CHAPTER

Inference of Graph Topology

1

Gonzalo Mateos[a,*], Santiago Segarra[**] and Antonio G. Marques[†]

[*] *University of Rochester, Dept. of Electrical and Computer Engineering, Rochester, NY, United States.*
[**] *Massachusetts Institute of Technology, Institute for Data, Systems, and Society, Cambridge, MA, United States.* [†] *King Juan Carlos University, Dept. of Signal Theory and Communications, Madrid, Spain*
[a] *Corresponding:* gmateosb@ece.rochester.edu

CHAPTER OUTLINE HEAD

Table 1.1 Notation

$\mathbf{x}$	vector with i-th entry x_i
$\mathbf{X}$	matrix with (i, j)-th entry X_{ij}
$\mathcal{I}$	set
$\mathbf{X}_{\mathcal{I}}$	submatrix of $\mathbf{X}$ formed by the rows indexed by $\mathcal{I}$
$(\cdot)^T$	matrix transpose
$(\cdot)^\dagger$	matrix pseudo-inverse
$\text{vec}(\cdot)$	matrix vectorization operator
$\sigma_{\min}(\cdot)$	minimum singular value of argument matrix
$\otimes$	Kronecker product
$\odot$	Khatri-Rao (columnwise Kronecker) product
$\text{tr}\{\cdot\}$	matrix trace
$\|\mathbf{x}\|_p$	vector ℓ_p-norm
$\|\mathbf{X}\|_p$	vector ℓ_p-norm of the vectorized form of $\mathbf{X}$
$\|\mathbf{X}\|_{M(p)}$	induced matrix ℓ_p-norm
$\|\mathbf{X}\|_F := \sqrt{\text{tr}\{\mathbf{X}^T\mathbf{X}\}}$	matrix Frobenius norm
$\text{diag}(\mathbf{x})$	diagonal matrix with (i, i)-th entry x_i
$\mathbf{I}$	identity matrix
$\mathbf{0}$	all-zero vector
$\mathbf{1}$	all-one vector

1.1 INTRODUCTION

Coping with the challenges found at the intersection of Network Science and Big Data necessitates fundamental breakthroughs in modeling, identification, and controllability of distributed network processes – often conceptualized as signals defined on graphs [1, 2, 3]. For instance, graph-supported signals can model vehicle trajectories over road networks [4]; economic activity observed over a network of production flows between industrial sectors [5, 6]; infectious states of individuals susceptible to an epidemic disease spreading on a social network [7]; gene expression levels defined on top of gene regulatory networks [8, 9, 10]; brain activity signals supported on brain connectivity networks [11, 12, 13]; and media cascades that diffuse on online social networks [14, 15], to name a few. There is an evident mismatch between our scientific understanding of signals defined over regular domains (time or space) and graph-valued signals. Knowledge about time series was developed over the course of decades and boosted by real needs in areas such as communications, speech, or control. On the contrary, the prevalence of network-related signal processing problems and the access to quality network data are recent events [1].

Under the assumption that the signals are related to the topology of the graph where they are supported, the goal of graph signal processing (GSP) is to develop algorithms that fruitfully leverage this relational structure, and can make inferences

about these relationships when they are only partially observed [5, 16, 10]. Most GSP efforts to date assume that the underlying network is known, and then analyze how the graph's algebraic and spectral characteristics impact the properties of the graph signals of interest. However, such assumption is often untenable in practice and arguably most graph construction schemes are largely informal, distinctly lacking an element of validation. In studies of e.g., functional brain connectivity or regulation among genes, inference of nontrivial pairwise interactions between signal elements (i.e., blood-oxygen-level dependent time series per voxel or gene expression levels, respectively) is often the goal per se.

In this chapter we present a framework to leverage information available from graph signals to learn the underlying undirected graph topology. The unknown graph represents direct relationships between signal elements, which one aims 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 operator that, while being consistent with the provided spectral information, endows the network with certain desired properties such as sparsity or minimum-energy edge weights.

1.2 GRAPH INFERENCE: A HISTORICAL OVERVIEW

Network topology inference is a prominent problem in Network Science [10, 17]. Since networks typically encode similarities between nodes, several topology inference approaches construct graphs whose edge weights correspond to nontrivial correlations or coherence measures between signal profiles at incident nodes. In this vein, informal (but widely popular) scoring methods rely on ad hoc thresholding of user-defined score functions such a Pearson product-moment correlation, Spearman rank correlation or mutual information. Formal hypothesis testing methods to assess non-trivial correlations have been proposed as well [10, Ch. 7.3.1]. Since in graph inference settings one performs a hypothesis test per node pair, the problem of multiple testing is prevalent and often addressed via e.g., false discovery rate (FDR) procedures.

Acknowledging that the observed correlations can be due to latent network effects, alternative statistical methods rely on inference of full partial correlations to eliminate potential confounding variables [10, Ch. 7.3.2]. Under Gaussianity assumptions, this line of work has well-documented connections with covariance selection [18] and sparse precision matrix estimation [19, 20, 21, 22, 23], as well as high-dimensional sparse linear regression [24]. Extensions to directed graphs include structural equation models (SEMs) [25, 14, 26], Granger causality [27, 17], or their nonlinear (e.g., kernelized) variants [28, 29].

Recent GSP-based network inference frameworks postulate instead 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 [30, 31, 32, 33, 34, 35]. For instance, network structure is estimated in [33] to unveil unknown relations among nodal time series adhering to an autoregressive model involving graph filter dynamics. Different from [32, 34, 33, 35] that operate on the graph domain, the goal here is to identify graphs that endow the given observations with desired spectral (frequency-domain) characteristics. Two works have recently explored this approach by identifying a graph-shift operator given its eigenvectors [30, 31], and rely on observations of stationary graph signals [36, 37, 38]. Different from [34, 35, 39, 40] that infer structure from signals assumed to be smooth over the sought graph, here the measurements are assumed related to the graph via filtering (e.g., modeling the diffusion of an idea or the spreading of a disease). Smoothness models are subsumed as special cases encountered when diffusion filters have a low-pass frequency response.

1.3 GRAPH INFERENCE FROM DIFFUSED SIGNALS

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 $\mathbf{D} := \text{diag}(\mathbf{A1})$. These are used to form the combinatorial Laplacian matrix $\mathbf{L}_c := \mathbf{D} - \mathbf{A}$ and the normalized Laplacian $\mathbf{L} := \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. More broadly, one can define a generic graph-shift operator (GSO) $\mathbf{S} \in \mathbb{R}^{N\times N}$ as any matrix whose off-diagonal sparsity pattern is equal to that of the adjacency matrix of $\mathcal{G}$ [2]. 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 $\mathbf{L}$.

1.3.1 STRUCTURE OF A NETWORK DIFFUSION PROCESS

The main focus of this chapter is on identifying graphs that explain the structure of a random signal. Formally, let $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$ be a graph signal in which the 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 $\mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right] = \mathbf{I}$. 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

$$\mathbf{x} \;=\; \alpha_0 \prod_{l=1}^{\infty}(\mathbf{I} - \alpha_l \mathbf{S})\mathbf{w} \;=\; \sum_{l=0}^{\infty}\beta_l \mathbf{S}^l \mathbf{w}. \tag{1.1}$$

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. [41]. The product and sum representations in (1.1) 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 (1.1). These include processes generated by graph filters with time-varying coefficients or those generated by the so-called *diffusion* Laplacian *kernels* [42], 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 (1.1) generates indirect relationships. In this context, the network topology inference problem 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

$$\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^T. \tag{1.2}$$

Further observe that while the diffusion expressions in (1.1) 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 (1.1) can be rewritten as

$$\mathbf{x} = \Big(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \Big) \mathbf{w} = \mathbf{H}\mathbf{w}, \tag{1.3}$$

for some particular $\mathbf{h}$ and L. Since a graph filter $\mathbf{H}$ is a polynomial on $\mathbf{S}$ [2], 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 arguments here, the filter representation in (1.3) can be used to show that *the eigenvectors of* $\mathbf{S}$ *are also eigenvectors of the covariance matrix* $\mathbf{C}_x := \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]$. To that end, notice that because $\mathbf{w}$ is white the said covariance is given by

$$\mathbf{C}_x = \mathbb{E}\left[\mathbf{H}\mathbf{w}(\mathbf{H}\mathbf{w})^T\right] = \mathbf{H}\mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right]\mathbf{H}^T = \mathbf{H}\mathbf{H}^T. \tag{1.4}$$

If we further use the spectral decomposition of the shift in (1.2) to express the filter as $\mathbf{H} = \sum_{l=0}^{L-1} h_l (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l)\mathbf{V}^T$, we can write the covariance matrix as

$$\mathbf{C}_x = \mathbf{V}\Big(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l \Big)^2 \mathbf{V}^T. \tag{1.5}$$

A consequence of (1.5) 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 (1.5), which includes indirect relationships between components, and $\mathbf{S}$ in (1.2), which includes exclusively direct relationships, is only on their *eigenvalues*. While the diffusion in (1.1) obscures the eigenvalues of $\mathbf{S}$, the eigenvectors $\mathbf{V}$ remain present in $\mathbf{C}_x$ as templates of the original spectrum.

