

# Identifying the Topology of Undirected Networks From Diffused Non-Stationary Graph Signals

RASOUL SHAFIPOUR <sup>1</sup>, SANTIAGO SEGARRA <sup>2</sup> (Member, IEEE),  
ANTONIO G. MARQUES <sup>3</sup> (Senior Member, IEEE), AND GONZALO MATEOS <sup>4</sup> (Senior Member, IEEE)

<sup>1</sup>Microsoft, Redmond, WA 98052 USA

<sup>2</sup>Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA

<sup>3</sup>Department of Signal Theory and Communications, King Juan Carlos University, 28943 Madrid, Spain

<sup>4</sup>Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA

CORRESPONDING AUTHOR: GONZALO MATEOS (e-mail: gmateosb@ece.rochester.edu).

This work was supported in part by NSF Awards CCF-1750428, CCF-1934962, CCF-2008555, and ECCS-1809356, in part by Spanish MINECO Grants KLINILYCS (TEC2016-75361-R) and SPGRAPH (PID2019-105032GB-I00), and in part by MIT IDSS seed grant. This paper was presented in part at the 2017 and 2018 ICASSP Conferences [1], [2].

**ABSTRACT** We address the problem of inferring an undirected graph from nodal observations, which are modeled as non-stationary graph signals generated by local diffusion dynamics that depend on the structure of the unknown network. Using the so-called graph-shift operator (GSO), which is a matrix representation of the graph, we first identify the eigenvectors of the shift matrix from observations of the diffused signals, and then estimate the eigenvalues by imposing desirable properties on the graph to be recovered. Different from the stationary setting where the eigenvectors can be obtained directly from the covariance matrix of the measurements, here we need to estimate first the unknown diffusion (graph) filter – a polynomial in the GSO that preserves the sought eigenbasis. To carry out this initial system identification step, we exploit different sources of information on the arbitrarily-correlated input signal driving the diffusion on the graph. We first explore the setting where the observations, the input information, and the unknown graph filter are linearly related. We then address the case where the relation is given by a system of matrix quadratic equations, which arises in pragmatic scenarios where only the second-order statistics of the inputs are available. While such a quadratic filter identification problem boils down to a non-convex fourth-order polynomial minimization, we discuss identifiability conditions, propose algorithms to approximate the solution, and analyze their performance. Numerical tests illustrate the effectiveness of the proposed topology inference algorithms in recovering brain, social, financial, and urban transportation networks using synthetic and real-world signals.

**INDEX TERMS** Network topology inference, graph learning, graph signal processing, (non-)stationary graph processes, network diffusion, system identification, semidefinite relaxation.

## I. INTRODUCTION

Consider a network represented as a weighted and undirected graph  $\mathcal{G}$ , consisting 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 in the *symmetric* adjacency matrix  $\mathbf{A} \in \mathbb{R}^{N \times N}$ . More broadly, one can define a generic *graph-shift operator* (GSO)  $\mathbf{S} \in \mathbb{R}^{N \times N}$  as any matrix having the same sparsity pattern than that of  $\mathcal{G}$  [3]. Although the choice

of  $\mathbf{S}$  can be adapted to the problem at hand, it is typically chosen as either  $\mathbf{A}$ , the Laplacian  $\mathbf{L} := \text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$ , or their normalized counterparts [4].

Our 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 zero-mean graph signal with covariance matrix  $\mathbf{C}_{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ , in which the  $i$ th element  $x_i$  denotes the signal value at node  $i$  of an *unknown graph*  $\mathcal{G}$  with shift operator  $\mathbf{S}$ . We state that the graph  $\mathbf{S}$  represents the structure

of the signal  $\mathbf{y} \in \mathbb{R}^N$  if there exists a linear diffusion process in the GSO  $\mathbf{S}$  that generates the signal  $\mathbf{y}$  from the input signal  $\mathbf{x}$ , that is

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

for some set of parameters  $\{\alpha_l\}$  or, equivalently,  $\{\beta_l\}$ . While  $\mathbf{S}$  encodes local one-hop interactions, each successive application of the shift in (1) percolates  $\mathbf{x}$  over  $\mathcal{G}$ ; see e.g. [5]. The product and sum representations in (1) are common equivalent models for the generation of random network processes. Indeed, any process that can be understood as the linear propagation of an input signal through a static graph can be written in the form in (1), including heat diffusion [6], consensus and the classic DeGroot model of opinion dynamics [7], as well as symmetric structural equation models (SEMs) [8] as special cases. When  $\mathbf{x}$  is white so that  $\mathbf{C}_x = \mathbf{I}_N$ , (1) is equivalent to saying that the graph process  $\mathbf{y}$  is *stationary* in  $\mathbf{S}$ ; see e.g., [9, Def. 1], [10], [11] and Section II-A for further details. Here though, we deal with more general non-stationary signals  $\mathbf{y}$  that adhere to linear diffusion dynamics as in (1), but where the input covariance  $\mathbf{C}_x$  can be arbitrary. This is, for instance, relevant to (geographically) correlated sensor network data, urban population mobility patterns, or models of opinion dynamics among polarized groups.

The justification to state that  $\mathbf{S}$  represents the structure of  $\mathbf{y}$  is that we can think of the edges of  $\mathbf{S}$  as direct (one-hop) relations between the elements of the signal. The diffusion described by (1) modifies the original correlation by inducing indirect (multi-hop) relations. In this context, our goal is to recover the fundamental relations dictated by  $\mathbf{S}$  from a set of independent samples of a non-stationary random signal  $\mathbf{y}$ , as well as realizations of  $\mathbf{x}$ , or more pragmatically, knowledge of  $\mathbf{C}_x$ . This additional information on the input  $\mathbf{x}$  is the price paid to accommodate the more general non-stationary generative models for  $\mathbf{y}$ , and is not needed when identifying the structure of stationary graph signals [12], since  $\mathbf{C}_x = \mathbf{I}_N$  in that case.

*Relation to prior work.* 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 [3], [4]. Many 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 an assumption is often untenable in practice and arguably most graph construction schemes are largely informal, distinctly lacking an element of validation.

Network topology inference is a prominent task in Network Science [8, Ch. 7], [13]. 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 [8], [14]. Acknowledging that

the observed correlations can be due to latent network effects, alternative methods rely on inference of full partial correlations [8, Ch. 7.3.2]. Under Gaussianity assumptions, there are well-documented connections with covariance selection [15] and sparse precision matrix estimation [16]–[21], as well as neighborhood-based regression [22]. Extensions to directed graphs include SEMs [23], [24], Granger causality [14], [25], or their nonlinear (kernelized) variants [26], [27]. 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 [12], [13], [28]–[32]. Different from [29]–[32] that operate on the graph domain, the goal here is to identify graphs that endow the given observations with desired spectral (frequency-domain) characteristics. Works have recently explored this approach by identifying a GSO given its eigenvectors [12], [28], or given an estimate of the invertible diffusion kernel that generates the Gaussian observations [33], but all rely on measurements of stationary graph signals. Different from [31], [32], [34], [35] that infer structure from signals assumed to be smooth over the sought graph, here the measurements are related to the graph via filtering. Network processes that can be represented as graph filters [cf. (1)] include heat diffusion [29], consensus and opinion formation [36], and consumption levels in a network of connected consumers [37]. Moreover, widely adopted smoothness models [31], [32], [34], [35] are subsumed as special cases where graph filters have a low-pass frequency response.

*Paper outline.* In Section II, we formulate the problem of identifying a GSO that explains the fundamental structure of a random signal diffused on a graph. While for stationary  $\mathbf{y}$  the sought GSO shares its eigenvectors with the signal's covariance matrix [9], [10], [12], in the general (non-stationary) setting dealt with here this no longer holds and we elaborate on the ensuing challenges (Section II-A). Still, the graph's eigenvectors are preserved by the polynomial graph filter that governs the underlying diffusion dynamics (1). This motivates a novel two-step network topology inference approach whereby we: i) identify the GSO's eigenbasis from a judicious *graph filter estimate*; and ii) rely on these *spectral templates* to estimate the GSO's eigenvalues such that the inferred graph exhibits desirable structural characteristics such as sparsity; see also Section II-B. The estimation of the graph diffusion filter in step i), which is not required when the signals are stationary [12], has merit on its own and is of interest beyond topology inference. Feasibility of this additional system identification task requires extra information on the excitation signal  $\mathbf{x}$ . Section III addresses the setup where direct observations of the inputs are available so that the unknown filter matrix and the input-output signal pairs are linearly related. The focus in Section IV shifts to scenarios where second-order statistical information is used, and the relationship between the input-output covariances and the filter is quadratic. In practice, the output covariance matrix is estimated from signals, and we provide a (topology recovery) performance analysis for the

setting where the diffusion graph filter is known to be positive semidefinite (PSD). Identifiability conditions are discussed for the general (non-PSD) case. Section V develops projected gradient descent and semidefinite relaxation-based algorithms with complementary strengths, to deal with the (non-convex) fourth-order polynomial minimization associated with the recovery problem in Section IV. Numerical tests with synthetic and real-world data corroborate the effectiveness of the novel approach in recovering the topology of social, brain, financial, and transportation networks (Section VI). Concluding remarks and future directions are outlined in Section VII.

Relative to its conference precursors [1], [2], in this journal paper we consider identification of undirected graphs from observations of diffused non-stationary processes in various settings through a unified presentation along with full-blown technical details (including extended discussions and unpublished proofs in the Appendices for all the theoretical results). Some noteworthy novel pieces include: (i) a theoretical upper bound in the functional dependence of the end-to-end GSO estimation error in terms of the sample size; (ii) a projected gradient algorithm for graph filter identification with a number of desirable features relative to the alternating-directions method of multipliers variant in [1]; and (iii) a comprehensive performance evaluation protocol. Relative to the preliminary numerical tests conducted in [1], [2], here we incorporate more meaningful figures of merit, we study the effect of input signal richness and different graph topologies in the recovery performance, we expand the comparison with competing approaches – both classical statistical methods and more recent GSP algorithms–, and we include a real-world firm-clustering application using Yahoo stock price data.

*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; while  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix. For a vector  $\mathbf{x}$ ,  $\text{diag}(\mathbf{x})$  is a diagonal matrix whose  $i$ th diagonal entry is  $x_i$ . The operators  $\otimes$ ,  $\odot$ , and  $\text{vec}(\cdot)$  stand for Kronecker product, Khatri-Rao (column-wise Kronecker) product, and matrix vectorization, respectively.  $\|\mathbf{X}\|_p$  denotes the  $\ell_p$  norm of  $\text{vec}(\mathbf{X})$  whereas  $\|\mathbf{X}\|_{M(p)}$  is the matrix norm induced by the vector  $\ell_p$  norm. Lastly,  $\ker(\mathbf{X})$  refers to the null space of  $\mathbf{X}$  and the spectral radius of matrix  $\mathbf{X}$  is denoted by  $\lambda_{\max}(\mathbf{X})$ .

## II. PROBLEM STATEMENT

Consider the generative model in (1), whereby the properties of the graph signal  $\mathbf{y}$  depend on those of the excitation input  $\mathbf{x}$  and the underlying graph  $\mathcal{G}$  represented by the GSO  $\mathbf{S}$ . Given i.i.d. realizations of the output and prior information on the input, the goal is to infer a parsimonious graph representation that explains the structure of  $\mathbf{y}$ . Alternatively, we can state that the goal is to recover the sparse GSO which encodes direct relationships between the elements of  $\mathbf{y}$  from observable indirect relationships generated by a diffusion process.

To formally state the problem, consider the symmetric GSO  $\mathbf{S}$  associated with the undirected graph  $\mathcal{G}$ . Define the eigenvector matrix  $\mathbf{V} := [\mathbf{v}_1, \dots, \mathbf{v}_N]$  and the eigenvalue matrix  $\mathbf{\Lambda} := \text{diag}(\lambda_1, \dots, \lambda_N)$  to write  $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ . Now observe that while the diffusion expressions in (1) are polynomials on the GSO of possibly infinite degree, the Cayley-Hamilton theorem asserts that they are equivalent to polynomials of degree smaller than  $N$ . Upon defining the vector of coefficients  $\mathbf{h} := [h_0, \dots, h_{L-1}]^T \in \mathbb{R}^L$  and the symmetric graph filter  $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l \in \mathbb{R}^{N \times N}$  [3], the generative model in (1) can be rewritten as

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

for some particular  $\mathbf{h}$  and  $L \leq N$ . With this notation in place, we can formally define our problem of interest.

*Problem 1:* Given a set  $\mathcal{Y} := \{\mathbf{y}^{(p)}\}_{p=1}^P$  of  $P$  i.i.d. samples of a graph non-stationary random signal  $\mathbf{y}$  adhering to the network diffusion model (2) possibly contaminated by additive noise, identify the sparse GSO  $\mathbf{S}$  assuming:

- a) knowledge of the respective input signals  $\{\mathbf{x}^{(p)}\}_{p=1}^P$ ; or
- b) sparsity of the input signals; or
- c) that the inputs signals are drawn from a zero-mean distribution with known covariance matrix.

Problems 1-a and 1-b assume knowledge about the realizations of the inputs that give rise to the observations in  $\mathcal{Y}$ . These problems are respectively treated in Sections III-A and III-B. In contrast, Problem 1-c assumes statistical knowledge of the inputs but does not assume knowledge of the specific realizations of  $\mathbf{x}$ . We tackle this more challenging version of the problem in Section IV. In particular, we divide our analysis between the cases where  $\mathbf{H}$  in (2) can be assumed PSD (Section IV-A) and the more general setting where this assumption need not hold (Section IV-B).

Fundamental to the topology inference approach towards solving Problem 1 is that, because  $\mathbf{H}$  is a polynomial on  $\mathbf{S}$ , then: i) all such graph filters (spanned by the unknown coefficients  $\mathbf{h}$ ) have the *same eigenvectors*; and ii) such eigenvectors are the same as those of the shift, namely  $\mathbf{V}$ . In other words, while the diffusion implicit in  $\mathbf{H}$  obscures the eigenvalues of the GSO, the eigenvectors  $\mathbf{V}$  are preserved as *spectral templates* of the underlying network topology. It is worth mentioning that the implicit assumption in the eigenvectors preservation is that  $\sum_{l=0}^{L-1} h_l \lambda_i^l \neq 0$  for  $i = 1, \dots, N$  and  $\lambda_k \neq \lambda_l$  for  $k \neq l$  [cf. (2)]. This is a pragmatic assumption that we carry in the analysis, otherwise the eigenvector  $\mathbf{v}_j$  corresponding to  $\sum_{l=0}^{L-1} h_l \lambda_j^l = 0$  or  $\mathbf{v}_m$  and  $\mathbf{v}_n$  corresponding to  $\lambda_m = \lambda_n$  would be obscured. Next, Section II-A describes how to leverage (2) to obtain the GSO eigenbasis from a set of nodal observations  $\mathcal{Y}$ , by first estimating the unknown graph filter  $\mathbf{H}$ . We show that the information in  $\mathcal{Y}$  is in general not enough to uniquely recover  $\mathbf{H}$ . Hence, we will resort to additional knowledge on the input signal  $\mathbf{x}$  (either realizations, sparsity properties, or second-order statistical information) and also possibly on the structure of the graph filter  $\mathbf{H}$ . Section II-B outlines how to

use the spectral templates  $\mathbf{V}$  to recover the desired GSO by estimating its eigenvalues  $\Lambda$  and, as a byproduct, the sparse graph shift  $\mathbf{S} = \mathbf{V}\Lambda\mathbf{V}^T$  itself.

### A. STATIONARY VERSUS NON-STATIONARY OBSERVATIONS

Consider that the  $P$  observations in  $\mathcal{Y}$  correspond to independent realizations of a process  $\mathbf{y}$  adhering to the generative model in (2). The goal is to use  $\mathcal{Y}$  to estimate the spectral templates  $\mathbf{V}$  of the filter  $\mathbf{H}$  that governs the diffusion in (2).

