad (16)$$

where $d(\cdot, \cdot)$ is a *convex* matrix distance whose form depends on the particular application. E.g., if $\|\mathbf{S} - \mathbf{S}'\|_F$ is chosen, the focus is more on the similarities across the entries of the shifts, while $\|\mathbf{S} - \mathbf{S}'\|_{M(2)}$ focuses on their spectrum. Additional conic constraints of the form $\|(\mathbf{S} - \mathbf{S}') \hat{\mathbf{v}}_k\|_2 \leq \lambda_k \epsilon_k$ enforcing that particular eigenvectors are well approximated can also be incorporated. From an application point of view, the formulation in (16) 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, for example, finding an \mathbf{S} that is both sparse and with the exact templates collected in $\hat{\mathbf{V}}$ is impossible (cf. Section II-C).

Solving (16) faces similar challenges to those in Section III. The ℓ_1 norm relaxation of (16) yields [cf. (10)]

$$\begin{aligned} & \hat{\mathbf{S}}_1^* := \underset{\{\mathbf{S}, \lambda, \mathbf{S}'\}}{\text{argmin}} \|\mathbf{S}\|_1 \\ & \text{s.t. } \mathbf{S}' = \sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T, \quad \mathbf{S} \in \mathcal{S}, \quad d(\mathbf{S}, \mathbf{S}') \leq \epsilon, \end{aligned} \quad (17)$$

where iteratively re-weighted schemes are also possible. 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 $\mathcal{S} = \mathcal{S}_A$ (cf. Proposition 2 and (33) for additional details).

When the interest is in recovering a normalized Laplacian [cf. (8)], a possible implementation is to enforce the constraint $\lambda_1 = 0$ *talis qualis* (17) entailing that one of the eigenvalues of \mathbf{S}' (and not \mathbf{S}) is equal to zero. However, the smallest eigenvalue of \mathbf{S} must be close to zero due to the constraint on the distance between \mathbf{S} and \mathbf{S}' . Alternatively, the objective can be augmented by also considering the nuclear norm $\|\mathbf{S}\|_*$ to further promote rank-deficiency on \mathbf{S} .

To assess the effect of the noise in recovering the sparsest \mathbf{S} , we define matrices $\hat{\mathbf{W}}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{Q}}$ which are counterparts of \mathbf{W} , \mathbf{R} and \mathbf{Q} defined prior to Theorem 1, but based on the noisy templates $\hat{\mathbf{V}}$ instead of \mathbf{V} . Further, we drop the non-negativity constraint in \mathcal{S}_A – to obtain $\tilde{\mathcal{S}}_A$ – 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.

Proposition 2: When $d(\mathbf{S}, \mathbf{S}') = \|\mathbf{S} - \mathbf{S}'\|_F$, and assuming that there exists at least one \mathbf{S}' such that $\tilde{d}(\mathbf{S}_0^*, \mathbf{S}') \leq \epsilon$, the solution $\hat{\mathbf{S}}_1^* := \text{vec}(\hat{\mathbf{S}}_1^*)$ to (17) for $\mathcal{S} = \tilde{\mathcal{S}}_A$ with scale ambiguity satisfies

$$\|\hat{\mathbf{S}}_1^* - \mathbf{s}_0^*\|_1 \leq C\epsilon, \quad \text{with } C = 2C_1 + 2C_2C_3, \quad (18)$$

if the same conditions stated in Theorem 1 hold but for $\hat{\mathbf{R}}$ instead of \mathbf{R} . Constants C_1 , C_2 , and C_3 are given by

$$C_1 = \frac{\sqrt{|\mathcal{K}|}}{\sigma_{\min}(\hat{\mathbf{R}}_{\mathcal{K}}^T)}, \quad C_2 = \frac{1 + \|\hat{\mathbf{R}}^T\|_{M(2)} C_1}{1 - \psi_{\hat{\mathbf{R}}}}, \quad C_3 = \|\hat{\mathbf{R}}^\dagger\|_{M(2)} N, \quad (19)$$

where $\sigma_{\min}(\cdot)$ denotes the minimum singular value of the argument matrix. An analogous result can be derived for the case $\mathcal{S} = \tilde{\mathcal{S}}_L$ (where the non-positivity constraint is dropped) whenever $\hat{\mathbf{Q}}$ satisfies the conditions in Theorem 2.

Proof: See Appendix B. ■

When given noisy versions $\hat{\mathbf{V}}$ of the spectral templates of our target GSO, Proposition 2 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} . The value of ϵ must be chosen based on the (expected) discrepancy between \mathbf{V} and $\hat{\mathbf{V}}$. The smaller the discrepancy, the smaller the value of ϵ needed to guarantee feasibility via $\tilde{d}(\mathbf{S}_0^*, \mathbf{S}') \leq \epsilon$. Indeed, when $\hat{\mathbf{V}}$ is obtained from the sample covariance $\hat{\mathbf{C}}_x$, we have that as the number of observed signals R increases, $\hat{\mathbf{C}}_x$ tends to the covariance \mathbf{C}_x . Hence, if \mathbf{C}_x has no repeated eigenvalues, $\hat{\mathbf{V}}$ will tend to the actual eigenvectors \mathbf{V} (see, e.g., [42, Th. 3.3.7]) and the tolerance ϵ in (17) can be set to a smaller value. In the limit when $R \rightarrow \infty$ and one has that $\hat{\mathbf{V}} = \mathbf{V}$ then the tolerance ϵ can be set to zero. This yields $\mathbf{S}' = \mathbf{S}_0^*$ and (18) guarantees perfect recovery under conditions A-1) and A-2) in Theorem 1 or L-1) and L-2) in Theorem 2.

B. Incomplete Spectral Templates

Thus far we have assumed that the entire set of eigenvectors $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ is known, either perfectly or corrupted by noise. However, it is conceivable that in a number of scenarios 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 where \mathbf{x} is bandlimited in the graph Fourier domain [43]. More generally, whenever \mathbf{C}_x contains repeated eigenvalues there is a rotational ambiguity in the definition of the associated eigenvectors. Hence, in this case, we keep the eigenvectors that can be unambiguously characterized and, for the eigenvectors with repeated eigenvalues, we include the rotational ambiguity as an additional constraint in our optimization problem.

Formally, assume that the K first eigenvectors $\mathbf{V}_K = [\mathbf{v}_1, \dots, \mathbf{v}_K]$ are those which are known. Then, the network topology inference problem with incomplete spectral templates can be formulated as [cf. (10)]

$$\begin{aligned} \bar{\mathbf{S}}_1^* &:= \operatorname{argmin}_{\{\mathbf{S}, \mathbf{S}_{\bar{K}}, \lambda\}} \|\mathbf{S}\|_1 \\ \text{s.t. } \mathbf{S} &= \mathbf{S}_{\bar{K}} + \sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \quad \mathbf{S} \in \mathcal{S}, \quad \mathbf{S}_{\bar{K}} \mathbf{V}_K = \mathbf{0}, \end{aligned} \quad (20)$$

where we already particularized the objective to the ℓ_1 convex relaxation. The formulation in (20) 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 $\operatorname{Im}(\mathbf{V}_K)$. Notice that, as a consequence, the rank of $\mathbf{S}_{\bar{K}}$ is at most $N - K$. As in the previous cases, \mathcal{S} incorporates a priori information about the GSO. Notice that the constraint in (20) automatically enforce symmetry on $\mathbf{S}_{\bar{K}}$, as wanted. 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 (20) is larger than that in (10). This is particularly important when the templates do not come from a preexisting shift but, rather, one has the freedom to choose \mathbf{S} provided it satisfies certain spectral properties. A practical example is the selection of the

topology of a sensor network aimed at implementing estimation tasks such as consensus averaging, which can be oftentimes written as rank-one transformations of the sensor observations (cf. Section II-C and [30]).