Identity (1.5) also shows that the problem of finding a GSO that generates $\mathbf{x}$ from a white input $\mathbf{w}$ with unknown coefficients [cf. (1.1)] is *underdetermined*. As long as the matrices $\mathbf{S}$ and $\mathbf{C}_x$ have the same eigenvectors, filter coefficients that generate $\mathbf{x}$ through a diffusion process on $\mathbf{S}$ exist. In fact, the covariance matrix $\mathbf{C}_x$ itself is a GSO that can generate $\mathbf{x}$ through a diffusion process and so is the precision matrix $\mathbf{C}_x^{-1}$. To sort out this ambiguity, which amounts to selecting the eigenvalues of the shift, we assume that the GSO of interest is optimal in some sense. This pursuit is the subject of the next section, where we formally state the graph inference problem.

Remark 1 (Graph stationarity meets topology inference)**.** Recently, a group of works has generalized the definition of stationarity to graph processes [36, 37, 38]. In a nutshell, 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 [36, Def. 2], or if its covariance matrix is simultaneously diagonalizable with $\mathbf{S}$ [36, Def. 3]. These are precisely the conditions in (1.3) and (1.5), 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. Advances dealing with identification of undirected networks from diffused non-stationary graph signals are outlined in Section 1.5.

1.3.2 OPTIMAL GRAPH SHIFT OPERATOR

Given estimates $\hat{\mathbf{V}}$ of the filter eigenvectors (e.g., obtained from observations $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^R$ via the eigenvectors of the *sample covariance* $\hat{\mathbf{C}}_x = \frac{1}{R}\sum_{r=1}^R \mathbf{x}_r\mathbf{x}_r^T$), recovery of $\mathbf{S}$ amounts to selecting its eigenvalues $\mathbf{\Lambda}$ and to that end we assume that the shift of interest is optimal in some sense. At the same time, we should account for the discrepancies between $\hat{\mathbf{V}}$ and the actual eigenvectors of $\mathbf{S}$, due to finite sample size constraints and noise corrupting the observations in $\mathcal{X}$. Accordingly, we seek a shift operator $\mathbf{S}$ that: (a) is optimal with respect to (often convex) criteria $f(\mathbf{S})$; (b) belongs to a convex set $\mathcal{S}$ that specifies the desired type of shift operator (e.g., the adjacency $\mathbf{A}$ or Laplacian $\mathbf{L}$); and (c) is close to $\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T$ as measured by a convex matrix distance $d(\cdot,\cdot)$. Formally, one can solve

$$\mathbf{S}^* := \underset{\mathbf{\Lambda},\mathbf{S}\in\mathcal{S}}{\operatorname{argmin}} \; f(\mathbf{S}), \quad \text{s. to } \; d(\mathbf{S}, \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T) \leq \epsilon, \tag{1.6}$$

which is a convex optimization problem provided $f(\mathbf{S})$ is convex, and ϵ is a tuning parameter chosen based on a priori information on the imperfections. Within the scope of the signal model (1.1), the formulation (1.6) entails a general class of network topology inference problems parametrized by the choices in (a)-(c) above.

Following a formal statement of the problem, we briefly outline the spectrum of alternatives for points (a)-(c), while concrete choices are made for the analysis and numerical tests in Sections 1.4 and 1.5.

Problem statement. Given a set $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^{R}$ of R independent samples of the random signal $\mathbf{x}$ adhering to (1.1), estimate the optimal description of the structure of $\mathbf{x}$ in the form of the graph-shift operator $\mathbf{S}^* \in \mathcal{S}$ defined in (1.6).

Criteria. The selection of $f(\mathbf{S})$ allows to incorporate physical characteristics of the desired graph into the formulation, while being consistent with the spectral templates $\hat{\mathbf{V}}$. For instance, the matrix pseudo-norm $f(\mathbf{S}) = \|\mathbf{S}\|_0$ which counts the number of nonzero entries in $\mathbf{S}$ can be used to minimize the number of edges towards identifying sparse graphs (e.g., of direct relations among signal elements); $f(\mathbf{S}) = \|\mathbf{S}\|_1$ is a convex proxy for the aforementioned edge cardinality function. Alternatively, the Frobenius norm $f(\mathbf{S}) = \|\mathbf{S}\|_F$ can be adopted to minimize the energy of the edges in the graph, or $f(\mathbf{S}) = \|\mathbf{S}\|_\infty$ can be selected to obtain shifts $\mathbf{S}$ associated with graphs of uniformly low edge weights. This can be meaningful when identifying graphs subject to capacity constraints.

Constraints. The constraint $\mathbf{S} \in \mathcal{S}$ in (1.6) 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

$$\mathcal{S}_{\mathrm{A}} := \{\mathbf{S} \,|\, S_{ij} \geq 0,\ \ \mathbf{S} \in \mathcal{M}_N,\ \ S_{ii} = 0,\ \ \textstyle\sum_j S_{j1} = 1\}. \tag{1.7}$$

The first condition in $\mathcal{S}_{\mathrm{A}}$ encodes the non-negativity of the weights whereas the second condition incorporates that $\mathcal{G}$ 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 rules out the trivial solution $\mathbf{S} = \mathbf{0}$. Other GSOs such as the normalized Laplacian $\mathbf{L}$ can be accommodated in this framework via minor adaptations to $\mathcal{S}$; see [30].

The form of the convex matrix distance $d(\cdot, \cdot)$ depends on the particular application. For instance, if $\|\mathbf{S} - \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T\|_F$ is chosen, the focus is more on the similarities across the entries of the shifts, while $\|\mathbf{S} - \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T\|_M(2)$ focuses on their spectrum.

1.4 ROBUST NETWORK TOPOLOGY INFERENCE

This section deals with robust network inference problems from imperfect (noisy or incomplete) spectral templates.

1.4.1 NOISY SPECTRAL TEMPLATES

We first address the case where knowledge of an approximate version of the spectral templates $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N]$ is available, e.g., from the eigenvectors of a *sample* covariance matrix $\hat{\mathbf{C}}_x$. For the particular case of sparse shifts, adopting $f(\mathbf{S}) = \|\mathbf{S}\|_1$ as criterion in (1.6) and $d(\mathbf{S}, \hat{\mathbf{V}}\boldsymbol{\Lambda}\hat{\mathbf{V}}^T) = \|\mathbf{S} - \hat{\mathbf{V}}\boldsymbol{\Lambda}\hat{\mathbf{V}}^T\|_F$ yields

$$\mathbf{S}_1^* := \underset{\boldsymbol{\Lambda}, \mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \ \|\mathbf{S}\|_1, \quad \text{s. to } \|\mathbf{S} - \hat{\mathbf{V}}\boldsymbol{\Lambda}\hat{\mathbf{V}}^T\|_F \leq \epsilon \tag{1.8}$$

Note that $\|\mathbf{S}\|_1$ in (1.8) refers to the ℓ_1 norm of the vectorized version of $\mathbf{S}$. Moreover, further uncertainties can be introduced in the definition of the feasible set $\mathcal{S}$, e.g. in the scale of the admissible graphs for the case of $\mathcal{S} = \mathcal{S}_{\mathrm{A}}$ (cf. Proposition 1 and [30] for additional details).