To gain insights, suppose first that  $\mathbf{x}$  is white so that  $\mathbf{C}_x = \mathbf{I}_N$  [12]. Then the covariance matrix of  $\mathbf{y} = \mathbf{H}\mathbf{x}$  is

$$\mathbf{C}_y := \mathbb{E}[\mathbf{y}\mathbf{y}^T] = \mathbb{E}[\mathbf{H}\mathbf{x}(\mathbf{H}\mathbf{x})^T] = \mathbf{H}\mathbb{E}[\mathbf{x}\mathbf{x}^T]\mathbf{H} = \mathbf{H}^2. \quad (3)$$

In obtaining the third equality we used that  $\mathbf{H}$  is symmetric, because it is a polynomial in the symmetric GSO  $\mathbf{S}$ . If the GSO is drawn from a distribution of random graphs (see the numerical tests in Section VI-A), then the expectation in (3) should be understood as conditioned on  $\mathbf{S}$ . Using the spectral decomposition of  $\mathbf{S} = \mathbf{V}\Lambda\mathbf{V}^T$  to express the filter as  $\mathbf{H} = \sum_{l=0}^{L-1} h_l(\mathbf{V}\Lambda\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l\Lambda^l)\mathbf{V}^T$ , we can diagonalize the covariance matrix as

$$\mathbf{C}_y = \mathbf{V} \left( \sum_{l=0}^{L-1} h_l \Lambda^l \right)^2 \mathbf{V}^T. \quad (4)$$

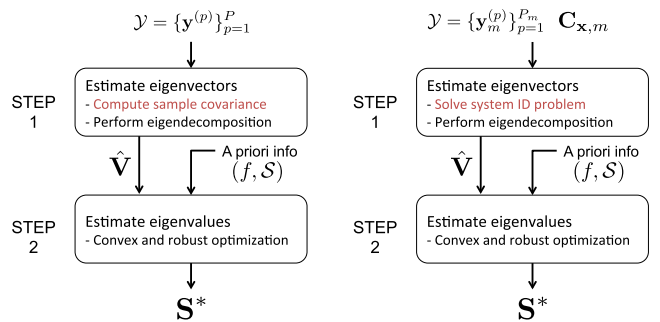
Such a covariance expression is precisely the requirement for a graph signal to be stationary in  $\mathbf{S}$  [9, Def. 2.b]. Remarkably, if  $\mathbf{y}$  is graph stationary, or equivalently if  $\mathbf{x}$  is white, (4) shows that the *eigenvectors* of the shift  $\mathbf{S}$ , the filter  $\mathbf{H}$ , and the covariance  $\mathbf{C}_y$  are *all the same*. As a result, to estimate  $\mathbf{V}$  from the observations  $\{\mathbf{y}^{(p)}\}_{p=1}^P$  one can form the *sample covariance*  $\hat{\mathbf{C}}_y = \frac{1}{P} \sum_{p=1}^P \mathbf{y}^{(p)}(\mathbf{y}^{(p)})^T$  and use its eigenvectors  $\hat{\mathbf{V}}$  as a consistent estimator of  $\mathbf{V}$ . The estimates  $\hat{\mathbf{V}}$  can be then used to recover  $\mathbf{S}$  [12], [28]; see also Section II-B.

In this context, the novel and broader focus of the present paper is on identifying the GSO  $\mathbf{S}$  that is considered to be the best possible description of the structure of a *non-stationary* signal  $\mathbf{y} = \mathbf{H}\mathbf{x}$  [cf. (2), where  $\mathbf{x}$  is not white]. We emphasize the distinction that graph non-stationarity of  $\mathbf{y}$  is meant with respect to the GSO  $\mathbf{S}$ . There is formally no temporal component to the problem here. Observations  $\{\mathbf{y}^{(p)}\}_{p=1}^P$  are i.i.d. and hence stationary in the usual sense.

For generic (non-identity) input covariance matrix  $\mathbf{C}_x$ , we face the challenge that the signal covariance [cf. (3)]

$$\mathbf{C}_y = \mathbf{H}\mathbf{C}_x\mathbf{H} \quad (5)$$

is no longer simultaneously diagonalizable with  $\mathbf{S}$ . This rules out using the eigenvectors of the sample covariance  $\hat{\mathbf{C}}_y$  as spectral templates of  $\mathbf{S}$ . Still, as argued following (2), the eigenvectors of the GSO coincide with those of the graph filter  $\mathbf{H}$  that governs the underlying diffusion dynamics. This motivates using realizations of observed signals together with additional information on the excitation inputs  $\mathbf{x}$  (either realizations of the graph signals, sparsity assumptions, or the covariance matrix  $\mathbf{C}_x$  [24]) to *identify the filter  $\mathbf{H}$* , with the ultimate goal of estimating its eigenvectors  $\mathbf{V}$ . This system



**FIGURE 1.** Schematic view of the two-step network inference method for (left) stationary and (right) non-stationary observations from diffusion processes. The main differences between both approaches lie in Step 1. For non-stationary processes, covariance matrices are no longer simultaneously diagonalizable with  $\mathbf{S}$ , thus requiring a more challenging system identification step in order to estimate  $\mathbf{H}$  (see Sections III-V). In both cases, the output of Step 1 is an estimate  $\hat{\mathbf{V}}$  of the eigenvectors of the sought shift. During Step 2, this estimate is combined with a priori information about the shift in an optimization problem to obtain the estimate  $\mathbf{S}^*$ , as described in Section II-B.

identification task in the graph setting is the subject dealt with in Sections III and IV, but before moving on, we close the loop by showing how to recover  $\mathbf{S}$  given its estimated eigenbasis  $\hat{\mathbf{V}}$ ; see also the comparative schematic in Fig. 1.

### B. USING THE SPECTRAL TEMPLATES TO RECOVER THE SHIFT

Given estimates  $\hat{\mathbf{V}}$  of the filter eigenvectors, recovery of  $\mathbf{S}$  amounts to estimating its eigenvalues  $\Lambda$ . Among all possible  $\Lambda$ , we seek eigenvalues that confer desirable structural properties to  $\mathbf{S}$ . In particular, the adopted criterion will be to find  $\Lambda$  in order to minimize the number of edges in  $\mathbf{S}$ , with the goal of identifying a sparse graph. 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 unavoidable errors in estimating the filter  $\mathbf{H}$ . Accordingly, we build on [12] and seek for the shift operator  $\mathbf{S}$  that: (a) is optimal with respect to sparsity-promoting criteria such as  $f(\mathbf{S}) = \|\mathbf{S}\|_1$ ; (b) belongs to a convex set  $\mathcal{S}$  that specifies the desired type of GSO (e.g., the adjacency  $\mathbf{A}$  or Laplacian  $\mathbf{L}$ ); and (c) is close to  $\hat{\mathbf{V}}\Lambda\hat{\mathbf{V}}^T$  as measured by the Frobenius-norm matrix distance. Formally, we solve the convex problem

$$\mathbf{S}^* := \underset{\Lambda, \mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \|\mathbf{S}\|_1, \quad \text{s. to } \|\mathbf{S} - \hat{\mathbf{V}}\Lambda\hat{\mathbf{V}}^T\|_F^2 \leq \epsilon, \quad (6)$$

where  $\epsilon$  is a tuning parameter chosen based on a priori information on the imperfections. 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 the convex set  $\mathcal{S}$  as follows

$$\mathcal{S}_A := \{\mathbf{S} \mid S_{ij} \geq 0, \mathbf{S} \in \mathcal{H}_N, S_{ii} = 0, \sum_j S_{j1} = 1\}. \quad (7)$$

The first condition in  $\mathcal{S}_A$  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  $\mathcal{H}_N$  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 solution  $\mathbf{S} = \mathbf{0}$ . Other GSOs (e.g., the Laplacian  $\mathbf{L}$  and its normalized variants) can be accommodated in our framework via minor modifications to  $\mathcal{S}$ . We refer to [12] for further details including a discussion of the identifiability and recovery performance associated with the problem in (7). Lastly, it should be noted that even when  $\mathbf{S}$  is a sparse matrix, the associated filter  $\mathbf{H}$  will, in general, be dense. Thus, sparsity cannot be leveraged in the identification of  $\mathbf{H}$ .

### III. (BI)LINEAR GRAPH FILTER IDENTIFICATION

Consider  $m = 1, \dots, M$  diffusion processes on  $\mathcal{G}$ , and assume that the observed non-stationary signal  $\mathbf{y}_m$  corresponds to an input  $\mathbf{x}_m$  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  $\mathbf{x}_m$  can be used to identify  $\mathbf{H}$  and, as a byproduct, its eigenvectors  $\mathbf{V}$ . We consider settings in which this extra information comes either from direct observation of  $\{\mathbf{x}_m\}_{m=1}^M$ , or through an assumption on input signal sparsity. In the context of online media-based marketing campaigns or rumor (fake news) diffusion over social networks, the observable input graph signal could correspond to the initial excitation instilled by those known (often paid) influencers [38]. In neuroscience, the observed inputs may represent controlled external stimuli aimed at exciting a few neural regions via e.g., transcranial magnetic stimulation [39], [40].

#### A. INPUT-OUTPUT SIGNAL REALIZATION PAIRS

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

$$\mathbf{y}_m = \mathbf{H}\mathbf{x}_m, \quad m = 1, \dots, M. \quad (8)$$

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

$$\mathbf{H}^* = \operatorname{argmin}_{\mathbf{H} \in \mathcal{H}_N} \sum_{m=1}^M \|\mathbf{y}_m - \mathbf{H}\mathbf{x}_m\|_2^2. \quad (9)$$

Because  $\mathbf{H}$  is symmetric, the free optimization variables in (9) correspond to, say, the lower triangular part of  $\mathbf{H}$ , meaning the entries on and below the main diagonal. These  $N_{\mathbf{H}} := N(N+1)/2$  non-redundant entries can be conveniently arranged in the so-termed half-vectorization of  $\mathbf{H}$ , i.e., a vector  $\operatorname{vech}(\mathbf{H}) \in \mathbb{R}^{N_{\mathbf{H}}}$  from which one can recover  $\operatorname{vec}(\mathbf{H}) \in \mathbb{R}^{N^2}$  via duplication. Indeed, there exists a unique duplication matrix  $\mathbf{D}_N \in \{0, 1\}^{N^2 \times N_{\mathbf{H}}}$  such that one can write  $\mathbf{D}_N \operatorname{vech}(\mathbf{H}) =$

$\operatorname{vec}(\mathbf{H})$  [41]. The MoorePenrose pseudoinverse of  $\mathbf{D}_N$ , denoted as  $\mathbf{D}_N^\dagger$ , possesses the property  $\operatorname{vech}(\mathbf{H}) = \mathbf{D}_N^\dagger \operatorname{vec}(\mathbf{H})$ . With this notation in place, several properties of the solution  $\mathbf{H}^*$  of (9) are stated next.

*Proposition 1:* Let  $M_r$  denote the rank of matrix  $\mathbf{X} \in \mathbb{R}^{N \times M}$ . Then, it holds that:

- a) The entries of the symmetric filter  $\mathbf{H}^*$  that solves (9) are

$$\operatorname{vech}(\mathbf{H}^*) = [(\mathbf{X}^T \otimes \mathbf{I}_N) \mathbf{D}_N]^\dagger \operatorname{vec}(\mathbf{Y}). \quad (10)$$

- b)  $\operatorname{rank}((\mathbf{X}^T \otimes \mathbf{I}_N) \mathbf{D}_N) \leq N_{\mathbf{H}} - \frac{(N-M_r+1)(N-M_r)}{2}$ .

- c) The minimizer of (9) is unique if and only if  $M_r = N$ .

*Proof:* See Appendix VII-A.

Proposition 1 asserts that if the excitation input set  $\{\mathbf{x}_m\}_{m=1}^M$  is sufficiently rich – i.e., if  $M \geq N$  and the excitation signals are linearly independent –, then the solution  $\mathbf{H}^*$  to (9) can be uniquely determined in closed form. Interestingly, the fact that  $\mathbf{H}$  has only  $N(N+1)/2$  different entries cannot be exploited to reduce the number  $M$  of input signals required for this uniqueness result. The reason is that the matrix  $(\mathbf{X}^T \otimes \mathbf{I}_N) \mathbf{D}_N$  is rank deficient if  $\mathbf{X}^T$  has a non-trivial null space. In other words, when using input-output pairs to estimate the filter  $\mathbf{H}$  one needs the same number of pairs for uniqueness in the solution to (9), regardless of whether the graph is symmetric or not.

As explained in Section II-B, once  $\mathbf{H}^*$  is estimated using (10), the next step is to decompose the filter as  $\mathbf{H}^* = \hat{\mathbf{V}} \hat{\mathbf{\Lambda}} \hat{\mathbf{V}}^T$  and use  $\hat{\mathbf{V}}$  as input for the GSO identification problem (6). Obtaining such an eigendecomposition is always possible since filter estimates  $\mathbf{H}^* \in \mathcal{H}_N$  are constrained to be symmetric.

#### B. SPARSE INPUT SIGNALS

It is not uncommon to encounter application domains in which the diffused graph signals adhere to linear network dynamics  $\mathbf{y} = \mathbf{H}\mathbf{x}$  and the input  $\mathbf{x}$  is sparse, having only a few nonzero entries [5], [7], [29], [42]. Sparsity in  $\mathbf{x}$  is well-motivated due to its practical relevance and modeling value – network processes such as  $\mathbf{y}$  are oftentimes the diffused version of few localized sources, hereby indexed by  $\operatorname{supp}(\mathbf{x}) := \{i : x_i \neq 0\}$  [29], [42]. For instance, opinion formation processes in social networks have been modeled using graph filters (see e.g., [5], [7]), and sparse  $\mathbf{x}$  could represent the initial opinion of those few influential actors that instilled the observed status-quo. Similar ideas are naturally relevant to linear network dynamics encountered with rumor spreading, adoption of new technologies, epidemic outbreaks [42], as well as with the identification of heat, pollutants, or seismic localized sources [29].

Given realizations of  $M$  diffusion processes  $\{\mathbf{y}_m\}_{m=1}^M$  arranged as columns of matrix  $\mathbf{Y} \in \mathbb{R}^{N \times M}$ , a possible *blind* formulation of the graph filter identification problem amounts to finding  $\mathbf{H} \in \mathcal{H}_N$  such that  $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|$  is small for some matrix norm of interest, where the unobserved matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]$  is assumed to be sparse. Different from Section III-A, the formulation is termed blind because input realizations are now unavailable and the resulting bilinear

problem entails finding the decomposition factors up to unavoidable scaling and permutation ambiguities. In the absence of noise, recent fundamental results and accompanying algorithms developed in [43] can be brought to bear here. Regarding identifiability, it is established therein that  $M = \mathcal{O}(N \log N)$  samples are sufficient to uniquely determine the decomposition with high probability, under the assumption that  $\mathbf{X}$  is generated by a sparsity-inducing Bernoulli-Gaussian or Bernoulli-Rademacher process [43, Theorem 1]. From an algorithmic standpoint, an efficient dictionary learning procedure called Exact Recovery of Sparsely-Used Dictionaries (ER-SpUD) is proposed that solves a sequence of linear programs with varying constraints. Under the aforementioned assumptions, the algorithm exactly recovers  $\mathbf{H}$  and  $\mathbf{X}$  with high probability. This result holds when: (i) the sparsity level measured by the expected number of nonzero elements in each column of  $\mathbf{X}$  is at most of order  $\sqrt{N}$ ; and (ii) the number of samples  $M$  is at least of order  $N^2 \log^2 N$ .

#### IV. QUADRATIC GRAPH FILTER IDENTIFICATION

In a number of applications, realizations of the excitation input process  $\mathbf{x}_m$  may be challenging to acquire, but information about the *statistical* description of  $\mathbf{x}_m$  could still be available. For instance, consider opinion formation models in social networks and suppose (e.g., socio-economic or demographic) features for each of the agents are measurable, while initial beliefs (namely the entries of  $\mathbf{x}_m$ ) are challenging to acquire. It is conceivable that agents with similar features will exhibit correlated beliefs (akin to homophily) [44], hence one could build a model of  $\mathbf{C}_{\mathbf{x},m} = \mathbb{E}[\mathbf{x}_m \mathbf{x}_m^T]$  based on said features. As a second example, consider financial networks comprising stocks as nodes and their interdependencies as links. Publicly-traded stock prices ( $\mathbf{y}_m$ ) are known to depend on stock purchases by investors ( $\mathbf{x}_m$ ), whose details are often hidden from the public for privacy and strategic reasons. On the other hand, each publicly-traded company may broadcast monthly statistical summaries of purchases of its stock ( $\mathbf{C}_{\mathbf{x},m}$ ); see for example the setting considered in [24] that also informs the experiments in Section VI-D. To be specific, assume that the excitation input processes are zero mean and their covariance  $\mathbf{C}_{\mathbf{x},m}$  is known. Further suppose that for each input process  $\mathbf{x}_m$ , we have access to a set of output observations  $\{\mathbf{y}_m^{(p)}\}_{p=1}^{P_m}$ , which are then used to estimate the output covariance as  $\hat{\mathbf{C}}_{\mathbf{y},m} = \frac{1}{P_m} \sum_{p=1}^{P_m} \mathbf{y}_m^{(p)} (\mathbf{y}_m^{(p)})^T$ . Since under (2) the *ensemble covariance* is  $\mathbf{C}_{\mathbf{y},m} = \mathbb{E}[\mathbf{y}_m \mathbf{y}_m^T] = \mathbf{H} \mathbf{C}_{\mathbf{x},m} \mathbf{H}$  [cf. (5)], the aim is to identify a filter  $\mathbf{H}$  such that matrices  $\hat{\mathbf{C}}_{\mathbf{y},m}$  and  $\mathbf{H} \mathbf{C}_{\mathbf{x},m} \mathbf{H}$  are close.

