Blind Identification of Graph Filters

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

***Abstract*—Network processes are often represented as signals defined on the vertices of a graph. To untangle the latent structure of such signals, one can view them as outputs of *linear graph filters* modeling underlying network dynamics. This paper deals with the problem of joint identification of a graph filter and its input signal, thus broadening the scope of classical blind deconvolution of temporal and spatial signals to the less-structured graph domain. Given a graph signal $\mathbf{y}$ modeled as the output of a graph filter, the goal is to recover the vector of filter coefficients $\mathbf{h}$, and the input signal $\mathbf{x}$ which is assumed to be sparse. While $\mathbf{y}$ is a bilinear function of $\mathbf{x}$ and $\mathbf{h}$, the filtered graph signal is also a linear combination of the entries of the lifted rank-one, row-sparse matrix $\mathbf{x}\mathbf{h}^T$. The blind graph-filter identification problem can thus be tackled via rank and sparsity minimization subject to linear constraints, an inverse problem amenable to convex relaxations offering provable recovery guarantees under simplifying assumptions. Numerical tests using both synthetic and real-world networks illustrate the merits of the proposed algorithms, as well as the benefits of leveraging multiple signals to aid the blind identification task.**

***Index Terms*—Graph signal processing, blind system identification, graph filter, network diffusion process.**

I. Introduction

COPING with the challenges found at the intersection of Network Science and Big Data necessitates broadening the scope beyond classical temporal signal analysis and processing, to also accommodate *signals defined on graphs* [3]–[5]. Under the assumption that the signal properties are related to the topology of the graph where they are supported, the goal of graph signal processing (GSP) is to develop algorithms that fruitfully leverage this relational structure, and can make inferences about these relationships when they are only partially observed [5]. A suitable way to accomplish these objectives is to rely on the so-called graph-shift operator, which is a matrix that reflects the local connectivity of the graph [4].

Manuscript received April 14, 2016; revised October 5, 2016; accepted November 1, 2016. Date of publication November 14, 2016; date of current version December 15, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yuichi Tanaka. This work was supported by the USA NSF CCF-1217963 and Spanish MINECO under Grant TEC2013-41604-R. Part of the results in this paper were presented at the 2015 CAMSAP Workshop [1], and at the 2016 ICASSP Conference [2].

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

G. Mateos is with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14611 USA (e-mail: gmateosb@ur.rochester.edu).

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

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2628343

We consider here that each node has a certain value, and these values are collected across nodes to form a graph signal. With this definition, graph filters – which are a generalization of classical time-invariant systems – are a specific class of operators whose input and output are graph signals (cf. Section II). Mathematically, graph filters are linear transformations that can be expressed as polynomials of the graph-shift operator [6]. The polynomial coefficients determine completely the transformation and are referred to as *filter coefficients*. Such linear transformations can be implemented via local interactions among nodes, and may be used to model e.g., diffusion or percolation dynamics in the network [7]–[9].

Contributions: This paper investigates the problem of blind identification of graph filters. Specifically, we are given a graph signal $\mathbf{y}$ which is assumed to be the output of a graph filter, and seek to *jointly identify* the *filter* coefficients $\mathbf{h}$ and the *input signal* $\mathbf{x}$ that gave rise to $\mathbf{y}$. This is the extension to graphs of the classical problem of blind system identification or blind deconvolution of signals in the time or spatial domains [10]. Since the inverse problem is ill-posed, we assume that the length of $\mathbf{h}$ is small and that $\mathbf{x}$ is sparse. This is the case when, e.g., a few seeding nodes inject a signal that is diffused throughout a network [11], [12]. While $\mathbf{y}$ is a bilinear function of $\mathbf{x}$ and $\mathbf{h}$, we show that the filtered graph signal is also a linear combination of the entries of the lifted rank-one, row-sparse matrix $\mathbf{x}\mathbf{h}^T$ [10], [13]. The blind graph-filter identification problem can thus be tackled via joint rank and sparsity minimization subject to linear constraints, an approach amenable to convex relaxation [14], [15]. Several alternatives are proposed to approach such a relaxation, including generalizations facilitating blind graph-filter identification when multiple outputs (each corresponding to a different input) are available; see also [16] for identifiability claims in a setting unrelated to graphs. Under simplifying assumptions, probabilistic recovery conditions are also derived. Together with the proof, effort is devoted towards building intuition on the obtained performance guarantees by identifying graph-related parameters that have a major impact on blind identification, as well as by distilling the fundamental differences relative to the time domain [13]. Numerical tests not only showcase the effectiveness of the proposed algorithm on synthetic and real graphs, but also illustrate that recovery is in practice possible under conditions far less restrictive than those stemming from the analysis in Section IV.

Envisioned applications: The dynamics of opinion formation in social networks can be modeled using distributed linear operations implemented by multi-agent systems; see e.g., [17], [18]. Interestingly, graph filters have been used to implement distributedly related linear transformations such as fast consensus [19], and projections onto the low-rank space of graph-bandlimited signals [20]. This motivates adopting the algorithms proposed here to identify those influential actors (i.e., the non-zero entries in $\mathbf{x}$) that instilled the observed status-quo.

Another example of interest is given by structural and functional brain networks, which are becoming increasingly central to the analysis of brain signals. Nodes correspond to regions of interest (ROIs) and their associated information (e.g., their level of neural activity) can be represented as a graph signal. Suppose that an observed brain signal corresponds to the linear combination of a diffused pattern of an originally sparse brain signal (i.e., generated by a few active ROIs). Blind identification amounts then to jointly estimating the desired original brain signal and the combination coefficients. In the analysis of epileptic seizure data for instance, estimating the (sparse) input state can help to identify the ROIs from where the seizure emanated, which may serve to guide surgical intervention [21]. While linear models of processes in the brain are admittedly simplistic, they can still offer informative insights [22]. In the same vein, we envision applications in marketing where e.g., social media advertisers want to identify a small set of initiators so that an online campaign can go viral; in healthcare policy implementing network analytics to infer hidden needle-sharing networks of injecting drug users [5]; or, in environmental monitoring using wireless sensor networks to localize heat or seismic sources [23].

Relation to prior work: Some of the ideas here were inspired by the work in [10], where matrix lifting is used for blind deconvolution of temporal and spatial signals. In the current paper, the linear operator mapping $\mathbf{x}\mathbf{h}^T$ to the output signal $\mathbf{y}$ depends on the spectral properties of the graph-shift operator [1], a departure from the random (Gaussian or partial Fourier) operators arising with the biconvex compressed sensing approach in [13]. Despite its practical interest, the setup where multiple output signals are observed (each one corresponding to a different sparse input) has received little attention in recent convex relaxation approaches to blind deconvolution [16].

Paper outline: Section II introduces notation and explains how graph signals and filters can be used to model linear diffusion processes. Section III formulates the problem of blind graph-filter identification and proposes several efficient convex relaxations. In particular, Section III-C discusses algorithms for the setup where multiple observed outputs are available. Section IV provides analytical results on the recovery performance, along with graph-specific parameters that affect the recovery guarantees. Numerical experiments illustrating the merits of our approach are presented in Section V and concluding remarks are given in Section VI.

Notation: Entries of a matrix $\mathbf{X}$ and a (column) vector $\mathbf{x}$ are denoted as X_{ij} and x_i. Operators $(\cdot)^T$, $(\cdot)^H$, $\mathbb{E}\,[\cdot]$, $\circ$, $\otimes$ and $\odot$ stand for matrix transpose, conjugate transpose (Hermitian), expectation, Hadamard (entry-wise), Kronecker, and Khatri-Rao (column-wise Kronecker) products, respectively. The complex conjugate of x is denoted by $\bar{x}$; $\text{diag}(\mathbf{x})$ is a diagonal matrix whose entry (i,i) is equal to x_i; and $|\cdot|$ is used for the cardinality of a set, and the magnitude of a scalar. The $n \times n$ identity matrix is represented by $\mathbf{I}_n$, while $\mathbf{0}_n$ stands for the $n \times 1$ vector of all zeros, and $\mathbf{0}_{n\times p} = \mathbf{0}_n\mathbf{0}_p^T$. The notation $\|\mathbf{X}\|_\infty$ and $\|\mathbf{X}\|$ stand for the entrywise largest absolute value and the largest singular value of $\mathbf{X}$, respectively. For a linear operator $\mathcal{X}$, $\|\mathcal{X}\| := \|\mathbf{X}\|$, where $\mathbf{X}$ is the matrix representation of $\mathcal{X}$. Otherwise, standard vector and matrix norm notation is used.

II. Graph Signals and Graph Filters

Often, networks have intrinsic value and are themselves the object of study. In other occasions, the network defines an underlying notion of proximity, but the object of interest is a signal defined over the graph, i.e., data associated with the nodes of the network. This is the matter addressed by GSP, where the notions of, e.g., frequency and filtering (reviewed next) are extended to signals supported on graphs [3], [24].

Graph signals and graph-shift operator: Let $\mathcal{G}$ denote a directed graph with a set of nodes $\mathcal{N}$ (with cardinality N) and a set of links $\mathcal{E}$, if i is connected to j then $(i,j) \in \mathcal{E}$. Since $\mathcal{G}$ is directed, local connectivity is captured by the set $\mathcal{N}_i := \{j\,|(j,i) \in \mathcal{E}\}$ which stands for the (incoming) neighborhood of i. For any given $\mathcal{G}$ we define the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ as a sparse matrix with non-zero elements A_{ji} if and only if $(i,j) \in \mathcal{E}$. The value of A_{ji} captures the strength of the connection from i to j.

The focus of the paper is on analyzing and modeling (graph) signals defined on $\mathcal{N}$. These signals can be represented as vectors $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$, where x_i represents the value of the signal at node i. Since the vectorial representation does not account explicitly for the structure of the graph, $\mathcal{G}$ can be endowed with the so-called *graph-shift operator* $\mathbf{S}$ [4], [6]. The shift $\mathbf{S} \in \mathbb{R}^{N\times N}$ is a matrix whose entry S_{ji} can be non-zero only if $i = j$ or if $(i,j) \in \mathcal{E}$. The sparsity pattern of the matrix $\mathbf{S}$ captures the local structure of $\mathcal{G}$, but we make no specific assumptions on the values of its non-zero entries. The intuition behind $\mathbf{S}$ is to represent a linear transformation that can be computed locally at the nodes of the graph. More rigorously, if $\mathbf{y}$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node i can compute y_i as linear combination of the signal values x_j at node i's neighbors $j \in \mathcal{N}_i$. For example, one can think of an individual's opinion formation process as one of weighing in the views of close friends regarding the subject matter. Typical choices for $\mathbf{S}$ are the adjacency matrix $\mathbf{A}$ [4], [6], and the graph Laplacian [3]. We assume henceforth that $\mathbf{S}$ is diagonalizable, so that $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ with $\mathbf{\Lambda} \in \mathbb{C}^{N\times N}$ being diagonal. In particular, $\mathbf{S}$ is diagonalizable when it is normal, i.e., it satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$. In that case we have that $\mathbf{V}$ is unitary, which implies $\mathbf{V}^{-1} = \mathbf{V}^H$, and leads to the decomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$.

Graph filters as models of network diffusion processes: The shift $\mathbf{S}$ can be used to define linear graph-signal *operators* of the form

$$\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l \tag{1}$$

which are called *graph filters* [4]. For a given input $\mathbf{x}$, the output of the filter is simply $\mathbf{y} = \mathbf{H}\mathbf{x}$. The coefficients of the filter are collected into $\mathbf{h} := [h_0, \ldots, h_{L-1}]^T$, with $L-1$ denoting the filter degree. Graph filters are of particular interest because they represent linear transformations that can be implemented in a distributed fashion [7], [12], e.g., with $L-1$ successive exchanges of information among neighbors.

Graph filters can be used to model linear diffusion dynamics that depend on the network topology. Formally, the signal at node i during the step $(l+1)$ of a linear diffusion process in

$\mathcal{G}$ can be written as

$$x_i^{(l+1)} = \alpha_{ii} x_i^{(l)} + \sum_{j \in \mathcal{N}_i} \alpha_{ij} x_j^{(l)} \tag{2}$$

where α_{ij} are the diffusion coefficients. Leveraging the GSP framework, (2) is equivalent to writing that the graph signal at iteration $l+1$ is the shifted version of the signal at the previous iteration $\mathbf{x}^{(l+1)} = \mathbf{S}\mathbf{x}^{(l)}$, where the entries of the shift operator $\mathbf{S}$ are $S_{ij} = \alpha_{ij}$ if either $i = j$ or $(j,i) \in \mathcal{E}$, and $S_{ij} = 0$ otherwise. Notice that if, for example, we set $\mathbf{S} = \mathbf{I}_N - \beta \mathbf{L}$ and say that the signal of interest is $\mathbf{y} := \mathbf{x}^{(\infty)}$, then $\mathbf{y}$ solves the heat diffusion equation. However, more complex diffusion dynamics, such as $\mathbf{y} = \Pi_{l=0}^{\infty}(\mathbf{I}_N - \beta_l \mathbf{S})\mathbf{x}^{(0)}$ and $\mathbf{y} = \sum_{l=0}^{\infty} \gamma_l \mathbf{x}^{(l)} = \sum_{l=0}^{\infty} \gamma_l \mathbf{S}^l \mathbf{x}^{(0)}$, could also be of interest.