To assess the effect of the noise in recovering the sparsest $\mathbf{S}$ (henceforth denoted as $\mathbf{S}_0^*$, the solution of (1.8) when $f(\mathbf{S}) = \|\mathbf{S}\|_0$), some additional notation must be introduced. Define $\hat{\mathbf{W}} := \hat{\mathbf{V}} \odot \hat{\mathbf{V}} \in \mathbb{R}^{N^2 \times N}$, where $\odot$ denotes the Khatri-Rao product. Let $\mathbf{s}_0^* := \mathrm{vec}(\mathbf{S}_0^*)$, denote by $\mathcal{D}$ the diagonal indices such that $(\mathbf{s}_0^*)_{\mathcal{D}} = \mathrm{diag}(\mathbf{S}_0^*)$ and partition its complement $\mathcal{D}^c$ into $\mathcal{K}$ and $\mathcal{K}^c$, with the former indicating the positions of the nonzero entries of $\mathbf{s}_{0\mathcal{D}^c}^* := (\mathbf{s}_0^*)_{\mathcal{D}^c}$, where matrix *calligraphic subscripts* select rows. Denoting by † the matrix pseudo-inverse, we define $\hat{\mathbf{M}} := (\mathbf{I} - \hat{\mathbf{W}}\hat{\mathbf{W}}^\dagger)_{\mathcal{D}^c} \in \mathbb{R}^{N^2 - N \times N^2}$, i.e., the orthogonal projector onto the kernel of $\hat{\mathbf{W}}^T$ constrained to the off-diagonal elements in $\mathcal{D}^c$. With $\mathbf{e}_1$ denoting the first canonical basis vector, we construct

$$\hat{\mathbf{R}} := [\hat{\mathbf{M}}, \ \mathbf{e}_1 \otimes \mathbf{1}_{N-1}] \ \in \mathbb{R}^{N^2 - N \times N^2 + 1}, \tag{1.9}$$

by horizontally concatenating $\hat{\mathbf{M}}$ and a column vector of size $|\mathcal{D}^c|$ with ones in the first $N-1$ positions and zeros elsewhere. Further, we drop the non-negativity constraint in $\mathcal{S}_{\mathrm{A}}$ – to obtain $\tilde{\mathcal{S}}_{\mathrm{A}}$ – and incorporate the scale ambiguity by augmenting $d(\mathbf{S}, \mathbf{S}')$ as $\tilde{d}(\mathbf{S}, \mathbf{S}') = (d(\mathbf{S}, \mathbf{S}')^2 + (\sum_j S_{j1} - 1)^2)^{1/2}$. With this notation, the following result on robust recovery of network topologies holds (see [30] for a proof).

Proposition 1. *Assuming that there exists at least one* $\mathbf{S}'$ *such that* $\tilde{d}(\mathbf{S}_0^*, \mathbf{S}') \leq \epsilon$*, the solution* $\hat{\mathbf{s}}_1^* := \mathrm{vec}(\hat{\mathbf{S}}_1^*)$ *to* (1.8) *for* $\mathcal{S} = \tilde{\mathcal{S}}_{\mathrm{A}}$ *with scale ambiguity satisfies*

$$\|\hat{\mathbf{s}}_1^* - \mathbf{s}_0^*\|_1 \leq C\epsilon, \quad \textit{with } \ C = 2C_1 + 2C_2C_3, \tag{1.10}$$

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

$$\psi := \|\mathbf{I}_{\mathcal{K}^c}(\delta^{-2}\hat{\mathbf{R}}\hat{\mathbf{R}}^T + \mathbf{I}_{\mathcal{K}^c}^T\mathbf{I}_{\mathcal{K}^c})^{-1}\mathbf{I}_{\mathcal{K}}^T\|_\infty < 1. \tag{1.11}$$

Constants C_1*,* C_2*, and* C_3 *are given by*

$$C_1 = \frac{\sqrt{|\mathcal{K}|}}{\sigma_{\min}(\hat{\mathbf{R}}_{\mathcal{K}}^T)}, \ C_2 = \frac{1 + \|\hat{\mathbf{R}}^T\|_2 C_1}{1 - \psi}, \ C_3 = \|\hat{\mathbf{R}}^\dagger\|_2 N, \tag{1.12}$$

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

When given noisy versions $\hat{\mathbf{V}}$ of the spectral templates of our target GSO, Proposition 1 quantifies the effect that the noise has on the recovery. More precisely, the recovered shift is guaranteed to be at a maximum distance from the desired shift bounded by the tolerance ϵ times a constant, which depends on $\hat{\mathbf{R}}$ and the support $\mathcal{K}$. This also implies that as the number of observed signals increases we recover the true GSO. In particular, as the number of observed signals increases, the sample covariance $\hat{\mathbf{C}}_x$ tends to the covariance $\mathbf{C}_x$ and, for the cases where the latter has no repeated eigenvalues, the noisy eigenvectors $\hat{\mathbf{V}}$ tend to the eigenvectors $\mathbf{V}$ of the desired shift; see, e.g., [43, Th. 3.3.7]. In particular, with better estimates $\hat{\mathbf{V}}$ the tolerance ϵ in (1.8) needed to guarantee feasibility can be made smaller, entailing a smaller discrepancy between the recovered $\mathbf{S}_1^*$ and the sparsest shift $\mathbf{S}_0^*$. In the limit when $\hat{\mathbf{V}} = \mathbf{V}$ and under no additional uncertainties, the tolerance ϵ can be made zero and (1.10) guarantees perfect recovery under conditions *1)* and *2)*.

1.4.2 INCOMPLETE SPECTRAL TEMPLATES

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

Formally, assume that the K first eigenvectors $\mathbf{V}_K = [\mathbf{v}_1, ..., \mathbf{v}_K]$ are those which are known. For simplicity of exposition, suppose as well that $\mathbf{V}_K$ is estimated error free. Then, the network topology inference problem with incomplete spectral templates can be formulated as

$$\bar{\mathbf{S}}_1^* := \underset{\mathbf{S}\in\mathcal{S},\mathbf{S}_{\bar{K}},\boldsymbol{\lambda}}{\operatorname{argmin}} \ \|\mathbf{S}\|_1, \quad \text{s. to } \ \mathbf{S} = \mathbf{S}_{\bar{K}} + \textstyle\sum_{k=1}^{K} \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \ \mathbf{S}_{\bar{K}}\mathbf{V}_K = \mathbf{0} \tag{1.13}$$

where we already particularized the objective to the ℓ_1-norm convex relaxation. The formulation in (1.13) enforces $\mathbf{S}$ to be partially diagonalized by the known spectral templates $\mathbf{V}_K$, while its remaining component $\mathbf{S}_{\bar{K}}$ is forced to belong to the orthogonal complement of range($\mathbf{V}_K$). Notice that, as a consequence, the rank of $\mathbf{S}_{\bar{K}}$ is at most $N - K$. An advantage of using only partial information of the eigenbasis as opposed to the whole $\mathbf{V}$ is that the set of feasible solutions in (1.13) is larger than that in (1.6). This is particularly important when the templates do not come from a prescribed GSO but, rather, one has the freedom to choose $\mathbf{S}$ provided it satisfies certain spectral properties (see [41] for examples in the context of distributed estimation).

GSO recovery guarantees can be derived for (1.13) [30]. To formally state these, define $\mathbf{W}_K := \mathbf{V}_K \odot \mathbf{V}_K$ and $\boldsymbol{\Upsilon} := [\mathbf{I}_{N^2}, \mathbf{0}_{N^2\times N^2}]$. Also, define matrices $\mathbf{B}^{(i,j)} \in \mathbb{R}^{N\times N}$

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

$$\mathbf{P}_1 := \begin{bmatrix} \mathbf{I} - \mathbf{W}_K \mathbf{W}_K^{\dagger} \\ \mathbf{I}_{\mathcal{D}} \\ \mathbf{B} \\ \mathbf{0}_{NK \times N^2} \\ (\mathbf{e}_1 \otimes \mathbf{1}_N)^T \end{bmatrix}^T, \quad \mathbf{P}_2 := \begin{bmatrix} \mathbf{W}_K \mathbf{W}_K^{\dagger} - \mathbf{I} \\ \mathbf{0}_{N \times N^2} \\ \mathbf{0}_{\binom{N}{2} \times N^2} \\ \mathbf{I} \otimes V_K^T \\ \mathbf{0}_{1 \times N^2} \end{bmatrix}^T, \tag{1.14}$$

and $\mathbf{P} := [\mathbf{P}_1^T, \mathbf{P}_2^T]^T$. With this notation and denoting by $\mathcal{J}$ the indices of the support of $\mathbf{s}_0^* = \text{vec}(\mathbf{S}_0^*)$, the following result is proved in [30].

