Network Topology Identification from Imperfect Spectral Templates

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

***Abstract*—This paper studies network topology inference, which is a cornerstone problem in statistical analysis of complex systems. The fresh look advocated here builds on recent advances in convex optimization and graph signal processing to identify the so-termed graph-shift operator (encoding the network topology) given only the eigenvectors of the shift. These *spectral templates* can be obtained, for example, from the covariance of a set of graph signals defined on the particular network. The novel idea is to find a graph shift that, while being consistent with the provided spectral information, endows the network structure with certain desired properties such as sparsity. To that end we develop efficient inference algorithms stemming from provably-tight convex relaxations of natural non-convex criteria. We initially propose algorithms along with theoretical performance guarantees for the case when the eigenbasis is perfectly known. We then investigate setups where an imperfect (noisy) eigenbasis is available as well as others when only a subset of the eigenvectors is known. Numerical tests showcase the effectiveness of the proposed algorithms in recovering real-world social networks.**

***Index Terms*—Network topology inference, graph signal processing, spectral graph theory, network diffusion processes**

I. Introduction

Advancing a holistic theory of networks necessitates breakthroughs in modeling, identification, and controllability of distributed network processes – often conceptualized as *signals defined on the vertices of a graph* [1], [2]. 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 [3], [4]. Instrumental to that end is the so-termed graph-shift operator (GSO) [4], a matrix capturing the graph's local topology and whose eigendecomposition defines the graph Fourier transform [4]. Most GSP works assume that the GSO (hence the graph) is known, and then analyze how the algebraic and spectral characteristics of the GSO affect the properties of the signals and filters defined on such a graph. We take here the reverse path and investigate how to use information available from graph signals and filters to infer the underlying graph topology; see also [5], [6]. By advocating a two-step approach, we first leverage results from GSP theory to estimate the GSO's eigenbasis, and then rely on these (possibly imperfect and incomplete) *spectral templates* to recover the GSO itself.

Network topology inference from a set of (graph-signal) observations is a prominent problem in Network Science [2], [7], [8]. Since networks can encode similarities and dependencies between nodes, a number of approaches construct graphs whose edge weights correspond to the correlation, or the coherence, between signal profiles at incident nodes [2, Ch. 7]. While these approaches are not without merit, they form links taking into account only pairwise interactions, ignoring latent network effects. Acknowledging these limitations, alternative methods rely on partial correlations [2], [9], Gaussian graphical models [10], [11], or, Granger causality [7], [12]. Differently, recent GSP-based network inference frameworks postulate that the network exists as a latent underlying structure, and that observations are generated as a result of a network process defined in such graph. For instance, network structure is estimated in [6] to unveil unknown relations among nodal time series adhering to an autoregressive model involving graph-filter dynamics. A factor analysis-based approach was put forth in [5] to infer graph Laplacians, seeking that input graph signals are smooth over the learned topologies. Different from [5], [6] that operate on the graph domain, the goal here is to identify graphs that endow the given observations with desired spectral (frequency-domain) characteristics.

We begin by surveying the required GSP background and showing how the spectral templates can be obtained in practice without having access to the GSO itself (Section II). We then formulate the first version of our GSO identification problem given spectral templates (Section III-A). The novel idea is to search among all feasible networks for the one that endows the resulting graph-signal transforms with prescribed spectral properties, while the inferred graph also exhibits desirable structural characteristics including sparsity. Conditions under which the feasible set reduces to a singleton are derived, convex relaxations leading to computationally-efficient algorithms are proposed, and theoretical performance guarantees are provided. We then introduce an inference method for the pragmatic case where knowledge of the spectral templates is imperfect (Section III-B). Last but not least, we propose a topology inference algorithm for the case where not all the eigenvectors of the GSO are known (Section III-C). Computer simulations highlight the effectiveness of the proposed algorithms in identifying both synthetic and real-world social networks (Section IV). Concluding remarks are given in Section V.

II. Filtered graph signals and spectral templates

Here we introduce basic GSP tools and explain how the graph topology may influence properties of graph signals.