According to the previous discussion, it is apparent that the steady-state signal $\mathbf{y}$ generated by a diffusion process can be viewed as the output of a graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l$ with input $\mathbf{x}^{(0)}$. Note also that the Cayley-Hamilton theorem guarantees that the aforementioned infinite-horizon processes can be equivalently described by a filter of degree $N-1$.

Frequency domain representation: Leveraging the spectral decomposition of $\mathbf{S}$, graph filters and signals can be represented in the frequency domain. To be precise, let us use the eigenvectors of $\mathbf{S}$ to define the $N \times N$ matrix $\mathbf{U} := \mathbf{V}^{-1}$, and the eigenvalues of $\mathbf{S}$ to define the $N \times L$ Vandermonde matrix $\mathbf{\Psi}$, where $\Psi_{ij} := (\Lambda_{ii})^{j-1}$. Using these conventions, the frequency representations of a *signal* $\mathbf{x}$ and of a *filter* $\mathbf{h}$ are defined as $\widehat{\mathbf{x}} := \mathbf{U}\mathbf{x}$ and $\widehat{\mathbf{h}} := \mathbf{\Psi}\mathbf{h}$, respectively [6]. Exploiting such representations, the output $\mathbf{y} = \mathbf{H}\mathbf{x}$ of a graph filter in the frequency domain is given by

$$\widehat{\mathbf{y}} = \text{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x} = \text{diag}\big(\widehat{\mathbf{h}}\big)\widehat{\mathbf{x}} = \widehat{\mathbf{h}} \circ \widehat{\mathbf{x}}. \tag{3}$$

Identity (3) is the counterpart of the celebrated convolution theorem for temporal signals, and follows from $\mathbf{H} = \mathbf{V}\big(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\big)\mathbf{U}$ [cf. (1)] and $\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l = \text{diag}(\mathbf{\Psi}\mathbf{h})$; see e.g., [1] for a detailed derivation. To establish further connections with the time domain, let us consider the directed cycle graph whose adjacency matrix $\mathbf{A}_{dc}$ is zero, except for entries $A_{ij} = 1$ whenever $i = \text{mod}_N(j) + 1$, where $\text{mod}_N(x)$ denotes the modulus (remainder) obtained after dividing x by N. If $\mathbf{S} = \mathbf{A}_{dc}$, one can verify that: i) $\mathbf{y} = \mathbf{H}\mathbf{x}$ can be found as the circular convolution of $\mathbf{h}$ and $\mathbf{x}$, and ii) both $\mathbf{U}$ and $\mathbf{\Psi}$ correspond to the Discrete Fourier Transform (DFT) matrix. Interestingly, while in the time domain $\mathbf{U} = \mathbf{\Psi}$, this is not true for general graphs.

III. Blind Identification of Graph Filters

The concepts introduced in the previous section can be used to formally state the problem. For given shift operator $\mathbf{S}$ and filter degree $L-1$, suppose that we observe the output signal $\mathbf{y} = \mathbf{H}\mathbf{x}$ [cf. (1)], where $\mathbf{x}$ is *sparse* having at most $S \ll N$ non-zero entries. For future reference introduce the ℓ_0 (pseudo) norm $\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})|$, where the support of $\mathbf{x}$ is $\text{supp}(\mathbf{x}) := \{i \,|\, x_i \neq 0\}$ and hence $\|\mathbf{x}\|_0 \leq S$. The present paper deals with *blind identification* of the graph filter (and its input signal), which amounts to estimating the sparse $\mathbf{x}$ and the filter coefficients $\mathbf{h}$ from the observed output signal $\mathbf{y}$; see Fig. 1. This problem is a natural extension to graphs of classical blind system identification, or blind deconvolution of signals in the temporal or spatial domains.

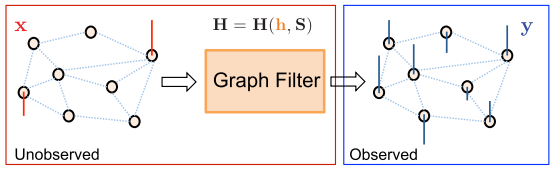

Fig. 1. Setup for the blind graph-filter identification problem. Given a graph signal $\mathbf{y}$ modeled as the output of a graph filter, the goal is to recover the vector of filter coefficients $\mathbf{h}$, and the input signal $\mathbf{x}$ that is assumed to be sparse.

Remark 1 (Sparse input): Sparsity in $\mathbf{x}$ is well-motivated due to its practical relevance and modeling value – network signals such as $\mathbf{y}$ are oftentimes the diffused version of few localized sources, hereby indexed by $\text{supp}(\mathbf{x})$. In addition, the non-sparse formulation with $S = N$ is ill-posed, since the number of unknowns $N + L$ in $\{\mathbf{x}, \mathbf{h}\}$ exceeds the number of observations N in $\mathbf{y}$. Alternatively, a low-dimensional subspace model for $\mathbf{x}$ could be also adopted to effectively reduce the degrees of freedom in the problem [1].

Given the observed filtered output $\mathbf{y}$, one can obtain its frequency-domain representation $\widehat{\mathbf{y}} = \mathbf{U}\mathbf{y} = \text{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x}$ [cf. (3)] and state the blind graph-filter identification problem as the following feasibility problem

$$\begin{aligned} &\text{find} \quad \{\mathbf{h}, \mathbf{x}\} \\ &\text{s. to} \quad \widehat{\mathbf{y}} = \text{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x}, \quad \|\mathbf{x}\|_0 \leq S. \end{aligned} \tag{4}$$

In other words, the goal is to find the solution to a set of bilinear equations subject to a sparsity constraint in $\mathbf{x}$.

A. Lifting the Bilinear Constraints

While very natural, (4) is in fact a difficult problem due to the non-convex ℓ_0-norm as well as the bilinear constraints. To deal with the latter, it is convenient to rewrite the first constraint in (4) as

$$\widehat{\mathbf{y}} = \big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \text{vec}\big(\mathbf{x}\mathbf{h}^T\big) \tag{5}$$

where $\odot$ denotes the Khatri-Rao (i.e., columnwise Kronecker) product, and $\text{vec}(\cdot)$ is the matrix vectorization operator. To establish (5), let $\mathbf{u}_i^T$ and $\boldsymbol{\psi}_i^T$ denote the i-th rows of $\mathbf{U}$ and $\mathbf{\Psi}$, respectively. It follows from (3) that $\widehat{y}_i = (\boldsymbol{\psi}_i^T \mathbf{h})(\mathbf{u}_i^T \mathbf{x}) = \big(\boldsymbol{\psi}_i^T \otimes \mathbf{u}_i^T\big)\text{vec}\big(\mathbf{x}\mathbf{h}^T\big)$, where $\otimes$ denotes the Kronecker product. Upon stacking the entries $\widehat{y}_i$ to form $\widehat{\mathbf{y}}$, the result follows by identifying $\boldsymbol{\psi}_i^T \otimes \mathbf{u}_i^T$ with the i-th row of $\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T$.

While (5) confirms that the filtered graph signal $\widehat{\mathbf{y}}$ is a *bilinear* function of $\mathbf{x}$ and $\mathbf{h}$, it also shows that $\widehat{\mathbf{y}}$ is a *linear* combination of the entries of the lifted rank-one, outer-product matrix $\mathbf{Z} := \mathbf{x}\mathbf{h}^T \in \mathbb{R}^{N \times L}$. In other words, there exists a linear mapping $\mathcal{M} : \mathbb{R}^{N \times L} \mapsto \mathbb{C}^N$ such that $\widehat{\mathbf{y}} = \mathcal{M}(\mathbf{Z})$. Note that $\mathcal{M}$ can be expressed in terms of a matrix multiplication with $\mathbf{M} := (\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \in \mathbb{C}^{N \times LN}$, since $\widehat{\mathbf{y}} = \mathbf{M}\text{vec}(\mathbf{Z})$ as per (5). In addition to being of rank one, note that the sparsity in $\mathbf{x}$ renders $\mathbf{Z}$ row-wise sparse, i.e., rows $\mathbf{z}_i^T$ indexed by $\{1, \ldots, N\} \setminus \text{supp}(\mathbf{x})$ are identically zero. Building on the ideas in [10], [13], one can thus pose the blind graph-filter identification problem as a *linear*

inverse problem, where the goal is to recover a row-sparse, rank-one $N \times L$ matrix $\mathbf{Z}$ from observations $\widehat{\mathbf{y}} = \mathcal{M}(\mathbf{Z})$. To this end, a natural formulation to tackle such inverse problem is

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \operatorname{rank}(\mathbf{Z}) \\ \text{s. to} \quad & \widehat{\mathbf{y}} = \left(\boldsymbol{\Psi}^T \odot \mathbf{U}^T\right)^T \operatorname{vec}(\mathbf{Z}), \quad \|\mathbf{Z}\|_{2,0} \leq S \end{aligned} \tag{6}$$

where $\|\mathbf{Z}\|_{2,0}$ is equal to the number of non-zero rows of $\mathbf{Z}$.

A basic question is whether (6) is equivalent to the original blind identification problem. To give a rigorous answer, some definitions are introduced next. For a given matrix $\mathbf{U}$, $\operatorname{spark}(\mathbf{U})$ is the smallest number n such that there exists a subgroup of n columns from $\mathbf{U}$ that are linearly dependent [25]. Given a set of row indices $\mathcal{I}$, define the complement set of indices $\mathcal{I}^c := \{1, \ldots, N\} \backslash \mathcal{I}$ and the matrix $\mathbf{U}_{\mathcal{I}}$ formed by the rows of $\mathbf{U}$ indexed by $\mathcal{I}$. Moreover, for a given graph-shift operator $\mathbf{S}$ – fixed $\mathbf{V}$, $\boldsymbol{\Psi}$, and $\mathbf{U}$ – define the set $\mathcal{O}_{\widehat{\mathbf{y}}}$ of matrix minimizers of (6) as a function of $\widehat{\mathbf{y}}$. Then, the following result on the validity of the matrix problem formulation in (6) holds.

Proposition 1: Let $\mathcal{I}_S$ be a set of row indices such that $\operatorname{spark}(\mathbf{U}_{\mathcal{I}_S}) \leq S$. Then, the set of minimizers of (6), satisfies

$$\mathcal{O}_{\widehat{\mathbf{y}}} = \left\{ \mathbf{x}\mathbf{h}^T \;\middle|\; \widehat{\mathbf{y}} = \mathbf{U} \sum_{l=0}^{L-1} h_l \mathbf{S}^l \mathbf{x}, \;\; \|\mathbf{x}\|_0 \leq S \right\} \tag{7}$$

for any $\widehat{\mathbf{y}}$ if and only if

$$\min_{\mathcal{I}_S} \left| \{\lambda_i\}_{i \in \mathcal{I}_S^c} \right| > L - 1. \tag{8}$$

Proof: If we show that (8) is violated if and only if there exists a rank-one matrix $\mathbf{Z} = \mathbf{x}\mathbf{h}^T$ such that $\left(\boldsymbol{\Psi}^T \odot \mathbf{U}^T\right)^T \operatorname{vec}(\mathbf{Z}) = \mathbf{0}_N$ and $\|\mathbf{Z}\|_{2,0} \leq S$, then Corollary 1 in [26] completes the proof. The above system of homogeneous equations can be written as $(\boldsymbol{\psi}_i^T \mathbf{h})(\mathbf{u}_i^T \mathbf{x}) = 0$ for $i = 1, \ldots, N$, where $\boldsymbol{\psi}_i^T$ denotes the i-th row of $\boldsymbol{\Psi}$ and similarly for $\mathbf{U}$. Since $\operatorname{spark}(\mathbf{U}_{\mathcal{I}_S}) \leq S$, there exists $\mathbf{x} \neq \mathbf{0}_N$ with $\|\mathbf{x}\|_0 \leq S$ such that $(\boldsymbol{\psi}_i^T \mathbf{h})(\mathbf{u}_i^T \mathbf{x}) = 0$ holds for $i \in \mathcal{I}_S$. Exploiting the Vandermonde structure of $\boldsymbol{\Psi}$, it follows that $\mathbf{h} \neq \mathbf{0}_L$ satisfying the equality for $i \in \mathcal{I}_S^c$ can be found if and only if (8) is violated. ■

Ideally, when solving (6) for some output $\widehat{\mathbf{y}}$ one should recover the set of outer products of all possible combinations of sparse inputs $\mathbf{x}$ and filter coefficients $\mathbf{h}$ that can give rise to such output [cf. (7)]. This is not true in general [26, Theorem 1], however, Proposition 1 states conditions on the graph-shift operator [cf. (8)] for the desired equivalence to hold. For the particular case of the directed cycle graph, we may select the support of $\mathbf{x}$ so that every choice of S rows of $\mathbf{U}$ forms a full-rank matrix. Consequently, the cardinality in (8) is equal to $N - S + 1$ entailing the following corollary.

Corollary 1: If $\mathbf{S} = \mathbf{A}_{dc}$ and the support of $\mathbf{x}$ consists of either S adjacent or S equally spaced nodes, then (7) holds if and only if $N > L + S - 2$.