Theoretical guarantees of recovery analogous to those presented in Section III-B can be derived for (20). To formally state these, the following notation must be introduced. Define $\mathbf{W}_K := \mathbf{V}_K \odot \mathbf{V}_K$ and $\Upsilon := [\mathbf{I}_{N^2}, \mathbf{0}_{N^2 \times N^2}]$. Also, define matrices $\mathbf{B}^{(i,j)} \in \mathbb{R}^{N \times N}$ for $i < j$ such that $B_{ij}^{(i,j)} = 1$, $B_{ji}^{(i,j)} = -1$, 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 $i, j \in \{1, 2, \dots, N\}$ where $i < j$. In this way, $\mathbf{B}\mathbf{s} = \mathbf{0}$ when \mathbf{s} is the vectorized form of a symmetric matrix. Further, we define the following matrices

$$\mathbf{P}_1 := \begin{bmatrix} \mathbf{I} - \mathbf{W}_K \mathbf{W}_K^\dagger \\ \mathbf{I}_{\mathcal{D}} \\ \mathbf{B} \\ \mathbf{0}_{N \times N^2} \\ (\mathbf{e}_1 \otimes \mathbf{1}_N)^T \end{bmatrix}^T, \quad \mathbf{P}_2 := \begin{bmatrix} \mathbf{W}_K \mathbf{W}_K^\dagger - \mathbf{I} \\ \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{\binom{N}{2} \times N^2} \\ \mathbf{I} \otimes V_K^T \\ \mathbf{0}_{1 \times N^2} \end{bmatrix}^T, \quad (21)$$

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

Theorem 3: If $\mathcal{S} = \mathcal{S}_A$ and (20) is feasible, then $\bar{\mathbf{S}}_1^* = \mathbf{S}_0^*$ provided that the two following conditions are satisfied:

- A-1) $\operatorname{rank}([\mathbf{P}_1^T, \mathbf{P}_2^T]) = |\mathcal{J}| + N^2$; and
- A-2) There exists a constant $\delta > 0$ such that

$$\eta_{\mathbf{P}} := \|\Upsilon_{\mathcal{J}^c} (\delta^{-2} \mathbf{P} \mathbf{P}^T + \Upsilon_{\mathcal{J}^c}^T \Upsilon_{\mathcal{J}^c})^{-1} \Upsilon_{\mathcal{J}}^T\|_{M(\infty)} < 1. \quad (22)$$

Proof: See Appendix C. ■

Theorem 3 provides sufficient conditions for the relaxed problem in (20) to recover the sparsest graph, even when incomplete information about the eigenvectors is available. In practice it is observed that for smaller number K of known spectral templates the value of $\eta_{\mathbf{P}}$ in (22) tends to be larger, indicating a less favorable setting for recovery. This observation is aligned with the results obtained in practice; see Fig. 2(c).

To state results similar to those in Theorem 3 but for the recovery of normalized Laplacians, we define $\tilde{\mathbf{U}}_K := \tilde{\mathbf{V}}_K \odot \tilde{\mathbf{V}}_K$ where $\tilde{\mathbf{V}}_K := [\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_K]$ and define the matrices

$$\mathbf{T}_1 := \begin{bmatrix} \mathbf{I} - \tilde{\mathbf{U}}_K \tilde{\mathbf{U}}_K^\dagger \\ \mathbf{I}_{\mathcal{D}} \\ \mathbf{B} \\ \mathbf{0}_{N \times N^2} \end{bmatrix}^T, \quad \mathbf{T}_2 := \begin{bmatrix} \tilde{\mathbf{U}}_K \tilde{\mathbf{U}}_K^\dagger - \mathbf{I} \\ \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{\binom{N}{2} \times N^2} \\ \mathbf{I} \otimes V_K^T \end{bmatrix}^T, \quad (23)$$

and $\mathbf{T} := [\mathbf{T}_1^T, \mathbf{T}_2^T]^T$. Under the assumption that the first eigenvector (i.e., the one whose associated eigenvalue is zero) is among the K eigenvectors known, the following result holds.

Theorem 4: If $\mathcal{S} = \mathcal{S}_L$ and (20) is feasible, then $\bar{\mathbf{S}}_1^* = \mathbf{S}_0^*$ provided that the two following conditions are satisfied:

- L-1) $\operatorname{rank}([\mathbf{T}_1^T, \mathbf{T}_2^T]) = |\mathcal{J}| + N^2$; and
- L-2) There exists a constant $\delta > 0$ such that

$$\eta_{\mathbf{T}} := \|\Upsilon_{\mathcal{J}^c} (\delta^{-2} \mathbf{T} \mathbf{T}^T + \Upsilon_{\mathcal{J}^c}^T \Upsilon_{\mathcal{J}^c})^{-1} \Upsilon_{\mathcal{J}}^T\|_{M(\infty)} < 1. \quad (24)$$

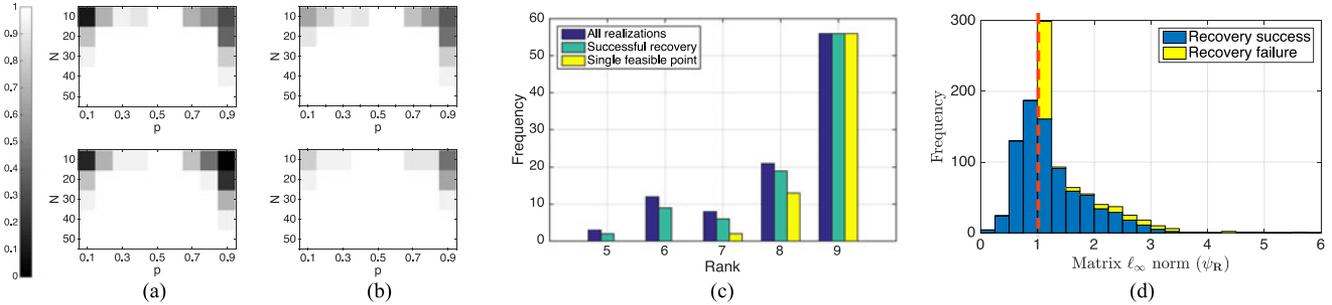


Fig. 1. (a) Proportion of topology inference problems with a unique feasible point for Erdős-Rényi graphs as a function of N and p for adjacency (top) and normalized Laplacian (bottom) matrices. (b) Recovery rate for the same set of graphs in (a) when implementing the iteratively re-weighted approach in (9). (c) Histogram of the rank of matrix \mathbf{U} for $N = 10$ and $p = 0.2$. (d) Experimental validation of Theorem 1. The total height of the bars represents the empirical frequency of the norm $\psi_{\mathbf{R}}$. Each bar is then split into two parts (colors) one representing the fraction of cases where recovery succeeds and the other one where it fails. Whenever $\psi_{\mathbf{R}} < 1$, perfect recovery is achieved.

Proof: The proof follows the same steps as those in Theorem 3 and, thus, is omitted. ■

Notice that scenarios that combine the settings in Sections IV-A and IV-B, i.e. where the knowledge of the K templates is imperfect, can be handled by combining the formulations in (17) and (20). This can be achieved upon implementing the following modifications to (20): considering the shift \mathbf{S}' as a new optimization variable, replacing the first constraint in (20) 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. (17)].

Regarding the computational complexity of our algorithms, notice that there are two major tasks to consider: (i) computing the eigenvectors which incurs $\mathcal{O}(N^3)$ complexity [44]; and (ii) solving iteratively the various sparsity minimization problems to recover the GSO, which incur $\mathcal{O}(N^3)$ complexity per iteration using naively e.g., the solvers in [45] or [46]. Simulations in Section V were run using the CVX package for Matlab [47], an off-the-shelf tool which may incur suboptimal complexity both in terms of the number of variables and constraints. Building on optimization advances for sparse recovery problems, custom-made scalable algorithms that exploit the structure of (10), (17) and (20) could be developed. This is an interesting direction that we are currently pursuing, but is beyond the scope of this paper.

V. NUMERICAL EXPERIMENTS

We test the proposed topology inference methods on different synthetic and real-world graphs. A comprehensive performance evaluation is carried out whereby we: (i) investigate the recovery of both adjacency and normalized Laplacian matrices; (ii) corroborate our main theoretical findings; (iii) assess the impact of imperfect information in the recovery; (iv) carry out comparisons with state-of-the-art methods; and (v) illustrate how our framework can promote sparsity on a given network.

A. Topology Inference From Noiseless Templates

Consider Erdős-Rényi (ER) graphs² [48] of varying size $N \in \{10, 20, \dots, 50\}$ and different edge-formation probabili-

²Although for conciseness the results in Section V-A are only presented for ER graphs, these results are consistent with those for other types of random graphs including, small-world and preferential attachment graphs [48].

ties $p \in \{0.1, 0.2, \dots, 0.9\}$. For each combination of N and p we generate 100 connected graphs and try to recover their adjacency \mathbf{A} and normalized Laplacian \mathbf{L} matrices from the corresponding spectral templates \mathbf{V} . In Fig. 1(a) we plot the proportion of instances where the corresponding optimization problems – problem (6) for $\mathcal{S} = \mathcal{S}_{\mathbf{A}}$ and $\mathcal{S} = \mathcal{S}_{\mathbf{L}}$ – have singleton feasibility sets. Notice that multiple solutions are more frequent when the expected number of neighbors of a given node is close to either 1 or N . For intermediate values of p , the rank of both $\mathbf{W}_{\mathcal{D}}$ and \mathbf{U} is typically $N - 1$, guaranteeing a single feasible point (cf. Proposition 1). Using the same set of graphs that those in Fig. 1(a), Fig. 1(b) shows the recovery rate when solving the iteratively re-weighted problem (9) for both the adjacency (top) and the normalized Laplacian (bottom). As expected, the rates in Fig. 1(b) dominate those in Fig. 1(a) since every instance with a unique feasible point is recovered successfully. Moreover, the improved rates observed in Fig. 1(b) are indicative of the beneficial effect that the weighted ℓ_1 norm objective has in the recovery. The aforementioned experiment was repeated for (weighted) random geometric graphs of varying size and connectivity range, and similar findings were observed.