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

$$\mathbf{C}_{\mathbf{y},m} = \mathbf{H} \mathbf{C}_{\mathbf{x},m} \mathbf{H}, \quad m = 1, \dots, M. \quad (11)$$

To gain some initial insights, consider first the case where  $M = 1$  and henceforth drop the subindex  $m$  so that we can

write (11) as (5). Given the eigendecomposition of the symmetric and PSD covariance matrix  $\mathbf{C}_{\mathbf{y}} = \mathbf{V}_{\mathbf{y}} \mathbf{\Lambda}_{\mathbf{y}} \mathbf{V}_{\mathbf{y}}^T$ , the *principal square root* of  $\mathbf{C}_{\mathbf{y}}$  is the unique symmetric and PSD matrix  $\mathbf{C}_{\mathbf{y}}^{1/2}$  which satisfies  $\mathbf{C}_{\mathbf{y}} = \mathbf{C}_{\mathbf{y}}^{1/2} \mathbf{C}_{\mathbf{y}}^{1/2}$ . It is given by  $\mathbf{C}_{\mathbf{y}}^{1/2} = \mathbf{V}_{\mathbf{y}} \mathbf{\Lambda}_{\mathbf{y}}^{1/2} \mathbf{V}_{\mathbf{y}}^T$ , where  $\mathbf{\Lambda}_{\mathbf{y}}^{1/2}$  stands for a diagonal matrix with the nonnegative square roots of the eigenvalues of  $\mathbf{C}_{\mathbf{y}}$ .

With this notation in place, introduce the matrices  $\mathbf{C}_{\mathbf{xyx}} := \mathbf{C}_{\mathbf{x}}^{1/2} \mathbf{C}_{\mathbf{y}} \mathbf{C}_{\mathbf{x}}^{1/2}$  and  $\mathbf{H}_{\mathbf{xx}} := \mathbf{C}_{\mathbf{x}}^{1/2} \mathbf{H} \mathbf{C}_{\mathbf{x}}^{1/2}$ . Clearly,  $\mathbf{C}_{\mathbf{xyx}}$  is both symmetric and PSD. Regarding the transformed filter  $\mathbf{H}_{\mathbf{xx}}$ , note that by construction we have that  $\mathbf{H}_{\mathbf{xx}}$  is symmetric. Moreover, if  $\mathbf{H}$  is assumed to be PSD (see Section IV-A for examples of such filters), then so will be  $\mathbf{H}_{\mathbf{xx}}$ . These properties will be instrumental towards characterizing the solutions of the matrix quadratic equation  $\mathbf{C}_{\mathbf{y}} = \mathbf{H} \mathbf{C}_{\mathbf{x}} \mathbf{H}$  in (5), which can be equivalently recovered from the solutions  $\mathbf{H}_{\mathbf{xx}}$  of

$$\mathbf{C}_{\mathbf{xyx}} = \mathbf{C}_{\mathbf{x}}^{1/2} \mathbf{C}_{\mathbf{y}} \mathbf{C}_{\mathbf{x}}^{1/2} = \mathbf{C}_{\mathbf{x}}^{1/2} \mathbf{H} \mathbf{C}_{\mathbf{x}} \mathbf{H} \mathbf{C}_{\mathbf{x}}^{1/2} = \mathbf{H}_{\mathbf{xx}}^2. \quad (12)$$

This relationship has the same quadratic form as its counterpart for the stationary case [cf. (3)], with the identifications  $\mathbf{C}_{\mathbf{xyx}} \leftrightarrow \mathbf{C}_{\mathbf{y}}$  and  $\mathbf{H}_{\mathbf{xx}} \leftrightarrow \mathbf{H}$ .

*Remark 1. (Input realizations versus covariances):* If input signals are available as in Section III-A, these could be used to estimate input covariances. However, tackling the problem when input-output realization pairs are given entails solving a linear matrix equation (8). On the other hand, filter identification based on covariance information requires solving a system of quadratic matrix equations (11), which is in general more challenging. Moreover, estimating covariances a priori typically requires a larger sample relative to the one needed to solve the linear system identification task directly.

#### A. POSITIVE SEMIDEFINITE GRAPH FILTERS

Let us suppose first that  $\mathbf{H}$  is PSD (henceforth denoted  $\mathbf{H} \in \mathcal{H}_N^+$ ), so that  $\mathbf{H}_{\mathbf{xx}}$  in (12) is PSD as well. Such filters arise, for example, with heat diffusion processes of the form  $\mathbf{y} = (\sum_{l=0}^{\infty} \beta^l \mathbf{L}^l) \mathbf{x}$  with  $\beta > 0$ , where the Laplacian shift  $\mathbf{L}$  is PSD and the filter coefficients  $h_l = \beta^l$  are all positive. Another example of a PSD filter arises when  $\mathbf{H}$  models a finite-horizon application of consensus dynamics, i.e.,  $\mathbf{y} = (\mathbf{I} - \alpha \mathbf{L})^L \mathbf{x}$  for small enough diffusion rate  $\alpha$ . In this setting, the solution of (12) is unique and given by the principal square root

$$\mathbf{H}_{\mathbf{xx}} = \mathbf{C}_{\mathbf{xyx}}^{1/2}. \quad (13)$$

Consequently, if  $\mathbf{C}_{\mathbf{x}}$  is nonsingular (so that the excitation inputs are not degenerate), the definition of  $\mathbf{H}_{\mathbf{xx}}$  can be used to recover  $\mathbf{H}$  via

$$\mathbf{H} = \mathbf{C}_{\mathbf{x}}^{-1/2} \mathbf{C}_{\mathbf{xyx}}^{1/2} \mathbf{C}_{\mathbf{x}}^{-1/2}. \quad (14)$$

The previous arguments demonstrate that the assumption  $\mathbf{H} \in \mathcal{H}_N^+$  gives rise to a strong identifiability result. Indeed, if  $\{\mathbf{C}_{\mathbf{y},m}\}_{m=1}^M$  are known perfectly, a PSD 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  $\{\hat{\mathbf{C}}_{y,m}\}_{m=1}^M$  respectively estimated with enough samples  $P_m$  to ensure that they are of full rank, for each  $m$  define  $\hat{\mathbf{C}}_{\text{xyx},m} := \mathbf{C}_{x,m}^{1/2} \hat{\mathbf{C}}_{y,m} \mathbf{C}_{x,m}^{1/2}$ . The quadratic equation (12) motivates solving the LS problem

$$\mathbf{H}^* = \operatorname{argmin}_{\mathbf{H} \in \mathcal{H}_N^+} \sum_{m=1}^M \|\hat{\mathbf{C}}_{\text{xyx},m}^{1/2} - \mathbf{C}_{x,m}^{1/2} \mathbf{H} \mathbf{C}_{x,m}^{1/2}\|_F^2. \quad (15)$$

Whenever the number of samples  $P_m$  – and accordingly the accuracy of the empirical covariances  $\hat{\mathbf{C}}_{y,m}$  – differs significantly across diffusion processes  $m = 1, \dots, M$ , it may be prudent to introduce non-uniform coefficients to downweigh those residuals in (15) with inaccurate covariance estimates. The following proposition offers insights on the solution to (15), and extensions to weighted LS criteria are straightforward.

*Proposition 2:* Define the matrices  $\bar{\mathbf{X}} = [\mathbf{C}_{x,1}^{1/2} \otimes \mathbf{C}_{x,1}^{1/2}, \dots, \mathbf{C}_{x,M}^{1/2} \otimes \mathbf{C}_{x,M}^{1/2}]^T$  and  $\bar{\mathbf{Y}} = [\hat{\mathbf{C}}_{\text{xyx},1}^{1/2}, \dots, \hat{\mathbf{C}}_{\text{xyx},M}^{1/2}]^T$ . Then, the filter  $\mathbf{H}^*$  that solves (15) can be found as

$$\operatorname{vec}(\mathbf{H}^*) = \bar{\mathbf{X}}^\dagger \operatorname{vec}(\bar{\mathbf{Y}}^T). \quad (16)$$

Moreover, if  $M = 1$  and matrix  $\mathbf{C}_{x,1}$  is nonsingular, the minimizer  $\mathbf{H}^*$  is unique and given by

$$\mathbf{H}^* = \mathbf{C}_{x,1}^{-1/2} \hat{\mathbf{C}}_{\text{xyx},1}^{1/2} \mathbf{C}_{x,1}^{-1/2}. \quad (17)$$

*Proof:* To show (16) one can follow steps similar to those for (10) in Proposition 1. The identifiability result for  $M = 1$  follows from the arguments leading to (14), noting that the cost in (15) vanishes if  $\mathbf{H}^*$  is selected to satisfy  $\hat{\mathbf{C}}_{\text{xyx},1}^{1/2} = \mathbf{C}_{x,1}^{1/2} \mathbf{H}^* \mathbf{C}_{x,1}^{1/2}$ . Left and right multiplying both sides of the equality by  $\mathbf{C}_{x,1}^{-1/2}$ , (17) follows. ■

Particularizing our discussion to the case of  $M = 1$ , (17) reveals that as the number of observations  $P$  increases, the estimated graph filter  $\mathbf{H}^*$  tends to the true one [cf. (14)]. Upon performing the eigendecomposition of  $\mathbf{H}^*$  and solving (6), we ultimately obtain an estimate  $\mathbf{S}^*$  of the underlying graph. Naturally, an increase in  $P$  leads to a better estimate  $\mathbf{S}^*$ .

The remainder of this section is devoted to characterizing this dependence analytically. To this end, some notation must be introduced. Define  $\mathbf{W} = \mathbf{V} \odot \mathbf{V}$  and the set  $\mathcal{D}$  containing the indices of  $\operatorname{vec}(\mathbf{S})$  corresponding to the diagonal entries of  $\mathbf{S}$ . Furthermore, 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 non-zero entries in the true GSO  $\mathbf{S}$ . Based on this notation we define the matrix  $\mathbf{M} = (\mathbf{I} - \mathbf{W}\mathbf{W}^\dagger)_{\mathcal{D}^c}$  where the set  $\mathcal{D}^c$  selects rows of  $\mathbf{I} - \mathbf{W}\mathbf{W}^\dagger$ , i.e.,  $\mathbf{M}$  is the orthogonal projector onto the kernel of  $\mathbf{W}^T$  constrained to the off-diagonal elements in  $\mathcal{D}^c$ . Finally, with  $\mathbf{e}_1$  denoting the first canonical basis vector, we construct the matrix

$$\mathbf{R} = [\mathbf{M}, \mathbf{e}_1 \otimes \mathbf{1}_{N-1}]. \quad (18)$$

With this notation in place, the following recovery result holds.

*Proposition 3:* Consider the estimate  $\mathbf{S}_P^*$  obtained by solving (6), where the eigenvectors  $\hat{\mathbf{V}}$  are procured from the eigendecomposition of  $\mathbf{H}^*$  [cf. (17)] computed based on  $P$  observations in  $\mathcal{Y} := \{\mathbf{y}^{(p)}\}_{p=1}^P$ . Assume that:

- A-1) The input process  $\mathbf{x}$  is sub-Gaussian;
- A-2)  $\mathbf{H}$  has no repeated eigenvalues;
- A-3)  $\operatorname{rank}(\mathbf{R}_{\mathcal{K}}) = |\mathcal{K}|$ ;
- A-4) There exists a constant  $\mu > 0$  such that

$$\|\mathbf{I}_{\mathcal{K}^c} (\mu^{-2} \mathbf{R}\mathbf{R}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T\|_{M(\infty)} < 1.$$

Under these assumptions, there exists a  $P_0$  such that for all  $P > P_0$  we have that, with probability at least  $1 - \delta$

$$\|\mathbf{S}_P^* - \mathbf{S}\|_1 \leq C_\delta P^{-1/4}, \quad (19)$$

where the term  $C_\delta$  contains the dependencies on all other variables except for  $P$ .

*Proof:* See Appendix B. ■

Proposition 3 characterizes an upper bound in the functional dependence of the end-to-end estimation error of the proposed two-step procedure in terms of the number of observations  $P$ . We first observe that the error vanishes with increasing  $P$ , rendering  $\mathbf{S}^*$  in (6) a consistent estimator under the conditions stated in Proposition 3. Moreover, the functional form in (19) provides a guideline for the design of observational studies and data collection, revealing that, e.g., doubling the number of observations implies modifying the error bound by a factor of  $2^{-1/4} \approx 0.84$ , leading to an error reduction of 16%.

## B. SYMMETRIC GRAPH FILTERS

Consider now a more general (but not less relevant) setting whereby  $\mathbf{H}$  is only assumed to be symmetric, and once more let  $M = 1$  to simplify notation. Widely adopted (symmetric) SEMs of the form  $\mathbf{y} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{x}$  are subsumed by (2) and will not be PSD in general. The same holds true for arbitrary polynomials of the adjacency matrix of unweighted, undirected graphs with no self loops. With the unitary matrix  $\mathbf{V}_{\text{yx}}$  denoting the eigenvectors of  $\mathbf{C}_{\text{yx}}$  and with  $\mathbf{b} \in \{-1, 1\}^N$  being a binary (signed) vector, one can conclude that solutions of (12) have the general form (see Appendix C)

$$\mathbf{H}_{\text{xx}} = \mathbf{C}_{\text{yx}}^{1/2} \mathbf{V}_{\text{yx}} \operatorname{diag}(\mathbf{b}) \mathbf{V}_{\text{yx}}^T. \quad (20)$$

If the input covariance matrix  $\mathbf{C}_x$  is nonsingular, all symmetric solutions  $\mathbf{H} \in \mathcal{H}_N$  of (12) [and hence of (5)] are given by

$$\mathbf{H} = \mathbf{C}_x^{-1/2} \mathbf{C}_{\text{yx}}^{1/2} \mathbf{V}_{\text{yx}} \operatorname{diag}(\mathbf{b}) \mathbf{V}_{\text{yx}}^T \mathbf{C}_x^{-1/2}. \quad (21)$$

In the absence of the PSD assumption, the problem for  $M = 1$  is non-identifiable. Inspection of (21) shows there are  $2^N$  possible solutions to the quadratic equation (5), which are parametrized by the binary vector  $\mathbf{b}$ . For the PSD setting in Section IV-A the solution is unique and corresponds to  $\mathbf{b} = \mathbf{1}$ .

For  $M > 1$ , the set of feasible solutions to the system of equations (11) is naturally given by

$$\mathcal{H}_M^{\text{sym}} = \bigcap_{m=1}^M \{ \mathbf{H} \in \mathcal{H}_N \mid \mathbf{b}_m \in \{-1, 1\}^N \text{ and} \\ \mathbf{H} = \mathbf{C}_{x,m}^{-1/2} \mathbf{C}_{yx,m}^{1/2} \mathbf{V}_{yx,m} \text{diag}(\mathbf{b}_m) \mathbf{V}_{yx,m}^T \mathbf{C}_{x,m}^{-1/2} \}. \quad (22)$$

It is thus conceivable that as  $M$  grows and, therefore, the number of equations increases, the cardinality of  $\mathcal{H}_M^{\text{sym}}$  shrinks and the problem is rendered identifiable (up to an unavoidable sign ambiguity because if  $\mathbf{H} \in \mathcal{H}_N$  is a solution of (11), so is  $-\mathbf{H}$ ). Next, we show that even with two excitation inputs having covariances  $\mathbf{C}_{x,1}$  and  $\mathbf{C}_{x,2}$  with *identical* eigenvectors, uniqueness can be established as long as their eigenvalues are sufficiently different.

*Proposition 4:* Consider the system of quadratic equations (11) for  $M = 2$  and suppose  $\mathbf{C}_{x,1} = \mathbf{U} \text{diag}(\lambda_1) \mathbf{U}^T$  and  $\mathbf{C}_{x,2} = \mathbf{U} \text{diag}(\lambda_2) \mathbf{U}^T$ . Then (11) has a unique symmetric solution  $\mathbf{H} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$  that is identifiable up to a sign ambiguity if the following conditions hold:

- A-1) All eigenvalues in  $\lambda_1$  are distinct;
- A-2)  $\lambda_{1,i} \lambda_{2,j} \neq \lambda_{1,j} \lambda_{2,i}$  for all  $i, j$ ;
- A-3)  $\mathbf{V}$  and  $\mathbf{U}$  do not share any eigenvector; and
- A-4)  $\text{rank}(\mathbf{H}) = N$ .

*Proof:* See Appendix D. ■

Conditions A-1) and A-2) encode a notion of richness on the excitation signals. In fact, condition A-2) is the specification for  $M = 2$  of a generalizable requirement based on the Kruskal rank of a matrix related to the eigenvalues of the excitation processes (see Appendix D). Under this perspective, it becomes apparent that a larger  $M$  facilitates the fulfillment of this more general requirement, leading to the expected conclusion that the more input processes we consider, the easier it becomes to identify  $\mathbf{H}$ . Moreover, from the proof arguments it follows that symmetry of  $\mathbf{H}$  is essential (see Lemma 2 in Appendix VII-D). Actually, if one lifts the symmetry assumption and all input covariances have the same eigenvectors, the problem remains non-identifiable even for high values of  $M$  (regardless of the input covariance eigenvalues).