Notice that condition (8) does not guarantee that the solution of (6) is unique, but rather that the outer product of the desired sparse signal and filter coefficients is contained in $\mathcal{O}_{\widehat{\mathbf{y}}}$. For instance, uniqueness necessarily requires $\operatorname{rank}(\boldsymbol{\Psi}) = L$, otherwise there is no hope to recover the actual $\mathbf{h}$ from observations $\widehat{\mathbf{y}}$ [cf. (4)].

B. Algorithmic Approach via Convex Relaxation

Albeit natural, problem (6) is challenging since both the rank and the ℓ_0-norm are in general NP-hard to optimize; see e.g., [27]. Over the last decade or so, convex relaxation approaches to tackle rank and/or sparsity minimization problems have enjoyed remarkable success, since they oftentimes entail no loss in optimality. The nuclear norm $\|\mathbf{Z}\|_* = \sum_k \sigma_k(\mathbf{Z})$, where $\sigma_k(\mathbf{Z})$ denotes the k-th singular value of $\mathbf{Z}$, is typically adopted as a convex surrogate to $\operatorname{rank}(\mathbf{Z})$ [14], [27]. Likewise, the $\ell_{2,1}$ mixed norm $\|\mathbf{Z}\|_{2,1} := \sum_{i=1}^N \|\mathbf{z}_i^T\|_2$ is the closest convex approximation of $\|\mathbf{Z}\|_{2,0}$ [28]. With τ denoting a tuning parameter to control the rank versus row-sparsity tradeoff, a convex heuristic is to solve

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_* + \tau \|\mathbf{Z}\|_{2,1} \\ \text{s. to} \quad & \widehat{\mathbf{y}} = \left(\boldsymbol{\Psi}^T \odot \mathbf{U}^T\right)^T \operatorname{vec}(\mathbf{Z}) \end{aligned} \tag{9}$$

hoping that the optimal solution is of rank one and has S non-zero rows, so that we can recover $\mathbf{x}$ and $\mathbf{h}$ up to scaling.

Recovery of simultaneously low-rank and row-sparse matrices from noisy compressive measurements was also considered in [29] for hyperspectral image reconstruction. Recent theoretical results on recovery of simultaneously structured matrix models suggest that minimizing only $\|\mathbf{Z}\|_1$ could as well suffice [15]; see also [13], the discussion at the end of this section and the performance guarantees in Section IV. Being convex, (9) is computationally appealing, in fact off-the-shelf interior point solvers are available. Customized scalable algorithms for large-scale graphs can be developed to minimize the composite, non-differentiable cost in (9). For instance the solver implemented to run the numerical tests in Section V leverages the alternating-direction method of multipliers (ADMM) [30]; see also [29] for a related proximal-splitting algorithm.

Refinement via iteratively-reweighted optimization: Instead of substituting $\|\mathbf{Z}\|_{2,0}$ in (6) by its closest convex approximation, namely $\|\mathbf{Z}\|_{2,1}$, letting the surrogate function to be non-convex can yield tighter approximations, and potentially improve the statistical properties of the estimator. In the context of sparse signal recovery for instance, the ℓ_0 norm of a vector was surrogated in [31] by the logarithm of the geometric mean of its elements.

Building on this last idea, consider replacing $\|\mathbf{Z}\|_{2,1}$ in (9) with $\sum_{i=1}^N \log(\|\mathbf{z}_i^T\|_2 + \delta)$, where δ is a small positive constant. Since the new surrogate term is concave, the overall minimization problem is non-convex and admittedly more complex to solve than (9). With k denoting iterations, local methods based on iterative linearization of $\log(\|\mathbf{z}_i^T\|_2 + \delta)$ around the current iterate $\mathbf{z}_i^T(k)$, can be adopted to minimize the resulting non-convex cost. Skipping details that can be found in [31], application of the majorization-minimization technique leads to an *iteratively-reweighted* version of (9), namely solve for $k = 0, 1, \ldots$

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_* + \sum_{i=1}^N w_i(k) \|\mathbf{z}_i^T\|_2 \\ \text{s. to} \quad & \widehat{\mathbf{y}} = \left(\boldsymbol{\Psi}^T \odot \mathbf{U}^T\right)^T \operatorname{vec}(\mathbf{Z}) \end{aligned} \tag{10}$$

with weights $w_i(k) := \tau / \left(\|\mathbf{z}_i^T(k-1)\|_2 + \delta\right)$. If the value of $\|\mathbf{z}_i^T(k-1)\|_2$ is small, then in the next iteration the

regularization term $w_i(k)\|\mathbf{z}_i^T\|_2$ has a large weight, promoting shrinkage of that entire row vector to zero. Numerical tests in Section V suggest that few iterations of the iteratively-reweighted procedure suffice to yield improved recovery of $\{\mathbf{x},\mathbf{h}\}$, when compared to (9).

Blind identification via linear programming: Since schemes aimed at finding sparse matrices can also lead to low-rank solutions [13], [15], the last proposed relaxation adopts a single-structure enforcing criterion under which, as in SparseLift [13], only the ℓ_1-norm of $\mathbf{Z}$ is minimized

$$\min_{\mathbf{Z}} \quad \|\mathbf{Z}\|_1$$
$$\text{s. to} \quad \widehat{\mathbf{y}} = \left(\mathbf{\Psi}^T \odot \mathbf{U}^T\right)^T \text{vec}\left(\mathbf{Z}\right). \tag{11}$$

The above optimization is a linear program that, if needed, can be modified to accommodate an iteratively-reweighted counterpart. While we know $\mathbf{Z}$ is low-rank and row-wise sparse, we admittedly relaxed the structural constraints and only encouraged $\mathbf{Z}$ to be entry-wise sparse, with no specific pattern preferred a priori. The reason for considering the simpler formulation in (11) is threefold. First, the absence of $\|\mathbf{Z}\|_*$ circumvents the need to perform a singular value decomposition (SVD) per iteration, as is customary with nuclear-norm minimization. Second, (11) eliminates the burden of selecting an adequate tuning parameter τ in (9). Third, as we establish in Section IV, under some conditions the simplification in (11) is enough to uniquely recover $\mathbf{Z}=\mathbf{x}\mathbf{h}^T$ with high probability.

Before closing this section, a couple remarks are in order.

Remark 2 (Noisy and partial observations): The proposed relaxations can be easily modified to account for noisy or partial observations of the graph signal $\mathbf{y}$. Following the standard approach for sparse recovery problems, when the observations $\mathbf{y}$ are noisy, it suffices to rewrite the filter output constraint as $\|\mathbf{V}\widehat{\mathbf{y}} - \mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T\text{vec}(\mathbf{Z})\|_2^2 \leq \epsilon^2$, where the specific norm and value of ϵ^2 will depend on the observation noise model. Moreover, it is not uncommon to encounter graph-based settings where one measures $\mathbf{y}$ in a subset of nodes only. This could happen because it is impossible to access parts of the network, or due to intentional sampling with the goal of reducing overall processing complexity. Accordingly, suppose that $C \leq N$ and define the partially observed signal $\mathbf{y}_c \in \mathbb{R}^C$ as $\mathbf{y}_c := \mathbf{C}\mathbf{y} = \mathbf{C}\mathbf{V}\widehat{\mathbf{y}}$, with $\mathbf{C}$ being a sampling matrix formed by a subset of rows of the $N \times N$ identity matrix. The graph filter output constraint should be now written as $\mathbf{y}_c = \mathbf{C}\mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T\text{vec}(\mathbf{Z})$, so that the matrix $\mathbf{C}\mathbf{V}$ is incorporated into the linear mapping $\mathcal{M}(\mathbf{Z})$. Because the input signal $\mathbf{x}$ is assumed to be sparse, it may still be feasible to recover $\{\mathbf{h},\mathbf{x}\}$ from partial observations $\mathbf{y}_c$; see also the numerical tests in Section V.

Remark 3 (A priori filter knowledge): In general, if the specific graph filter form is known a priori, this additional information can be incorporated in the problem formulation and thus aid the recovery. For instance, if the frequency response of the unknown filter follows a specific linear parametric model we may write $\widehat{\mathbf{h}} = \mathbf{B}\boldsymbol{\alpha}$, for parameters $\boldsymbol{\alpha} \in \mathbb{R}^{L'}$, $L' < L$ and some basis matrix $\mathbf{B}$. The signal model (5) can be rewritten as $\widehat{\mathbf{y}} = \left(\mathbf{B}^T \odot \mathbf{U}^T\right)^T \text{vec}\left(\mathbf{x}\boldsymbol{\alpha}^T\right)$ and the corresponding problem formulations in (6), (9), (10), and (11) can be modified accordingly. This parametric description of $\widehat{\mathbf{h}}$ effectively reduces the size of the problem, which in general entails a better recovery performance; see Sections IV and V. Other models of prior knowledge on $\mathbf{h}$ or $\widehat{\mathbf{h}}$ could be relevant as well. For example, one might know that the observed output is a linear combination of L' of the first L shifted versions of the input. In this case, a sparse prior can be incorporated on the vector of filter coefficients $\mathbf{h}$ by augmenting the objective function in (9) with $\tau_2\|\mathbf{Z}^T\|_{2,1}$. In this way, we are simultaneously enforcing row and column sparsity on $\mathbf{Z}$ to account for the sparse priors on $\mathbf{x}$ and $\mathbf{h}$, respectively.

C. Multiple Output Signals

Jointly processing multiple output signals (when available) can aid the blind identification task, and this is the subject of the present section. Suppose now that we have access to a collection of P (possibly time-indexed) output signals $\{\mathbf{y}_p\}_{p=1}^P$, each one corresponding to a different sparse input $\mathbf{x}_p$ fed to the *common* graph filter $\mathbf{H}$ we wish to identify. Although each of the P identification problems could be solved separately (and naively) as per Section III, the recovery performance can be improved by tackling them jointly.

While extending the feasibility problem in (4) to this new setup is straightforward [each output gives rise to a couple constraints as in (4)], generalizing the formulation in (9) requires more work. To this end, consider the $NP \times 1$ supervector of stacked output signals $\tilde{\mathbf{y}} := [\widehat{\mathbf{y}}_1^T, ..., \widehat{\mathbf{y}}_P^T]^T$, and likewise for the unobserved inputs $\tilde{\mathbf{x}} := [\mathbf{x}_1^T, ..., \mathbf{x}_P^T]^T$. Next, introduce the unknown rank-one matrices $\mathbf{Z}_p := \mathbf{x}_p\mathbf{h}^T$, $p = 1, ..., P$, and stack them: (i) vertically in $\tilde{\mathbf{Z}}_v := [\mathbf{Z}_1^T, ..., \mathbf{Z}_P^T]^T = \tilde{\mathbf{x}}\mathbf{h}^T \in \mathbb{R}^{NP\times L}$; and (ii) horizontally in $\tilde{\mathbf{Z}}_h := [\mathbf{Z}_1, ..., \mathbf{Z}_P] \in \mathbb{R}^{N\times PL}$. Note that $\tilde{\mathbf{Z}}_v$ is a rank-one matrix. Further, when all $\mathbf{x}_p$ share a common support, then so will all the row-sparse matrices $\mathbf{Z}_p$ (and hence $\tilde{\mathbf{Z}}_h$). These observations motivate the following convex formulation [cf. (9)]

$$\min_{\{\mathbf{Z}_p\}_{p=1}^P} \quad \|\tilde{\mathbf{Z}}_v\|_* + \tau\|\tilde{\mathbf{Z}}_h\|_{2,1} \tag{12}$$
$$\text{s. to} \quad \tilde{\mathbf{y}} = \left(\mathbf{I}_P \otimes \left(\left(\mathbf{\Psi}^T \odot \mathbf{U}^T\right)^T\right)\right)\text{vec}\left(\tilde{\mathbf{Z}}_h\right)$$

where all P lifted bilinear constraints have been compactly expressed in terms of $\tilde{\mathbf{y}}$ and $\text{vec}\left(\tilde{\mathbf{Z}}_h\right)$ using a Kronecker product.

When the sparse support is not the same for all $\mathbf{x}_p$, matrix $\tilde{\mathbf{Z}}_h$ is not row-sparse. In that case, $\|\tilde{\mathbf{Z}}_h\|_{2,1}$ in (12) must be replaced with $\sum_{p=1}^P \|\mathbf{Z}_p\|_{2,1}$, possibly adjusting individual tuning parameters τ_p per signal. Either way, an efficient ADMM solver can be implemented for the multiple signal setting as well, and extensive numerical tests indicated that iteratively-reweighing as in Section III-B can yield markedly improved recovery performance (cf. Section V).

IV. Exact Recovery via Convex Optimization

Here we show that under some technical conditions, both the filter coefficients $\mathbf{h}$ as well as the sparse input signal $\mathbf{x}$ in (5) can be exactly recovered by solving the convex problem (11). Assumptions delineating the analysis' scope are first outlined in Section IV-A, after which the main result is formally stated in

Section IV-B followed by a discussion of the recovery conditions and their dependence on the graph. The proof follows closely the ideas in [13] and the key steps are given in Appendix A, with an emphasis on the novel aspects introduced by the GSP context dealt with here.

A. Assumptions and Scope of the Analysis

Two main assumptions are made to facilitate the analysis.