As indicated by Proposition 1, the rate of recovery is intimately related to the ranks of $\mathbf{W}_{\mathcal{D}}$ and \mathbf{U} for the adjacency and normalized Laplacian cases, respectively. Fig. 1(c) further illustrates this relation via a histogram of the rank of \mathbf{U} for the 100 graphs with $N = 10$ and $p = 0.2$. Given that ER graphs are unweighted, we have that $\text{rank}(\mathbf{U}) < N - 1$ for a non-negligible fraction of the realizations (cf. discussion after Proposition 1). Nevertheless, for more than half of the instances, the rank of \mathbf{U} is equal to 9 (blue bar) and, as stated in Proposition 1, for all these graphs there is a unique feasible point (yellow bar) that is successfully recovered (cyan bar). We see that, as the rank of \mathbf{U} degrades, uniqueness is no longer guaranteed but for most cases the true graph can still be recovered following the iteratively re-weighted scheme proposed. Only in 8 of the cases where $\text{rank}(\mathbf{U}) < 9$ the recovery was not successful, entailing a recovery rate of 0.92, as reported in the corresponding entry ($N = 10, p = 0.2$) of the bottom plot in Fig. 1(b).

Finally, in order to corroborate the conditions for noiseless recovery stated in Theorem 1, we draw ER random graphs of size $N = 20$ and edge-formation probability $p = 0.25$. For each

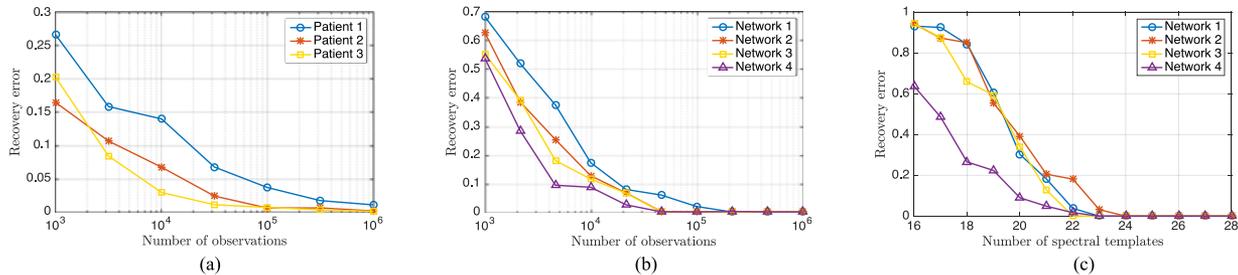


Fig. 2. (a) Brain graph recovery error for three patients as a function of the number of synthetic signals observed in the estimation of the spectral templates. (b) Recovery error for four social networks as a function of the number of synthetic signals observed in the estimation of the spectral templates. (c) Recovery error for four social networks (with $N = 32$ nodes) as a function of K , the number of spectral templates that are known.

graph, we make sure that the associated \mathbf{W}_D matrix has rank strictly smaller than $N - 1$ (to rule out the cases where the feasible set is a singleton), and that condition A-1) in Theorem 1 is satisfied. In Fig. 1(d) we plot the number of successes and failures in recovering the adjacency as a function of ψ_R in (13). We consider 1000 realizations and for each of them the constant δ in (13) is chosen to minimize ψ_R . Fig. 1(d) clearly depicts the result of Theorem 1 in that, for all cases where $\psi_R < 1$, relaxation (10) achieves perfect recovery. Equally important, it is clear that the bound stated in (13) is tight since a large proportion of the realizations with ψ_R equal to 1 or just above this value lead to failed recoveries.

B. Topology Inference From Noisy and Incomplete Templates

In this set of experiments we consider imperfect spectral templates, real-world graphs, and synthetic signals. We start with the identification of unweighted and undirected graphs corresponding to human brains [49]. Each graph consists of $N = 66$ nodes, which represent brain regions of interest (ROIs). An edge between two ROIs exists if the density of anatomical connections is greater than a threshold, which is chosen as the largest one that entails a connected graph [49]. We test the recovery from noisy spectral templates $\hat{\mathbf{V}}$ [cf. (17)] obtained from sample covariances of synthetic signals generated through diffusion processes (cf. Section II). We model such processes as graph filters of random degree between 3 and 7, and with independent and normally distributed coefficients. Denoting by $\hat{\mathbf{V}}_i$ the noisy spectral templates of patient $i \in \{1, 2, 3\}$ and by $\hat{\mathbf{A}}_i$ the adjacency matrices recovered, Fig. 2(a) plots the recovery error as a function of the number of signals observed in the computation of the sample covariance. The error is quantified as the proportion of edges misidentified, i.e., $\|\mathbf{A}_i - \hat{\mathbf{A}}_i\|_0 / \|\mathbf{A}_i\|_0$, and each point in Fig. 2(a) is the average across 50 realizations. Notice that for an increasing number of observed signals we see a monotonous decrease in the recovery error. For example, when going from 10^4 to 10^5 observations the error is (approximately) divided by seven, when averaged across patients. This is reasonable since a larger number of observations gives rise to a more reliable estimate of the covariance matrix entailing less noisy spectral templates. Traditional methods like graphical lasso [17] fail to recover \mathbf{S} from the sample covariance of filtered white signals. For example, when signals are generated using a filter of the form $\mathbf{H} = h_0 \mathbf{I} + h_1 \mathbf{S}$, graphical lasso performs significantly worse than the method based on spectral templates. More

precisely, when 10^5 signals are observed, the recovery error of graphical lasso averaged over 50 realizations and with optimal tuning parameters is 0.303, 0.350, and 0.270 for patients 1, 2, and 3, respectively. Such errors are between 5 and 50 times larger than those reported in Fig. 2(a). Further comparisons of our method with graphical lasso and other existing alternatives are provided in Section V-C.

We repeat the previous experiment on four social networks defined on a common set of $N = 32$ nodes, which represent students from the University of Ljubljana³. Links for each of the networks capture different types of interactions among the students, and were built by asking each student to select a group of preferred college mates for different situations, e.g., to discuss a personal issue or to invite to a birthday party (see footnote 2 for further details). The considered graphs are unweighted and symmetric, and the edge between i and j exists if either student i picked j in the questionnaire or vice versa. As done for the brain graphs, we test the recovery performance for noisy spectral templates $\hat{\mathbf{V}}$ obtained from sample covariances. Fig. 2(b) plots the reconstruction error as a function of the number of observed signals for the different networks studied. As was observed in Fig. 2(a), we see a monotonous decrease in recovery error for all the analyzed networks.

Finally, we illustrate the recovery performance in the presence of incomplete spectral templates by solving (20) for the four networks in Fig. 2(b). More specifically, in Fig. 2(c) we plot the recovery error as a function of the number K of eigenvectors available. Each point in the plot is the average across 50 realizations in which different K eigenvectors were selected from the $N = 32$ possible ones. As expected, the performance for all four networks improves with the number of spectral templates known. The performance improvement is sharp and precipitous going from a large error of over 0.85 for three of the networks when 17 spectral templates are known to a perfect recovery for all the networks when 24 eigenvectors are given. Moreover, notice that network 4 is consistently the easiest to identify both for noisy [cf. Fig. 2(b)] and incomplete [cf. Fig. 2(c)] spectral templates. For example, when given 19 spectral templates the error associated with network 4 is 0.224 whereas the average across the other three networks is 0.584. This could be partially explained by the fact that network 4 is the sparsest, containing

³Access to the data and additional details are available at <http://vladoviki.fmf.uni-lj.si/doku.php?id=pajek:data:pajek:students>

TABLE I
PERFORMANCE COMPARISON BETWEEN SPECTRAL TEMPLATES (SPECTEMP), KALOFOLIAS [12], AND DONG *et al.* [8]

	Inverse Laplacian			Diffusion			Exponential		
	SpecTemp	Kalofolias	Dong <i>et al.</i>	SpecTemp	Kalofolias	Dong <i>et al.</i>	SpecTemp	Kalofolias	Dong <i>et al.</i>
Erdős-Rényi									
F-measure	0.896	0.791	0.818	0.924	0.868	0.828	0.703	0.651	0.667
edge error	0.108	0.152	0.168	0.071	0.149	0.177	0.276	0.318	0.332
degree error	0.058	0.071	0.105	0.040	0.055	0.111	0.162	0.201	0.222
Barabási-Albert									
F-measure	0.926	0.855	0.873	0.945	0.845	0.894	0.814	0.732	0.798
edge error	0.143	0.173	0.209	0.135	0.154	0.235	0.310	0.314	0.393
degree error	0.108	0.124	0.169	0.109	0.092	0.188	0.240	0.244	0.282

16.7% of all possible edges while this indicator attains 32.5%, 31.3%, and 27.2% for the rest of the networks. Hence, the sparsity promoting objectives in (17) and (20) could be more effective in recovering network 4. Nevertheless, a formal analysis of which classes of graphs are inherently more robust for identification when given imperfect spectral templates is left as future work.