## V. ALGORITHMS

Building on the findings in Section IV-B, here we propose two algorithms with complementary strengths to tackle the quadratic filter identification problem when the only assumption is for  $\mathbf{H}$  to be symmetric (undirected graph), but not (necessarily) PSD.

### A. PROJECTED GRADIENT DESCENT

Going back to the beginning of Section IV, given realizations  $\{\mathbf{y}_m^{(p)}\}_{p=1}^{P_m}$  of the diffusion processes the goal is to identify a symmetric filter  $\mathbf{H} \in \mathcal{H}_N$  that drives  $\{\mathbf{H} \mathbf{C}_{x,m} \mathbf{H}\}_{m=1}^M$  close to the empirical covariances  $\{\hat{\mathbf{C}}_{y,m}\}_{m=1}^M$ . Such quadratic functions of  $\mathbf{H}$  can be formed under perfect knowledge on the input covariances  $\{\mathbf{C}_{x,m}\}_{m=1}^M$ .

---

### Algorithm 1: Graph Filter Identification Using PGD.

---

- 1: **Input:**  $\{\mathbf{C}_{x,m}, \hat{\mathbf{C}}_{y,m}\}_{m=1}^M$ , step size  $\eta > 0$ , tol.  $\delta > 0$ .
  - 2: **Initialize**  $k = 0$  and  $\mathbf{H}_0 \in \mathcal{H}_N$  at random.
  - 3: **repeat**
  - 4:  $\nabla \varepsilon(\mathbf{H}_k) = \sum_{m=1}^M \mathbf{H}_k \mathbf{C}_{x,m} \mathbf{H}_k^T \mathbf{H}_k \mathbf{C}_{x,m} - \hat{\mathbf{C}}_{y,m} \mathbf{H}_k \mathbf{C}_{x,m}$ .
  - 5:  $\tilde{\mathbf{H}}_k = ((\mathbf{H}_k - \eta \nabla \varepsilon(\mathbf{H}_k)) + (\mathbf{H}_k - \eta \nabla \varepsilon(\mathbf{H}_k))^T) / 2$ .
  - 6:  $\mathbf{H}_{k+1} = \mathbf{H}_k + \alpha_k (\tilde{\mathbf{H}}_k - \mathbf{H}_k)$ ,  $\alpha_k$  chosen via line search.
  - 7:  $k \leftarrow k + 1$ .
  - 8: **until**  $\|\mathbf{H}_k - \mathbf{H}_{k-1}\|_F \leq \delta$
  - 9: **Return**  $\hat{\mathbf{H}} = \mathbf{H}_k$
- 

Accordingly, adopting a constrained LS criterion yields a graph filter estimate

$$\mathbf{H}^* = \underset{\mathbf{H} \in \mathcal{H}_N}{\text{argmin}} \frac{1}{4} \sum_{m=1}^M \|\hat{\mathbf{C}}_{y,m} - \mathbf{H} \mathbf{C}_{x,m} \mathbf{H}^T\|_F^2. \quad (23)$$

Weighted variants of the criterion could also be pertinent here, as discussed following (15). Problem (23) is a non-convex fourth-order polynomial optimization, which can potentially have multiple solutions. Since finding  $\mathbf{H}^*$  is challenging, we seek algorithms capable of finding stationary solutions.

A viable approach is to rely on projected gradient descent (PGD) to obtain the provably convergent iterations tabulated under Algorithm 1, where  $\varepsilon(\mathbf{H})$  denotes the cost function in (23) [45, Prop. 2.3.1]. The updates entail multiplications and additions of  $N \times N$  matrices, and accordingly the computational complexity per iteration is  $\mathcal{O}(MN^3)$ . The factor  $M$  can be shaved off by parallelizing the computation of the gradient in Algorithm 1. Taking into account all these desirable features, Algorithm 1 markedly improves upon its precursor in [1]. Since multiple stationary points exist, we typically run Algorithm 1 for  $I$  random initializations. Among the  $I$  estimated filters we select the one whose eigenvectors lead to the sparsest graph shift  $\mathbf{S}$  when solving (6); see also the numerical tests in Section VI.

*Remark 2. (Combining multiple sources of information):* The formulations in (8) and (23) can be combined if both input covariances and pairs of input-output realizations are available. It is also relevant in scenarios where the inputs are not zero mean but their first and second moments are known. Defining  $\hat{\boldsymbol{\mu}}_{y,m} := \frac{1}{P_m} \sum_{p=1}^{P_m} \mathbf{y}_m^{(p)}$  and  $\boldsymbol{\mu}_{x,m} := \mathbb{E}[\mathbf{x}_m]$ , a natural cost would be  $\tilde{\varepsilon}(\mathbf{H}) = \varepsilon(\mathbf{H}) + \nu \sum_{m=1}^M \|\hat{\boldsymbol{\mu}}_{y,m} - \mathbf{H} \boldsymbol{\mu}_{x,m}\|^2$ , where  $\varepsilon(\mathbf{H})$  is the cost function of (23) and  $\nu$  is a tuning parameter. PGD iterations similar to those in Algorithm 1 can be derived to minimize the cost  $\tilde{\varepsilon}(\mathbf{H})$ .

### B. SEMIDEFINITE RELAXATION

Here we show that the graph filter identification task can also be tackled using semidefinite relaxation (SDR) [46], a convexification technique which has been successfully applied to a wide variety of non-convex quadratically-constrained



quadratic programs (QCQP). To that end, we first cast the filter identification problem as a Boolean quadratic program (BQP), see Appendix E for a proof.

*Proposition 5:* For  $m = 1, \dots, M$  consider matrices  $\mathbf{A}_m := (\mathbf{C}_{x,m}^{-1/2} \mathbf{V}_{\text{yx},m}) \odot (\mathbf{C}_{x,m}^{-1/2} \mathbf{C}_{\text{yx},m}^{1/2} \mathbf{V}_{\text{yx},m}) \in \mathbb{R}^{N^2 \times N}$  and unknown binary vectors  $\mathbf{b}_m \in \{-1, 1\}^N$ . Define  $\Psi \in \mathbb{R}^{N^2(M-1) \times NM}$  as

$$\Psi := \begin{bmatrix} \mathbf{A}_1 & -\mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & -\mathbf{A}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{M-1} & -\mathbf{A}_M \end{bmatrix} \quad (24)$$

and  $\mathbf{b} := [\mathbf{b}_1^T, \dots, \mathbf{b}_M^T]^T \in \{-1, 1\}^{NM}$ . If  $\text{rank}(\Psi) = NM - 1$ , then the filter can be exactly recovered (up to a sign) as  $\text{vec}(\mathbf{H}^*) = \mathbf{A}_1 \mathbf{b}_1^*$ , where  $\mathbf{b}_1^*$  is the first  $N \times 1$  sub-vector of the solution to the following BQP problem

$$\mathbf{b}^* = \underset{\mathbf{b} \in \{-1, 1\}^{NM}}{\text{argmin}} \mathbf{b}^T \Psi^T \Psi \mathbf{b}. \quad (25)$$

Problem (25) offers a natural formulation for the pragmatic setting whereby  $\{\mathbf{C}_{y,m}\}_{m=1}^M$  are replaced by sample estimates, and one would aim at minimizing the residuals  $\sum_{m=1}^{M-1} \|\hat{\mathbf{A}}_m \mathbf{b}_m - \hat{\mathbf{A}}_{m+1} \mathbf{b}_{m+1}\|_2^2 = \|\hat{\Psi} \mathbf{b}\|_2^2$  in the LS sense. Given a solution of (25) with  $\hat{\Psi}$  replacing  $\Psi$ ,  $\mathbf{H} \in \mathcal{H}_N$  can be estimated as [cf. (21)]

$$\hat{\mathbf{H}} = \frac{1}{M} \sum_{m=1}^M \mathbf{C}_{x,m}^{-1/2} \hat{\mathbf{C}}_{\text{yx},m}^{1/2} \hat{\mathbf{V}}_{\text{yx},m} \text{diag}(\mathbf{b}_m^*) \hat{\mathbf{V}}_{\text{yx},m}^T \mathbf{C}_{x,m}^{-1/2}. \quad (26)$$

Even though the BQP is a classical NP-hard combinatorial optimization problem [46], via SDR one can obtain near-optimal solutions with provable approximation guarantees. To derive the SDR of (25), first introduce the  $NM \times NM$  symmetric PSD matrices  $\mathbf{W} := \Psi^T \Psi$  and  $\mathbf{B} := \mathbf{b} \mathbf{b}^T$ . By construction, the binary matrix  $\mathbf{B}$  has rank one and its diagonal entries are  $B_{ii} = b_i^2 = 1$ . Conversely, any matrix  $\mathbf{B} \in \mathbb{R}^{NM \times NM}$  that satisfies  $\mathbf{B} \succeq \mathbf{0}$ ,  $B_{ii} = 1$ , and  $\text{rank}(\mathbf{B}) = 1$  necessarily has the form  $\mathbf{B} = \mathbf{b} \mathbf{b}^T$ , for some  $\mathbf{b} \in \{-1, 1\}^{NM}$ . Using these definitions, one can write  $\mathbf{b}^T \Psi \mathbf{b} = \text{trace}(\mathbf{b}^T \Psi \mathbf{b}) = \text{trace}(\mathbf{W} \mathbf{b} \mathbf{b}^T) = \text{trace}(\mathbf{W} \mathbf{B})$  and accordingly (25) is equivalent to

$$\begin{aligned} & \underset{\mathbf{B}}{\text{minimize}} \text{trace}(\mathbf{W} \mathbf{B}) \\ & \text{subject to: } \mathbf{B} \succeq \mathbf{0}, \text{rank}(\mathbf{B}) = 1, \\ & B_{ii} = 1, \quad i = 1, \dots, NM. \end{aligned} \quad (27)$$

The only source of non-convexity in (27) is the rank constraint. We follow the usual practice of relaxing the above problem by dropping the rank constraint [46] to obtain the convex SDR

$$\begin{aligned} & \mathbf{B}^* = \underset{\mathbf{B}}{\text{argmin}} \text{trace}(\mathbf{W} \mathbf{B}) \\ & \text{s. to } \mathbf{B} \succeq \mathbf{0}, \\ & B_{ii} = 1, \quad i = 1, \dots, NM, \end{aligned} \quad (28)$$

---

**Algorithm 2:** Graph Filter Identification Using SDR.

---

- 1: **Input:**  $\mathbf{W} = \Psi^T \Psi \in \mathcal{H}_{NM}$  and  $L \in \mathbb{N}$ .
  - 2: Solve the SDP in (28) to obtain  $\mathbf{B}^*$ .
  - 3: **for**  $l = 1, \dots, L$  **do**
  - 4:     Draw  $\mathbf{z}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^*)$ .
  - 5:     Round  $\hat{\mathbf{b}}_l = \text{sgn}(\mathbf{z}_l)$ .
  - 6: **end for**
  - 7: Determine  $l^* = \text{argmin}_{l=1, \dots, L} \hat{\mathbf{b}}_l^T \mathbf{W} \hat{\mathbf{b}}_l$ .
  - 8: **Return**  $\hat{\mathbf{b}}_{l^*}$
- 

which coincides with the bidual (dual of the dual) problem of (25). Problem (28) is a semidefinite program (SDP) and can be solved using an off-the-shelf interior-point method [47].

It is immediate that a rank-one optimal solution  $\mathbf{B}^* = \mathbf{b}^* (\mathbf{b}^*)^T$  of (28) solves the original BQP as well; however, in general  $\text{rank}(\mathbf{B}^*) \neq 1$ . To generate a feasible solution of (25) from  $\mathbf{B}^*$ , we adopt the so-termed Gaussian randomization procedure [46], [48]. The overall method is tabulated under Algorithm 2 and the quality of the rounded solutions is evaluated via computer simulations in Section VI. The computational complexity is discussed under Remark 3.

Interestingly, it is possible to derive theoretical approximation guarantees for the feasible solutions generated via the SDR scheme in Algorithm 2. Leveraging a result in [48], a guarantee for the BQP (25) follows immediately.

*Corollary 1:* Let  $\mathbf{b}^*$  be the solution of (25) and  $\hat{\mathbf{b}}_{l^*}$  be the output of Algorithm 2. For  $\gamma = (1 - \frac{2}{\pi}) \lambda_{\max} NM$ , where  $\lambda_{\max}$  is the largest eigenvalue of  $\mathbf{W}$ , then

$$\gamma + \frac{2}{\pi} (\mathbf{b}^*)^T \mathbf{W} \mathbf{b}^* \geq \mathbb{E} [(\hat{\mathbf{b}}_{l^*})^T \mathbf{W} \hat{\mathbf{b}}_{l^*}] \geq (\mathbf{b}^*)^T \mathbf{W} \mathbf{b}^*. \quad (29)$$

Notice that although the bounds in (29) offer guarantees in terms of the expected objective value, particular realizations  $\hat{\mathbf{b}}_{l^*}$  tend to fall within those bounds with high probability if  $L$  is chosen sufficiently large.

All in all, the recipe to estimate the graph filter via the SDR approach entails the following steps. First we calculate  $\{\hat{\mathbf{A}}_m\}_{m=1}^M$  from  $\{\hat{\mathbf{C}}_{y,m}, \mathbf{C}_{x,m}\}_{m=1}^M$  using the expression in the statement of Proposition 5, and form  $\hat{\Psi}$  as in (24) to finally obtain  $\hat{\mathbf{W}} = \hat{\Psi}^T \hat{\Psi}$ . Next, a feasible solution  $\hat{\mathbf{b}}_{l^*}$  to the BQP is obtained after running Algorithm 2 with  $\hat{\mathbf{W}}$  and an appropriate choice of  $L$  as inputs. Finally,  $\hat{\mathbf{H}}$  is estimated using (26).

*Remark 3. (SDR versus PGD):* Although SDR has well-documented merits when dealing with non-convex BQPs (and other QCQPs) in applications such as MIMO detection [49] and transmit beamforming [50], it has so far not been explored for graph filter identification or network topology inference. The relaxation entails dropping a rank constraint after ‘‘lifting’’ a (binary) vector-valued problem with  $NM$  variables to a matrix-valued one with  $NM(NM + 1)/2$  variables. This incurs an increase in memory and computational cost, since the complexity of a general-purpose interior point method to solve

the resulting SDP is  $\mathcal{O}(N^7 M^7 \log(1/\epsilon))$ , for a prescribed solution accuracy  $\epsilon > 0$  [46]. The additional cost could hinder the applicability of the SDR approach in Algorithm 2 to problems involving very large networks. In those scenarios, the PGD iterations in Algorithm 1 can still find stationary solutions with lower memory requirements and  $\mathcal{O}(MN^3)$  complexity per iteration. While nothing can be said a priori on the quality of the aforementioned stationary points, the more costly SDR-based solutions offer a quantifiable approximation guarantees as asserted in Corollary 1. Even though the focus here is not on pushing algorithmic scalability to the limit, the recovery performance of the proposed methods could serve as a benchmark for faster (possibly randomized) algorithms that can handle larger graphs. Since the desired solution of (28) has rank one, we envision future algorithmic improvements using large-scale SDP solvers based on e.g., the Burer-Monteiro factorization [51].

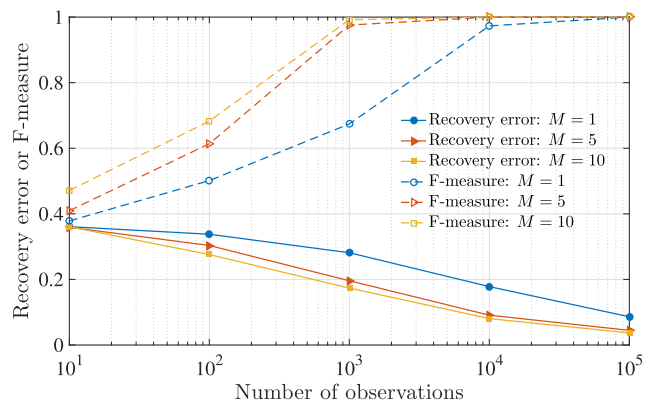
## VI. NUMERICAL TESTS

We study the recovery of synthetic and real-world graphs to assess the performance of the proposed network topology inference algorithms. To this end, we adopt the formulation in (6) to recover a sparse adjacency matrix that is close in the Frobenius-norm sense to being diagonalized by the estimated eigenbasis  $\hat{\mathbf{V}}$ . A comprehensive numerical evaluation is carried out whereby we: (i) study the graph inference performance in controlled synthetic settings (Section VI-A); (ii) carry out comparisons with some state-of-the-art algorithms (Section VI-B); (iii) use this framework to gain insights about urban mobility patterns in New York City from data of Uber pickups in 2015 (Section VI-C); and (iv) cluster companies using a graph obtained from time series of their daily opening and closing stock prices (Section VI-D).