(as1): The graph-shift operator $\mathbf{S}$ is normal, i.e., it satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$, and its eigenvalues are all distinct.
(as2): The frequency representation of the observed graph signal $\mathbf{y}$ adheres to the model $\widehat{\mathbf{y}} = \text{diag}(\mathbf{\Psi h})\tilde{\mathbf{U}}\mathbf{x}$, where $\tilde{\mathbf{U}}$ is a random $N \times N$ matrix obtained by concatenating N rows sampled independently and uniformly with replacement from $\mathbf{U}$.

Under (as1) $\mathbf{V}$ is unitary, which implies $\mathbf{U} := \mathbf{V}^{-1} = \mathbf{V}^H$ and leads to the decomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$. Normality is for instance satisfied when $\mathcal{G}$ is undirected and the graph-shift operator is chosen to be the adjacency matrix or the graph Laplacian. Furthermore, all the eigenvalues of $\mathbf{S}$ being distinct ensures that matrix $\mathbf{\Psi}$ is full rank independently of L, which is required for uniqueness as discussed in the end of Section III-A. Under (as1), we can assume that $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_L$ without loss of generality. To see this, consider, e.g., the SVD of $\mathbf{\Psi} = \mathbf{P}\mathbf{\Sigma}\mathbf{R}^H$ and rewrite the frequency response of the filter as $\mathbf{\Psi h} = \mathbf{P}\mathbf{\Sigma}\mathbf{R}^H\mathbf{h} := \mathbf{P h}'$, where $\mathbf{P}$ satisfies $\mathbf{P}^H\mathbf{P} = \mathbf{I}_L$ by definition and $\mathbf{h}$ can be recovered from $\mathbf{h}'$ due to the full-rank condition of $\mathbf{\Psi}$. Consequently, in the statement of Theorem 1 and its proof we assume that $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_L$.

Regarding the probabilistic model for the observations in (as2), this type of models are customary towards establishing recovery guarantees in the context of, e.g., compressed sensing and low-rank matrix completion [32], or even blind deconvolution of temporal signals [10], [13]. For instance, instrumental to the proof arguments in [13], is that rows of the matrix representation of operator $\mathcal{M}$ are independent. This way, one can bring to bear matrix Bernstein inequalities to bound the norm of relevant operators constructed from sums of these rows [33]. A direct consequence is that the recovery results obtained here are probabilistic in nature, namely Theorem 1 asserts that $\{\mathbf{x}, \mathbf{h}\}$ can be recovered with high probability over the measure induced by the aforementioned matrix randomization procedure. All in all, tractability is the main reason behind (as2), which resembles the random Fourier model in [13] but is more general since rows are sampled from a unitary matrix $\mathbf{U}$ – not necessarily the DFT matrix.

In any case, we would like to stress that the focus here is only on establishing that a convex relaxation can succeed for blind identification of graph filters, and that the graph structure plays a key role on the recovery performance. We are not after the tightest guarantees, and the success probability bounds obtained are admittedly loose. In fact, Theorem 1 deals with recovery of $\{\mathbf{x}, \mathbf{h}\}$ using the simplified convex formulation (11), despite the fact that (9) exhibits slightly better performance than (11) in practice; see also the numerical tests in Section V. Nevertheless, in theory both schemes are equivalent (at least order-wise) [15], while the optimality conditions and corresponding construction of dual certificates for (11) are markedly simpler.

B. Main Result

Given an arbitrary matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ and a positive integer $k \in \mathbb{N}$, we define the function $\rho_{\mathbf{A}}(k)$ as

$$\rho_{\mathbf{A}}(k) := \max_{l\in\{1,\ldots,M\}} \max_{\Omega\in\Omega_k^N} \|\mathbf{a}_{l,\Omega}\|_2^2 \tag{13}$$

where Ω_k^N represents the set of all k-subsets of $\{1,\ldots,N\}$, and $\mathbf{a}_{l,\Omega}$ is the orthogonal projection of the l-th row of $\mathbf{A}$ onto the index set Ω. In words, $\rho_{\mathbf{A}}(k)$ is the largest squared-norm of any vector formed by selecting k elements from a row of $\mathbf{A}$. This extends the concept of mutual coherence between the basis of Kronecker deltas on the graph and $\mathbf{A}$ [34]; see also the related cumulative coherence notion in [35]. The above definition allows us to formalize the following main result.

Theorem 1: For a given graph-shift operator $\mathbf{S}$, assume that an S-sparse graph signal $\mathbf{x}_0 \in \mathbb{R}^N$ when passed through a filter with coefficients $\mathbf{h}_0 \in \mathbb{R}^L$ results in a signal with frequency representation $\widehat{\mathbf{y}} \in \mathbb{C}^N$ adhering to the model in (as2). Also, denote by $\mathbf{U} \in \mathbb{C}^{N\times N}$ and $\mathbf{\Psi} \in \mathbb{C}^{N\times L}$ the GFT for signals and filters associated with $\mathbf{S}$, respectively, where $\mathbf{U}$ is normalized such that $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$. Define

$$\alpha := \frac{3\log(2)\left(120\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)} + 8\sqrt{\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)}}\right)^{-1}}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(4\gamma\sqrt{2LS})\log(2SN^2)} \tag{14}$$

where $\gamma := \sqrt{2N(\log(2LN)+1)+1}$. Under (as1)-(as2), if $\alpha \geq 1$ then the unique solution to (11) is the rank-one matrix $\mathbf{Z}_0 := \mathbf{x}_0\mathbf{h}_0^T$, with probability at least

$$P_{\text{rec}} \geq 1 - N^{-\alpha+1}. \tag{15}$$

Proof: See Appendix A. ■

We want to emphasize three differences between the above theorem and [13, Th. 3.1]. First and foremost, Theorem 1 provides probabilistic guarantees of recovery for blind identification in *arbitrary* graphs [cf. (as1)], whereas the results in [13] only apply for the cases where $\tilde{\mathbf{U}}$ is random Fourier or Gaussian distributed. Our generalization is reflected through the function $\rho_{\mathbf{A}}$, which as detailed after Lemma 1 also provides intuition about which graph topologies favor blind recovery. Secondly, we provide exact expressions for the constants throughout the proof – some of them embedded in expression (14). Accordingly, bounds for the probability of recovery are derived for finite values of L, S, and N, instead of order-wise asymptotic results. Lastly, by using $\rho_{\mathbf{\Psi}}$ to describe properties of $\mathbf{\Psi}$, the bounds obtained in Theorem 1 are tighter than those in [13, Th. 3.1], even for the case where $\tilde{\mathbf{U}}$ is random Fourier.

Leveraging the facts that both $\rho_{\mathbf{U}}$ and $\rho_{\mathbf{\Psi}}$ are non-decreasing functions by definition [cf. (13)], we can lower bound α in (14) to obtain an alternative expression which is more restrictive but simpler to understand, namely

$$\alpha_1 := \frac{3\log(2)/128}{LS\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(4\gamma\sqrt{2LS})\log(2SN^2)} \leq \alpha. \tag{16}$$

Inspection of (16) clearly shows that the recovery performance depends on S and L through the functions $\rho_{\mathbf{U}}$ and $\rho_{\mathbf{\Psi}}$, respectively. First notice that from $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$ and $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_L$ it follows that $\rho_{\mathbf{U}}(S) \leq N$ for all S and $\rho_{\mathbf{\Psi}}(L) \leq 1$ for all L, respectively. Moreover, the following lemma further characterizes the behavior of these functions.

Lemma 1: The functions $\rho_{\mathbf{U}}$ and $\rho_{\mathbf{\Psi}}$ as defined in (13) satisfy

$$S \leq \rho_{\mathbf{U}}(S) \leq S\rho_{\mathbf{U}}(1), \tag{17}$$

$$L/N \leq \rho_{\mathbf{\Psi}}(L) \leq L\rho_{\mathbf{\Psi}}(1) \tag{18}$$

for all graph shifts $\mathbf{S}$. Moreover, (17) and (18) are satisfied with equalities when $\mathbf{S} = \mathbf{A}_{dc}$.

Proof: Since (17) and (18) can be shown using similar arguments, we focus only on proving (17). To show the rightmost inequality in (17), we leverage the definition of $\rho_{\mathbf{A}}$ in (13) to write

$$\begin{aligned}\rho_{\mathbf{U}}(S) &\leq \max_{l\in\{1,\ldots,N\}} \max_{\Omega\in\Omega_S^N} S \max_{i\in\Omega} |u_{li}|^2 \\ &= S \max_{l\in\{1,\ldots,N\}} \max_{i\in\{1,\ldots,N\}} |u_{li}|^2 = S\rho_{\mathbf{U}}(1) \end{aligned} \tag{19}$$

where the first equality follows from the fact that maximizing over all S-subsets first and then maximizing over a particular entry $i \in \Omega$ is equivalent to an initial maximization over $i \in \{1,\ldots,N\}$. To show the leftmost inequality in (17), we again rely on (13) to write

$$\begin{aligned}\frac{N}{S}\rho_{\mathbf{U}}(S) &= \max_{l\in\{1,\ldots,N\}} \frac{N}{S} \max_{\Omega\in\Omega_S^N} \|\mathbf{u}_{l,\Omega}\|_2^2 \\ &\geq \max_{l\in\{1,\ldots,N\}} \|\mathbf{u}_l\|_2^2 = N \end{aligned} \tag{20}$$

where the last equality follows from the fact that $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$. Finally, whenever $\mathbf{S} = \mathbf{A}_{dc}$ notice that $\mathbf{U}$ is a DFT matrix with unit-magnitude elements and $\mathbf{\Psi}$ consists of columns from a normalized DFT matrix (with elements of magnitude $1/\sqrt{N}$), thus equalities in (17) and (18) follow. ■

In order to increase the probability of recovery, it is desirable to obtain large values of α [cf. (14) and (15)]. Consequently, functions $\rho_{\mathbf{U}}$ and $\rho_{\mathbf{\Psi}}$ indicate how the recovery performance decreases with increasing S – number of non-zero entries of the input – and L – number of filter coefficients. In particular, the closer $\rho_{\mathbf{U}}$ and $\rho_{\mathbf{\Psi}}$ are to their lower bounds in (17) and (18), the better – more specifically, the slower the recovery performance deteriorates with increasing S and L. With reference to the theoretical bound in (15), Lemma 1 implies that blind identification in time and, more generally, in circulant graphs corresponds to the most favorable setting. This follows since circulant graphs are diagonalized by the DFT matrix and, thus, the lower bound for $\rho_{\mathbf{U}}$ is achieved [cf. (17)]. The behavior of $\rho_{\mathbf{U}}$ for different graphs is depicted in Fig. 2.

Regarding the behavior of $\rho_{\mathbf{\Psi}}$, since $\mathbf{\Psi}$ depends not only on the graph but also on the normalization procedure chosen to achieve $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_L$ [cf. discussion after (as1)], the interpretation of $\rho_{\mathbf{\Psi}}$ and its dependence on the graph structure is more involved. To be more precise, when the matrix $\mathbf{P}$ of left singular vectors plays the role of $\mathbf{\Psi}$, both the Vandermonde structure of $\mathbf{\Psi}$ as well as the one-to-one correspondence between rows and eigenvalues of the graph are lost. Hence, in the ensuing section we limit our numerical analysis to the effect of $\rho_{\mathbf{U}}$ on the recovery performance.

V. Numerical Results

Four types of experiments are conducted to illustrate the performance of our blind graph-filter identification approach.

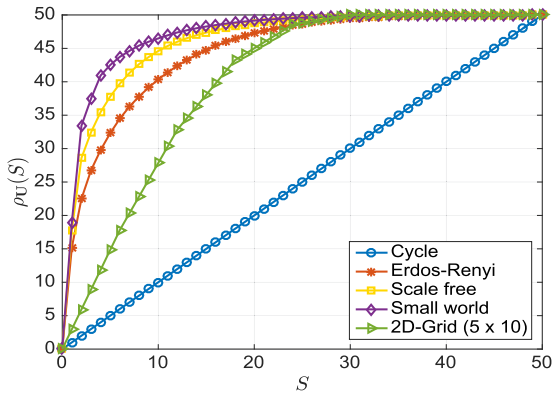

Fig. 2. $\rho_{\mathbf{U}}(S)$ for five types of graphs with $N = 50$ nodes. The graph parameters were chosen so that the expected number of edges is the same for the three random topologies considered. The values of $\rho_{\mathbf{U}}$ reported are the average among 100 realizations. We can see $\rho_{\mathbf{U}}$ achieving its lower bound for the directed cycle (cf. Lemma 1).

First, we evaluate the effectiveness of the different relaxations proposed in Section III in random and real-world graphs. Second, we compare the performance of our method with alternative approaches when solving a blind identification problem in a brain graph. Third, we assess the sensitivity of recovery with respect to the graph-dependent parameters identified in Section IV. Lastly, we illustrate how our method can be used to identify the sources of contagion in an epidemic model.

In the aforementioned experiments, we solve blind graph-filter identification problems for different graphs $\mathcal{G}$ while varying the parameters L, S, P, and N. The obtained signal and filter-coefficient estimates will be denoted by $\{\tilde{\mathbf{x}}, \tilde{\mathbf{h}}\}$. Unless otherwise stated, we select the graph-shift operator $\mathbf{S} = \mathbf{A}$ equal to the adjacency matrix of $\mathcal{G}$. The "true" vectors $\mathbf{x}_0$ and $\mathbf{h}_0$ are drawn from standard multivariate Gaussian distributions and are normalized to unit norm. In this way, the "true" filter models a generalized network diffusion process as introduced in Section II. Given $\mathbf{x}_0$ and $\mathbf{h}_0$, synthetic observations $\mathbf{y}$ are generated with frequency components given by (5). The root-mean-square error RMSE $:= \|\tilde{\mathbf{x}}\tilde{\mathbf{h}}^T - \mathbf{x}_0\mathbf{h}_0^T\|_F$ is adopted as a figure of merit to assess recovery performance.