C. Performance Comparison

We compare the performance of the presented method based on spectral templates (we refer to it as SpecTemp for conciseness) with broadly used statistical approaches as well as recent GSP-based algorithms.

Comparison with baseline statistical methods: We analyze the performance of SpecTemp in comparison with two broadly used methods, namely, (thresholded) correlation [4, Ch. 7.3.1] and graphical lasso [17]. Our goal is to recover the adjacency matrix of an undirected and unweighted graph with no self-loops from the observation of filtered graph signals. For the implementation of SpecTemp, we use the eigendecomposition of the sample covariance of the observed signals in order to extract noisy spectral templates $\hat{\mathbf{V}}$. We then solve problem (17) for $\mathcal{S} = \mathcal{S}_A$, where ϵ is selected as the smallest value that admits a feasible solution. We include as a priori knowledge that each node has at least one neighbor, i.e., we replace the constraint $\sum_j S_{j1} = 1$ by the constraint $\mathbf{S}\mathbf{1} \geq \beta\mathbf{1}$ for a small positive constant β in the set \mathcal{S}_A . For the correlation-based method, we keep the absolute value of the sample correlation of the observed signals, force zeros on the diagonal and set all values below a certain threshold to zero. This threshold is determined during a training phase, as explained in more detail in the next paragraph. Lastly, for graphical lasso we follow the implementation in [17] based on the sample covariance and select the tuning parameter ρ (see [17]) during the training phase. We then force zeros on the diagonal and keep the absolute values of each entry. Leveraging that the sought graphs are unweighted, for SpecTemp and graphical lasso a fixed threshold of 0.3 is used so that, after recovery, every edge with weight smaller than the threshold is set to zero.

We test the recovery of adjacency matrices $\mathbf{S} = \mathbf{A}$ of ER graphs with $N = 20$ nodes and edge probability $p = 0.2$. We vary the number of observed signals from 10^1 to 10^6 in pow-

ers of 10. Each signal is generated by passing white Gaussian noise through a graph filter \mathbf{H} . Two different types of filters are considered. As a first type we consider a *general* filter $\mathbf{H}_1 = \mathbf{V}\text{diag}(\hat{\mathbf{h}}_1)\mathbf{V}^T$, where the entries of $\hat{\mathbf{h}}_1$ are independent and chosen randomly between 0.5 and 1.5. The second type is a *specific* filter of the form $\mathbf{H}_2 = (\delta_{\mathcal{H}}\mathbf{I} + \mathbf{S})^{-1/2}$, where the constant $\delta_{\mathcal{H}}$ is chosen so that $\delta_{\mathcal{H}}\mathbf{I} + \mathbf{S}$ is positive definite to ensure that \mathbf{H}_2 is real and well-defined. Following the discussion in Section II, this implies that the precision matrix of the filtered signals is given by $\mathbf{C}_x^{-1} = \mathbf{H}_2^{-2} = \delta_{\mathcal{H}}\mathbf{I} + \mathbf{S}$, which coincides with \mathbf{S} in the off-diagonal elements. For each combination of filter type and number of observed signals, we generate 10 ER graphs that are used for training and 20 ER graphs that are used for testing. Based on the 10 training graphs, the optimal threshold for the correlation method and parameter ρ for graphical lasso are determined and then used for the recovery of the 20 testing graphs. Given that for SpecTemp we are fixing ϵ beforehand, no training is required.

As figure of merit we use the F-measure [50], i.e. the harmonic mean of edge precision and edge recall, that solely takes into account the support of the recovered graph while ignoring the weights. In Fig. 3(a) we plot the performance of the three methods as a function of the number of filtered graph signals observed for filters \mathbf{H}_1 and \mathbf{H}_2 , where each point is the mean F-measure over the 20 testing graphs.

When considering a general graph filter \mathbf{H}_1 , SpecTemp clearly outperforms the other two. For instance, when 10^5 signals are observed, our average F-measure is 0.81 while the measures for correlation and graphical lasso are 0.29 and 0.25, respectively. Moreover, of the three methods, our approach is the only consistent one, i.e., achieving perfect recovery with increasing number of observed signals. Although striking at a first glance, the deficient performance of the baseline statistical methods was expected. For general filters \mathbf{H}_1 , neither the correlation nor the precision matrices are sparse or share the support of the GSO to be recovered \mathbf{S} . When analyzing the specific case of graph filters \mathbf{H}_2 , where the precision matrix exactly coincides with the desired graph-shift operator, graphical lasso outperforms both our method and the correlation-based method. This is not surprising since graphical lasso was designed for the recovery of sparse precision matrices. Notice however that for large number of observations SpecTemp, without assuming any specific filter model, also achieves perfect recovery and yields

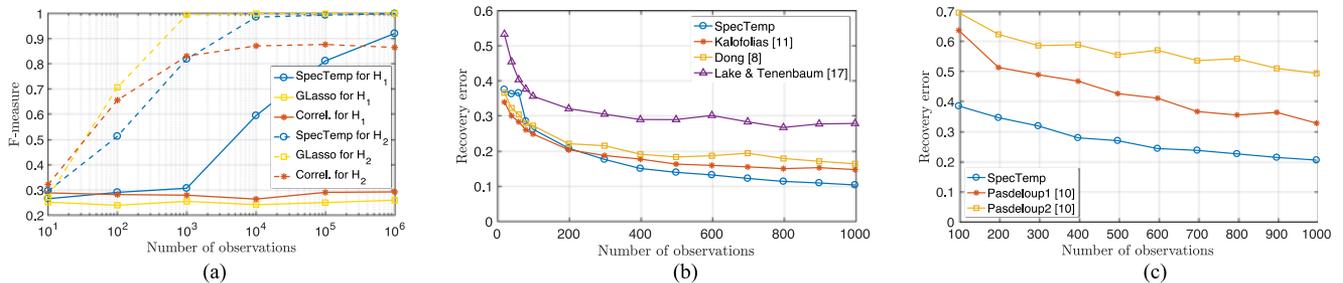


Fig. 3. (a) Performance comparison between spectral templates (SpecTemp), graphical lasso, and correlation-based recovery. For general filters, SpecTemp outperforms the other two. (b) Comparison of edge recovery error for the combinatorial Laplacian as a function of the number of signals observed for SpecTemp, Kalofolias [11], Dong *et al.* [8], as well as Lake and Tenenbaum [17]. (c) Comparison of edge recovery error for the normalized Laplacian as a function of the number of signals observed for SpecTemp and two alternatives of Padeloup *et al.* [10].

an F-measure equal to 1. Consequently, if a practitioner knows a priori that the sought graph is (close to) the precision matrix and Gaussian signal assumptions are tenable, then graphical lasso will be the preferred method. However, for the general case in which this information is unavailable, SpecTemp is a more prudent alternative.

Comparison with GSP methods: We compare the recovery using SpecTemp with two classes of algorithms in the GSP literature: i) methods designed to identify the (combinatorial) Laplacian of a graph when given a set of smooth graph signals ([8] and [12]); and ii) methods designed to recover the normalized Laplacian of a graph by observing stationary signals on the graph ([10]).