### A. PERFORMANCE ASSESSMENT

Throughout this section, we infer networks from the observation of diffusion processes that are synthetically generated via graph filtering as in (2). We consider the filter  $\mathbf{H}_1 = \sum_{l=0}^2 h_l \mathbf{S}^l$ , where the coefficients  $\{h_l\}$  are drawn uniformly at random on  $[0,1]$ . As a measure of recovery error, we adopt  $\|\mathbf{S}^* - \mathbf{S}\|_F / \|\mathbf{S}\|_F$  (averaged over independent realizations of the experiment), where  $\mathbf{S}^*$  is the solution of (6) and  $\mathbf{S}$  denotes the ground-truth GSO. To directly assess edge-support recovery, we also compute the F-measure defined as the harmonic mean of edge precision and recall (precision is the percentage of correct edges in  $\mathbf{S}^*$ , and recall is the fraction of edges in  $\mathbf{S}$  that are successfully retrieved).

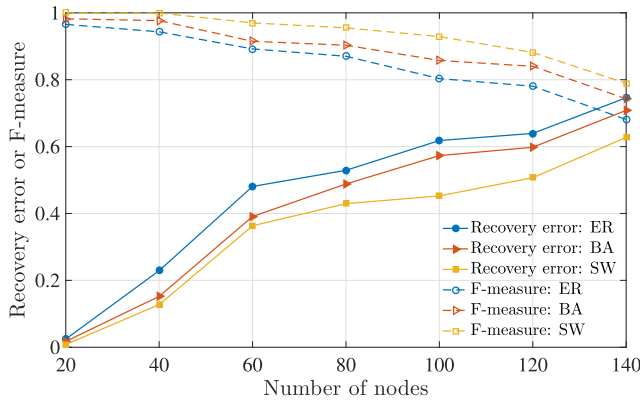
*Inference of PSD graph shifts.* We consider the karate club social network studied by Zachary [52], which is represented by a graph  $\mathcal{G}$  consisting of  $N = 34$  nodes or members of the club and undirected edges symbolizing friendships among them. Denoting by  $\mathbf{L}_n$  the normalized Laplacian of  $\mathcal{G}$ , we define the graph-shift operator  $\mathbf{S} = \mathbf{I} - \gamma \mathbf{L}_n$  with  $\gamma = 1/\lambda_{\max}(\mathbf{L}_n)$ , modeling the diffusion of opinions between the members of the club. A graph signal  $\mathbf{y}$  can be regarded



**FIGURE 2.** Recovery error and F-measure in recovering Zachary's karate club social network as a function of the number  $P$  of opinion profiles observed, and parametrized by the number of topics  $M$ . As expected, the graph estimate becomes increasingly accurate with increasing values of  $P$  and  $M$ .

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  $\mathbf{L}_n$  – 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  $\mathbf{y}_m$  is obtained by diffusing through the network an initial belief  $\mathbf{x}_m$  (cf. (2)). More precisely, for each topic  $m = 1, \dots, M$ , we model  $\mathbf{x}_m$  as a zero-mean Gaussian process with known covariance  $\mathbf{C}_{\mathbf{x},m}$ . The input covariances are generated as  $\mathbf{C}_{\mathbf{x},m} = \mathbf{U}_m |\mathbf{\Lambda}_m| \mathbf{U}_m^T$ , where the diagonal matrix  $|\mathbf{\Lambda}_m|$  has diagonal entries equal to the absolute values of i.i.d. samples drawn from a standard normal distribution. Matrix  $\mathbf{U}_m$  collects the eigenvectors of a symmetric matrix with i.i.d. standard normal entries. We are then given a set  $\{\mathbf{y}_m^{(p)}\}_{p=1}^P$  of independent opinion profiles generated from different sources  $\{\mathbf{x}_m^{(p)}\}_{p=1}^P$  diffused through filter  $\mathbf{H}_1$  of unknown nonnegative coefficients. From these  $P$  opinion profiles we first form an estimate  $\hat{\mathbf{C}}_{\mathbf{y},m}$  of the output covariance. Leveraging that  $\mathbf{S}$  is PSD and  $h_l \geq 0$  for  $l = 0, 1, 2$  (cf. Section IV-A), then we estimate the filter  $\mathbf{H}_1$  via (16) and solve (6) using the eigenbasis  $\hat{\mathbf{V}}$  of the estimated filter. Set  $\mathcal{S}$  is modified accordingly for the recovery of a normalized Laplacian instead of an adjacency matrix; see [12].

In Fig. 2 we plot the recovery error as a function of the number of observations  $P$  for three different values of  $M \in \{1, 5, 10\}$ . As  $P$  increases, the estimate  $\hat{\mathbf{C}}_{\mathbf{y},m}$  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  $P$ . Moreover, for a fixed number of observations  $P$ , the error in the estimation of  $\hat{\mathbf{C}}_{\mathbf{y},m}$  can be partially overcome by observing multiple diffusion processes, thus, larger values of  $M$  lead to smaller graph recovery errors. Similar trends can be observed for the F-measure in Fig. 2, corroborating the efficacy of the approach in identifying the GSO support.

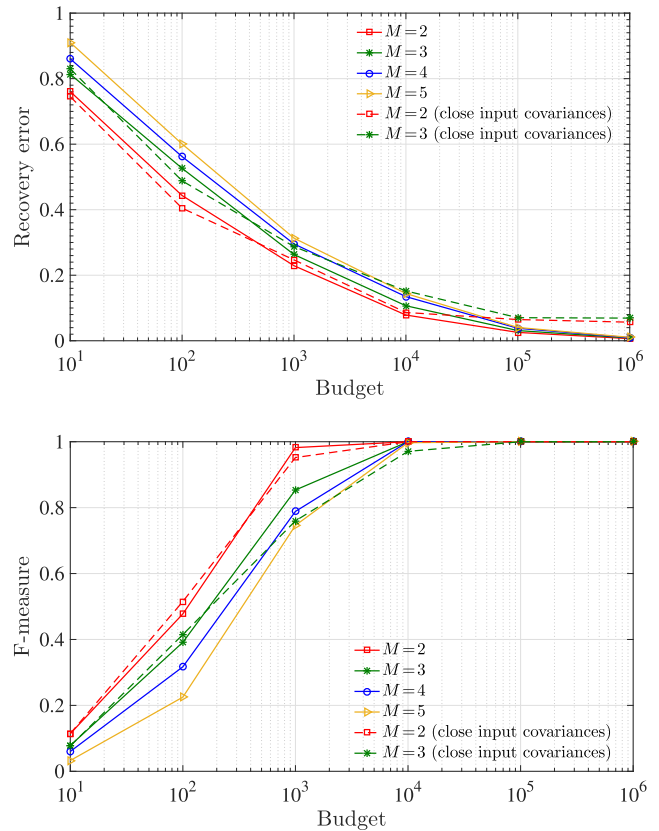


**FIGURE 3.** Recovery error and F-measure in estimating random graphs from  $M = 10$  processes and  $P = 10^6$  observations for each process. As  $N$  increases, the output covariances become less accurate and the performance of Algorithm 1 deteriorates. The SW graph with rewiring probability 0.1, which is close to a regular graph, is the easiest to identify while the purely random ER graph is the hardest.

*Inference of random graphs.* Here we evaluate the performance of Algorithm 1 on three different types of random graphs for a varying number of nodes  $N$ . We consider the recovery of the adjacency matrix  $\mathbf{S} = \mathbf{A}$  for: i) Erdős-Rényi (ER) graphs with edge probability  $p = 0.3$ ; ii) Barabási-Albert (BA) preferential attachment graphs generated from  $\mu_0 = 0.5N$  initially placed nodes, where each new node is connected to  $\mu = 0.3N$  existing ones; and iii) small world (SW) graphs with rewiring probability 0.1 and mean degree of  $0.3N$ . The parameter choices yield roughly equal mean degrees across all three graph models.

We consider  $M = 10$  different processes in the respective graphs, where we observe  $P = 10^6$  outputs for each process. The given output signals  $\{\mathbf{y}_m^{(p)}\}_{p=1}^P$  are generated using the same procedure described under *Inference of PSD graph shifts*. Then we estimate the filter  $\mathbf{H}_1$  via Algorithm 1 and solve (6) using the eigenbasis  $\hat{\mathbf{V}}$  of the estimated filter. Fig. 3 depicts both the relative recovery error and the F-measure versus the number of nodes  $N$  for the ER, BA, and SW random graphs. As  $N$  increases for a fixed number of observations, the estimate  $\hat{\mathbf{C}}_{y,m}$  becomes less reliable for larger graphs and as a result, the performance deteriorates. However, for  $N \sim 20 - 80$ , we observe a reasonable recovery of the sought graphs. These observations are valid across all graph models. We also find that the SW graph with rewiring probability 0.1, which is close to a regular graph, is the easiest to identify while the purely random ER graph is the hardest. For such unweighted graphs, this is aligned with findings in [12]. The preferential attachment BA graph which relies on a copying procedure that introduces correlation [8, Ch. 6.4.1], falls in between the SW and ER ends of the spectrum.

*Inference with a fixed signal budget.* In practice, we estimate output covariances from observed signals via sample averaging. In particular, assume that we estimate the covariance of each of the  $M$  processes by observing  $P$  independent graph



**FIGURE 4.** (top) Error in recovering a collaboration network using SDR for varying budgets (number of processes  $M$  times the signals observed per process  $P$ ). (bottom) Counterpart of the top plot but for the F-measure showing the success in recovering the GSO support. For a fixed budget it is better to allocate the sensing resources in learning  $M = 2$  processes as accurately as possible, rather than having a coarser estimate of more ( $M > 2$ ) processes.

signals. For the cases where the total budget of signals  $M \times P$  is fixed, this numerical test studies the trade-off between  $M$  and  $P$  as it pertains to recovery performance. Is it better to have accurate estimates of a few processes' covariances (larger  $P$  and smaller  $M$ ), or instead, coarser estimates of more processes (smaller  $P$  and larger  $M$ )? In order to answer this question, we run an experiment whose goal is to recover the adjacency matrix of the collaboration network of  $N = 31$  scientists working in the field of Network Science [53], and model the diffusion of ideas among the scientists via simple linear dynamics as per  $\mathbf{H}_1$ . The input signals are i.i.d., generated using the same procedure described under *Inference of PSD graph shifts*. We observe outputs  $\{\mathbf{y}_m^{(p)}\}_{p=1}^P$  and implement Algorithm 2 for  $L = 10$  random draws to recover the collaboration graph from  $M$  processes, each inferred from  $P$  signals such that  $M \times P$  is fixed.

Fig. 4 (top) depicts the recovery error as a function of the total budget ( $M \times P$ ) of observed signals, parametrized by  $M \in \{2, 3, 4, 5\}$ . The corresponding counterparts for the F-measure are shown in Fig. 4 (bottom). As expected, the performance enhances (i.e., increase in F-measure and decrease in

recovery error) for increasing budget for all values of  $M$ , since we can better estimate the covariances of all the processes. More interestingly, for a fixed budget it is better to consider only  $M = 2$  processes. Given that SDR can often recover the filter with perfect knowledge of  $M = 2$  covariances, it is better to focus on just two processes and obtain the most accurate estimates of the associated covariances.

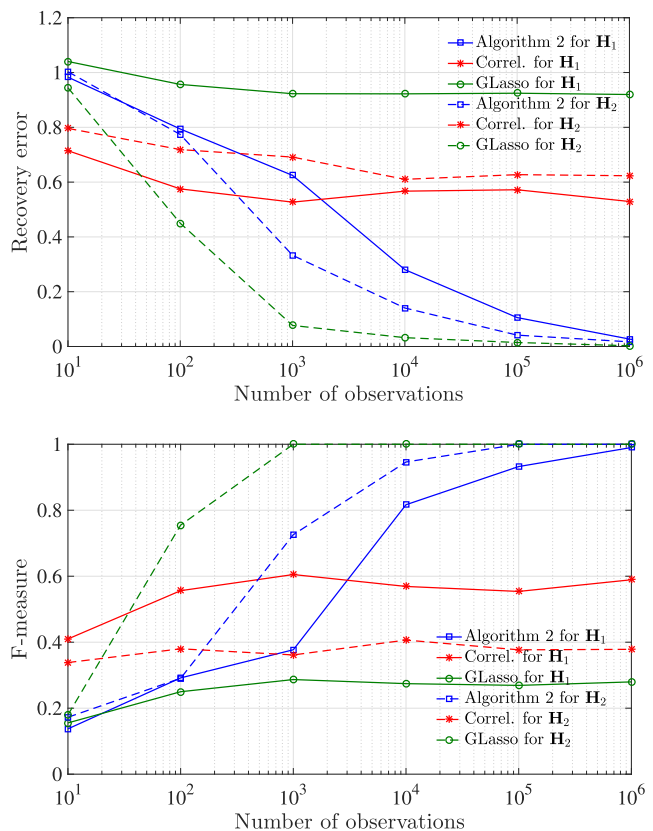
Moreover, to emphasize the importance of the richness of the input processes, we repeat the experiment for a setting where the input covariances are similar to each other. To generate these covariances, we replace  $\{\mathbf{C}_{x,m}\}_{m=2}^M$  with  $\mathbf{C}_{x,1} + (10^{-8}) \times \{\mathbf{C}_{x,m}\}_{m=2}^M$  for  $M = 2, 3$ ; see the dashed lines in Fig. 4 for the associated recovery performance. Notice that for  $(M \times P) \leq 10^4$  the recovery error and the F-measure are comparable to (though slightly larger than) the corresponding counterparts in the original setting. More interestingly, for larger values of the budget the recovery error for the setting of similar covariance matrices saturates. Intuitively, we know that for  $M = 1$  the problem is non-identifiable (cf. Section IV-B). For practical purposes, the setting with similar covariance matrices behaves as having only one input process, thus resulting in an insurmountable error even for increasing values of  $M$ .

## B. PERFORMANCE COMPARISON

We compare the performance of the proposed approach with baseline statistical approaches as well as with relevant counterparts in the GSP literature.

*Comparison with statistical approaches.* We analyze the performance of Algorithm 2 in comparison with two workhorse statistical methods, namely, (thresholded) correlation [8, Ch. 7.3.1] and graphical lasso [16]. Our goal is to recover the adjacency matrix of an undirected and unweighted graph with no self-loops. To that end, we are given (a varying number of) observed graph signals, which are modeled as the output of  $M = 2$  different diffusion processes. We test the recovery of adjacency matrices  $\mathbf{S} = \mathbf{A}$  of ER graphs with  $N = 20$  nodes and edge probability  $p = 0.2$ . The zero-mean Gaussian inputs have covariance matrices  $\mathbf{C}_{x,m} = \mathbf{U}_m |\mathbf{\Lambda}_m| \mathbf{U}_m^T$ , where the diagonal matrix  $|\mathbf{\Lambda}_m|$  is generated as in *Inference of PSD graph shifts*. We consider two types of filters and respectively two generative mechanisms for  $\mathbf{U}_m$ : (i) the second-order filter  $\mathbf{H}_1$  already defined, while  $\mathbf{U}_m$  collects the eigenvectors of a symmetric matrix with i.i.d. standard normal entries; and (ii) filter  $\mathbf{H}_2 = (\kappa \mathbf{I} + \mathbf{S})^{-1/2} \mathbf{C}_x^{-1/2}$ , where  $\kappa$  is selected to ensure that  $\kappa \mathbf{I} + \mathbf{S}$  is positive definite and  $\mathbf{C}_x$  is the average of  $\mathbf{C}_{x,1}$  and  $\mathbf{C}_{x,2}$ . Under (ii),  $\mathbf{U}_1 = \mathbf{U}_2 = \mathbf{V}$  are the eigenvectors of  $\mathbf{S}$  to ensure that  $\mathbf{H}_2$  is a polynomial on  $\mathbf{S}$ , i.e., a graph filter.

We vary the number of diffused observed signals from  $P = 10$  to  $10^6$  in powers of 10. Output signals are generated by passing the inputs through a graph filter. 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



**FIGURE 5.** Performance comparison in recovering random graphs between the proposed Algorithm 2, graphical lasso, and correlation-based methods for general second-order filters ( $\mathbf{H}_1$ ) and specific filters ( $\mathbf{H}_2$ ) as a function of the number of observations for a fixed  $M = 2$ . We adopt two performance measures, namely, (top) relative recovery error, and (bottom) the F-measure on the detected edges.

threshold for the correlation method and the regularization parameter for graphical lasso are determined and then used for the recovery of the 20 testing graphs. In Fig. 5 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 of the recovery error (top) or the mean of the F-measure (bottom) over the 20 testing graphs. When considering a general second-order graph filter  $\mathbf{H}_1$ , our proposed algorithm outperforms the other two baseline statistical methods. This is expected since the graph recovered by graphical lasso corresponds to the penalized maximum-likelihood estimate of the precision matrix  $\mathbf{C}_y^{-1} = \mathbf{H}_1^{-1} \mathbf{C}_x^{-1} \mathbf{H}_1^{-1}$ , that bears no direct relation with  $\mathbf{S}$ . However, for the specific case of graph filter  $\mathbf{H}_2$ , where the sought GSO matches the precision matrix in the off-diagonal entries, graphical lasso outperforms the other two methods as expected. Indeed, while Algorithm 2 still succeeds in learning the topology of the random graph in this setting, Fig. 5 (bottom) shows it does so with two orders of magnitude more observations than graphical lasso.