Recovery performance: Defining successful recovery when the RMSE is smaller than 0.01, we empirically estimate the successful recovery rate for Erdős-Rényi graphs ($N = 50$ nodes, edge formation probability $p = 0.1$) [36] as a function of L and S by averaging the success counts over 20 realizations for each parameter combination; see Fig. 3(a)–(d)(top). In each experiment we assume the true filter length L – but not the sparsity level S – to be known. We assess the recovery performance for different convex relaxations of increasing effectiveness: (a) ℓ_1 minimization [cf. (11)]; (b) $\ell_{2,1}$ plus nuclear norm [cf. (9)]; (c) reweighted $\ell_{2,1}$ plus nuclear norm [cf. (10)]; and (d) reweighted $\ell_{2,1}$ plus nuclear norm with $P = 5$ output observations [cf. (12)]. As expected, the difficulty of the problem increases when either L or S increase, depicted in the figures by the darker area around the bottom-right corners. Moreover, when going from one figure to the next, the growing white regions portray the benefits of leveraging additional signal structure in the algorithms. In particular, when moving from (a) to (b), we observe the benefit of incorporating the row-sparse and low-rank features of $\mathbf{Z}$ into

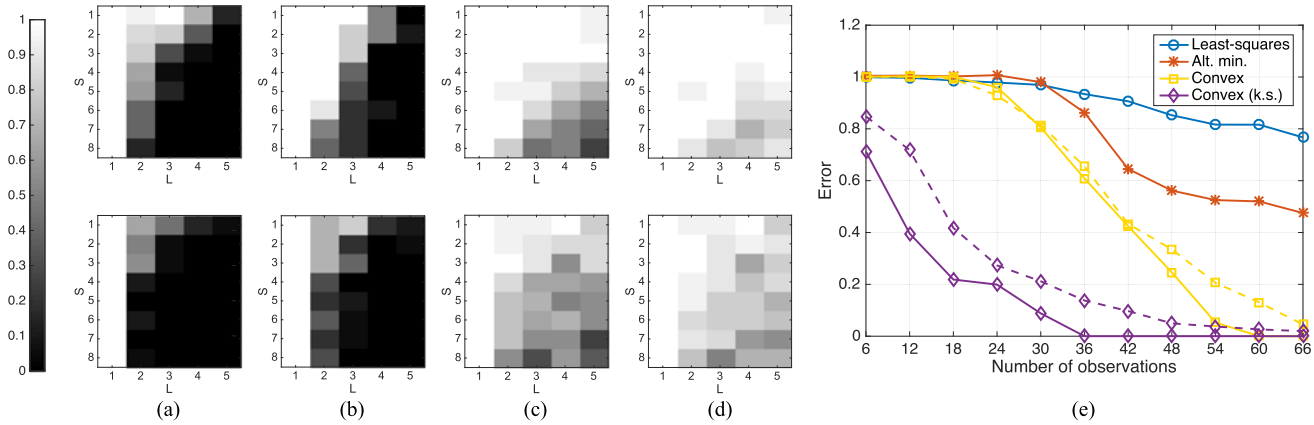

Fig. 3. (a)–(d) Rate of recovery of $\mathbf{x}_0\mathbf{h}_0^T$ as a function of S (sparsity in $\mathbf{x}_0$) and L (filter length) in 50-node Erdős-Rényi graphs (top) and a brain graph (bottom) for different recovery algorithms: (a) ℓ_1 minimization [cf. (11)]; (b) $\ell_{2,1}$ plus nuclear norm minimization [cf. (9)]; (c) reweighted $\ell_{2,1}$ plus nuclear norm [cf. (10)]; and (d) reweighted $\ell_{2,1}$ plus nuclear norm with $P = 5$ output observations [cf. (12)]. (e) Recovery errors for several methods as a function of the number of observations in the brain network for $L = S = 3$. Dashed lines represent recovery with noisy observations.

the model as opposed to merely considering a sparse model of $\mathbf{Z}$. Notice, however, that the performance of these two approaches is comparable [15]. When going from (b) to (c), we see the conspicuous performance improvement entailed by considering the iteratively-reweighted scheme to promote row-sparsity in $\mathbf{Z}$. Lastly, when comparing (c) to (d), we gauge the benefits of observing multiple ($P = 5$) output signals, especially for large values of S and L. In particular, when $S = 8$ and $L = 5$ we go from a success rate of 0.25 in (c) to a success rate of 0.90 in (d). For this latter setting, exact recovery is achieved consistently for most combinations of L and S.

We now consider a weighted undirected graph of the human brain, consisting of $N = 66$ nodes or regions of interest (ROIs) and whose edge weights are given by the density of anatomical connections between regions [37]. The level of activity of each ROI can be represented by a graph signal $\mathbf{x}$, thus successive applications of $\mathbf{S}$ model a linear evolution of the brain activity pattern. Supposing we observe a linear combination (filter) of the evolving states of an originally sparse brain signal, then blind identification amounts to jointly estimating which regions were originally active, the activity in these regions, and the coefficients of the linear combination. We mimic the recovery rate analysis performed for Erdős-Rényi graphs; see Fig. 3(a)–(d)(bottom). As expected, the success rates increase gradually when going from (a) to (d), as we consider more sophisticated algorithms. Furthermore, when comparing the bottom plots in Figs. 3(a)–(d) with their top counterparts, it is immediate that, for fixed L and S, recovery in the brain network is more challenging than in Erdős-Rényi graphs. This can be explained by the marked structure of the brain network where nodes are divided into two weakly connected hemispheres. Hence, the output signals in one hemisphere are not very informative about the input signals in the opposite hemisphere, rendering recovery more difficult.

In order to assess the recovery related to different graph-shift operators associated with the same underlying graph, we repeat the experiment depicted in Fig. 3(d) for Erdős-Rényi graphs ($N = 30$ nodes, edge formation probability $p = 0.15$) and for three different shifts: the adjacency $\mathbf{A}$, the Laplacian $\mathbf{L}$, and the normalized Laplacian $\mathbf{L}_{\text{n}}$; see Fig. 4(a). From the figure it is clear that the recovery rates associated with the three analyzed

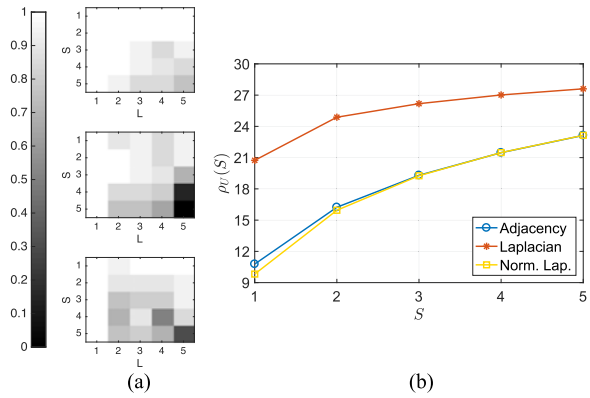

Fig. 4. (a) Experiment analogous to the one depicted in Fig. 3(d) for three different shifts based on the same graph: the adjacency $\mathbf{A}$ (top), the Laplacian $\mathbf{L}$ (middle), and the normalized Laplacian $\mathbf{L}_{\text{n}}$ (bottom) matrices. (b) The mean values of $\rho_{\text{U}}(S)$ for $S = 1, \ldots, 5$ for the three graph-shift operators considered.

shifts are comparable, with slightly lower performance corresponding to the Laplacian $\mathbf{L}$. This can be partially explained by the values of $\rho_{\text{U}}(S)$ for $S = 1, \ldots, 5$; see Fig. 4(b). The figure shows that the values of ρ_{U} associated with the Laplacian uniformly dominate the other two, thus, entailing a more challenging scenario for recovery (cf. discussion around Fig. 2).

Comparison with alternative methods: So far we have assumed that we observe the entire output signal $\mathbf{y}$ when trying to infer $\mathbf{x}$ and $\mathbf{h}$. Nevertheless, it can be the case that we can only sample a subset of the nodes of the graph and try to recover $\{\mathbf{x}_0, \mathbf{h}_0\}$ from this reduced number of observations. Specifically, in Fig. 3(e) we fix $L = S = 3$, $P = 1$, and analyze the error behavior (median errors across 50 realizations) as a function of the number of accessible values y_i of the output for different recovery algorithms. Apart from our convex relaxation approach, we consider a naive least squares (LS) baseline where we solve (5) via a pseudoinverse. Moreover, we consider an alternating minimization (AM) algorithm entailing two steps per iteration: i) given $\mathbf{x}$, vector $\mathbf{h}$ is found as the LS solution of (3); and ii) given $\mathbf{h}$, vector $\mathbf{x}$ is found by minimizing $\|\mathbf{x}\|_1$ subject to (3) followed by a thresholding operation to retain S non-zero values. These two steps are repeated until convergence, and

the algorithm is initialized with the LS estimate of $\mathbf{h}$. Finally, we consider as a benchmark our convex method when the support of $\mathbf{x}_0$ is known (k.s.) [1].

Our proposed method clearly outperforms the naive LS and AM approaches. Say for 60 observations (6 less than the total number of nodes), our method achieves a median error of 0 while the median errors for AM and LS are 0.52 and 0.82, respectively. Outperforming LS is not surprising since this algorithm is agnostic to the sparsity in $\mathbf{x}_0$ and to the fact that $\mathbf{x}_0\mathbf{h}_0^T$ is a rank-one matrix. Further, notice that our method outperforms AM even though the latter assumes that the value of S – but not the support of $\mathbf{x}_0$ – is known. The big gap between the yellow and the purple curves represents the performance penalty due to $\text{supp}(\mathbf{x}_0)$ being unknown. For situations where partial information about the support is available, e.g. we know a priori that the input signal is null on a subset of nodes, intermediate curves (not shown in the figure) are obtained.

Finally, we include two dashed curves to assess the performance of our approach (with known and unknown support), when the partial observations of $\mathbf{y}$ are noisy. More specifically, we define $\widetilde{\mathbf{y}}$ as a perturbed version of $\mathbf{y}$ given by $\widetilde{\mathbf{y}} = \mathbf{y} + \sigma\, \mathbf{y} \circ \mathbf{r}$ where σ controls the magnitude of the perturbation and $\mathbf{r}$ is a random vector whose entries are drawn independently from a standard normal distribution. In particular, the dashed curves in Fig 3(e) correspond to $\sigma = 0.01$. As expected, the median reconstruction errors from noisy observations are larger than their noiseless counterparts, however, our approach can be seen to be robust to noise. For intermediate number of observations, the performance of the noisy scheme with unknown support is within 5% of the noiseless one and, once we observe the whole 66 nodes of the output signal, the error of the noisy approach is in the order of σ and comparable to that obtained by the noisy scheme with known support.

Recovery dependence on graph structure: Theorem 1 reveals that the recovery performance of scheme (11) depends on the graph structure. In particular, when recovering an S-sparse input signal, the value of $\rho_{\mathbf{U}}(S)$ plays an important role, with lower values of $\rho_{\mathbf{U}}(S)$ leading to larger values of α and, hence, larger probabilities of recovery [cf. (14) and (15)]. Even though Theorem 1 refers to theoretical bounds on the performance, we see that in practice the value of $\rho_{\mathbf{U}}(S)$ correlates with the recovery success when implementing (11); see Fig. 5(a). More specifically, we generate 50 Erdős-Rényi graphs of size $N = 50$ and probability of edge appearance p drawn uniformly from $[0.05, 0.15]$. For each of these graphs we simulate 50 blind-identification problems with $L = 3$ and $S = 3$, solve them using (11), and record the reconstruction RMSE and whether the recovery was successful, i.e., RMSE smaller than 0.01. In Fig. 5(a) we plot for each graph the proportion of unsuccessful recoveries (blue circles) and the mean RMSE (orange stars) as a function of $\rho_{\mathbf{U}}(3)$. First notice that the values of $\rho_{\mathbf{U}}(3)$ are clearly larger than their theoretical lower bound of 3 (cf. Lemma 1 and Fig. 2). Most importantly, blind identification is more challenging on graphs with high values of $\rho_{\mathbf{U}}(3)$. For example, the average failure rate for the graphs with $\rho_{\mathbf{U}}(3) \leq 25$ is 0.51 whereas for the rest of the graphs the average is 0.64. Similarly, the average RMSE for the former class of graphs is 0.12, whereas the graphs with larger values of $\rho_{\mathbf{U}}(3)$ achieve an average RMSE of 0.20.