Small modifications can be made to our framework to accommodate the recovery of a combinatorial Laplacian, thus permitting a fair comparison with [8] and [12]. More precisely, in solving (17) we use the set of admissible shifts given by

$$\mathcal{S}_{L_c} := \{\mathbf{S} \mid S_{ij} \leq 0 \text{ for } i \neq j, \mathbf{S} \in \mathcal{M}_+^N, \mathbf{S}\mathbf{1} = \mathbf{0}\}. \quad (25)$$

Moreover, in order to account for the smoothness of the observed signals in the unknown graph we sort the eigenvectors $\hat{\mathbf{v}}_k$ of the sample covariance in increasing order of their corresponding eigenvalues, and we require the recovered eigenvalues λ to satisfy $\lambda_i \geq \lambda_{i+k} + \delta_\lambda$ for all i , and fixed k and δ_λ . In this way, we assign the frequencies with larger presence in the observed signals to low eigenvalues in the recovered Laplacian. Unless otherwise noted, we set $\delta_\lambda = 0.1$ and $k = 3$.

We compare the three methods of interest on two different types of graphs and three different signal generation models. We consider the recovery of the combinatorial Laplacian $\mathbf{S} = \mathbf{L}_c$ of ER graphs with $N = 20$ nodes and edge probability $p = 0.3$ as well as Barabási-Albert preferential attachment graphs [48] with $N = 20$ generated from $m_0 = 4$ initially placed nodes, where each new node is connected to $m = 3$ existing ones. Following [12] we consider three models for smooth graph signals: i) multivariate normal signals with covariance given by the pseudo-inverse of \mathbf{L}_c , i.e., $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{L}_c^\dagger)$; ii) white signals filtered through an autoregressive (diffusion) process, that is $\mathbf{x}_2 = (\mathbf{I} + \mathbf{L}_c)^{-1} \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; and iii) white signals passed through an exponential filter, $\mathbf{x}_3 = \exp(-\mathbf{L}_c) \mathbf{w}$. For each of the six settings considered (two graphs combined with three signal types) we generate 10 training graphs, 100 test-

ing graphs, and for every graph we generate 1000 graph signals. The training set is used to set the parameters in [8] and [12], and in our case it serves the purpose of selecting the best ϵ [cf. (17)]. To increase the difficulty of the recovery task, every signal \mathbf{x} is perturbed as $\hat{\mathbf{x}} = \mathbf{x} + \sigma \mathbf{x} \circ \mathbf{z}$, for $\sigma = 0.1$ and where each entry of \mathbf{z} is an independent standard normal random variable. We focus on three performance measures, namely, the F-measure as explained in the previous experiment, the ℓ_2 relative error of recovery of the edges, and the ℓ_2 relative error of recovery of the degrees. The performance achieved by each method in the testing sets is summarized in Table I. In all but one case, our method attains the largest F-measures and the smallest errors for all the graphs and signal types considered.

Finally, for the particular cases of ER graphs and signals \mathbf{x}_1 (inverse Laplacian), we replicate the above procedure varying the number of observed signals R from 20 to 100 in steps of 20 and from 100 to 1000 in steps of 100. For SpecTemp, we increase k when the number of observations decreases to account for the noisier ordering of the eigenvectors in the sample covariance. In this experiment we use $k = 5$ for $R \leq 400$, $k = 4$ for $400 < R < 800$ and $k = 3$ for $R \geq 800$. In Fig. 3(b) we plot the associated ℓ_2 edge recovery errors. We also present the performance of [18] to serve as a baseline for Laplacian recovery based on smooth signals. However, for the experiment at hand, the GSP-based methods achieve better performance, ratifying the findings in [8]. For small number of observations $R \leq 60$ both [12] and [8] outperform SpecTemp, whereas for $60 < R \leq 200$ only [12] recovers a better graph than SpecTemp. For larger number of observed signals, SpecTemp outperforms the alternative methods. This can be attributed to the fact that SpecTemp assumes no *specific* model on the smoothness of the signal other than decreasing energy for increasing frequencies. Thus, when enough signals are observed, our more model agnostic, data-driven approach exhibits a performance advantage. Similar results were found for the other combinations of graph and signal models tabulated under Table I.

The last experiment in this section compares SpecTemp with the related approach in [10], which considers signals that are stationary in a normalized Laplacian GSO. More precisely, [10] focuses on the recovery of a *diffusion matrix* $\mathbf{T} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$. Since the normalized Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{T}$ has the same eigenvectors than \mathbf{T} , recovering one is equivalent to recover-

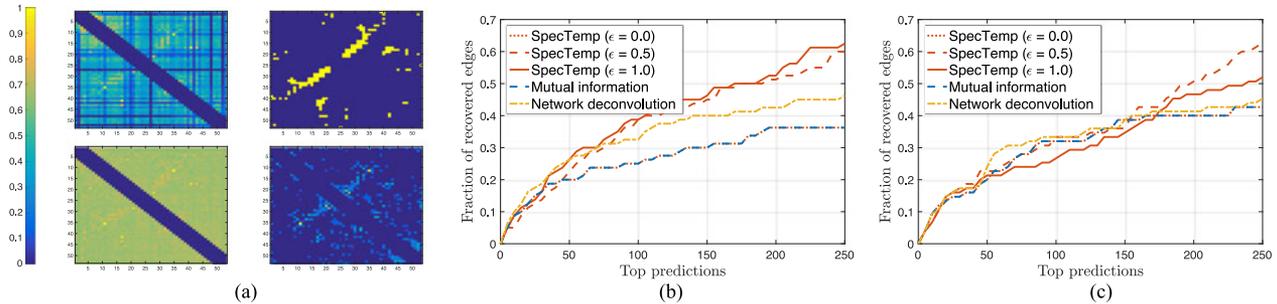


Fig. 4. (a) Real and inferred contact networks between amino-acid residues for protein BPT1 BOVIN. Mutual information of the co-variation of amino-acid residues (top left), ground truth contact network (top right), contact network inferred by network deconvolution (bottom left), contact network inferred 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.

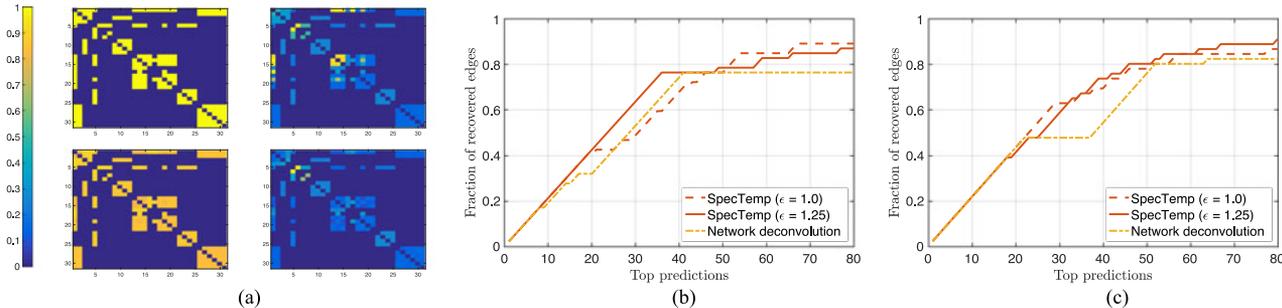


Fig. 5. (a) Real and inferred collaboration networks between 31 network scientists. Unweighted co-authorship network (top left), true weighted collaboration network (top right), weighted collaboration network inferred by network deconvolution (bottom left), weighted collaboration network inferred by our method based on spectral templates (bottom right). (b) Fraction of the important collaboration edges between scientists recovered for each method as a function of the number of edges considered. (c) Counterpart of (b) for a different collaboration network between 28 network scientists.

ing the other. However, the fundamental difference between [10] and SpecTemp is the (explicit) consideration of imprecise eigenvectors. In [10] there is no analogous treatment to the one we propose in (17). This precludes the approach in [10] from enforcing both zeros in the diagonal (simplicity) and sparsity of \mathbf{T} simultaneously.

Specifically, we run the comparison considering ER random graphs ($N = 10$ and $p = 0.2$), setting the shift to the normalized Laplacian, and varying the number of observed filtered signals R from 100 to 1,000. We test three different recovery algorithms: i) SpecTemp, i.e., solving problem (17) for $\mathcal{S} = \mathcal{S}_L$; ii) Padeloup1, i.e., solving [10, problem (15)]; and iii) Padeloup2, i.e., solving [10, problem (16)]. Fig. 3(c) depicts the obtained recovery rates. It is apparent that SpecTemp markedly outperforms the other two methods, possibly due to the fact that our formulation models the actual and the imperfect shift as two different entities and accounts explicitly for the imperfections.