Proposition 2. *Whenever $\mathcal{S} = \mathcal{S}_{\text{A}}$ and assuming problem* (1.13) *is feasible, $\tilde{\mathbf{S}}_1^* = \mathbf{S}_0^*$ if the two following conditions are satisfied:*
1) $\text{rank}([\mathbf{P}_{1\mathcal{J}}^T, \mathbf{P}_2^T]) = |\mathcal{J}| + N^2$*; and*
2) There exists a constant $\delta > 0$ such that

$$\eta := \|\boldsymbol{\Upsilon}_{\mathcal{J}^c}(\delta^{-2}\mathbf{P}\mathbf{P}^T + \boldsymbol{\Upsilon}_{\mathcal{J}^c}^T \boldsymbol{\Upsilon}_{\mathcal{J}^c})^{-1} \boldsymbol{\Upsilon}_{\mathcal{J}}^T\|_{\infty} < 1. \tag{1.15}$$

The proposition provides sufficient conditions for the relaxed problem in (1.13) to recover the sparsest graph, even when not all the eigenvectors are known. In practice it is observed that for smaller number K of known spectral templates the value of η in (1.15) tends to be larger, indicating a less favorable setting for recovery.

Notice that scenarios that combine the settings in Secs. 1.4.1 and 1.4.2, i.e., where the knowledge of the K templates is imperfect, can be handled by combining the formulations in (1.8) and (1.13). This can be achieved upon implementing the following modifications to (1.13): considering the shift $\mathbf{S}'$ as a new optimization variable, replacing the first constraint in (1.13) 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. (1.8)].

Laplacian graph shift operators. Counterparts to the optimizations in (1.8) and (1.13) as well as for the recovery guarantees in Propositions 1 and 2 can be derived for the case of (normalized) Laplacian operators. This requires changing the definition of $\mathcal{S}$ and accounting for the fact that the Laplacian has a zero eigenvalue; see the numerical tests in the following section and [30] for further details.

1.4.3 NUMERICAL TESTS

Here we test the proposed topology inference methods on different synthetic and real-world graphs. A comprehensive performance evaluation is carried out through comparisons with state-of-the-art methods and a test case to illustrate how our framework can promote sparsity on a given network.

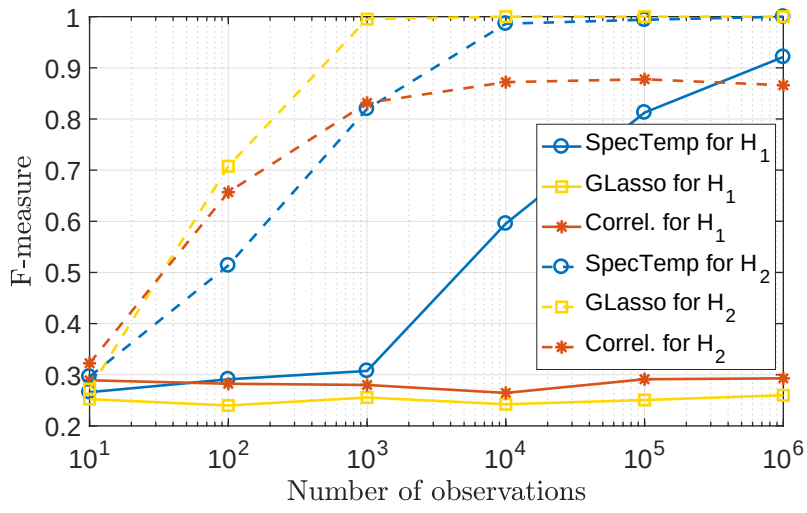


FIGURE 1.1 Comparison with baseline statistical methods

Performance comparison between the proposed SpecTemp approach (1.8), graphical lasso [19], and correlation-based recovery. For general filters, SpecTemp outperforms the competing alternatives.

Comparison with baseline statistical methods. First we analyze the performance of the topology inference algorithm (1.8) (henceforth referred to as SpecTemp) in comparison with two workhorse statistical methods, namely, (thresholded) correlation networks [10, Ch. 7.3.1] and graphical lasso [19]. The goal is to recover the adjacency matrix of an undirected and unweighted graph with no self-loops from the observation of filtered graph signals $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^R$. For the implementation of SpecTemp, we use the eigendecomposition of the sample covariance $\hat{\mathbf{C}}_x$ in order to extract noisy spectral templates $\hat{\mathbf{V}}$. We then solve problem (1.8) for $\mathcal{S} = \mathcal{S}_{\mathrm{A}}$, where ϵ is selected as the smallest value that admits a feasible solution. 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; see [30] for additional details. Lastly, for graphical lasso we follow the implementation in [19] based on $\hat{\mathbf{C}}_x$ and select the tuning parameter ρ (see [19]) during the training phase. We then force zeros on the diagonal and keep the absolute values of each entry.

We test the recovery of adjacency matrices $\mathbf{S} = \mathbf{A}$ of Erdős-Rényi (ER) random 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 powers 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 $\mathbf{H}_1 = \mathbf{V}\mathrm{diag}(\widehat{\mathbf{h}}_1)\mathbf{V}^T$, where the entries of $\widehat{\mathbf{h}}_1$ 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 cho-

sen so that $\delta_{\mathcal{H}}\mathbf{I} + \mathbf{S}$ is positive definite to ensure that $\mathbf{H}_2$ is real and well-defined. According to (1.4), 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, 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. 1.1 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, the proposed approach in (1.8) 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 SpecTemp and the correlation-based method. This is not surprising since graphical lasso was designed for the recovery of sparse precision matrices and is optimal (in the maximum-likelihood sense) for Gaussian signals. Notice however that for large number of observations SpecTemp, without assuming any specific filter model, also achieves perfect recovery and yields 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. Here we compare the network topology inference approach (1.6) with the state-of-the art algorithms in [34] and [35], which are designed to identify the Laplacian of a graph from observations of smooth graph signals. We select $f(\mathbf{S}) = \|\mathbf{S}\|_1$ and $d(\mathbf{S}_1, \mathbf{S}_2) = \|\mathbf{S}_1 - \mathbf{S}_2\|_F^2$ in (1.6), resulting in the SpecTemp formulation in (1.8). We study the recovery of the combinatorial Laplacian $\mathbf{S} = \mathbf{L}_c$ of Barabási-Albert preferential attachment graphs [44], with $N = 20$ generated from $m_0 = 4$ initially placed nodes, where each new node is connected to $m = 3$ existing ones. Following [35] we adopt two models for smooth graph signals: i) multivariate normal signals with covariance given by the pseudo-inverse

TABLE 1.2 Comparison with GSP methods

Performance comparison between the proposed SpecTemp approach (1.8), Kalofolias [35], and Dong et al [34].

	Inverse Laplacian			Diffusion		
Barabási-Albert	Proposed	Kalofolias	Dong et al	Proposed	Kalofolias	Dong et al
F-measure	**0.926**	0.855	0.873	**0.945**	0.845	0.894
edge error	**0.143**	0.173	0.209	**0.135**	0.154	0.235
degree error	**0.108**	0.124	0.169	0.109	**0.092**	0.188

of $\mathbf{L}_c$, i.e., $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{L}_c^\dagger)$; and ii) white signals filtered through an autoregressive (diffusion) process, that is $\mathbf{x}_2 = (\mathbf{I} + \mathbf{L}_c)^{-1}\mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For both settings we generate 10 training graphs, 100 testing graphs, and for every graph we generate $R = 1000$ graph signals. The training set is used to set the parameters in [34] and [35], and it serves the purpose of selecting the best ϵ [cf. (1.8)]. 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 $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. We assess performance via the F-measure, 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 1.2. In all but one case, SpecTemp attains the highest F-measures and the lowest errors for both signal models. Similar results were found for ER graphs; see [30].