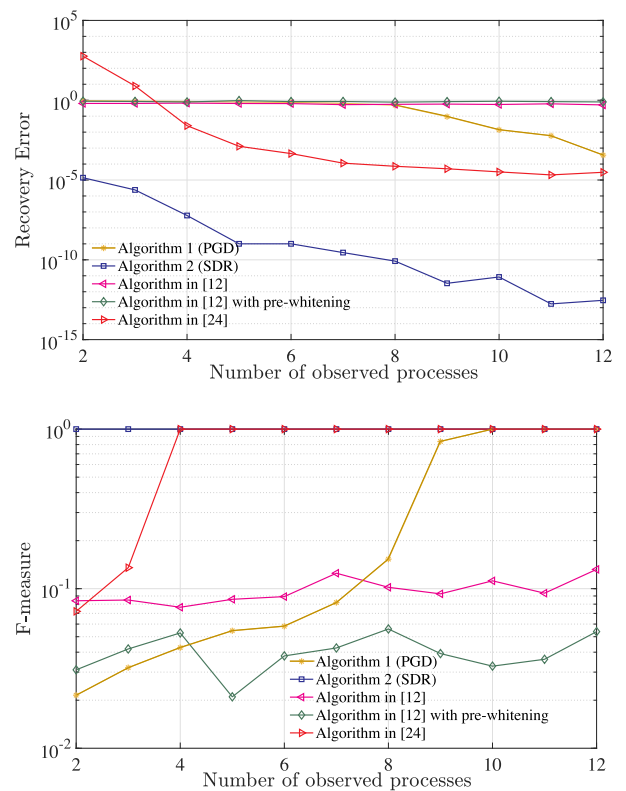
*Comparison with GSP methods.* We finally compare the recovery performance of the proposed algorithms with the SEM

inference method based on parallel factor (PARAFAC) tensor decomposition in [24], the algorithm in [12] which assumes that the observed graph processes are stationary, and a variant of the latter using a whitening transformation. We study a symmetric structural brain graph with  $N = 66$  nodes or neural regions and edge weights given by the density of anatomical connections between regions [54]. Graph filter-based diffusion models with impulsive sources over the brain connectivity network [cf. (2)] were adopted and validated to model the progression of brain atrophy [55]. To compare with [24, Algorithm 1], we draw zero-mean Gaussian inputs with diagonal covariance matrices  $\mathbf{C}_{x,m} = |\mathbf{\Lambda}_m|$ , where  $|\mathbf{\Lambda}_m|$  is generated as in *Inference of PSD graph shifts*. We then generate output signals from the diffusion filter  $\mathbf{H}_3 = (\mathbf{I} - \mathbf{A})^{-1}$  (corresponding to a symmetric SEM), thus, output covariances satisfy  $\mathbf{C}_{y,m} = \mathbf{H}_3 \mathbf{C}_{x,m} \mathbf{H}_3$  for each  $m$ . Assuming perfect knowledge of the second-order statistics  $\{\mathbf{C}_{x,m}, \mathbf{C}_{y,m}\}_{m=1}^M$ , we first estimate the graph filter using either the PGD approach in Algorithm 1 or the SDR approach in Algorithm 2, and then solve (6) to recover the adjacency matrix of the brain network.

Fig. 6 depicts the recovery error (top) and the F-measure (bottom) versus  $M$  for: (i) Algorithms 1 and 2; (ii) a baseline GSP-based method that exploits signal stationarity [12]; (iii) a variant of the latter where we adopt a whitening transformation as in (12) and re-run the algorithm in [12] using the eigenvectors of  $\mathbf{C}_{yx}$  as spectral templates; and (iv) the tensor-based approach in [24]. First, we notice that the performance of both the PGD and SDR algorithms as well as [24, Algorithm 1] improves for increasing  $M$ . Moreover, the SDR approach uniformly (in  $M$ ) outperforms all other methods. Recall that this gain in performance comes at the price of a higher computational complexity (relative to the PGD approach) as explained in Remark 3 and is due to exploiting the structure of the solution in formulating the topology inference problem. Also, [24, Algorithm 1] outperforms the PGD algorithm, since it is tailored for SEMs matching this simulation setup. Finally, both proposed methods outperform the algorithm in [12] (with or without pre-whitening) for all  $M$ . This is expected due to the model mismatch suffered when incorrectly assuming that the observed graph processes are stationary, and accordingly using the output covariance eigenvectors (or eigenvectors of  $\mathbf{C}_{yx}$ , if whitened) as spectral templates of  $\mathbf{S}$ .

### C. UNVEILING URBAN MOBILITY PATTERNS FROM UBER PICKUPS

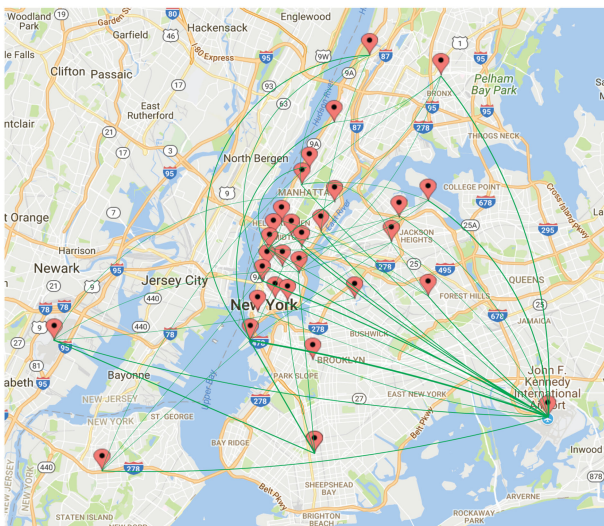
We implement our SDR graph topology inference method (Algorithm 2) in order to detect mobility patterns in New York City from Uber pickups data.<sup>1</sup> We have access to times and locations of pickups from January 1st to June 29th 2015 for 263 known location IDs. For simplicity, we cluster the locations into  $N = 30$  zones based on their geographical proximity; these are shown as red pins in Fig. 7. These zones represent the nodes of the graph to be recovered. The total number of



**FIGURE 6.** (top) Error in recovering a brain network as a function of the number  $M$  of graph processes observed for different recovery algorithms. We assume perfect second-order statistical knowledge of the processes and a graph filter corresponding to a symmetric SEM. SDR outperforms PGD at the expense of higher computational complexity (cf. Remark 3). Also, [24, Algorithm 1] has a superior performance relative to the PGD algorithm due to the specific filter and set up. The mismatched stationarity assumption in [12] explains its worse performance, even when using a whitening transformation as in (12). (bottom) Counterpart of the top plot but depicting the F-measure as a figure of merit, where similar trends are observed.

pickups aggregated by zone during a specific time horizon can be regarded as graph signals defined on the unknown graph. More specifically, we consider  $M = 2$  graph processes: weekday ( $m = 1$ ) and weekend ( $m = 2$ ) pickups. Moreover, we consider that the pickups from 6am to 11am constitute the inputs of our process whereas the pickups from 3pm to 8pm comprise the outputs of our process. To be more precise, for a specific day we aggregate all the pickups within 6–11am to form an input signal  $\mathbf{x}$  and similarly we group all the pickups within 3–8pm to generate the associated output signal  $\mathbf{y}$ . If the day considered is a weekday, we think of this pair as being generated from process  $m = 1$ , and if it is a weekend we consider the pair coming from process  $m = 2$ . We repeat this procedure for all the days included in the period of study, and estimate input-output covariance pairs  $\{\hat{\mathbf{C}}_{x,m}, \hat{\mathbf{C}}_{y,m}\}_{m=1}^2$ . We then run Algorithm 2 to infer an underlying graph filter  $\hat{\mathbf{H}}$  and solve (6) given the estimated eigenbasis of  $\hat{\mathbf{H}}$  to find a sparse mobility pattern. The modeling presumption is that throughout the day, the population diffuses over an unknown graph of mobility patterns we seek to identify. By looking

<sup>1</sup>Dataset [Online]. Available: <https://github.com/fivethirtyeight/uber-tlc-foil-response>



**FIGURE 7. New York City's mobility pattern inferred from 2015 Uber pickups data. Most edges connect Manhattan with the other boroughs indicating that Uber is widely used to commute to/from the suburbs. Airports (Kennedy, Newark, and LaGuardia) can also be distinguished as high degree nodes.**

at aggregates over a large number of trips, the population flows and mobility patterns in the city can be reasonably well approximated by linear diffusion dynamics; see also [29].

The recovered graph is depicted in Fig. 7, where the weights of the recovered edges are represented by the line widths in the figure. Given the nature of the input and output processes considered, the graph obtained is a sparse description of the mobility pattern of people throughout the day. Notice that most connections occur between Manhattan and the other boroughs (Queens, Bronx, Staten Island, Brooklyn, and Newark), while only a few edges connect zones within Manhattan. This indicates that people use Uber to commute from their homes in the suburbs to their work (or leisure activities at the weekends) in the city. These findings are consistent with exploratory research of this same dataset [56] as well as a recent New York Times article that writes: “*The ride-hail app has increasingly shifted its focus to the city’s other four boroughs, where frustration over subway overcrowding and delays and fewer taxi options have made it the ride of choice for many. As a result, Uber is booming in the other boroughs, with half of all Uber rides now starting outside Manhattan...*” [57]. Lastly, observe that the JFK, Newark and, LaGuardia airports are strongly connected with Manhattan and the other boroughs, as expected.

#### D. CLUSTERING FIRMS FROM HISTORICAL STOCK PRICES

We implement our SDR topology inference approach in another real-world setting where we consider historical stock price data from Yahoo.<sup>2</sup> Opening and closing prices are obtained from 18th May 2012 to 30th December 2016 for the top 50 firms that have had the most wealth creation to

shareholders in aggregate among all companies with common stock in the Center for Research in Security Prices (CRSP) database since July 1926 until December 2016 [58, Table 5].<sup>3</sup> We consider the  $N = 48$  firms analyzed as nodes in a graph and the stock prices of the 1160 days studied as graph signals defined on the graph. Our goal is then to recover the edges of this graph from the observed stock prices. We consider  $M = 4$  graph processes for each three-month period of stock market activity. Moreover, the opening and closing prices of each day are respectively considered as inputs and outputs of the graph processes. A similar procedure to the one in Section VI-C is used to estimate input-output covariance pairs  $\{\hat{\mathbf{C}}_{x,m}, \hat{\mathbf{C}}_{y,m}\}_{m=1}^4$ . We then run Algorithm 2 to infer an underlying graph filter  $\hat{\mathbf{H}}$  and solve (6) given the estimated eigenbasis of  $\hat{\mathbf{H}}$  to reveal a sparse graph of inter-dependencies among the  $N$  companies.

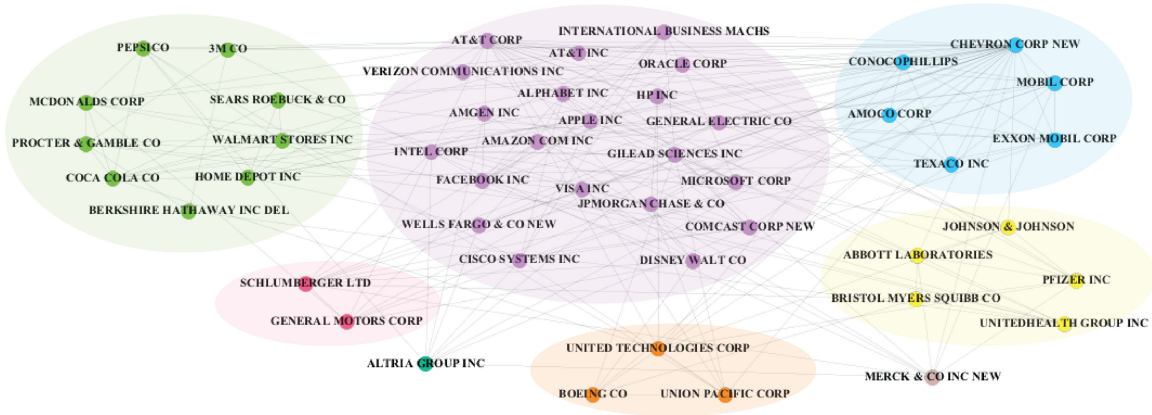
The recovered graph is shown in Fig. 8. In order to (indirectly) test the validity and usefulness of the obtained graph, we run the Louvain community detection algorithm [59] and analyze the resulting clusters, depicted in different colors in Fig. 8. Notice that firms that broadly belong to the same economic sector are indeed clustered into the same communities. For example, the blue community represents the firms that are mostly related to energy and oil industry, the purple community mostly relates to technology, telecommunication, and finance, the yellow community highly relates to healthcare and pharmaceuticals, the orange community represents the aviation industry, and the green one embodies the food, drink, and retail sectors.

#### VII. CONCLUDING SUMMARY

We studied the problem of inferring an undirected network from observations of *non-stationary* signals diffused on the unknown graph. Relative to the stationary setting, the main challenge is that the GSO eigenvectors differ from those of the signal covariance matrix. To overcome this hurdle, we leverage that the sought eigenbasis is preserved by the polynomial graph filter that governs the diffusion process. As a result, the novel approach is to first identify the GSO eigenvectors from a judicious graph filter estimate, and then we rely on these spectral templates to estimate the eigenvalues by imposing desirable properties (such as edge sparsity) on the graph to be recovered. We propose different estimators of the symmetric diffusion filter depending on whether: i) explicit realizations; or, ii) second-order statistical information is available from the input-output graph signal pair. These estimators arise as solutions of systems of linear and quadratic matrix equations, respectively. We thus investigate identifiability properties of the resulting problems and bound the topology recovery error for the special case of PSD diffusion filters. We also develop PGD and SDR algorithms with complementary strengths to compute near-optimal solutions of the said system of quadratic

<sup>3</sup>Of the 50 firms mentioned, DuPont was excluded from the analysis for having excessive missing data and Pfizer and Warner Lambert were merged into a single one (since Pfizer bought Warner Lambert in 2000), resulting in a total of  $N = 48$  firms analyzed.

<sup>2</sup>Data [Online]. Available: <https://finance.yahoo.com/lookup?s=API>



**FIGURE 8.** The sparse network topology inferred from Yahoo's stock price data. Firms that represent similar sectors of the economy tend to be clustered together. For example, the blue community collects firms from the energy sector whereas the orange cluster relates to the aviation industry.

equations. The overall network topology inference pipeline is validated via comprehensive numerical tests on real social and structural brain networks. Moreover, we show how the developed graph inference tools can be utilized to unveil mobility patterns in New York City from data of Uber pickups and to reveal inter-dependencies between firms from their historical stock prices.

The network topology inference problem addressed here is admittedly challenging, and our solutions are not without their limitations. The heavily data-driven nature of the approach along with the generality of the postulated model come at the price of increased sample and computational complexities. For these reasons, the developed tools are better suited for problems where the number of nodes is not excessively large and the observations are abundant. This is often the case for gene-regulatory networks, molecular contact structures, functional brain connectomes, power grids, and transportation networks, just to name a few examples. As future work, it is of interest to expand the scope of the proposed topology inference framework to accommodate i) sampling whereby observations are only acquired from a limited subset of nodes; and ii) directed graphs that represent the structure of signals (possibly) generated via nonlinear network interactions. In addition to scalable SDP solvers, adaptive algorithms that can track slowly-varying graphs, thus providing memory and computational savings by processing the signals on-the-fly, are subject of ongoing investigation.

## APPENDIX

### A. PROOF OF PROPOSITION 1

To show that a) is true, note first that the cost in (9) can be compactly rewritten as  $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2$ . Using the Kronecker product and the matrix vectorization operator, we can further rewrite it as  $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2 = \|\text{vec}(\mathbf{Y}) - (\mathbf{X}^T \otimes \mathbf{I}_N)\text{vec}(\mathbf{H})\|_2^2$ . Moreover, the redundant entries in  $\text{vec}(\mathbf{H})$  can be removed using the duplication matrix  $\mathbf{D}_N$ , to yield  $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2 = \|\text{vec}(\mathbf{Y}) - (\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N\text{vech}(\mathbf{H})\|_2^2$ . This LS cost

can be minimized using the Moore-Penrose pseudoinverse as  $\text{vech}(\mathbf{H}^*) = [(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N]^\dagger \text{vec}(\mathbf{Y})$ , so (10) follows.

In order to prove that the b) holds true, we denote by  $\{\mathbf{v}_i\}_{i=1}^{N-M_r}$  a basis of the null space  $\ker(\mathbf{X}^T)$ . We use these vectors to form all non-repeated symmetric matrices of the form  $\mathbf{V}_{ij} = \mathbf{v}_i\mathbf{v}_j^T + \mathbf{v}_j\mathbf{v}_i^T$ , and then collect the  $N_{\mathbf{H}}$  distinct entries of those symmetric matrices into vector  $\tilde{\mathbf{v}}_{ij} = \mathbf{D}_N^\dagger \text{vec}(\mathbf{V}_{ij}) = \mathbf{D}_N^\dagger(\mathbf{v}_i \otimes \mathbf{v}_j + \mathbf{v}_j \otimes \mathbf{v}_i)$ .