Graphs. Let $\mathcal{G}$ denote an undirected graph with a set of nodes $\mathcal{N}$ (with cardinality N) and a set of links $\mathcal{E}$, such that if node

Work in this paper is supported by the NSF award CCF-1217963 and Spanish MINECO TEC2013-41604-R. [†]Dept. of Electrical and Systems Eng., Univ. of Pennsylvania. [‡]Dept. of Signal Theory and Comms., King Juan Carlos Univ. [*]Dept. of Electrical and Comp. Eng., Univ. of Rochester. Emails: {ssegarra,aribeiro}@seas.upenn.edu, antonio.garcia.marques@urjc.es, gmateosb@ece.rochester.edu.

i is connected to j, then both (i,j) and (j,i) belong to $\mathcal{E}$. The set $\mathcal{N}_i := \{j\,|(j,i) \in \mathcal{E}\}$ stands for the neighborhood of i. For any given $\mathcal{G}$, its adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ is a sparse matrix with nonzero elements $A_{ij} = A_{ji}$ if and only if $(i,j) \in \mathcal{E}$. The values of A_{ji} can be binary, or real in the weighted case to capture the strength of the link from i to j.

Graph signals and shift operator. Graph signals defined on the nodes of $\mathcal{G}$ are functions $f : \mathcal{N} \mapsto \mathbb{R}$, equivalently represented as vectors $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$, where x_i denotes the signal value at node i. Since $\mathbf{x}$ does not account explicitly for the structure of the graph where the signal is defined, $\mathcal{G}$ is endowed with the graph-shift operator $\mathbf{S}$ [4]. The shift $\mathbf{S} \in \mathbb{R}^{N\times N}$ is a matrix whose entry S_{ij} can be nonzero only if $i = j$ or if $(i,j) \in \mathcal{E}$. The sparsity pattern of $\mathbf{S}$ captures the local structure of $\mathcal{G}$, but we make no specific assumptions on the values of its nonzero entries. The shift $\mathbf{S}$ can also be understood as 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 provided that it has access to the value of x_j at $j \in \mathcal{N}_i$. Typical choices for $\mathbf{S}$ are the adjacency matrix $\mathbf{A}$ [4], the (normalized) Laplacian $\mathbf{L}$ [3], and their respective generalizations [13]. We assume henceforth that $\mathbf{S}$ is symmetric, so that $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ with $\mathbf{\Lambda} \in \mathbb{R}^{N\times N}$ being diagonal, but our results hold for any normal GSO. This spectral decomposition can be leveraged to represent signals in the *frequency domain* as $\widehat{\mathbf{x}} := \mathbf{V}^T\mathbf{x}$.

Graph filters. The shift $\mathbf{S}$ can be used to define linear, shift-invariant graph-signal *operators* of the form

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

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 in $\mathbf{h} := [h_0, \ldots, h_{L-1}]^T$, with $L-1$ denoting the filter degree. Graph filters are of particular interest because they represent linear transformations that can be implemented locally [14], [15]. Moreover, since $\mathbf{H}$ is a polynomial of $\mathbf{S}$, it is diagonalized by $\mathbf{V}$. Hence, it can be written as $\mathbf{H} = \mathbf{V}\text{diag}(\widehat{\mathbf{h}})\mathbf{V}^T$, where $\widehat{\mathbf{h}} \in \mathbb{R}^N$ is the so-called *frequency response* of the filter [4], [15].

Network diffusion processes. Graph filters can be used to model network diffusion processes. Specifically, the signal at node i during the step $(l+1)$ of a linear diffusion process can be written as [14], [16]

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

where α_{ij} are the diffusion coefficients; see e.g., [16]. Leveraging the GSP framework, (2) implies that the graph signal $\mathbf{x}^{(l+1)} = \mathbf{S}\mathbf{x}^{(l)}$ at iteration $l+1$ is the shifted version of $\mathbf{x}^{(l)}$, for a shift $\mathbf{S}$ with entries $S_{ij} = \alpha_{ij}$ if either $i = j$ or $(i,j) \in \mathcal{E}$, and $S_{ij} = 0$ otherwise. For instance, if we set $\mathbf{S} = \mathbf{I} - \beta\mathbf{L}$ and let the signal of interest be $\mathbf{x} := \mathbf{x}^{(\infty)}$, then $\mathbf{x}$ solves the heat diffusion equation. However, more complex diffusion dynamics such as $\mathbf{x} = \Pi_{l=0}^{\infty}(\mathbf{I} - \beta_l\mathbf{S})\mathbf{x}^{(0)}$ and $\mathbf{x} = \sum_{l=0}^{\infty} \gamma_l \mathbf{x}^{(l)} = \sum_{l=0}^{\infty} \gamma_l \mathbf{S}^l\mathbf{x}^{(0)}$, could also be of interest.