Network deconvolution. The network deconvolution problem is the identification of an adjacency matrix $\mathbf{S} = \mathbf{A}$ that encodes direct dependencies when given an adjacency $\mathbf{T}$ that includes indirect relationships. The problem is a generalization of channel deconvolution and can be solved by making $\mathbf{T} = \mathbf{S}\,(\mathbf{I} - \mathbf{S})^{-1}$ [45]. This solution assumes a diffusion as in (1.1) but for the particular case of a single-pole-single-zero graph filter. A more general approach is to assume that $\mathbf{T}$ can be written as a polynomial of $\mathbf{S}$ but be agnostic to the form of the filter. This leads to problem formulation (1.8) with $\mathbf{V}$ given by the eigenvectors of $\mathbf{T}$. Note that here matrix $\mathbf{T}$ is not necessarily an empirical covariance matrix.

In this context, our goal is to identify the structural properties of proteins from a mutual information graph of the co-variation between the constitutional amino-acids [46]; see [45] for details. For example, for a particular protein, we want to recover the structural graph in Fig. 1.2 (a - left) when given the graph of mutual information in Fig. 1.2 (a - right). 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 [45] is illustrated in Fig. 1.2 (b - left) whereas the one recovered using SpecTemp is depicted in Fig. 1.2 (b - right). Comparing both recovered graphs, SpecTemp leads to a sparser graph that follows more closely the desired structure to be recovered. To quantify this latter assertion, in Fig. 1.2 (c) we plot the fraction of the real contact edges recovered for each method as a function of

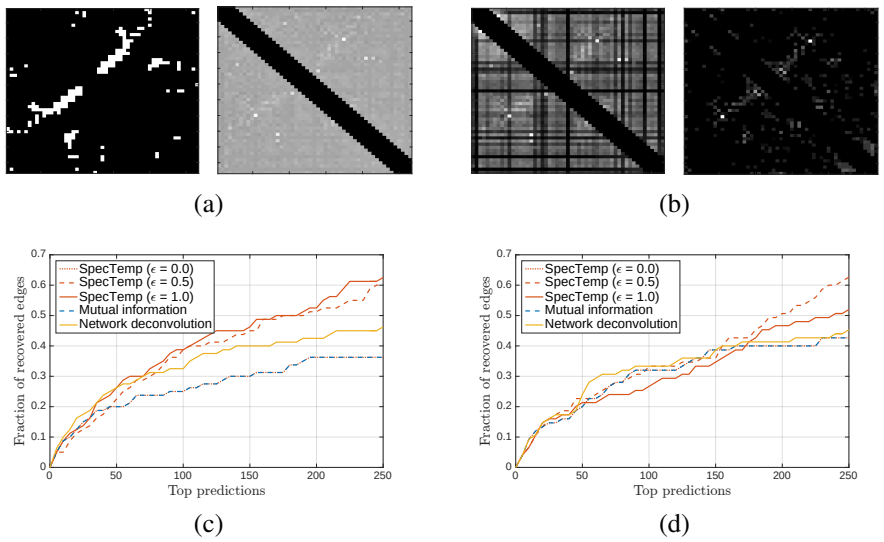


FIGURE 1.2 Identifying the structural properties of proteins

(a) Real and (b) inferred contact networks between amino-acid residues for protein BPT1 BOVIN. Ground truth contact network (a - left), mutual information of the co-variation of amino-acid residues (a - right), contact network inferred by network deconvolution (b - left), contact network inferred the proposed SpecTemp approach (1.8) (b - right). (c) Fraction of the real contact edges between amino-acids recovered for each method as a function of the number of edges considered. (d) Counterpart of (c) for protein YES HUMAN.

the number of edges considered, as done in [45]. 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 100 top edge predictions achieve a fraction of recovered edges equal to 0.4. As claimed in [45], network deconvolution improves the estimation when compared to raw mutual information data. Nevertheless, from Fig. 1.2 (c) it follows that SpecTemp outperforms network deconvolution. Notice that when $\epsilon = 0$ [cf. (1.8)] we are forcing the eigenvectors of $\mathbf{S}$ to coincide exactly with those of the matrix of mutual information $\mathbf{S}'$. However, since $\mathbf{S}'$ is already a valid adjacency matrix, we end up recovering $\mathbf{S} = \mathbf{S}'$. By contrast, for larger values of ϵ the additional flexibility in the choice of the eigenvectors allows us to recover shifts $\mathbf{S}$ that more closely resemble the ground truth. For example, when considering the top 200 edges, the mutual information and the network deconvolution methods recover 36% and 43% of the desired edges, respectively, while our method for $\epsilon = 1$ achieves a recovery of 53%. In Fig. 1.2 (d) we present this same analysis for a different protein and similar results can be appreciated.

1.5 NON-STATIONARY DIFFUSION PROCESSES

We now deal with more general *non-stationary* signals $\mathbf{x}$ that adhere to linear diffusion dynamics (1.1) in $\mathcal{G}$, but where the input covariance $\mathbf{C}_w = \mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right]$ can be arbitrary. In other words, we relax the assumption of $\mathbf{w}$ being white, which led to the stationary signal model dealt with so far (cf. Remark 1). Such a broader model is for instance relevant to (geographically) correlated sensor network data, or to models of opinion dynamics, where (even before engaging in discussion) the network agents can be partitioned into communities according to their standing on the subject matter.

For generic (non-identity) input covariance matrix $\mathbf{C}_w$, we face the challenge that the signal covariance [cf. (1.4)]

$$\mathbf{C}_x = \mathbf{H}\mathbf{C}_w\mathbf{H}^T \tag{1.16}$$

is no longer simultaneously diagonalizable with $\mathbf{S}$. This rules out using the eigenvectors of the sample covariance $\hat{\mathbf{C}}_x$ as spectral templates of $\mathbf{S}$. Still, as argued following (1.3) the eigenvectors of the GSO coincide with those of the graph filter $\mathbf{H}$ that governs the underlying diffusion dynamics. This motivates using snapshot observations $\mathcal{X} := \{\mathbf{x}_r\}_{r=1}^R$ together with additional information on the excitation input $\mathbf{w}$ (either realizations of the graph signal, sparsity assumptions, or its covariance matrix $\mathbf{C}_\mathbf{x}$ [26]) to *identify the filter* $\mathbf{H}$, with the ultimate goal of estimating its eigenvectors $\mathbf{V}$ [47]. These spectral templates are then used as inputs to the GSO identification problem (1.6), exactly as in the stationary setting of Sections 1.3 and 1.4. Accordingly, focus is henceforth placed on the graph filter (i.e., system) identification task.

1.5.1 LINEAR GRAPH FILTER IDENTIFICATION

Consider $m = 1, \ldots, M$ diffusion processes on $\mathcal{G}$, and assume that the observed non-stationary signal $\mathbf{x}_m$ corresponds to an input $\mathbf{w}_m$ diffused by an unknown graph filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$, which encodes the structure of the network via $\mathbf{S}$. In this section we show how additional knowledge about *realizations* of the input signals $\mathbf{w}_m$ can be used to identify $\mathbf{H}$ and, as a byproduct, its eigenvectors $\mathbf{V}$.