*Lemma 1:* Define the set  $\mathcal{V}^{\text{ker}} = \{\tilde{\mathbf{v}}_{ij} : i \leq j, 1 \leq j \leq N - M_r\}$ , then it holds that: i) any of the  $(N - M_r + 1)(N - M_r)/2$  elements of  $\mathcal{V}^{\text{ker}}$  belongs to  $\ker[(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N]$ ; and ii) the elements in  $\mathcal{V}^{\text{ker}}$  are linearly independent.

*Proof:* To establish i), we need to show that if  $\tilde{\mathbf{v}}_{ij} \in \mathcal{V}^{\text{ker}}$ , then  $[(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N]\tilde{\mathbf{v}}_{ij} = \mathbf{0}$ . Indeed, we have that

$$\begin{aligned} [(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N]\tilde{\mathbf{v}}_{ij} &= (\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N \mathbf{D}_N^\dagger \text{vec}(\mathbf{V}_{ij}) \\ &= (\mathbf{X}^T \otimes \mathbf{I}_N)\text{vec}(\mathbf{V}_{ij}) \\ &= (\mathbf{X}^T \otimes \mathbf{I}_N)(\mathbf{v}_i \otimes \mathbf{v}_j + \mathbf{v}_j \otimes \mathbf{v}_i) \\ &= (\mathbf{X}^T \mathbf{v}_i \otimes \mathbf{I}_N \mathbf{v}_j) + (\mathbf{X}^T \mathbf{v}_j \otimes \mathbf{I}_N \mathbf{v}_i) \\ &= \mathbf{0}, \end{aligned}$$

where the last equality follows because  $\mathbf{v}_i, \mathbf{v}_j \in \ker(\mathbf{X}^T)$ .

To prove ii) we first define a matrix  $\tilde{\mathbf{V}}^{\text{ker}}$  whose columns correspond to the vectors in  $\mathcal{V}^{\text{ker}}$  and write  $\mathbf{V}^{\text{ker}} = \mathbf{D}_N \tilde{\mathbf{V}}^{\text{ker}}$ . Notice that if  $\mathbf{V}^{\text{ker}}$  is full column rank, then all the columns in  $\tilde{\mathbf{V}}^{\text{ker}}$  must be linearly independent since  $\mathbf{D}_N$  is just a replication operator. Hence, we need to show that vectors  $\mathbf{v}_{ij}$  and  $\mathbf{v}_{i'j'}$  (both columns of  $\mathbf{V}^{\text{ker}}$ ) are orthogonal unless both  $i = i'$  and  $j = j'$ . To see why this is true, note that  $\mathbf{v}_{ij} = \mathbf{v}_i \otimes \mathbf{v}_j + \mathbf{v}_j \otimes \mathbf{v}_i$  and so

$$\begin{aligned} \mathbf{v}_{ij}^T \mathbf{v}_{i'j'} &= [(\mathbf{v}_i^T \otimes \mathbf{v}_j^T) + (\mathbf{v}_j^T \otimes \mathbf{v}_i^T)] [(\mathbf{v}_{i'} \otimes \mathbf{v}_{j'}) + (\mathbf{v}_{j'} \otimes \mathbf{v}_{i'})] \\ &= (\mathbf{v}_i^T \mathbf{v}_{i'} \otimes \mathbf{v}_j^T \mathbf{v}_{j'}) + (\mathbf{v}_i^T \mathbf{v}_{j'} \otimes \mathbf{v}_j^T \mathbf{v}_{i'}) \\ &\quad + (\mathbf{v}_j^T \mathbf{v}_{i'} \otimes \mathbf{v}_j^T \mathbf{v}_{j'}) + (\mathbf{v}_j^T \mathbf{v}_{j'} \otimes \mathbf{v}_i^T \mathbf{v}_{i'}) = \mathbf{0} \end{aligned}$$

since in all four summands there is at least one inner product that is zero, hence their sum vanishes as well. ■

From Lemma 1, we conclude that the dimension of  $\ker(\mathbf{X}^T)$  is at least the cardinality of  $\mathcal{V}^{\ker}$ . Since  $|\mathcal{V}^{\ker}| = (N - M_r + 1)(N - M_r)/2$ , it follows that  $\text{rank}[(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N]$  is at most  $N_{\mathbf{H}} - (N - M_r + 1)(N - M_r)/2$ .

Finally, to see that c) is true, first notice that whenever  $N < M_r$ , then b) guarantees that  $\text{rank}((\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N) < N_{\mathbf{H}}$ . Consequently, the Moore-Penrose minimizer is not the unique LS minimizer. By contrast, if  $N = M_r$ , then  $\mathbf{X}$  becomes full rank and, consequently,  $(\mathbf{X}^T \otimes \mathbf{I}_N)$  is also full rank. Given that  $\mathbf{D}_N$  is a full column rank matrix,  $(\mathbf{X}^T \otimes \mathbf{I}_N)\mathbf{D}_N$  is also full column rank and the Moore-Penrose minimizer becomes the unique LS minimizer, as wanted. ■

### B. PROOF OF PROPOSITION 3

Throughout this proof, the unspecified norm  $\|\cdot\|$  applied to a matrix refers to the induced  $\ell_2$  norm, also denoted as operator norm. We first quantify the error in estimating  $\mathbf{H}$  attributable to having access to a finite number  $P$  of signals. From (17) it follows that

$$\begin{aligned} \|\mathbf{H} - \mathbf{H}^*\| &= \|\mathbf{C}_{\mathbf{x}}^{-1/2}\mathbf{C}_{\mathbf{y}\mathbf{x}}^{1/2}\mathbf{C}_{\mathbf{x}}^{-1/2} - \mathbf{C}_{\mathbf{x}}^{-1/2}\hat{\mathbf{C}}_{\mathbf{y}\mathbf{x}}^{1/2}\mathbf{C}_{\mathbf{x}}^{-1/2}\| \\ &\leq \|\mathbf{C}_{\mathbf{x}}^{-1}\| \|(\mathbf{C}_{\mathbf{x}}^{1/2}\mathbf{C}_{\mathbf{y}}\mathbf{C}_{\mathbf{x}}^{1/2})^{1/2} - (\mathbf{C}_{\mathbf{x}}^{1/2}\hat{\mathbf{C}}_{\mathbf{y}}\mathbf{C}_{\mathbf{x}}^{1/2})^{1/2}\| \\ &\leq \|\mathbf{C}_{\mathbf{x}}^{-1}\| \|\mathbf{C}_{\mathbf{x}}^{1/2}\mathbf{C}_{\mathbf{y}}\mathbf{C}_{\mathbf{x}}^{1/2} - \mathbf{C}_{\mathbf{x}}^{1/2}\hat{\mathbf{C}}_{\mathbf{y}}\mathbf{C}_{\mathbf{x}}^{1/2}\|^{1/2}, \end{aligned}$$

where the first inequality follows from the definition of  $\mathbf{C}_{\mathbf{y}\mathbf{x}}$  in (12) and the submultiplicative property of the matrix norm, while the second inequality follows from [60]. A new application of the submultiplicative property readily yields

$$\|\mathbf{H} - \mathbf{H}^*\| \leq \|\mathbf{C}_{\mathbf{x}}^{-1}\| \|\mathbf{C}_{\mathbf{x}}\|^{1/2} \|\mathbf{C}_{\mathbf{y}} - \hat{\mathbf{C}}_{\mathbf{y}}\|^{1/2}.$$

Finally, relying on the sub-Gaussianity of the input [cf. A-1)], we may bound the distance between the sample covariance of  $\mathbf{y}$  and its true covariance with probability  $1 - \delta$  as [61]

$$\|\mathbf{H} - \mathbf{H}^*\| \leq \|\mathbf{C}_{\mathbf{x}}^{-1}\| \|\mathbf{C}_{\mathbf{x}}\|^{1/2} C_{\delta}' P^{-1/4}. \quad (30)$$

In order to quantify the distance between the true eigenvectors  $\mathbf{V}$  and those  $\hat{\mathbf{V}}$  estimated from  $\mathbf{H}^*$  we rely on the Davis-Kahan theorem [62] to obtain, for all  $i$ ,

$$\|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2 \leq \frac{\sqrt{8} \|\mathbf{H} - \mathbf{H}^*\|}{\gamma}, \quad (31)$$

where  $\gamma > 0$  is the minimum difference between two eigenvalues of  $\mathbf{H}$ . Assumption A-2) guarantees that, indeed,  $\gamma$  is bounded away from zero.

Our goal is to leverage assumptions A-3) and A-4) to claim that the conditions needed to invoke [12, Prop. 2] hold for large enough  $P$ . However, to leverage [12, Prop. 2] in establishing the sought bound (19), we need to find a matrix  $\mathbf{S}' = \hat{\mathbf{V}}\mathbf{\Lambda}'\hat{\mathbf{V}}^T$  and a constant  $\epsilon$  such that

$$\|\mathbf{S} - \mathbf{S}'\|_F^2 + \left( \sum_j S'_{1j} - 1 \right)^2 \leq \epsilon^2. \quad (32)$$

Notice that, without loss of generality, we assumed that  $\sum_j S_{1j} = 1$  for the true GSO. Let us propose the matrix  $\mathbf{S}' = \hat{\mathbf{V}}\mathbf{\Lambda}'\hat{\mathbf{V}}^T$ , where  $\mathbf{\Lambda}'$  denote the eigenvalues of the true  $\mathbf{S}$ , and compute the  $\epsilon$  needed for (32) to hold. We have that

$$\begin{aligned} \|\mathbf{S} - \mathbf{S}'\|_F &= \left\| \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T - \sum_i \lambda_i \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^T \right\|_F \\ &\leq \sum_i \lambda_i \|\mathbf{v}_i \mathbf{v}_i^T - \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^T\|_F \\ &\leq 2N \sum_i \lambda_i \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2, \end{aligned} \quad (33)$$

where the first inequality is an immediate application of the triangle inequality and the second one follows from expanding the Frobenius norm. Regarding the second term in (32) we can leverage the fact that  $\sum_j S_{1j} = 1$  to obtain

$$\begin{aligned} \sum_j S'_{1j} - 1 &= \sum_j (S'_{1j} - S_{1j}) = \sum_{ij} \lambda_i ([\hat{\mathbf{v}}_i]_1 [\hat{\mathbf{v}}_i]_j - [\mathbf{v}_i]_1 [\mathbf{v}_i]_j) \\ &\leq 2N \sum_i \lambda_i \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2, \end{aligned} \quad (34)$$

where the second equality stems from the eigendecomposition of  $\mathbf{S}'$  and  $\mathbf{S}$  whereas the inequality follows the same reasoning as in (33). Combining (33) and (34), we have that (32) is satisfied if we select

$$\epsilon = \sqrt{8}N \sum_i \lambda_i \|\mathbf{v}_i - \hat{\mathbf{v}}_i\|_2 \leq C_{\delta}'' P^{-1/4}, \quad (35)$$

where the inequality is derived from (30) and (31) by absorbing all quantities that do not depend on  $P$  into the constant  $C_{\delta}''$ . Having identified an  $\mathbf{S}'$  and an associated  $\epsilon$  such that (32) is satisfied, we may rely on [12, Prop. 2] to establish that for all  $P > P_0$ , with high probability,

$$\|\mathbf{S}_P^* - \mathbf{S}\|_1 \leq C_{P_0} \epsilon \leq C_{\delta} P^{-1/4}, \quad (36)$$

where the final inequality follows by substituting  $\epsilon$  as in (35) and merging all constants into  $C_{\delta}$ . ■

### C. DERIVATION OF (20)

Recall that the matrix  $\mathbf{C}_{\mathbf{y}\mathbf{x}}$  is symmetric and PSD, and hence, it can be eigendecomposed as  $\mathbf{C}_{\mathbf{y}\mathbf{x}} = \mathbf{V}_{\mathbf{y}\mathbf{x}} \text{diag}(\boldsymbol{\lambda}_{\mathbf{y}\mathbf{x}}) \mathbf{V}_{\mathbf{y}\mathbf{x}}^T$  with  $\mathbf{V}_{\mathbf{y}\mathbf{x}} \mathbf{V}_{\mathbf{y}\mathbf{x}}^T = \mathbf{I}$  and  $[\boldsymbol{\lambda}_{\mathbf{y}\mathbf{x}}]_i \geq 0$  for all  $i$ . Our goal is to show that, if  $\mathbf{H}_{\mathbf{x}\mathbf{x}}$  is symmetric, then any solution to

$$\mathbf{C}_{\mathbf{y}\mathbf{x}} = \mathbf{H}_{\mathbf{x}\mathbf{x}} \mathbf{H}_{\mathbf{x}\mathbf{x}}^T. \quad (37)$$

has the form given in (20), that is

$$\mathbf{H}_{\mathbf{x}\mathbf{x}} = \mathbf{C}_{\mathbf{y}\mathbf{x}}^{1/2} \mathbf{V}_{\mathbf{y}\mathbf{x}} \text{diag}(\mathbf{b}) \mathbf{V}_{\mathbf{y}\mathbf{x}}^T, \quad (38)$$

where  $\mathbf{C}_{\mathbf{y}\mathbf{x}}^{1/2}$  is the principal square root of  $\mathbf{C}_{\mathbf{y}\mathbf{x}}$ , and  $\mathbf{b} \in \{-1, 1\}^N$  is a binary signed vector.

Focus first on (38) and observe that  $\mathbf{C}_{\mathbf{y}\mathbf{x}}^{1/2} = \mathbf{V}_{\mathbf{y}\mathbf{x}} \text{diag}(\bar{\boldsymbol{\lambda}}_{\mathbf{y}\mathbf{x}}) \mathbf{V}_{\mathbf{y}\mathbf{x}}^T$ , with  $[\bar{\boldsymbol{\lambda}}_{\mathbf{y}\mathbf{x}}]_i \geq 0$  and  $[\bar{\boldsymbol{\lambda}}_{\mathbf{y}\mathbf{x}}]_i^2 = [\boldsymbol{\lambda}_{\mathbf{y}\mathbf{x}}]_i$  for all  $i$ . Hence, (38) can be alternatively written as



$$\begin{aligned} \mathbf{H}_{\mathbf{xx}} &= \mathbf{V}_{\mathbf{yx}} \text{diag}(\bar{\lambda}_{\mathbf{yx}}) \mathbf{V}_{\mathbf{yx}} \mathbf{V}_{\mathbf{yx}}^T \text{diag}(\mathbf{b}) \mathbf{V}_{\mathbf{yx}}^T \\ &= \mathbf{V}_{\mathbf{yx}} \text{diag}(\bar{\lambda}_{\mathbf{yx}} \odot \mathbf{b}) \mathbf{V}_{\mathbf{yx}}^T. \end{aligned} \quad (39)$$

We now shift focus to (37). The symmetry of  $\mathbf{H}_{\mathbf{xx}}$  implies that: i) the eigendecomposition  $\mathbf{H}_{\mathbf{xx}} = \mathbf{V}_{\mathbf{Hx}} \text{diag}(\lambda_{\mathbf{Hx}}) \mathbf{V}_{\mathbf{Hx}}^T$  is guaranteed to exist; and ii) the right-hand-side of (37) can be written as  $\mathbf{H}_{\mathbf{xx}} \mathbf{H}_{\mathbf{xx}}^T = \mathbf{H}_{\mathbf{xx}}^2 = \mathbf{V}_{\mathbf{Hx}} \text{diag}^2(\lambda_{\mathbf{Hx}}) \mathbf{V}_{\mathbf{Hx}}^T$ . Since this latter matrix has to be equal to  $\mathbf{V}_{\mathbf{yx}} \text{diag}(\lambda_{\mathbf{yx}}) \mathbf{V}_{\mathbf{yx}}^T$ , the left-hand-side of (37), we readily have that: a) the eigenvectors of  $\mathbf{C}_{\mathbf{yx}}$  coincide with the eigenvectors of  $\mathbf{H}_{\mathbf{xx}}$ ; and b) the eigenvalues of the two matrices of interest satisfy  $\text{diag}(\lambda_{\mathbf{yx}}) = \text{diag}^2(\lambda_{\mathbf{Hx}})$ . This is exactly the form in (39), with  $[\lambda_{\mathbf{yx}}]_i = |[\lambda_{\mathbf{Hx}}]_i|$ ,  $[\mathbf{b}]_i = \text{sign}([\lambda_{\mathbf{Hx}}]_i)$  and  $\mathbf{V}_{\mathbf{yx}} = \mathbf{V}_{\mathbf{Hx}}$ . ■

#### D. PROOF OF PROPOSITION 4

From (11) it follows that

$$\mathbf{C}_{y,1} = \mathbf{H} \mathbf{C}_{x,1} \mathbf{H} = \mathbf{H} \mathbf{U} \text{diag}(\lambda_1) \mathbf{U}^T \mathbf{H} = \mathbf{Q} \text{diag}(\lambda_1) \mathbf{Q}^T,$$

where we have implicitly defined  $\mathbf{Q} := \mathbf{H} \mathbf{U}$ . Notice that the basis  $\mathbf{U}$  is completely determined since all eigenvalues in  $\lambda_1$  are distinct [cf. A-1)]. Similarly, for the second diffusion process we obtain that  $\mathbf{C}_{y,2} = \mathbf{Q} \text{diag}(\lambda_2) \mathbf{Q}^T$ . Furthermore, if we define the matrix  $\mathbf{R}_x = [\lambda_1, \lambda_2]^T \in \mathbb{R}^{2 \times N}$ , and the  $N \times N \times 2$  tensor  $\underline{\mathbf{C}}_y$  (with slices along the third mode given by  $\mathbf{C}_{y,1}$  and  $\mathbf{C}_{y,2}$ ), then the partial symmetric PARAFAC decomposition of  $\underline{\mathbf{C}}_y$  factors into the matrices  $\mathbf{Q}$ ,  $\mathbf{Q}$ , and  $\mathbf{R}_x$ ; see [24], [63].