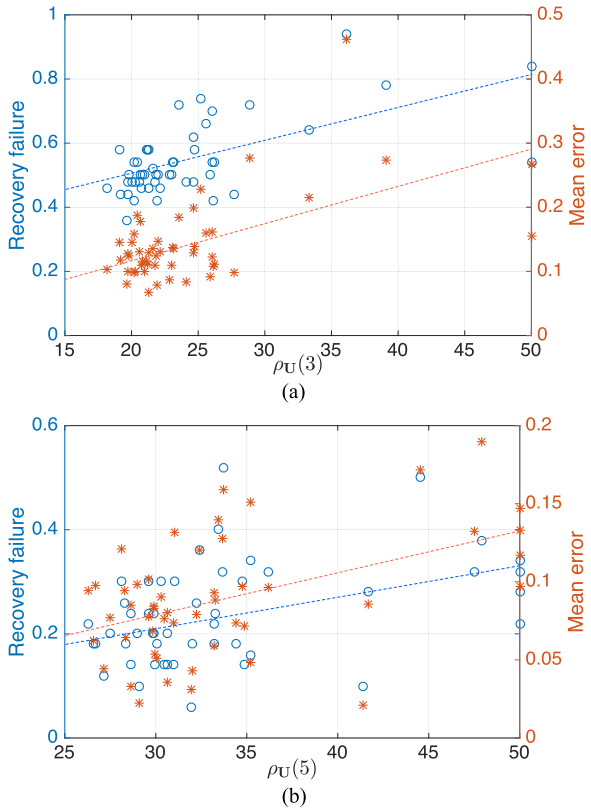

Fig. 5. (a) Recovery failure rate and mean reconstruction error of the ℓ_1 minimization [cf. (11)] as a function of $\rho_{\mathbf{U}}(3)$ for $L = 3$ and $S = 3$. Each point corresponds to an Erdős-Rényi random graph of size $N = 50$ and $p \in [0.05, 0.15]$. The failure rates and mean errors are computed based on blind identification of 50 graph filters on each graph. The best linear fits are drawn in dashed lines. (b) Counterpart of (a) for the reweighted $\ell_{2,1}$ plus nuclear norm [cf. (10)] as a function of $\rho_{\mathbf{U}}(5)$ for $L = 4$ and $S = 5$.

Although Theorem 1 specifies theoretical bounds of recovery for ℓ_1 minimization, in practice the performance of more sophisticated schemes such as the reweighted $\ell_{2,1}$ plus nuclear norm minimization also depends on the value of $\rho_{\mathbf{U}}(S)$; see Fig. 5(b). We mimic the recovery rate analysis for ℓ_1 minimization but in this case we consider a more challenging blind-identification problem ($L = 4$ and $S = 5$) so that the failure rates are significant [cf. Fig. 5(b)]. As can be seen from the figure, the values of $\rho_{\mathbf{U}}(S)$ influence the recovery performance. For graphs with values $\rho_{\mathbf{U}}(5) \leq 35$ for instance, the average failure rate and average RMSE are 0.22 and 0.08, respectively, whereas these values are 0.30 and 0.12 for the remaining graphs. The same trends of increasing recovery failure rates and mean errors with increasing $\rho_{\mathbf{U}}$ have been found in other random graphs including small-world graphs and Barabási-Albert scale-free graphs [36]. Note finally that the fact of the recovery in actual graphs being dependent on $\rho_{\mathbf{U}}$, which is one of the parameters in (14), serves as an indirect validation of the practical value of Theorem 1 and the preceding assumptions.

Epidemics: We consider the social network of Zachary's karate club [38] represented by a graph $\mathcal{G}$ – with adjacency $\mathbf{A}$ – consisting of 34 nodes or members of the club and 78 undirected edges symbolizing friendships among members. On $\mathcal{G}$ we simulate an N-intertwined SIS epidemic model [39], which

is a susceptible-infected-susceptible (SIS) dynamical process where contagion can only occur among friends (i.e., neighbors) in the social network. To be more precise, each individual at any point in time can be in one of two states: susceptible or infected. At each time point, an infected individual heals with probability ω, thus becoming susceptible. By contrast, a susceptible person i becomes infected with probability $\beta\,|\mathcal{I}_i|$, where $|\mathcal{I}_i|$ is the number of neighbors of i that are infected at that point in time. Our goal is to use the blind graph-filter identification methods to identify the sources of the epidemic outbreak. Even though this model entails non-linear and stochastic dynamics, its mean-field approximation can be estimated by a linear function [39], [40]. Formally, denoting by $\mathbf{p}_t$ a vector collecting the probability of infection of each individual at time t, then for small values of $\mathbf{p}_t$ we have that

$$\mathbf{p}_t \approx [\mathbf{I}_N - \upsilon_t(\omega\mathbf{I}_N - \beta\mathbf{A})]\,\mathbf{p}_{t-1} \tag{21}$$

where υ_t is a time-varying step-size that arises after discretization of the dynamics. We generally let υ_t vary with time, representing that some time instants could be very active, with multiple contagions and recoveries, whereas in other time instants the epidemic remains more dormant. Equivalently, a time-varying υ_t can be interpreted as having time-varying healing and infection probabilities but maintaining a fixed ratio between them. Defining $\mathbf{S} := \omega\mathbf{I}_N - \beta\mathbf{A}$, we leverage (21) to write the probabilities of infection at time T as

$$\mathbf{p}_T = \prod_{t=0}^{T-1}(\mathbf{I}_N - \upsilon_t\mathbf{S})\mathbf{p}_0. \tag{22}$$

Notice that the above matrix product is a polynomial in $\mathbf{S}$ – hence a graph filter – of degree T whose coefficients are a function of $\{\upsilon_t\}_{t=0}^{T-1}$, which we assume to be unknown. Moreover, since we are interested in identifying the original sources of infection, the initial probability vector $\mathbf{p}_0$ is what we want to estimate. If we were to observe $\mathbf{p}_T$, then the problem would be precisely an instance of blind graph-filter identification. Given that observing the probabilities of infection at time T seems impractical, we consider a case where we observe W epidemic outbreaks in the same population and estimate $\mathbf{p}_T$ from these. Notice that the epidemic outbreaks need not refer to diseases but could model the spread of rumors or the adoption of new technologies. We want to emphasize that the number and identities of the sources in each of these outbreaks are generally different, but what remains constant is the probability $\mathbf{p}_0$ with which the sources are chosen. Nevertheless, we assume that the nodes that *could* start an epidemic are just a subset of the total (of cardinality S), so that $\mathbf{p}_0$ is sparse. For the simulations here, we consider $T = 3$, $W = 500$, and S drawn at random from $\{3, 4, 5\}$. After estimating $\mathbf{p}_T$ from the realizations, we implement our reweighted $\ell_{2,1}$ plus nuclear norm minimization [cf. (10)] to obtain $\tilde{\mathbf{p}}_0$, our estimate of $\mathbf{p}_0$, which we use to identify the potential sources of contagion. In particular, we quantify the localization error as $\|\text{supp}(\tilde{\mathbf{p}}_0) - \text{supp}(\mathbf{p}_0)\|_0/S$, which specifies the proportion of misidentified sources. Assuming that in each realization we can only access the state (infected or susceptible) of a subset of nodes, in Fig. 6 we plot the localization error – mean across 20 realizations – as a function of the number of nodes observed. In addition, we consider the incorporation of different levels of a priori information. Specifically, we analyze scenarios where we know that the S sources belong to a subset of potential nodes of cardinality Q, so that when $Q = N$ no a priori information is considered.

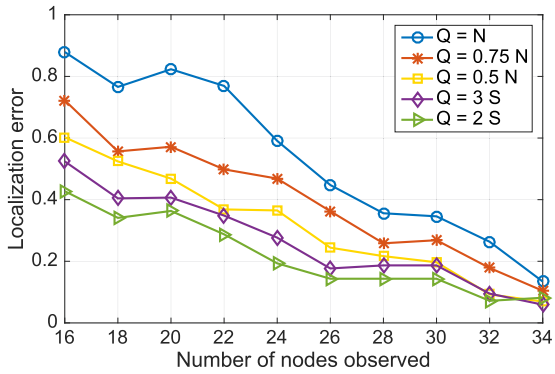

Fig. 6. Localization error for epidemic sources as a function of the number of observed nodes and parametrized by the number of potential sources Q.

Fig. 6 shows that, independently of the value of Q, the localization error decreases when the number of observed nodes increases. Intuitively, the larger the number of observed nodes, the more complete the estimation of $\mathbf{p}_T$ is, hence, entailing a more reliable estimate $\tilde{\mathbf{p}}_0$ of the epidemic sources. For $Q = N$, if only 16 nodes are observed the localization is very poor with an average error of 0.88, but when all 34 nodes are observed this error decreases to 0.13. Furthermore, Fig. 6 illustrates the benefit of a priori information. For instance, when only 24 nodes are observable and we have no a priori information, the average error is 0.59. However, this error can be reduced to 0.37 and 0.19 by constraining the potential sources to subsets of cardinality $Q = N/2$ and $Q = 2S$, respectively.

VI. Conclusion

We formulated and studied the problem of blind graph-filter identification, an extension of blind deconvolution of time (or spatial) domain signals to graphs. While the graph signal observations are bilinear functions of the filter coefficients and the sparse input signal, leveraging the frequency interpretation of graph signals and graph filters it was possible to show that they are also linearly related to the entries of a lifted rank-one, row-sparse matrix of the unknowns. Accordingly, the blind graph-filter identification problem was tackled via rank and sparsity minimization subject to linear constraints, an inverse problem amenable to convex relaxations offering provable recovery guarantees under simplifying assumptions. The probabilistic guarantees derived offer intuition as to how the recovery performance depends on the sparsity level, filter order, as well as on spectral properties of the underlying graph via suitably defined coherence parameters. A particular upshot of the analysis is that the most favorable recovery setting arises with circulant graphs, which include blind deconvolution in the temporal domain as a special case. Numerical tests validated the theoretical claims and demonstrated that the proposed approach offers satisfactory recovery performance, even for settings well beyond the scope of the analysis. Looking forward and building on the insights gained here, it is of interest to develop spectral graph theoretical tools to facilitate identifying the key relationships between

the (e.g., topological) properties of graphs and the resulting recovery performance. For a given graph, it is also worth to formally characterize whether there are filters or signal supports for which recovery is easier or more challenging. Future research includes also estimating the shift operator by bringing to bear methods of network topology inference [5]; see also [9], [41] for recent related approaches to estimate the graph structure from the observation of graph signals.

APPENDIX A
PROOF OF THEOREM 1

We reiterate that the proof in this section follows closely the ideas used to establish [13, Theorem 3.1]; see also [10]. Thus the emphasis will be on the differences that arise in the graph setting dealt with here.

Preliminary definitions and notations: Going back to the discussion in Section III, for $\mathbf{Z} = \mathbf{x}\mathbf{h}^T$ recall the linear mapping $\mathcal{M} : \mathbb{C}^{N\times L} \mapsto \mathbb{C}^N$ such that

$$\widehat{\mathbf{y}} = \mathcal{M}(\mathbf{Z}) := \{\mathbf{u}_i^T \mathbf{Z}\boldsymbol{\psi}_i\}_{i=1}^N, \tag{23}$$

where $\mathbf{u}_i^T$ and $\boldsymbol{\psi}_i^T$ denote the i-th rows of $\mathbf{U}$ and $\boldsymbol{\Psi}$, respectively. Whenever clear from context, as in (23), notation $\{\cdot\}$ will be used for the element-wise definition of a vector. Accordingly, one can write $\widehat{\mathbf{y}} = \mathbf{M}\text{vec}(\mathbf{Z})$ for $\mathbf{M} := \left(\boldsymbol{\Psi}^T \odot \mathbf{U}^T\right)^T \in \mathbb{C}^{N\times LN}$, where

$$\mathbf{M}^H = [\mathbf{m}_1, \ldots, \mathbf{m}_N] \in \mathbb{C}^{LN\times N}, \quad \mathbf{m}_i = \overline{\boldsymbol{\psi}}_i \otimes \overline{\mathbf{u}}_i. \tag{24}$$

By using the inner product defined on $\mathbb{C}^{N\times L}$ as $\langle \mathbf{X}, \mathbf{Z}\rangle := \text{tr}(\mathbf{X}\mathbf{Z}^H)$, the adjoint operator $\mathcal{M}^* : \mathbb{C}^N \mapsto \mathbb{C}^{N\times L}$ as well as $\mathcal{M}^*\mathcal{M} : \mathbb{C}^{N\times L} \mapsto \mathbb{C}^{N\times L}$ take the form

$$\mathcal{M}^*(\mathbf{z}) = \sum_{i=1}^N z_i \overline{\mathbf{u}}_i \boldsymbol{\psi}_i^H, \quad \mathcal{M}^*\mathcal{M}(\mathbf{Z}) = \sum_{i=1}^N \overline{\mathbf{u}}_i \mathbf{u}_i^T \mathbf{Z}\boldsymbol{\psi}_i\boldsymbol{\psi}_i^H. \tag{25}$$

Based on the aforementioned definitions, the following useful formula holds

$$\begin{aligned}\text{vec}(\mathcal{M}^*\mathcal{M}(\mathbf{Z})) &= \mathbf{M}^H\mathbf{M}\text{vec}(\mathbf{Z}) \\ &= \sum_{i=1}^N \left(\overline{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i^T \otimes \overline{\mathbf{u}}_i\mathbf{u}_i^T\right)\text{vec}(\mathbf{Z}).\end{aligned} \tag{26}$$

Consider a subset of measurements $\Gamma_p \subset \{1, \ldots, N\}$ with $|\Gamma_p| = Q$, and define the row-sampled operator $\mathcal{M}_p : \mathbb{C}^{N\times L} \mapsto \mathbb{C}^Q$ as $\mathcal{M}_p(\mathbf{Z}) := \{\mathbf{u}_i^T\mathbf{Z}\boldsymbol{\psi}_i\}_{i\in\Gamma_p}$; as well as $\mathcal{M}_p^*\mathcal{M}_p(\mathbf{Z}) := \sum_{i\in\Gamma_p} \overline{\mathbf{u}}_i\mathbf{u}_i^T\mathbf{Z}\boldsymbol{\psi}_i\boldsymbol{\psi}_i^H$. The linear operator $\mathcal{M}_p$ has matrix representation $\mathbf{M}_p \in \mathbb{C}^{Q\times LN}$, where $\mathbf{M}_p^H$ has columns $\{\mathbf{m}_i\}_{i\in\Gamma_p}$. Accordingly, one can write $\mathcal{M}_p(\mathbf{Z}) = \mathbf{M}_p\text{vec}(\mathbf{Z})$ and $\text{vec}(\mathcal{M}_p^*\mathcal{M}_p(\mathbf{Z})) = \mathbf{M}_p^H\mathbf{M}_p\text{vec}(\mathbf{Z})$.