D. Network Sparsification

With reference to the network sparsification problem outlined in Section II-C, our goal here is to identify the structural properties of proteins from a mutual information graph of the co-variation between the constitutional amino-acids [28]; see [38] for details. For example, for a particular protein, we want to recover the structural graph in the top right of Fig. 4(a) when given the graph of mutual information in the top left 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 (NetDec) [38] is illustrated in the bottom left corner of Fig. 4(a) whereas the one recovered using SpecTemp is depicted in the bottom right corner of the figure. Comparing both recovered graphs, SpecTemp leads to a sparser graph that follows more closely the desired structure to be recovered. We show this in Fig. 4(b) by plotting the sensitivity of the top edge predictions, i.e., the fraction of the real contact edges recovered, as in [38]. For example, 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 sensitivity of the 100 top edge predictions is 0.4. As claimed in [38], NetDec improves the estimation when compared to raw mutual information data. Nevertheless, from Fig. 4(b) it follows that SpecTemp outperforms network deconvolution. Notice that when $\epsilon = 0$ [cf. (17)] we force 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. E.g., 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. 4(c) we present this same analysis for a different protein and similar results can be appreciated.

E. Collaboration Networks

We now illustrate how SpecTemp can be used to unveil the most relevant collaborations in a social network encoding co-authorship information. To that end, consider the network of 31 scientists working in the field of network science presented in [29], whose *unweighted* adjacency matrix \mathbf{A} is depicted in the *top left* of Fig. 5(a). In \mathbf{A} , an unweighted edge exists between two authors if they have co-authored at least one paper. The importance of these relationships [*top right* of Fig. 5(a)] has been assessed using additional information beyond network connectivity, including number of co-authored papers and total number of authors in each publication [29]; see [38] for details. Our goal is to use as input the unweighted connectivity graph \mathbf{A} (top left) to estimate the weighted importance (top right). Intuitively, our method is a candidate to carry out this task because one can think of more meaningful collaborations as direct influences between authors, whereas indirect relations would correspond to weak collaborations diluted over many co-authors. The network deconvolution approach in [38] seeks this same objective and outputs the network in the *bottom left* corner of Fig. 5(a). The network (weighted graph) inferred using our SpecTemp algorithm is depicted in the *bottom right* corner of the figure. This latter graph was obtained by solving (17), where the eigenbasis $\hat{\mathbf{V}}$ corresponds to \mathbf{A} and where we force the edges not in \mathbf{A} to attain zero value in the recovered \mathbf{S} . The graph obtained by SpecTemp resembles more closely the true weights of the collaborations. We quantify this in Fig. 5(b) utilizing a measure similar to the one in Section V-D. More precisely, we separate the edges into two halves depending on their true weight, namely, *important* edges (top half) and *unimportant* edges (bottom half). In Fig. 5(b) we plot the fraction of *important* edges recovered when considering the top edges as predicted by SpecTemp and NetDec. As was the case in Section V-D, if we set $\epsilon = 0$ we recover exactly \mathbf{A} since it is a valid adjacency matrix. However, by varying ϵ we can achieve better performance than the one in [38]. E.g. for $\epsilon = 1.25$ we see that the SpecTemp curve dominates the NetDec one, indicating the superior explanatory power of our approach. Moreover, notice that the SpecTemp curve is strictly increasing for the first 36 top predictions, meaning that all of the top 36 edges predicted by SpecTemp were in fact *important* edges. This increase in performance was found to be robust to the choice of ϵ . For instance, for $\epsilon = 1.0$ the proposed method still dominates NetDec for most values of top predictions considered. In Fig. 5(c) we repeat the analysis with another collaboration network – the connected component with 28 authors in the network considered in [38] – and observe similar results.

VI. CONCLUSIONS

With $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ being the shift operator associated with the graph \mathcal{G} , we studied the problem of identifying \mathbf{S} (hence the topology of \mathcal{G}) using a two-step approach where we first obtain the eigenvectors \mathbf{V} , and then use \mathbf{V} as input to find the eigenvalues $\mathbf{\Lambda}$. The problem of finding $\mathbf{\Lambda}$ given \mathbf{V} was formulated as a sparse recovery optimization. Efficient algorithms based on convex relaxations were developed, and theoretical

conditions under which exact and robust recovery is guaranteed were derived for the cases where \mathbf{S} represents the adjacency or the normalized Laplacian of \mathcal{G} . In identifying \mathbf{V} , our main focus was on using as input a set of graph signal realizations. Under the assumption that such signals resulted from diffusion dynamics on the graph or, equivalently, that they were (graph) stationary in \mathbf{S} , it was shown that \mathbf{V} could be estimated from the eigenvectors of the sample covariance of the available set. As a consequence, several well-established methods for topology identification based on local and partial correlations can be viewed as particular instances of the approach here presented.

APPENDIX A

PROOF OF THEOREM 1

Recalling that $\mathbf{s} = \text{vec}(\mathbf{S})$, problem (10) for the case where $\mathcal{S} = \mathcal{S}_A$ can be reformulated as

$$\min_{\{\mathbf{s}, \boldsymbol{\lambda}\}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{s} = \mathbf{W}\boldsymbol{\lambda}, \quad \mathbf{s}_D = \mathbf{0}, \quad (\mathbf{e}_1 \otimes \mathbf{1}_N)^T \mathbf{s} = 1, \quad (26)$$

where the last equality imposes that the first column of \mathbf{S} must sum up to 1 [cf. (7)]. Notice that the non-negativity constraint in \mathcal{S}_A is ignored in (26). However, if we show that (26) can recover the sparse solution \mathbf{s}_0^* , then the same solution would be recovered by the more constrained problem (10). Notice that we may solve for $\boldsymbol{\lambda}$ in closed form as $\boldsymbol{\lambda}^* = \mathbf{W}^\dagger \mathbf{s}$. Consequently, (26) becomes

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad (\mathbf{I} - \mathbf{W}\mathbf{W}^\dagger)\mathbf{s} = \mathbf{0}, \quad \mathbf{s}_D = \mathbf{0}, \quad (\mathbf{e}_1 \otimes \mathbf{1}_N)^T \mathbf{s} = 1. \quad (27)$$

Leveraging the fact that $\mathbf{I} - \mathbf{W}\mathbf{W}^\dagger$ is symmetric, the first equality in (27) can be rewritten as [cf. (11)]

$$(\mathbf{I} - \mathbf{W}\mathbf{W}^\dagger)_D^T \mathbf{s}_D + \mathbf{M}^T \mathbf{s}_{D^c} = \mathbf{0}, \quad (28)$$

and the second equality in (27) forces the first term of (28) to be zero. With these considerations, we may restate (27) as

$$\min_{\mathbf{s}_{D^c}} \|\mathbf{s}_{D^c}\|_1 \quad \text{s.t.} \quad \mathbf{R}^T \mathbf{s}_{D^c} = \mathbf{b}, \quad (29)$$

where \mathbf{b} is a binary vector of length $N^2 + 1$ with all its entries equal to 0 except for the last one that is a 1. Problem (29) takes the form of classical basis pursuit [51]. Notice that the system of linear equations in (29) is overdetermined since $\mathbf{R}^T \in \mathbb{R}^{N^2+1 \times |D^c|}$, however, feasibility of (26) guarantees that the mentioned system of equations is compatible. The following two conditions are required for the solution of (29) to coincide with the sparse solution $\mathbf{s}_{0D^c}^*$ (cf. [41]):

- $\ker(\mathbf{I}_{\mathcal{K}^c}) \cap \ker(\mathbf{R}^T) = \{\mathbf{0}\}$; and
- There exists a vector $\mathbf{y} \in \mathbb{R}^{|D^c|}$ such that $\mathbf{y} \in \text{Im}(\mathbf{R})$, $\mathbf{y}_{\mathcal{K}} = \text{sign}((\mathbf{s}_{0D^c}^*)_{\mathcal{K}})$, and $\|\mathbf{y}_{\mathcal{K}^c}\|_\infty < 1$.

The remainder of the proof is devoted to showing that if conditions A-1) and A-2) in the statement of the theorem hold true, then a) and b) are satisfied.

To see that A-1) implies a) notice that the nullspace of $\mathbf{I}_{\mathcal{K}^c}$ is spanned by the columns of $\mathbf{I}_{\mathcal{K}}^T$. Hence, for a) to hold we need the $|\mathcal{K}|$ columns of \mathbf{R}^T in positions \mathcal{K} to form a full column rank matrix. In condition A-1) we require $\mathbf{R}_{\mathcal{K}}$ to be full row rank, which is an equivalent property.