Recall that the *Kruskal rank* of a matrix  $\mathbf{A} \in \mathbb{R}^{N \times M}$  (denoted by  $\text{kr}(\mathbf{A})$ ) is defined as the maximum number  $k$  such that any combination of  $k$  columns of  $\mathbf{A}$  constitutes a full rank submatrix. In this way, from condition A-2) it follows that  $\text{kr}(\mathbf{R}_x) = 2$  and from the invertibility of  $\mathbf{H}$  [cf. A-4)] it follows that  $\text{kr}(\mathbf{Q}) = N$ . Leveraging established results on the uniqueness of PARAFAC tensor decompositions (see [24, Theorem 1]), it holds that a PARAFAC decomposition of  $\underline{\mathbf{C}}_y$  recovers  $\mathbf{Q}$  and  $\mathbf{R}_x$  up to scaling and rotation ambiguities. However, given that we know  $\mathbf{R}_x$  a priori, part of those ambiguities can be resolved; see e.g. [24, Lemma 1]. To be more precise, it holds that we can recover  $\mathbf{Q}'$ , where  $\mathbf{Q}' = \mathbf{Q} \text{diag}(\bar{\mathbf{b}})$  for some unknown  $\bar{\mathbf{b}} \in \{-1, 1\}^N$ . However, the following lemma establishes how to uniquely recover  $\bar{\mathbf{b}}$ , and the uniqueness of  $\mathbf{H} = \mathbf{Q} \mathbf{U}^T = \mathbf{Q}' \text{diag}(\bar{\mathbf{b}}) \mathbf{U}^T$  follows.

**Lemma 2:** Vector  $\bar{\mathbf{b}} \in \{-1, 1\}^N$  can be found as the only vector (up to sign) such that  $\mathbf{Q}' \text{diag}(\bar{\mathbf{b}}) \mathbf{U}^T$  is symmetric.

*Proof:* Combining the facts that  $\mathbf{H} = \mathbf{Q}' \text{diag}(\bar{\mathbf{b}}) \mathbf{U}^T$  for the true  $\bar{\mathbf{b}}$  and that  $\mathbf{H}$  is symmetric, we can conclude that  $\mathbf{Q}' \text{diag}(\bar{\mathbf{b}}) \mathbf{U}^T$  is symmetric. Thus, we are left to show that no other  $\mathbf{b}' \in \{-1, 1\}^N$  leads to a symmetric  $\mathbf{Q}' \text{diag}(\mathbf{b}') \mathbf{U}^T$ . To show this, begin by defining the symmetric matrix  $\mathbf{P} := \mathbf{U} \text{diag}(\bar{\mathbf{b}}) \text{diag}(\mathbf{b}') \mathbf{U}^T$ . Hence it follows that if indeed  $\mathbf{Q}' \text{diag}(\mathbf{b}') \mathbf{U}^T = (\mathbf{Q}' \text{diag}(\bar{\mathbf{b}}) \mathbf{U}^T)^T$  then

$$\mathbf{H} \mathbf{U} \text{diag}(\bar{\mathbf{b}}) \text{diag}(\mathbf{b}') \mathbf{U}^T = \mathbf{U} \text{diag}(\bar{\mathbf{b}}) \text{diag}(\mathbf{b}') \mathbf{U}^T \mathbf{H}. \quad (40)$$

This means that  $\mathbf{H}$  and  $\mathbf{P}$  must commute and this requires them to be simultaneously diagonalizable. However, since  $\mathbf{U}$  and  $\mathbf{V}$  (the eigenbasis of  $\mathbf{H}$ ) do not share any eigenvector [cf. A-3)],

this can only happen if  $\mathbf{P} = \mathbf{I}$  or  $\mathbf{P} = -\mathbf{I}$ . Hence, we must have that  $\mathbf{b}' = \bar{\mathbf{b}}$  or  $\mathbf{b}' = -\bar{\mathbf{b}}$  and identifiability of  $\bar{\mathbf{b}}$  is guaranteed. ■

#### E. PROOF OF PROPOSITION 5

Using properties of the Khatri-Rao product, feasible graph filters  $\mathbf{H} \in \mathcal{H}_M^{\text{sym}}$  satisfy the system of equations [cf. (22)]

$$\mathbf{A}_m \mathbf{b}_m = \text{vec}(\mathbf{H}), \quad m = 1, \dots, M, \quad (41)$$

with matrices  $\mathbf{A}_m$  defined in the statement of the proposition and for some binary vectors  $\mathbf{b}_m \in \{-1, 1\}^N$ . Based on (24), the equations in (41) can be compactly and equivalently rewritten as

$$\boldsymbol{\Psi} \mathbf{b} = \mathbf{0}, \quad \mathbf{A}_1 \mathbf{b}_1 = \text{vec}(\mathbf{H}). \quad (42)$$

Under the assumption that the covariances  $\{\mathbf{C}_{y,m}\}_{m=1}^N$  are perfectly known, then the filter  $\mathbf{H}$  can be uniquely identified (up to a sign ambiguity) provided  $\text{rank}(\boldsymbol{\Psi}) = NM - 1$ . This way,  $\ker(\boldsymbol{\Psi})$  has dimensionality one and the basis vector  $\mathbf{b}^* = [(\mathbf{b}_1^*)^T, \dots, (\mathbf{b}_M^*)^T]^T \in \{-1, 1\}^{NM}$  of the aforementioned null space can be used to recover the filter via  $\mathbf{A}_1 \mathbf{b}_1^* = \text{vec}(\mathbf{H})$ . Moreover, the desired solution of (42) can be equivalently obtained from the BQP

$$\mathbf{b}^* = \underset{\mathbf{b} \in \{-1, 1\}^{NM}}{\text{argmin}} \|\boldsymbol{\Psi} \mathbf{b}\|_2^2,$$

which is identical to (25). ■

#### ACKNOWLEDGMENT

The authors would like to thank the authors of [24] for sharing the Matlab code of their algorithm, and the anonymous reviewers for constructive comments on an earlier version of this paper.

#### REFERENCES

- [1] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 5870–5874.
- [2] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying undirected network structure via semidefinite relaxation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2018, pp. 4049–4053.
- [3] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
- [4] 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.
- [5] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, May 2017.
- [6] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. New York, NY, USA: Cambridge Univ. Press, 2008.
- [7] M. H. DeGroot, "Reaching a consensus," *J. Amer. Stat. Assoc.*, vol. 69, pp. 118–121, 1974.
- [8] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. New York, NY, USA: Springer, 2009.
- [9] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Aug. 2017.
- [10] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.

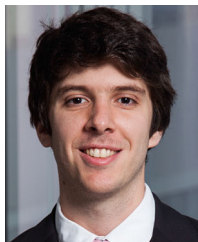
- [11] B. Girault, "Stationary graph signals using an isometric graph translation," in *Proc. Eur. Signal Process. Conf.*, Aug. 2015, pp. 1516–1520.
- [12] S. Segarra, A. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Aug. 2017.
- [13] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, May 2019.
- [14] O. Sporns, *Discovering the Human Connectome*. Boston, MA, USA: MIT Press, 2012.
- [15] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.
- [16] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [17] O. Banerjee, L. E. Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, 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] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 825–841, Sep. 2017.
- [21] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. Palomar, "Structured graph learning via Laplacian spectral constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 11651–11663.
- [22] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Stat.*, vol. 34, pp. 1436–1462, 2006.
- [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] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3675–3687, Jul. 2017.
- [25] 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.
- [26] 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.*, Mar. 2016, pp. 6315–6319.
- [27] 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, Feb. 2017.
- [28] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 3, pp. 481–496, Sep. 2018.
- [29] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 484–499, Sep. 2017.
- [30] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.
- [31] 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, Aug. 2016.
- [32] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. Int. Conf. Artif. Intel. Statist. J. Mach. Learn. Res.*, 2016, pp. 920–929.
- [33] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from filtered signals: Graph system and diffusion kernel identification," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 2, pp. 360–374, Jun. 2019.
- [34] 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.*, Mar. 2017, pp. 6508–6512.
- [35] M. G. Rabbat, "Inferring sparse graphs from smooth signals with theoretical guarantees," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 6533–6537.
- [36] Y. Zhu, M. T. Schaub, A. Jadbabaie, and S. Segarra, "Network inference from consensus dynamics with unknown parameters," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 300–315, Apr. 2020.
- [37] H. Wai, S. Segarra, A. E. Ozdaglar, A. Scaglione, and A. Jadbabaie, "Blind community detection from low-rank excitations of a graph filter," *IEEE Trans. Signal Process.*, vol. 68, pp. 436–451, Dec. 2019.
- [38] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," in *Proc. ICLR Workshop Representation Learn. Graphs Manifolds*, May 2019, pp. 1–13.
- [39] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE Trans. Signal Process.*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.
- [40] S. Mortaheb et al., "A graph signal processing approach to study high density EEG signals in patients with disorders of consciousness," in *Proc. IEEE Int. Conf. Eng. Med. Biol.*, 2019, pp. 4549–4553.
- [41] J. R. Magnus and H. Neudecker, "The elimination matrix: Some lemmas and applications," *SIAM. J. Algebr. Discrete Methods*, vol. 1, no. 4, pp. 422–449, Dec. 1980.
- [42] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters," *IEEE Trans. Signal Process.*, vol. 65, no. 5, pp. 1146–1159, Mar. 2017.
- [43] D. A. Spielman, H. Wang, and J. Wright, "Exact recovery of sparsely-used dictionaries," in *Proc. 25th Annu. Conf. Learn. Theory, Ser. Proc. Mach. Learn. Res.*, Edinburgh, Scotland, Jun. 2012, pp. 37.1–37.18.
- [44] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001.
- [45] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Belmont, MA, USA: Athena Scientific, 1999.
- [46] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [47] Y. Ye, *Interior Point Algorithms: Theory and Analysis* (Wiley Inter-science Series in Discrete Mathematics and Optimization). New York, NY, USA: Wiley, 1997.
- [48] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optim. Methods Softw.*, vol. 9, no. 1–3, pp. 141–160, 1998.
- [49] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1442–1449, Aug. 2001.
- [50] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.
- [51] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program.*, vol. 95, no. 2, pp. 329–357, Feb. 2003.
- [52] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977.
- [53] M. E. Newman, "Scientific collaboration networks. II. shortest paths, weighted networks, and centrality," *Phys. Rev. E*, vol. 64, no. 1, 2001, Art. no. 016132.
- [54] P. Hagmann et al., "Mapping the structural core of human cerebral cortex," *PLoS Biol.*, vol. 6, no. 7, 2008, Paper e159.
- [55] C. Hu, X. Hua, J. Ying, P. M. Thompson, G. E. Fakhri, and Q. Li, "Localizing sources of brain disease progression with network diffusion model," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1214–1225, Oct. 2016.
- [56] C. Bialik, A. Flowers, R. Fischer-Baum, and D. Mehta, *Uber is Serving New York's Outer Boroughs More Than Taxis Are. FiveThirtyEight*. Aug. 2015. [Online]. Available: <https://fivethirtyeight.com/features/uber-is-serving-new-yorks-outer-boroughs-more-than-taxis-are/>
- [57] W. Hu, Uber, *Surging Outside Manhattan, Taxis in New York City*. *The New York Times*. Oct. 12, 2017. [Online]. Available: <https://www.nytimes.com/2017/10/12/nyregion/uber-taxis-new-york-city.html>
- [58] H. Bessembinder, "Do stocks outperform treasury bills?," *J. Financial Econ.*, vol. 129, no. 3, pp. 440–457, 2018.
- [59] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statist. Mech. Theory Exp.*, vol. 2008, no. 10, 2008, Art. no. P10008.
- [60] K. M. Audenaert and J. S. Aujla, "On norm sub-additivity and super-additivity inequalities for concave and convex functions," *Linear Multilinear Algebra*, vol. 60, no. 11/12, pp. 1369–1389, 2012.

- [61] R. Vershynin, "How close is the sample covariance matrix to the actual covariance matrix?," *J. Theor. Probability*, vol. 25, no. 3, pp. 655–686, 2012.
- [62] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the davis-kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2014.
- [63] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the CAMDECOMP/PARAFAC decomposition," *Linear Algebra Appl.*, vol. 420, no. 2, pp. 540–552, Jan. 2007.



**RASOUL SHAFIPOUR** received the B.Sc. degree in electrical engineering (communication systems) from the Sharif University of Technology, Tehran, Iran, in 2015, and the M.Sc. degree in 2017 and the Ph.D. degree in 2020 in electrical engineering from the University of Rochester, Rochester, NY, USA. From January to May 2019, he was a Research Intern with Baidu Research, Seattle, WA, USA. Since March 2020, he has been a Data and Applied Scientist with Microsoft. His research interests include machine learning, signal processing,

network science, optimization, and algorithms. He obtained the sixth rank in the Iranian nation-wide university entrance exam among more than 270000 participants in 2010. He was a recipient of the University of Rochester's Donald M. and Janet C. Barnard fellowship and a Best Student Paper Award at the 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing.



**SANTIAGO SEGARRA** (Member, IEEE) 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 from the University of Pennsylvania (Penn), Philadelphia, PA, USA, in 2014 and 2016, respectively. From September 2016 to June 2018, he was a Postdoctoral Research Associate with the Institute for Data, Systems, and

Society, the Massachusetts Institute of Technology, Cambridge, MA, USA. Since July 2018, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. His research interests include network theory, data analysis, machine learning, and graph signal processing. He was the recipient of 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 2017 Penn's Joseph and Rosaline Wolf Award for Best Doctoral Dissertation in electrical and systems engineering, the 2020 IEEE Signal Processing Society Young Author Best Paper Award, and five best conference paper awards.



**ANTONIO G. MARQUES** (Senior Member, IEEE) received the telecommunications engineering degree and the Doctorate degree (with highest honors) from the Carlos III University of Madrid, Getafe, Spain, in 2002 and 2007, respectively. In 2007, he became a Faculty with the Department of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as a Full Professor. From 2005 to 2015, he held different visiting positions with the University of Minnesota, Minneapolis. In 2015, 2016 and 2017 he was a Visiting Scholar with the University of Pennsylvania, Philadelphia, PA, USA.

His current research focuses on signal processing, machine learning, data science and artificial intelligence over graphs, and nonlinear and stochastic optimization of wireless, power and transportation networks. He held the number of posts with the IEEE, including as an Associate Editor and the Technical/General Chair of different conferences, and he is currently a Member of the IEEE Signal Processing Theory and Methods Technical Communication. His work has been awarded in several journals, conferences and workshops, with recent ones including Asilomar 2015, IEEE SSP 2016, IEEE SAM 2016 and IEEE SPS Y.A. Best Paper Award (2020, senior author). He was the recipient of the 2020 EURASIP Early Career Award and is a Member of the EURASIP and the ELLIS society.



**GONZALO MATEOS** (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Universidad de la República, Montevideo, Uruguay, in 2005, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2009 and 2011, respectively. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA, and an Asaro Biggar Family Fellow in Data Science. In 2013, he was a Visiting

Scholar with the CS Department, Carnegie Mellon University, Pittsburgh, PA, USA. From 2004 to 2006, he was a Systems Engineer with Asea Brown Boveri, Montevideo, Uruguay.

His research interests include statistical learning from complex data, network science, decentralized optimization, and graph signal processing. He is currently the Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. He was the recipient of the NSF CAREER Award in 2018, the 2017 and 2020 IEEE Signal Processing Society Young Author Best Paper Awards (as senior coauthor), the 2019 IEEE Signal Processing Society Outstanding Editorial Board Award, and the Best Paper Awards at SPAWC 2012, SSP 2016, and ICASSP 2018 and 2019. His doctoral work has been recognized with the 2013 University of Minnesota's Best Dissertation Award (Honorable Mention) across all physical sciences and engineering areas.