Regarding the support of the sparse vector $\mathbf{x}_0$, we can assume without loss of generality that the first S entries of $\mathbf{x}_0$ are nonzero and so are the first S rows of $\mathbf{Z}_0 = \mathbf{x}_0\mathbf{h}_0^T$. While slightly abusing notation, it is convenient to denote as Ω the supports of both $\mathbf{x}_0$ and $\mathbf{Z}_0$. Next, let $\mathbf{x}_\Omega$ and $\mathbf{Z}_\Omega$ denote the orthogonal projections of $\mathbf{x}$ and $\mathbf{Z}$ onto Ω, respectively. Leveraging these definitions, we can define $\mathcal{M}_\Omega$ and $\mathcal{M}_\Omega^*$ as the restriction of $\mathcal{M}$ and $\mathcal{M}^*$ to Ω, such that

$$\begin{aligned}\mathcal{M}_\Omega(\mathbf{Z}) &= \{\mathbf{u}_i^T\mathbf{Z}_\Omega\boldsymbol{\psi}_i\}_{i=1}^N = \{\mathbf{u}_{i,\Omega}^T\mathbf{Z}_\Omega\boldsymbol{\psi}_i\}_{i=1}^N, \\ \mathcal{M}_\Omega^*(\mathbf{z}) &= \sum_{i=1}^N z_i\overline{\mathbf{u}}_{i,\Omega}\boldsymbol{\psi}_i^H.\end{aligned} \tag{27}$$

Naturally, one can be interested in projected *and* row-sampled operators of the form $\mathcal{M}_{p,\Omega} = \{\mathbf{u}_{i,\Omega}^T\mathbf{Z}_\Omega\boldsymbol{\psi}_i\}_{i\in\Gamma_p}$.

Instrumental to our proof will be the following form of the non-commutative Bernstein inequality for matrices [33].

Theorem 2 (Th. 4.4 [13]): Consider a finite sequence $\{\mathbf{Z}_i\}_{i\in\Gamma_p}$ of independent, centered random matrices with dimension $M \times M$. Assume that $\|\mathbf{Z}_i\| \le R$ and introduce the random matrix

$$\boldsymbol{\Sigma} = \sum_{i\in\Gamma_p} \mathbf{Z}_i$$

with variance parameter

$$\sigma^2 = \max\left\{\left\|\sum_{i\in\Gamma_p}\mathbb{E}\left[\mathbf{Z}_i\mathbf{Z}_i^H\right]\right\|, \left\|\sum_{i\in\Gamma_p}\mathbb{E}\left[\mathbf{Z}_i^H\mathbf{Z}_i\right]\right\|\right\}. \tag{28}$$

Then for all $t \ge 0$, it holds that

$$\text{P}\left(\|\boldsymbol{\Sigma}\| \ge t\right) \le 2M\exp\left(-\frac{t^2/2}{\sigma^2 + Rt/3}\right). \tag{29}$$

Optimality conditions: A key step in the proof is Proposition 2, which states four conditions guaranteeing that the unique solution to (11) is indeed $\mathbf{Z}_0 = \mathbf{x}_0\mathbf{h}_0^T$. See [13, Proposition 4.2] for a proof. In what follows, we denote by $\Omega^\perp$ the orthogonal complement of Ω.

Proposition 2: The rank-one matrix $\mathbf{Z}_0 = \mathbf{x}_0\mathbf{h}_0^T$ is the unique minimizer of (11), if there exists $\mathbf{Y} \in \text{range}(\mathcal{M}^*)$ and a scalar γ that satisfy conditions:

a) $\|\text{sign}(\mathbf{Z}_0) - \mathbf{Y}_\Omega\|_\text{F} \le 1/(4\sqrt{2}\gamma)$, b) $\|\mathbf{Y}_{\Omega^\perp}\|_\infty \le 1/2$;

and the operator $\mathcal{M}$ satisfies conditions:

c) $\|\mathcal{M}_\Omega^*\mathcal{M}_\Omega - \mathbf{I}_\Omega\| \le 1/2$, d) $\|\mathcal{M}\| \le \gamma$.

The rest of the section is devoted to establish each of the four conditions *a)-d)* given above. Although Proposition 2 entails deterministic conditions, as customary when dealing with sparse recovery algorithms we show that conditions *a)-d)* can be satisfied with a certain probability. More specifically, it turns out that all conditions hold with probability at least $1 - N^{-\alpha+1}$, giving rise to the statement in Theorem 1 [cf. (15)].

We begin by showing condition *c)* in Proposition 2. We do so by first showing a more general result in Lemma 2, which is a modification of [13, Lemma 4.7].

Lemma 2: For any fixed $0 < \delta \le 1$ and partition $\{\Gamma_p\}_{p=1}^P$ of $\{1, 2, \ldots, N\}$ with $|\Gamma_p| = Q$ for all p, and defining $\mathbf{T}_p = \sum_{i\in\Gamma_p}\overline{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i^T$ we have that

$$\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_\text{F}=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p\|_\text{F} \le \frac{\delta Q}{N}, \tag{30}$$

with probability at least $1 - N^{-\alpha+1}$ if $\alpha \ge 1$ and

$$\alpha \le \frac{Q\delta^2(5/2 + 2\delta/3)^{-1}}{\rho_\Psi(L)\rho_\text{U}(S)N\log(2NLS)}. \tag{31}$$

Proof: Define $\boldsymbol{\Upsilon}_i \in \mathbb{C}^{LN\times LN}$ as

$$\boldsymbol{\Upsilon}_i := (\overline{\boldsymbol{\psi}}_i \boldsymbol{\psi}_i^T) \otimes (\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^T - \mathbf{I}_{N,\Omega}), \tag{32}$$

and notice that every $\boldsymbol{\Upsilon}_i$ is a centered random matrix since $\mathbb{E}\left[\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^T\right] = \mathbf{I}_{N,\Omega}$ due to the normalization of $\mathbf{U}$ (cf. Theorem 1). Leveraging (26), it follows that $\text{vec}\left(\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p\right) = \sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\text{vec}(\mathbf{Z}_\Omega)$ and, from the definition of matrix induced norm, we obtain that

$$\sup_{\|\mathbf{Z}_\Omega\|_{\text{F}}=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p\|_{\text{F}} = \Big\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\Big\|. \tag{33}$$

Thus, our goal is to use Theorem 2 to find a probabilistic bound on $\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\|$. To this end, we need to find suitable expressions for R and σ^2. For computing R, notice that

$$\|\boldsymbol{\Upsilon}_i\| = |\boldsymbol{\psi}_i^T\overline{\boldsymbol{\psi}}_i|\|\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^T - \mathbf{I}_{N,\Omega}\| \leq \rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S), \tag{34}$$

where we used the definition of $\rho_{\mathbf{A}}$ in (13). Since (34) is true for all i, we have that $R \leq \rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S)$ as wanted. In order to compute σ^2, it suffices to consider $\boldsymbol{\Upsilon}_i\boldsymbol{\Upsilon}_i^H$ since $\boldsymbol{\Upsilon}_i$ is a Hermitian matrix for all i [cf. (28)]. Thus, we have

$$\begin{aligned}\sigma^2 &= \Big\|\sum_{i\in\Gamma_p}\mathbb{E}\left[\boldsymbol{\Upsilon}_i\boldsymbol{\Upsilon}_i^H\right]\Big\| \qquad (35)\\ &\leq \rho_{\boldsymbol{\Psi}}(L)\Big\|\sum_{i\in\Gamma_p}(\overline{\boldsymbol{\psi}}_i\boldsymbol{\psi}_i^T)\otimes\mathbb{E}\left[(\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^T - \mathbf{I}_{N,\Omega})^2\right]\Big\|.\end{aligned}$$

To compute the required expected value, expand the square and use the facts that $\mathbf{u}_{i,\Omega}^T\overline{\mathbf{u}}_{i,\Omega} \leq \rho_{\mathbf{U}}(S)$ and $\mathbb{E}\left[\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^T\right] = \mathbf{I}_{N,\Omega}$. Substituting these into (35) and recalling the definition of $\mathbf{T}_p$ from the statement of Lemma 2, it follows that

$$\sigma^2 \leq \rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S)\|\mathbf{T}_p\| \leq \rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S)\frac{5Q}{4N}, \tag{36}$$

where the last inequality follows from [13, equation (4.7)]. We use the results in (34) and (36) to apply Theorem 2 for the case where $t = \delta Q/N$. In using Theorem 2, $\boldsymbol{\Upsilon}_i$ can be interpreted as an $LS\times LS$ matrix for all i since this is the dimension of its support [cf. (32)]. Thus, we obtain

$$\text{P}\left(\Big\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\Big\| \geq \frac{\delta Q}{N}\right) \leq 2LS\exp\left(-\frac{\delta^2 Q/2N}{\rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S)(\frac{5}{4}+\frac{\delta}{3})}\right). \tag{37}$$

If, in particular, for some constant $\alpha \geq 1$ we choose

$$Q \geq \frac{\alpha}{\delta^2}\left(\frac{5}{2}+\frac{2\delta}{3}\right)\rho_{\boldsymbol{\Psi}}(L)\rho_{\mathbf{U}}(S)N\log(2NLS), \tag{38}$$

the right-hand side in (37) is not larger than $N^{-\alpha}$. Using a union bound on (37) to find the probability of $\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\| \geq \delta Q/L$ for *some* p, and then considering its complementary event, we get that

$$\text{P}\left(\max_{1\leq p\leq P}\Big\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\Big\| \leq \frac{\delta Q}{N}\right) \geq 1 - PN^{-\alpha} \geq 1 - N^{-\alpha+1}. \tag{39}$$

From (33) and (39), statement (30) follows whereas the expression for α in (31) is obtained from (38). ■

Condition *c)* in Proposition 2 follows by specializing Lemma 2 for $\delta = 1/2$ and $P = 1$. Notice that the latter equality implies that $Q = N$ and $\mathbf{T}_1 = \mathbf{I}_L$. As discussed after the statement of Theorem 1, a novel component introduced in the proof of Lemma 2 compared to that of [13, Lemma 4.7] is the appearance of $\rho_{\mathbf{U}}(S)$ in the lower bound for Q [cf. (38)]. Notice that in [13], $\rho_{\mathbf{U}}(S) = S$ since $\mathbf{U}$ is assumed to be random Fourier, thus, every element has unit magnitude. In our case, different graphs give rise to less favorable (larger) bounds on Q since $\rho_{\mathbf{U}}(S) \geq S$.

Our next step is to prove condition *d)* in Proposition 2. We attain this in the following lemma, a restatement of [13, Lemma 4.9].

Lemma 3: For $\mathcal{M}$ defined in (23) and $\alpha \geq 1$ then it holds that

$$\|\mathcal{M}\| \leq \gamma := \sqrt{2N(\log(2LN)+1)+1}, \tag{40}$$

with probability at least $1 - N^{-\alpha}$ if $\alpha \leq (\rho_{\boldsymbol{\Psi}}(L)\log(N))^{-1}$.