The Cayley-Hamilton theorem guarantees that the aforementioned infinite-horizon processes can be equivalently described by a filter of degree N. Accordingly, several works have recognized that the steady-state signal $\mathbf{x}$ generated by a diffusion process can be modeled as the output of a graph filter $\mathbf{H} = \sum_{l=0}^{N-1} h_l\mathbf{S}^l$ with input (seed) $\mathbf{x}^{(0)}$ [14], [16]. This key insight is used next to relate the statistical and spectral properties of $\mathbf{x}$ and $\mathbf{H}$ (hence the graph topology via $\mathbf{S}$).

Spectral templates from diffusion processes. Consider the diffused (steady-state) signal $\mathbf{x}$ given by $\mathbf{x} = \mathbf{H}\mathbf{x}^{(0)}$. Under the assumption that the seed signal $\mathbf{x}^{(0)}$ is white (with identity covariance matrix) and zero-mean, the covariance matrix of the output $\mathbf{C}_\mathbf{x} := \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]$ is given by

$$\mathbf{C}_\mathbf{x} = \mathbf{H}\mathbb{E}\big[\mathbf{x}^{(0)}(\mathbf{x}^{(0)})^T\big]\mathbf{H}^T = \mathbf{H}\mathbf{H}^T = \mathbf{V}\text{diag}(|\widehat{\mathbf{h}}|^2)\mathbf{V}^T, \tag{3}$$

where the last equality leverages the frequency interpretation of a graph filter. Identity (3) reveals that the eigenvectors of the covariance matrix $\mathbf{C}_\mathbf{x}$ and those of $\mathbf{S}$ are the same. Thus, if $\mathbf{C}_\mathbf{x}$ is known, the spectral templates $\mathbf{V}$ can be readily obtained. More commonly, if $\mathbf{C}_\mathbf{x}$ is unknown but we have access to a set of M diffused signals $\{\mathbf{x}_m\}_{m=1}^M$, we may approximate $\mathbf{C}_\mathbf{x}$ with the sample covariance $\hat{\mathbf{C}}_\mathbf{x} = 1/M \sum_{m=1}^M \mathbf{x}_m\mathbf{x}_m^T$ and estimate the eigenvectors of $\mathbf{S}$, also termed *spectral templates*. These will serve as inputs to the algorithms in the ensuing section.[1]

III. Shift inference from spectral templates

Given $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$, our goal is to find a graph shift $\mathbf{S}$ that is diagonalized by $\mathbf{V}$. Since the postulated problem has infinitely many solutions, we further impose conditions on $\mathbf{S}$ promoting desirable properties such as sparsity, or a priori information on the graph of interest such as non-negative edge weights. Note that by definition, $\mathbf{S}$ encodes the local structure of the graph it represents, thus, its recovery implies a successful identification of the graph topology of interest.

A. Sparsest shift from perfect spectral templates

With $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_N]^T$ collecting the unknown eigenvalues of $\mathbf{S}$, we seek to identify $\mathbf{S}$ by solving

$$\min_{\{\mathbf{S},\boldsymbol{\lambda}\}} \ \|\mathbf{S}\|_0 \quad \text{s. to } \mathbf{S} = \textstyle\sum_{k=1}^N \lambda_k \mathbf{v}_k\mathbf{v}_k^T, \ \mathbf{S} \in \mathcal{S}, \tag{4}$$

where the objective function aims at recovering a sparse $\mathbf{S}$. The first constraint encodes the definition of a general graph shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. The second constraint has a twofold motivation: a) incorporating a priori knowledge about $\mathbf{S}$ into $\mathcal{S}$; and b) preventing the trivial solution $\mathbf{S} = \mathbf{0}$. As can be seen from (4), when all eigenvectors $\{\mathbf{v}_k\}_{k=1}^N$ are given the design of $\mathbf{S}$ amounts to finding the N eigenvalues in $\boldsymbol{\lambda}$. If we let $\mathbf{S} = \mathbf{A}$ represent the adjacency matrix of an undirected graph with non-negative weights and no self-loops, we can explicitly write $\mathcal{S}$ as follows, where ij span all values in $\mathcal{N}\times\mathcal{N}$ and $i'j'$ only a subset $\mathcal{I}$ of these,

$$\mathcal{S} := \{\mathbf{S} \,|\, S_{ij} \in [0,1], \ \mathbf{S} \in \mathcal{M}^N, \ S_{ii} = 0, \ S_{i'j'} = \alpha_{i'j'}\}. \tag{5}$$

The first condition in (5) encodes the fact that $\mathbf{S}$ is an adjacency matrix with non-negative weights, thus, every entry of

[1] Diffusion processes are used here as a relevant example that illustrates how to obtain $\mathbf{V}$ without having access to $\mathbf{S}$. Recent GSP results show that there are a number of scenarios where that is also the case, with examples ranging from stationary graph processes [17] to the design of network operators [15].