Input-output signal realization pairs. Suppose first that realizations of M input-output pairs $\{\mathbf{w}_m, \mathbf{x}_m\}_{m=1}^M$ are available, which can be arranged in the data matrices $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_M]$ and $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_M]$. The goal is to identify a symmetric filter $\mathbf{H} \in \mathcal{M}_N$ such that the observed signals $\mathbf{x}_m$ and the predicted ones $\mathbf{H}\mathbf{w}_m$ are close in some sense. In the absence of measurement noise this simply amounts to solving a system of M linear matrix equations

$$\mathbf{x}_m = \mathbf{H}\mathbf{w}_m, \quad m = 1, \ldots, M. \tag{1.17}$$

When noise is present, using the workhorse least-squares (LS) criterion the filter can be estimated as

$$\hat{\mathbf{H}} = \underset{\mathbf{H}\in\mathcal{M}_N}{\operatorname{argmin}} \sum_{m=1}^{M} \|\mathbf{x}_m - \mathbf{H}\mathbf{w}_m\|_2^2. \tag{1.18}$$

Because $\mathbf{H}$ is symmetric, the free optimization variables in (1.18) correspond to, say, the lower triangular part of $\mathbf{H}$, meaning the entries on and below the main diagonal. These $N_H := N(N+1)/2$ non-redundant entries can be conveniently arranged in the so-termed half-vectorization of $\mathbf{H}$, i.e., a vector $\text{vech}(\mathbf{H}) \in \mathbb{R}^{N_H}$ from which one can recover $\text{vec}(\mathbf{H}) \in \mathbb{R}^{N^2}$ via duplication. Indeed, there exists a unique duplication matrix $\mathbf{D}_N \in \{0,1\}^{N^2\times N_H}$ such that one can write $\mathbf{D}_N\text{vech}(\mathbf{H}) = \text{vec}(\mathbf{H})$. The Moore-Penrose pseudoinverse of $\mathbf{D}_N$, denoted as $\mathbf{D}_N^\dagger$, possesses the property $\text{vech}(\mathbf{H}) = \mathbf{D}_N^\dagger\text{vec}(\mathbf{H})$. With this notation in place, several properties of the solution $\hat{\mathbf{H}}$ of (1.18) are stated next.

Proposition 3. *Regarding the graph filter problem* (1.18)*, it holds that:*
a) The entries of the symmetric solution $\hat{\mathbf{H}}$ *are given by*

$$\text{vech}(\hat{\mathbf{H}}) = \left[(\mathbf{W}^T \otimes \mathbf{I}_N)\mathbf{D}_N\right]^\dagger \text{vec}(\mathbf{X}). \tag{1.19}$$

b) $\text{rank}\left((\mathbf{W}^T \otimes \mathbf{I}_N)\mathbf{D}_N\right) \leq N_H - (N - \text{rank}(\mathbf{W}) + 1)(N - \text{rank}(\mathbf{W}))/2.$
c) The minimizer of (1.18) *is unique if and only if* $\text{rank}(\mathbf{W}) = N$.

Proposition 3 asserts that if the excitation input set $\{\mathbf{w}_m\}_{m=1}^M$ is sufficiently rich – i.e., if $M \geq N$ and the excitation signals are linearly independent –, the entries of the diffusion filter $\mathbf{H}$ can be found as the solution of an LS problem. Interestingly, the fact of $\mathbf{H}$ having only $N_H = N(N+1)/2$ different entries cannot be exploited to reduce the number of input signals required to identify $\mathbf{H}$. The reason being that the matrix $(\mathbf{W}^T \otimes \mathbf{I}_N)\mathbf{D}_N$ is rank deficient if $\mathbf{W}^T$ has a nontrivial nullspace. Symmetry, however, can be exploited to enhance the estimation performance in overdetermined scenarios with noisy observations.

Once $\hat{\mathbf{H}}$ is estimated using (1.19), the next step is to decompose the filter as $\hat{\mathbf{H}} = \hat{\mathbf{V}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}^T$ and use $\hat{\mathbf{V}}$ as input for the GSO identification problem (1.6). Note that obtaining such an eigendecomposition is always possible since filter estimates $\hat{\mathbf{H}} \in \mathcal{M}_N$ are constrained to be symmetric.

1.5.2 QUADRATIC GRAPH FILTER IDENTIFICATION

In a number of applications, realizations of the excitation input $\mathbf{w}_m$ may be challenging to acquire, but information about the *statistical* description of $\mathbf{w}_m$ could still be available. To be specific, assume that the excitation inputs are zero mean and their covariance $\mathbf{C}_{w,m} = \mathbb{E}\left[\mathbf{w}_m\mathbf{w}_m^T\right]$ is known. Further suppose that for each input $\mathbf{w}_m$, we have access to a set of observations $\{\mathbf{x}_m^{(r)}\}_{r=1}^{R_m}$, which are then used to *estimate the out-*

put covariance as $\hat{\mathbf{C}}_{x,m} = \frac{1}{R_m}\sum_{r=1}^{R_m} \mathbf{x}_m^{(r)}(\mathbf{x}_m^{(r)})^T$. Since under (1.3) the true *covariance* is $\mathbf{C}_{x,m} = \mathbb{E}\left[x_m x_m^T\right] = \mathbf{H}\mathbf{C}_{w,m}\mathbf{H}^T$ [cf. (1.16)], the aim is to identify a filter $\mathbf{H}$ such that matrices $\hat{\mathbf{C}}_{x,m}$ and $\mathbf{H}\mathbf{C}_{w,m}\mathbf{H}^T$ are close.

Assuming for now perfect knowledge of the signal covariances, the above rationale suggests studying the solutions of the following system of matrix quadratic equations

$$\mathbf{C}_{x,m} = \mathbf{H}\mathbf{C}_{w,m}\mathbf{H}^T, \quad m = 1, \ldots, M. \tag{1.20}$$

To gain some initial insights, consider first the case where $M = 1$ and henceforth drop the subindex m so that we can write (1.20) as (1.16). Given the eigendecomposition of the symmetric and positive semidefinite (PSD) covariance matrix $\mathbf{C}_x = \mathbf{V}_x\mathbf{\Lambda}_x\mathbf{V}_x^T$, the *principal square root* of $\mathbf{C}_x$ is the unique symmetric and PSD matrix $\mathbf{C}_x^{1/2}$ which satisfies $\mathbf{C}_x = \mathbf{C}_x^{1/2}\mathbf{C}_x^{1/2}$. It is given by $\mathbf{C}_x^{1/2} = \mathbf{V}_x\mathbf{\Lambda}_x^{1/2}\mathbf{V}_x^T$, where $\mathbf{\Lambda}_x^{1/2}$ stands for a diagonal matrix with the nonnegative square roots of the eigenvalues of $\mathbf{C}_x$.

With this notation in place, introduce the matrices $\mathbf{C}_{wxw} := \mathbf{C}_w^{1/2}\mathbf{C}_x\mathbf{C}_w^{1/2}$ and $\mathbf{H}_{ww} := \mathbf{C}_w^{1/2}\mathbf{H}\mathbf{C}_w^{1/2}$. Clearly, $\mathbf{C}_{wxw}$ is both symmetric and PSD. Regarding the transformed filter $\mathbf{H}_{ww}$, note that by construction we have that $\mathbf{H}_{ww}$ is symmetric. Moreover, if $\mathbf{H}$ is assumed to be PSD, then so will be $\mathbf{H}_{ww}$. These properties will be instrumental towards characterizing the solutions of the matrix quadratic equation $\mathbf{C}_x = \mathbf{H}\mathbf{C}_w\mathbf{H}^T$ in (1.16), which can be recovered from the solutions $\mathbf{H}_{ww}$ of

$$\mathbf{C}_{wxw} = \mathbf{C}_w^{1/2}\mathbf{C}_x\mathbf{C}_w^{1/2} = \mathbf{C}_w^{1/2}\mathbf{H}\mathbf{C}_w\mathbf{H}\mathbf{C}_w^{1/2} = \mathbf{H}_{ww}\mathbf{H}_{ww}. \tag{1.21}$$

Positive semidefinite graph filters. Suppose that $\mathbf{H}$ is PSD (henceforth denoted $\mathbf{H} \in \mathcal{M}_N^{++}$), so that $\mathbf{H}_{ww}$ in (1.21) is PSD as well. Such filters arise, for example, in heat diffusion processes of the form $\mathbf{x} = (\sum_{l=0}^{\infty}\beta^l\mathbf{L}_c^l)\mathbf{w}$ with $\beta > 0$, where the Laplacian GSO $\mathbf{L}_c$ is PSD and the filter coefficients $h_l = \beta^l$ are all positive. In this setting, the solution of (1.21) is unique and given by the principal square root

$$\mathbf{H}_{ww} = \mathbf{C}_{wxw}^{1/2}. \tag{1.22}$$

Consequently, if $\mathbf{C}_w$ is nonsingular the definition of $\mathbf{H}_{ww}$ can be used to recover $\mathbf{H}$ via

$$\mathbf{H} = \mathbf{C}_w^{-1/2}\mathbf{C}_{wxw}^{1/2}\mathbf{C}_w^{-1/2}. \tag{1.23}$$

The previous arguments demonstrate that the assumption $\mathbf{H} \in \mathcal{M}_N^{++}$ gives rise to a strong identifiability result. Indeed, if $\{\mathbf{C}_{x,m}\}_{m=1}^M$ are known perfectly, the graph filter is identifiable even for $M = 1$.