Proof: The proof follows the same steps as those in the proof of [13, Lemma 4.9]. The only step that needs to be checked is whether $\mathbb{E}\left[(\overline{\mathbf{u}}_i\mathbf{u}_i^T - \mathbf{I}_N)^2\right] = (N-1)\mathbf{I}_N$ still holds in our context. To see this, expand the square and use the facts that $\mathbb{E}\left[\overline{\mathbf{u}}_i\mathbf{u}_i^T\right] = \mathbf{I}_N$ and that $\mathbf{u}_i^T\overline{\mathbf{u}}_i = N$ for all i, which are immediate implications of the normalization of $\mathbf{U}$ (cf. Theorem 1). Finally, to obtain the expressions presented in the statement of the proposition notice that k, L and $\mu_{\max}^2 k/L$ in [13] are equal to L and N and $\rho_{\boldsymbol{\Psi}}(L)$ here, respectively. ■

Construction of dual certificate: We construct the inexact dual certificate $\mathbf{Y}$ mentioned in conditions *a)* and *b)* of Proposition 2 via the celebrated golfing scheme [42]. The goal of the scheme is to generate a sequence of random matrices $\mathbf{Y}_p$ in $\text{range}(\mathcal{M}^*)$ such that the sequence converges to $\text{sign}(\mathbf{Z}_0)$ [cf. *a)*] while keeping each entry in $\Omega^\perp$ small [cf. *b)*]. We initialize $\mathbf{Y}_0 = \mathbf{0}$ and set the following recursion

$$\mathbf{Y}_p := \mathbf{Y}_{p-1} + \frac{N}{Q}\mathcal{M}_p^*\mathcal{M}_p(\text{sign}(\mathbf{Z}_0) - \mathbf{Y}_{p-1,\Omega}) \tag{41}$$

where the successive operators $\mathcal{M}_p$ are based on different blocks Γ_p of a partition $\{\Gamma_p\}_{p=1}^P$ of $\{1,2,\ldots,N\}$ such that $|\Gamma_p| = Q$ for all p. We define our desired dual certificate as $\mathbf{Y} := \mathbf{Y}_P$ and the sequence of residuals $\mathbf{W}_p := \mathbf{Y}_{p,\Omega} - \text{sign}(\mathbf{Z}_0)$. From this definition, it follows that $\mathbf{W}_0 = \mathbf{Y}_0 - \text{sign}(\mathbf{Z}_0) = -\text{sign}(\mathbf{Z}_0)$ implying that $\|\mathbf{W}_0\|_F = \sqrt{LS}$ from the sparsity level in $\mathbf{Z}_0$. Furthermore, Lemma 2 can be leveraged to show that $\|\mathbf{W}_p\|_{\text{F}} \leq 2^{-1}\|\mathbf{W}_{p-1}\|_{\text{F}}$ with probability at least $1 - N^{-\alpha+1}$ as long as (31) is satisfied for $\delta = 1/4$ (cf. Lemma 4.6 in [13]). Combining these two observations, we obtain that $\|\mathbf{W}_P\|_{\text{F}} = \|\mathbf{Y}_\Omega - \text{sign}(\mathbf{Z}_0)\|_{\text{F}} \leq 2^{-P}\sqrt{LS}$. By equating this upper bound with that in condition *a)* of Proposition 2, we obtain a lower bound on the values of P that guarantee fulfillment of the condition, namely

$$P \geq \frac{\log(4\gamma\sqrt{2LS})}{\log(2)}. \tag{42}$$

Finally, in order to show condition *b)* in Proposition 2 we begin by leveraging (41) and the definition of $\mathbf{W}_p$ to write

$$\mathbf{Y}_P = \mathbf{Y} = -\frac{N}{Q}\sum_{p=1}^P\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1}). \tag{43}$$

In order to show that $\|\mathbf{Y}_{\Omega^\perp}\|_\infty \leq 1/2$, it suffices to show that $\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty \leq 2^{-p-1}Q/N$ and then use the expression of a geometric sum in (43). We show this latter claim via the following lemma, which is a modified version of [13, Theorem 4.11].

Lemma 4: Under the assumption that $\|\mathbf{W}_p\|_{\mathrm{F}} \leq 2^{-p}\sqrt{LS}$, it holds that

$$\mathrm{P}\left(\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty \leq \frac{Q}{2^{p+1}N}\right) \geq 1 - N^{-\alpha+1} \quad (44)$$

for all p if $\alpha \geq 1$ and

$$\alpha \leq \frac{3Q\left(120\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)} + 8\sqrt{\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)}}\right)^{-1}}{N\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(2SN^2)}. \quad (45)$$

Proof: The proof follows similar steps as those in the proof of [13, Theorem 4.11]. However, in our case, matrix $\mathbf{U}$ is not random Fourier, thus, different bounds must be used. In particular, we leverage the definition of $\rho_{\mathbf{A}}$ in (13) to state that [cf. (24)] $\|\mathbf{m}_i\|_\infty \leq \|\boldsymbol{\psi}_i\|_\infty\|\mathbf{u}_i\|_\infty = \sqrt{\rho_{\mathbf{\Psi}}(1)\rho_{\mathbf{U}}(1)}$ and, similarly, that $\|\mathbf{m}_i\|_2 \leq \sqrt{\rho_{\mathbf{\Psi}}(L)\rho_{\mathbf{U}}(S)}$. These bounds lead to the following expressions for R and σ^2

$$R \leq 2^{-p+1}\sqrt{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)LS}, \quad (46)$$

$$\sigma^2 \leq 2^{-2p}\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)5QLS/N. \quad (47)$$

After this, we apply the Bernstein inequality (cf. Theorem 2) and implement a union bound mimicking the procedure in (37)–(39) to attain the statement of the lemma. ■

Thus far, we have found requirements on $\alpha \geq 1$ that ensure the fulfillment of conditions *a)-d)* in Proposition 2 with high probability. Consequently, we need to satisfy the most restrictive of the requirements on α to guarantee simultaneous fulfillment of conditions *a)-d)* and, hence, ensure recovery of $\mathbf{Z}_0 = \mathbf{x}_0\mathbf{h}_0^T$. These requirements are re-stated below, in order of appearance:

$$\alpha \leq \begin{cases} \frac{3}{34\,\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(2NLS)}, & (48a) \\ \frac{1}{\rho_{\mathbf{\Psi}}(L)\log(N)}, & (48b) \\ \frac{3\log(2)/128}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(4\gamma\sqrt{2LS})\log(2NLS)}, & (48c) \\ \frac{3\log(2)\left(120\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)}+8\sqrt{\frac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{\Psi}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)}}\right)^{-1}}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{\Psi}}(L)\log(4\gamma\sqrt{2LS})\log(2N^2S)}. & (48d) \end{cases}$$

Recall that (48a) is obtained by specializing (31) to $Q = N$ and $\delta = 1/2$ [cond. *c)*] whereas (48b) comes from Lemma 3 [cond. *d)*]. Expression (48c) is obtained by particularizing (31) to $\delta = 1/4$, noting that $Q = N/P$ and using (42) to bound P [cond. *a)*]. Finally, (48d) is derived from (45) by again substituting $Q = N/P$ [cond. *b)*].

Notice that fulfillment of (48c) immediately implies fulfillment of (48a) and (48b). By leveraging the facts that $N \geq L$, $\rho_{\mathbf{U}}(1)S \geq \rho_{\mathbf{U}}(S)$, and $\rho_{\mathbf{\Psi}}(1)L \geq \rho_{\mathbf{\Psi}}(L)$ (cf. Lemma 1), it follows that (48d) implies (48c). This makes (48d) the most stringent of the requirements giving rise to (14) in Theorem 1.

REFERENCES

[1] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters with sparse inputs," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, Dec. 2015, pp. 449–452.

[2] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Blind identification of graph filters with multiple sparse inputs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 4099–4103.

[3] 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, Mar. 2013.

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

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

[6] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[7] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *Proc. 53rd Allerton Conf. Commun. Control Comput.*, Sep. 2015, pp. 1406–1413.

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

[9] J. Mei and J. Moura, "Signal processing on graphs: Estimating the structure of a graph," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*, 2015, pp. 5495–5499.

[10] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, Mar. 2014.

[11] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4363–4378, Aug. 2016.

[12] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Interpolation of graph signals using shift-invariant graph filters," in *Proc. Eur. Signal Process. Conf.*, Aug. 2015, pp. 210–214.

[13] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, 2015, Art. no. 115002.

[14] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Control Conf.*, 2001, vol. 6, pp. 4734–4739.

[15] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, May 2015.

[16] Y. Li, K. Lee, and Y. Bresler, "A unified framework for identifiability analysis in bilinear inverse problems with applications to subspace and sparsity models," *arXiv:1501.06120 [cs.IT]*, 2015.

[17] S. Patterson and B. Bamieh, "Interaction-driven opinion dynamics in online social networks," in *Proc. 1st Workshop Social Media Analytics*, Washington, DC, USA, Jul. 2010, pp. 98–105.

[18] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulations," *J. Artif. Soc. Social Simul.*, vol. 5, pp. 1–33, 2002.

[19] A. Sandryhaila and J. M. F. Moura, "Finite-time distributed consensus through graph filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 1080–1084.

[20] S. Safavi and U. Khan, "Revisiting finite-time distributed algorithms via successive nulling of eigenvalues," *IEEE Signal Process. Lett.*, vol. 22, pp. 54–57, 2015.

[21] M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch, "Emergent network topology at seizure onset in humans," *Epilepsy Res.*, vol. 79, pp. 173–86, 2008.

[22] E. E. Papalexakis, A. Fyshe, N. D. Sidiropoulos, P. P. Talukdar, T. Mitchell, and C. Faloutsos, "Good-enough brain model: Challenges, algorithms and discoveries in multi-subject experiments," in *Proc. ACM Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2014, pp. 95–104.

[23] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2365–2382, Jun. 2009.

[24] A. Sandryhaila and J. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[25] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, Feb. 2003.

[26] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv:1402.2637*, 2014.

[27] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[28] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.

[29] M. Golbabaee and P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 2741–2744.
[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, 2011.
[31] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, Dec. 2008.
[32] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer, 2013.
[33] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, pp. 1–230, 2015.
[34] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmon. Anal.*, vol. 40, no. 2, pp. 260–291, 2016.
[35] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[36] B. Bollobás, *Random Graphs*. New York, NY, USA: Springer, 1998.
[37] P. Hagmann *et al.*, "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, 2008, Art. no. e159.
[38] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropological Res.*, vol. 33, no. 4, pp. 452–473, 1977.
[39] P. Van Mieghem, "The N-intertwined SIS epidemic network model," *Computing*, vol. 93, no. 2, pp. 147–169, 2011.
[40] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, Mar. 2014.
[41] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," in *Proc. IEEE Workshop Stat. Signal Process.*, Palma de Mallorca, Spain, Jun. 2016, pp. 1–5.
[42] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, Mar. 2011.

Santiago Segarra (M'16) received the B.Sc. degree in industrial engineering with highest honors (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Argentina, in 2011, the M.Sc. degree in electrical engineering from the University of Pennsylvania (Penn), Philadelphia, PA, USA, in 2014, and the Ph.D. degree in electrical and systems engineering from Penn in 2016. Since 2016, he has been working as a Postdoctoral Research Associate with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include network theory, data analysis, machine learning, and graph signal processing. Dr. Segarra received the ITBA's 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, the Best Student Paper Awards at the 2015 Asilomar Conference and the 2016 IEEE Statistical Signal Processing Workshop, and the Best Paper Award at the 2016 IEEE Sensor Array, and Multichannel Signal Processing Workshop.

Gonzalo Mateos (M'12) received the B.Sc. degree in electrical engineering from the Universidad de la República (UdelaR), Montevideo, Uruguay, in 2005 and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota (UofM), Minneapolis, MN, USA, in 2009 and 2011, respectively. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, University of Rochester, NY, USA, where he is also affiliated with the Georgen Institute for Data Science. During 2013, he was a Visiting Scholar with the CS Department, Carnegie Mellon University. From 2003 to 2006, he was an Assistant with the Department of Electrical Engineering, UdelaR. From 2004 to 2006, he worked as a Systems Engineer at Asea Brown Boveri (ABB), Uruguay. His research interests include the areas of statistical learning from big data, network science, wireless communications, and graph signal processing. His current research focuses on algorithms, analysis, and application of statistical signal processing tools to dynamic network health monitoring, social, power grid, and big data analytics. His doctoral work has been recognized with the 2013 UofM's Best Dissertation Award (Honorable Mention) across all Physical Sciences and Engineering areas.

Antonio G. Marques (SM'13) received the Telecommunications Engineering degree and the Doctorate degree, both with highest honors, from the Carlos III University of Madrid, Madrid, Spain, in 2002 and 2007, respectively. In 2007, he became a Faculty of the Department of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain, where he currently develops his research and teaching activities as an Associate Professor. From 2005 to 2015, he held different visiting positions at the University of Minnesota, Minneapolis. From 2015 to 2016, he was a Visitor Scholar at the University of Pennsylvania, Philadelphia. His research interests include the areas of signal processing, networking and communications. His current research focuses on stochastic optimization of wireless and power networks, signal processing for graphs, and nonlinear network optimization. Dr. Marques has served the IEEE in a number of posts (he is currently an Associate Editor of the Signal Processing Letters) and his work has been awarded in several conferences and workshops.

Alejandro Ribeiro (M'07) received the B.Sc. degree in electrical engineering from the Universidad de la República Oriental del Uruguay, Montevideo, Uruguay, in 1998 and the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, in 2005 and 2007, respectively. From 1998 to 2003, he was a member of the technical staff at Bellsouth Montevideo. After his M.Sc. and Ph.D studies, in 2008 he joined the University of Pennsylvania (Penn), Philadelphia, PA, USA, where he is currently the Rosenbluth Associate Professor at the Department of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena. His focus is on structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control. Dr. Ribeiro received the 2014 O. Hugo Schuck Best Paper Award, the 2012 S. Reid Warren, Jr. Award presented by Penn's undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and Paper Awards at the 2016 SSP Workshop, 2016 SAM Workshop, 2015 Asilomar SSC Conference, ACC 2013, ICASSP 2006, and ICASSP 2005. Dr. Ribeiro is a Fulbright scholar and a Penn Fellow.