The next step is to show that condition A-2) implies b). For this, consider the following ℓ_2 -norm minimization problem

$$\min_{\{\mathbf{y}, \mathbf{z}\}} \delta^2 \|\mathbf{z}\|_2^2 + \|\mathbf{y}\|_2^2 \quad \text{s.t. } \mathbf{y} = \mathbf{R}\mathbf{z}, \quad \mathbf{y}_{\mathcal{K}} = \text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}}), \quad (30)$$

where δ is a positive tuning constant. The inclusion of the term $\delta^2 \|\mathbf{z}\|_2^2$ in the objective guarantees the existence of a closed-form expression for the optimal solution, while preventing numerical instability when solving the optimization. We will show that the solution \mathbf{y}^* to problem (30) satisfies the requirements imposed in condition b). The two constraints in (30) enforce the fulfillment of the first two requirements in b), hence, we are left to show that $\|\mathbf{y}_{\mathcal{K}^c}^*\|_\infty < 1$. Since the values of $\mathbf{y}_{\mathcal{K}}$ are fixed, the constraint $\mathbf{y} = \mathbf{R}\mathbf{z}$ can be rewritten as $\mathbf{I}_{\mathcal{K}}^T \text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}}) = -\mathbf{I}_{\mathcal{K}^c}^T \mathbf{y}_{\mathcal{K}^c} + \mathbf{R}\delta^{-1} \delta \mathbf{z}$. Then, by defining the vector $\mathbf{t} := [\delta \mathbf{z}^T, -\mathbf{y}_{\mathcal{K}^c}^T]^T$ and the matrix $\Phi := [\delta^{-1} \mathbf{R}^T, \mathbf{I}_{\mathcal{K}^c}]$, (30) can be rewritten as

$$\min_{\mathbf{t}} \|\mathbf{t}\|_2^2 \quad \text{s.t. } \mathbf{I}_{\mathcal{K}}^T \text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}}) = \Phi^T \mathbf{t}. \quad (31)$$

The minimum-norm solution to (31) is given by $\mathbf{t}^* = (\Phi^T)^{\dagger} \mathbf{I}_{\mathcal{K}}^T \text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}})$ from where it follows that

$$\mathbf{y}_{\mathcal{K}^c}^* = -\mathbf{I}_{\mathcal{K}^c} (\delta^{-2} \mathbf{R}\mathbf{R}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T \text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}}). \quad (32)$$

Condition a) guarantees the existence of the inverse in (32). Since $\|\text{sign}((\mathbf{s}_{0^{\mathcal{D}^c}}^*)_{\mathcal{K}})\|_\infty = 1$, we may bound the ℓ_∞ norm of $\mathbf{y}_{\mathcal{K}^c}^*$ as $\|\mathbf{y}_{\mathcal{K}^c}^*\|_\infty \leq \|\mathbf{I}_{\mathcal{K}^c} (\delta^{-2} \mathbf{R}\mathbf{R}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T\|_{M(\infty)} = \psi_{\mathbf{R}}$. Hence, condition A-2) in the theorem guarantees $\|\mathbf{y}_{\mathcal{K}^c}^*\|_\infty < 1$ as wanted, concluding the proof.

Notice that in this proof we ignored the non-negativity constraints in \mathcal{S}_A . Nonetheless, the sufficient conditions A-1) and A-2) are valid for the formulation considering all constraints in \mathcal{S}_A , as claimed in the statement of Theorem 1.

APPENDIX B PROOF OF PROPOSITION 2

We reformulate (17) in vector form for the case $\mathcal{S} = \tilde{\mathcal{S}}_A$ with scale ambiguity to obtain

$$\min_{\{\mathbf{s}, \lambda, \mathbf{s}'\}} \|\mathbf{s}\|_1 \quad \text{s.t. } \mathbf{s}' = \hat{\mathbf{W}} \lambda, \quad \mathbf{s}_{\mathcal{D}} = \mathbf{0},$$

$$\|\mathbf{s} - \mathbf{s}'\|_2^2 + ((\mathbf{e}_1 \otimes \mathbf{1}_N)^T \mathbf{s} - 1)^2 \leq \epsilon^2. \quad (33)$$

Substituting the first equality constraint in (33) into the inequality constraint, then solving for λ as $\lambda^* = \hat{\mathbf{W}}^{\dagger} \mathbf{s}$, and finally using the second equality constraint to reduce the optimization variables to $\mathbf{s}_{\mathcal{D}^c}$, we may restate (33) as [cf. (29)]

$$\min_{\mathbf{s}_{\mathcal{D}^c}} \|\mathbf{s}_{\mathcal{D}^c}\|_1 \quad \text{s.t. } \|\hat{\mathbf{R}}^T \mathbf{s}_{\mathcal{D}^c} - \mathbf{b}\|_2 \leq \epsilon, \quad (34)$$

where \mathbf{b} is, as in the proof of Theorem 1, a binary vector with all its entries equal to 0 except for the last one that is equal to 1. Notice that (34) takes the form of a basis pursuit problem with noisy observations [51]. Expressions (18) and (19) can be derived by applying the second claim in [41, Th. 2] to problem (34). In order to do so, a few factors must be taken into consideration. First, since $\hat{\mathbf{R}}^T$ is not full row rank (since it is a tall matrix), constant C_3 depends on the ℓ_2 norm of $\hat{\mathbf{R}}^{\dagger}$. Moreover, in order to make constants C_1 , C_2 , and C_3 independent of the

dual certificate $\mathbf{y} \in \mathbb{R}^{|\mathcal{D}^c|}$ – see condition b) within the proof of Theorem 1 – we have used that $\|\mathbf{y}\|_2 \leq N$ and $\|\mathbf{y}_{\mathcal{K}^c}\|_\infty \leq \psi_{\mathbf{R}}$, where the first one follows from the fact that the magnitude of every element in \mathbf{y} is at most 1 and the second one was shown after (32).

A similar procedure can be used to show the result pertaining the case where $\mathcal{S} = \tilde{\mathcal{S}}_L$.

APPENDIX C PROOF OF THEOREM 3

With $\mathbf{s} = \text{vec}(\mathbf{S})$ and $\mathbf{s}_{\bar{K}} = \text{vec}(\mathbf{S}_{\bar{K}})$, we reformulate (20) for $\mathcal{S} = \mathcal{S}_A$ as

$$\min_{\{\mathbf{s}, \mathbf{s}_{\bar{K}}, \lambda\}} \|\mathbf{s}\|_1$$

$$\text{s.t. } \mathbf{s} = \mathbf{s}_{\bar{K}} + \mathbf{W}_K \lambda, \quad \mathbf{s}_{\mathcal{D}} = \mathbf{0}, \quad \mathbf{B}\mathbf{s} = \mathbf{0},$$

$$(\mathbf{e}_1 \otimes \mathbf{1}_N)^T \mathbf{s} = 1, \quad (\mathbf{I} \otimes \mathbf{V}_K^T) \mathbf{s}_{\bar{K}} = \mathbf{0}. \quad (35)$$

The first and last constraints in (35) correspond to the first and last constraints in (20) written in vector form. The second constraint in (35) imposes that \mathbf{S} has no self-loops, the third one imposes symmetry on \mathbf{S} , and the fourth one normalizes the first column of \mathbf{S} to sum up to 1 [cf. (7)]. Notice that the non-negativity constraint in \mathcal{S}_A is ignored in (35); however, if we show that (35) can recover the sparse solution \mathbf{s}_0^* , then the same solution would be recovered by the more constrained problem (20). Using the first constraint to solve for λ , we obtain $\lambda = \mathbf{W}_K^{\dagger} (\mathbf{s} - \mathbf{s}_{\bar{K}})$. Moreover, defining the concatenated variable $\mathbf{t} := [\mathbf{s}^T, \mathbf{s}_{\bar{K}}^T]^T$, it follows from the definitions of Υ and \mathbf{P} that (35) can be reformulated as

$$\min_{\mathbf{t}} \|\Upsilon \mathbf{t}\|_1 \quad \text{s.t. } \mathbf{P}^T \mathbf{t} = \mathbf{b}, \quad (36)$$

where \mathbf{b} is a vector with every entry equal to 0 except for the last one which is equal to 1. We utilize existing results on ℓ_1 -analysis [41] to state that the solution to (36) coincides with the sparsest solution if:

- $\ker(\Upsilon_{\mathcal{J}^c}) \cap \ker(\mathbf{P}^T) = \{\mathbf{0}\}$; and
- There exists a vector $\mathbf{y} \in \mathbb{R}^{N^2}$ such that $\Upsilon^T \mathbf{y} \in \text{Im}(\mathbf{P})$, $\mathbf{y}_{\mathcal{J}} = \text{sign}(\mathbf{s}_{0^{\mathcal{J}}})$, and $\|\mathbf{y}_{\mathcal{J}^c}\|_\infty < 1$.