However, in pragmatic settings where only empirical covariances are available, then observation of multiple ($M > 1$) diffusion processes can improve the performance of the system identification task. Given empirical covariances $\{\hat{\mathbf{C}}_{x,m}\}_{m=1}^M$ respectively estimated with enough samples R_m to ensure they are full rank, for each m define $\hat{\mathbf{C}}_{wxw,m} := \mathbf{C}_{w,m}^{1/2}\hat{\mathbf{C}}_{x,m}\mathbf{C}_{w,m}^{1/2}$. The quadratic equation (1.23) motivates solving

the LS problem

$$\hat{\mathbf{H}} = \underset{\mathbf{H}\in\mathcal{M}_N^{++}}{\operatorname{argmin}} \sum_{m=1}^{M} \|\hat{\mathbf{C}}_{wxw,m}^{1/2} - \mathbf{C}_{w,m}^{1/2}\mathbf{H}\mathbf{C}_{w,m}^{1/2}\|_F^2. \tag{1.24}$$

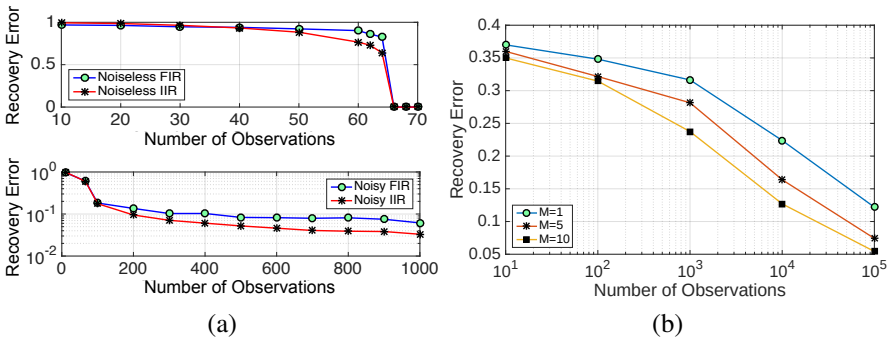
Whenever the number of samples R_m – and accordingly the accuracy of the empirical covariances $\hat{\mathbf{C}}_{x,m}$ – differs significantly across diffusion processes $m = 1, \ldots, M$, it may be prudent to introduce non-uniform coefficients to downweigh those residuals in (1.24) with inaccurate covariance estimates.

1.5.3 NUMERICAL TESTS

Here we study the recovery of two real-world graphs to assess the performance of the proposed network topology inference algorithms from non-stationary diffusion processes.

Brain graph. Consider a brain graph $\mathcal{G}$ with $N = 66$ nodes or neural regions and edge weights given by the density of anatomical connections between regions [48]. Denoting by $\mathbf{S} = \mathbf{A}$ the weighted adjacency of the brain graph, we consider two types of filters $\mathbf{H}_1 = \sum_{l=0}^{2} h_l \mathbf{A}^l$ and $\mathbf{H}_2 = (\mathbf{I} + \alpha\mathbf{A})^{-1}$, where the coefficients h_l and α are drawn uniformly on $[0, 1]$. We then generate M random input-output pairs $\{\mathbf{x}_m, \mathbf{w}_m\}_{m=1}^{M}$ (cf. Sec. 1.5.1), where signals are filtered by either $\mathbf{H}_1$ or $\mathbf{H}_2$, and estimate the filter using (1.19). Problem (1.8) with $\hat{\mathbf{V}}$ given by the eigenvectors of the estimated filter is then solved in order to infer the brain graph. In Fig. 1.3 (a - top) we plot the recovery error $\|\mathbf{S}_1^* - \mathbf{S}\|_F / \|\mathbf{S}\|_F$ as a function of M for both types of filters. First, notice that the performance is roughly independent of the filter type. More importantly, for $M \geq N$, the optimal filter estimation is unique (cf. Proposition 3) and leads to perfect recovery. We also consider the case where the observation of the output signals $\mathbf{x}_m$ is noisy; see Fig. 1.3(a - bottom). For this latter case, even though the estimation improves with increasing M, a larger number of observations is needed to guarantee successful recovery of the brain graph.

Social network. We consider the social network of a karate club studied by Zachary [49], represented by a graph $\mathcal{G}$ consisting of $N = 34$ nodes or members of the club and undirected edges symbolizing friendships among them. Denoting by $\mathbf{L}$ the normalized Laplacian of $\mathcal{G}$, we define the GSO $\mathbf{S} = \mathbf{I} - \alpha\mathbf{L}$ with $\alpha = 1/\lambda_{\max}(\mathbf{L})$, modeling the diffusion of opinions between the members of the club. A signal $\mathbf{x}$ can be regarded as a unidimensional opinion of each club member regarding a specific topic, and each application of $\mathbf{S}$ can be seen as an opinion update. Our goal is to recover $\mathbf{L}$ – hence, the social structure of the Karate club – from the observations of opinion profiles. We consider M different processes in the graph – corresponding, e.g., to opinions on M different topics – and assume that an opinion profile $\mathbf{x}_m$ is generated by the diffusion through the network of an initial signal $\mathbf{w}_m$. More precisely, for each topic $m = 1, \ldots, M$, we model $\mathbf{w}_m$ as a zero-mean process with known



(a) (b)

FIGURE 1.3 Inference of graph topology from non-stationary signals

(a) Brain network recovery error for FIR and IIR filters versus number of observed signals in noiseless (top) and noisy (bottom) settings. (b) Error in recovering a social network as a function of the number of opinion profiles observed and parametrized by the number of topics M.

covariance $\mathbf{C}_{w,m}$. We are then given a set $\{\mathbf{x}_m^{(r)}\}_{r=1}^R$ of opinion profiles generated from different sources $\{\mathbf{w}_m^{(r)}\}_{r=1}^R$ diffused through a filter of unknown nonnegative coefficients $\boldsymbol{\beta}$. From these R opinion profiles we build an estimate $\hat{\mathbf{C}}_{x,m}$ of the output covariance and, leveraging the fact that $\mathbf{S}$ is PSD and $\boldsymbol{\beta} \geq 0$ (cf. Sec. 1.5.2), we estimate the unknown filter $\hat{\mathbf{H}}$ by solving (1.24). Lastly, we use the eigenvectors $\hat{\mathbf{V}}$ of $\hat{\mathbf{H}}$ to solve (1.8), where $\mathcal{S}$ is modified accordingly for the recovery of a normalized Laplacian; see [30]. In Fig. 1.3 (b) we plot the shift recovery error as a function of the number of observations R and for three different values of M. As R increases, the estimate $\hat{\mathbf{C}}_{x,m}$ becomes more reliable entailing a better estimation of the underlying filter and, ultimately, leading to more accurate eigenvectors $\hat{\mathbf{V}}$. Hence, we observe a decreasing error with increasing R. Moreover, for a fixed R, the error in the estimation of $\hat{\mathbf{C}}_{x,m}$ can be partially overcome by observing multiple processes, thus, larger values of M lead to smaller errors.

1.6 DISCUSSION

With $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ being the shift operator associated with an undirected graph $\mathcal{G}$, we studied the problem of identifying $\mathbf{S}$ (hence the topology of $\mathcal{G}$) using a two-step approach where: (i) we first estimate the eigenvectors $\mathbf{V}$; and (ii) we then use $\mathbf{V}$ as input to find the eigenvalues $\boldsymbol{\Lambda}$ robustly via a convex optimization problem. Under the assumption that observed signals $\mathcal{X} = \{\mathbf{x}_r\}_{r=1}^R$ 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 $\mathcal{X}$. As a consequence, several well-established methods for topology identification based on Pearson and partial correlations can be viewed as particular instances of the approach

here presented. Contrasting with the stationary setting where $\mathbf{S}$ and the covariance matrix of the observed signals are simultaneously diagonalizable, for general (non-stationary) diffusion processes they are not. There is a workaround that entails estimating the unknown diffusion (graph) filter – a polynomial in the shift operator which preserves the sought eigenbasis $\mathbf{V}$. To carry out this initial system identification step, extra information is required on the input signal driving the diffusion process on the graph. Numerical tests showcase the effectiveness of the developed topology inference framework in recovering synthetic and real-world graphs.