As was the case for Theorem 1, the proof now reduces to showing that conditions A-1) and A-2) in the statement of the theorem imply the above conditions a) and b).

From the specific form of Υ , the kernel of $\Upsilon_{\mathcal{J}^c}$ is a space of dimension $|\mathcal{J}| + N^2$ spanned by the set of canonical basis vectors \mathbf{e}_i of length $2N^2$ where $i \in \mathcal{J} \cup \{N^2 + 1, N^2 + 2, \dots, 2N^2\}$. Thus for a) to hold we need the matrix formed by the columns of \mathbf{P}^T indexed by $\mathcal{J} \cup \{N^2 + 1, \dots, 2N^2\}$ to be full column rank, as stated in condition A-1).

Finally, the procedure to show that A-2) implies b) follows the same steps as those detailed in the proof of Theorem 1 – from (30) onwards – and, thus, is omitted here.

REFERENCES

- [1] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *Proc. IEEE Workshop Statist. Signal Process.*, Palma de Mallorca, Spain, Jun. 26–29, 2016, pp. 54–58.

- [2] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from imperfect spectral templates," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 6–9, 2016, pp. 1465–1469.
- [3] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamic Processes on Complex Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [4] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. New York, NY, USA: Springer-Verlag, 2009.
- [5] D. Shuman, S. 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, May 2013.
- [6] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [7] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
- [8] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Dec. 1, 2016.
- [9] J. Mei and J. Moura, "Signal processing on graphs: Estimating the structure of a graph," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 5495–5499.
- [10] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," arXiv:1605.02569v3, 2016.
- [11] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 20–25, 2016, pp. 6508–6512.
- [12] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2016, pp. 920–929.
- [13] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," arXiv:1603.04667, 2016.
- [14] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 1, 2017.
- [15] B. Girault, "Stationary graph signals using an isometric graph translation," in *Proc. Eur. Signal Process. Conf.*, Aug. 2015, pp. 1516–1520.
- [16] O. Sporns, *Discovering the Human Connectome*. Boston, MA, USA: MIT Press, 2012.
- [17] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [18] B. M. Lake and J. B. Tenenbaum, "Discovering structure by learning sparse graph," in *Proc. Annu. Cogn. Sci. Conf.*, 2010, pp. 778–783.
- [19] M. Slawski and M. Hein, "Estimation of positive definite M-matrices and structure learning for attractive gaussian Markov random fields," *Linear Algebra Appl.*, vol. 473, pp. 145–179, 2015.
- [20] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Statist.*, vol. 34, pp. 1436–1462, 2006.
- [21] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 20–25, 2016, pp. 6350–6354.
- [22] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations," *PLoS Comput. Biol.*, vol. 9, Jun. 2013, Art. no. e1003068.
- [23] 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, no. 4, pp. 563–575, Aug. 2014.
- [24] 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. 9849–9854, 2004.
- [25] G. V. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 20–25, 2016, pp. 6315–6319.
- [26] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2503–2516, May 15, 2017.
- [27] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," arXiv:1611.01456, 2016.
- [28] D. S. Marks *et al.*, "Protein 3d structure computed from evolutionary sequence variation," *PLoS ONE*, vol. 6, no. 12, 2011, Art. no. e28766.
- [29] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E*, vol. 64, Jun. 2001, Art. no. 016132.
- [30] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Aug. 2017.
- [31] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*. New York, NY, USA: Springer-Verlag, 2003, pp. 144–158.
- [32] S. Boyd, "Convex optimization of graph Laplacian eigenvalues," *Proc. Int. Congr. Math.*, vol. 3, no. 1–3, pp. 1311–1319, 2006.
- [33] T. Biyikoğlu, J. Leydold, and P. Stadler, *Laplacian Eigenvectors of Graphs: Perron-Frobenius and Faber-Krahn Type Theorems*. Berlin, Germany: Springer-Verlag, 2007.
- [34] F. Chung, *Spectral Graph Theory*, vol. 92. Providence, RI, USA: American Mathematical Soc., 1997.
- [35] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4363–4378, Aug. 2016.
- [36] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Interpolation of graph signals using shift-invariant graph filters," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 210–214.
- [37] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *Proc. Allerton Conf. Commun. Control Comput.*, 2015, pp. 1406–1413.
- [38] S. Feizi, D. Marbach, M. Medard, and M. Kellis, "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nature Biotechnol.*, vol. 31, no. 8, pp. 726–733, 2013.
- [39] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.
- [40] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, Dec. 2008.
- [41] H. Zhang, M. Yan, and W. Yin, "One condition for solution uniqueness and robustness of both ℓ_1 -synthesis and ℓ_1 -analysis minimizations," *Adv. Comput. Math.*, vol. 42, pp. 1381–1399, 2013.
- [42] J. Ortega, *Numerical Analysis: A Second Course* (ser. Classics in Applied Mathematics). Philadelphia, PA, USA: SIAM, 1990.
- [43] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.
- [44] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
- [45] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, Jan. 2011.
- [46] Z. Tan, Y. C. Eldar, A. Beck, and A. Nehorai, "Smoothing and decomposition for analysis sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1762–1774, Apr. 2014.
- [47] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. [Online]. Available: <http://cvxr.com/cvx>
- [48] B. Bollobás, *Random Graphs*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [49] P. Hagmann *et al.*, "Mapping the structural core of human cerebral cortex," *PLoS Biol.*, vol. 6, no. 7, 2008, Art. no. e159.
- [50] C. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [51] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.



Santiago Segarra (M'16) received the B.Sc. degree in industrial engineering with highest honors (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina, in 2011, and the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical and systems engineering, both from the University of Pennsylvania, Philadelphia, PA, USA, in 2014 and 2016, respectively. Since 2016, he has been working as a Postdoctoral Research Associate with the Institute for Data, Systems, and Society at the Massachusetts Institute of Technology.

His research interests include network theory, data analysis, machine learning, and graph signal processing. He received the ITBA's 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, the Best Student Paper Awards at the 2015 Asilomar Conference and the 2016 IEEE Statistical Signal Processing Workshop, and the Best Paper Award at the 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop.



Antonio G. Marques (SM'13) received the Telecommunications Engineering degree and the Doctorate degree, both with highest honors, from the Carlos III University of Madrid, Madrid, Spain, in 2002 and 2007, respectively. In 2007, he became a Faculty of the Department of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as an Associate Professor. From 2005 to 2015, he held different visiting positions at the University of Minnesota, Minneapolis. In 2015 and

2016, he was a visitor scholar at the University of Pennsylvania, Philadelphia. His research interests include the areas of signal processing, networking, and communications, and is currently focusing on stochastic optimization of wireless and power networks, signal processing for graphs, and nonlinear network optimization. He has served the IEEE in a number of posts. He is currently an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and his work has been awarded in several conferences and workshops.



Alejandro Ribeiro (M'07) received the B.Sc. degree in electrical engineering from the Universidad de la República Oriental del Uruguay, Montevideo, Uruguay, in 1998, and the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, in 2005 and 2007, respectively. From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. After his M.Sc. and Ph.D. studies, in 2008 he joined the University of Pennsylvania (Penn), Philadelphia, where

he is currently the Rosenbluth Associate Professor in the Department of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena. His focus is on structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control. He received the 2014 O. Hugo Schuck Best Paper Award, the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and paper awards at the 2016 SSP Workshop, 2016 SAM Workshop, 2015 Asilomar SSC Conference, ACC 2013, ICASSP 2006, and ICASSP 2005. He is a Fulbright Scholar and a Penn Fellow.



Gonzalo Mateos (SM'17) received the B.Sc. degree in electrical engineering from the Universidad de la República (UdelaR), Montevideo, Uruguay, in 2005, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota (UofM), Minneapolis, MN, USA, in 2009 and 2011, respectively. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Rochester, where he is also affiliated with the Georgen Institute for Data Science. During 2013, he was a visiting scholar with the CS Department,

Carnegie Mellon University. From 2004 to 2006, he was a Systems Engineer at Asea Brown Boveri, Uruguay. From 2003 to 2006, he was an Assistant with the Department of Electrical Engineering, UdelaR. His research interests include the areas of statistical learning from Big Data, network science, wireless communications, and graph signal processing, and is currently focusing on algorithms, analysis, and application of statistical signal processing tools to dynamic network health monitoring, social, power grid, and Big Data analytics. His doctoral work has been recognized with the 2013 UofM's Best Dissertation Award (Honorable Mention) across all Physical Sciences and Engineering areas.