ACKNOWLEDGMENTS

The work was supported in part by the Spanish MINECO grants OMICRON (TEC2013-41604-R) and KLINILYCS (TEC2016-75361-R).

REFERENCE

1. Shuman D, Narang S, Frossard P, Ortega A, Vandergheynst P, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process Mag 2013; 30(3):83–98.
2. Sandryhaila A, Moura J, Discrete signal processing on graphs. IEEE Trans Signal Process 2013; 61(7):1644–1656.
3. Sandryhaila A, Moura JMF, Big Data analysis with signal processing on graphs. IEEE Signal Process Mag 2014; 31(5):80–90.
4. Deri JA, Moura JMF, New york city taxi analysis with graph signal processing. In: Proc. IEEE Global Conf. on Signal and Information Process., 2016, pp. 1275–1279.
5. Marques AG, Segarra S, Leus G, Ribeiro A, Sampling of graph signals with successive local aggregations. IEEE Trans Signal Process 2016; 64(7):1832–1843.
6. Romero D, Ma M, Giannakis GB, Kernel-based reconstruction of graph signals. IEEE Trans Signal Process 2017; 65(3):764–778.
7. Barrat A, Barthelemy M, Vespignani A, Dynamical Processes on Complex Networks. New York, NY: Cambridge University Press, 2008.
8. Mittler R, Vanderauwera S, Gollery M, Breusegem FV, Reactive oxygen gene network of plants. Trends in Plant Science 2004; 9(10):490 – 498.
9. Segarra S, Huang W, Ribeiro A, Diffusion and superposition distances for signals supported on networks. IEEE Trans Signal Inform Process Networks 2015; 1(1):20–32.
10. Kolaczyk ED, Statistical Analysis of Network Data: Methods and Models. New York, NY: Springer, 2009.
11. Sporns O, Networks of the Brain. MIT press, 2011.
12. Huang W, Goldsberry L, Wymbs NF, Grafton ST, Bassett DS, Ribeiro A, Graph frequency analysis of brain signals. IEEE J Sel Topics Signal Process 2016; 10(7):1189–1203.
13. Ménoret M, Farrugia N, Pasdeloup B, Gripon V, Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction. In: Proc. IEEE Global Conf. on Signal and Information Process., 2017.

14. Baingana B, Mateos G, Giannakis GB, Proximal-gradient algorithms for tracking cascades over social networks. IEEE J Sel Topics Signal Process 2014; 8:563–575.
15. Gomez-Rodriguez M, Song L, Diffusion in social and information networks: Research problems, probabilistic models and machine learning methods. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015, 2015, pp. 2315–2316.
16. Chen S, Varma R, Sandryhaila A, Kovačević J, Discrete signal processing on graphs: Sampling theory. IEEE Trans Signal Process 2015; 63(24):6510–6523.
17. Sporns O, Discovering the Human Connectome. Boston, MA: MIT Press, 2012.
18. Dempster AP, Covariance selection. Biometrics 1972; 28(1):157–175.
19. Friedman J, Hastie T, Tibshirani R, Sparse inverse covariance estimation with the graphical lasso. Biostatistics 2008; 9(3):432–441.
20. Banerjee O, Ghaoui LE, dAspremont A, Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. Journal of Machine Learning Research 2008; 9:485–516.
21. Lake BM, Tenenbaum JB, Discovering structure by learning sparse graph. In: Annual Cognitive Sc. Conf., 2010, pp. 778 – 783.
22. Slawski M, Hein M, Estimation of positive definite M-matrices and structure learning for attractive gaussian markov random fields. Linear Algebra and its Applications 2015; 473:145–179.
23. Egilmez HE, Pavez E, Ortega A, Graph learning from data under laplacian and structural constraints. IEEE J Sel Topics Signal Process 2017; 11(6):825–841.
24. Meinshausen N, Buhlmann P, High-dimensional graphs and variable selection with the lasso. Ann Stat 2006; 34:1436–1462.
25. Cai X, Bazerque JA, Giannakis GB, Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations. PLoS, Computational Biology 2013; .
26. Shen Y, Baingana B, Giannakis GB, Tensor decompositions for identifying directed graph topologies and tracking dynamic networks. IEEE Trans Signal Process 2017; 65(14):3675–3687.
27. Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL, Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality. PNAS 2004; 101:9849–9854.
28. Karanikolas GV, Giannakis GB, Slavakis K, Leahy RM, Multi-kernel based nonlinear models for connectivity identification of brain networks. In: Proc. Int. Conf. Acoustics, Speech, Signal Process., Shanghai, China, Mar. 20-25, 2016.
29. Shen Y, Baingana B, Giannakis GB, Kernel-based structural equation models for topology identification of directed networks. IEEE Trans Signal Process 2017; 65(10):2503–2516.
30. Segarra S, Marques A, Mateos G, Ribeiro A, Network topology inference from spectral templates. IEEE Trans Signal Inf Process Netw 2017; 3(3):467–483.
31. Pasdeloup B, Gripon V, Mercier G, Pastor D, Rabbat MG, Characterization and inference of graph diffusion processes from observations of stationary signals. IEEE Trans Signal Inf Process Netw 2017; PP(99).
32. Thanou D, Dong X, Kressner D, Frossard P, Learning heat diffusion graphs. IEEE Trans Signal Inf Process Netw 2017; 3(3):484–499.
33. Mei J, Moura JMF, Signal processing on graphs: Causal modeling of unstructured data. IEEE Trans Signal Process 2017; 65(8):2077–2092.
34. Dong X, Thanou D, Frossard P, Vandergheynst P, Learning Laplacian matrix in smooth graph

signal representations. IEEE Trans Signal Process 2016; 64(23):6160–6173.
35. Kalofolias V, How to learn a graph from smooth signals. In: Intl. Conf. Artif. Intel. Stat. (AISTATS), J Mach. Learn. Res., 2016, pp. 920–929.
36. Marques AG, Segarra S, Leus G, Ribeiro A, Stationary graph processes and spectral estimation. IEEE Trans Signal Process 2017; 65(22):5911–5926.
37. Perraudin N, Vandergheynst P, Stationary signal processing on graphs. IEEE Trans Signal Process 2017; 65(13):3462–3477.
38. Girault B, Stationary graph signals using an isometric graph translation. In: Proc. of European Signal Process. Conf., 2015, pp. 1516–1520.
39. Chepuri SP, Liu S, Leus G, Hero AO, Learning sparse graphs under smoothness prior. In: Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, LA, Mar. 5-9, 2017, pp. 6508–6512.
40. Rabbat MG, Inferring sparse graphs from smooth signals with theoretical guarantees. In: Proc. Int. Conf. Acoustics, Speech, Signal Process., New Orleans, LA, Mar. 5-9, 2017, pp. 6533–6537.
41. Segarra S, Marques AG, Ribeiro A, Optimal graph-filter design and applications to distributed linear network operators. IEEE Trans Signal Process 2017; 65(15):4117–4131.
42. Smola AJ, Kondor R, Kernels and regularization on graphs. In: Learning theory and kernel machines, Springer, 2003; pp. 144–158.
43. Ortega J, Numerical Analysis: A Second Course. Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1990.
44. Bollobás B, Random Graphs. Cambridge University Press, 2001.
45. Feizi S, Marbach D, Medard M, Kellis M, Network deconvolution as a general method to distinguish direct dependencies in networks. Nat Biotech 2013; 31(8):726–733.
46. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al., Protein 3d structure computed from evolutionary sequence variation. PLoS ONE 2011; 6(12):e28766.
47. Shafipour R, Segarra S, Marques AG, Mateos G, Network topology inference from non-stationary graph signals. In: Proc. Int. Conf. Acoustics, Speech, Signal Process., 2017, pp. 5870–5874.
48. Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, et al., Mapping the structural core of human cerebral cortex. PLoS Biol 2008; 6(7):e159.
49. Zachary WW, An information flow model for conflict and fission in small groups. J of Anthropological Research 1977; 33(4):pp. 452–473.