$\mathbf{S}$ must lie between 0 and an arbitrary normalization constant which we fix to 1. The second condition incorporates the fact that the unknown graph is undirected, hence, $\mathbf{S}$ must belong to the set $\mathcal{M}^N$ of real and symmetric $N \times N$ matrices. The third condition in $\mathcal{S}$ encodes that the graph has no self-loops, thus, each diagonal entry of $\mathbf{S}$ must be null. Finally, the last condition in (5) incorporates potential a priori knowledge on $\mathbf{S}$ by assuming that we know the nonzero values $0 < \alpha_{i'j'} \leq 1$ that $\mathbf{S}$ takes for (possibly multiple) indices $i'j'$. We assume that at least one nonzero entry of $\mathbf{S}$ is known a priori, which conveniently rules out the undesirable solution $\mathbf{S} = \mathbf{0}$ of problem (4). For cases where we have no a priori knowledge of specific entries of $\mathbf{S}$, the trivial solution $\mathbf{S} = \mathbf{0}$ can be discarded by, e.g., imposing $\mathbf{S1} \geq \boldsymbol{\beta}$ which assumes that every node has a minimum degree of connectivity with its neighbors.

An interesting property of the proposed optimization in (4) is that its feasible set is generally small. To be rigorous, some notation must be introduced first. Define the matrix $\mathbf{W} := \mathbf{V} \odot \mathbf{V} \in \mathbb{R}^{N^2 \times N}$, where $\odot$ denotes the Khatri-Rao product. Notice that from the definition of $\mathbf{S}$ we can write $\mathbf{s} := \text{vec}(\mathbf{S})$ as $\mathbf{s} = \mathbf{W}\boldsymbol{\lambda}$. Hence, each row of $\mathbf{W}$ represents the N weighting coefficients that map $\boldsymbol{\lambda}$ to the corresponding entry of $\mathbf{S}$. Further, define the set $\mathcal{D}$ containing the indices of $\mathbf{s}$ corresponding to the diagonal entries of $\mathbf{S}$, and the set $\mathcal{I}$ of indices corresponding to the known entries of $\mathbf{S}$ [cf. (5)]. Those sets can be used to select the corresponding rows of $\mathbf{W}$ to form $\mathbf{W}_{\mathcal{D}} \in \mathbb{R}^{N \times N}$ and $\mathbf{W}_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}| \times N}$. We finally construct $\tilde{\mathbf{W}}$ by vertically concatenating matrix $[\mathbf{0}_N, \mathbf{W}_{\mathcal{D}}]$ and matrix $[\boldsymbol{\alpha}, \mathbf{W}_{\mathcal{I}}]$, where $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{I}|}$ collects the values $\alpha_{i'j'}$ in (5). Using these conventions, the following result holds.

Proposition 1 *Assume that* (4) *is feasible and denote by* Q *the number of singular values of* $\tilde{\mathbf{W}}$ *that are zero, then it holds that:*
a) The nullspace of $\tilde{\mathbf{W}}$ *has dimension at least one, i.e.,* $Q \geq 1$.
b) If $Q = 1$*, the feasible set of* (4) *is a singleton.*

Proof (sketch): The key of the proof resides in noting that, from the definitions of $\mathcal{D}$ and $\mathcal{I}$, we may write $\mathbf{W}_{\mathcal{D}}\boldsymbol{\lambda} = \text{diag}(\mathbf{S}) = \mathbf{0}$ and $\mathbf{W}_{\mathcal{I}}\boldsymbol{\lambda} = \boldsymbol{\alpha}$ for all feasible $\boldsymbol{\lambda}$. Moreover, if we define $\tilde{\boldsymbol{\lambda}} := [-1, \boldsymbol{\lambda}^T]^T$, the previous two equalities imply that $\tilde{\mathbf{W}}\tilde{\boldsymbol{\lambda}} = \mathbf{0}$. Thus, feasibility implies $Q \geq 1$. If $Q = 1$, $\tilde{\boldsymbol{\lambda}}$ is unique up to a scaling factor, however, the unique $\boldsymbol{\lambda}$ (and hence $\mathbf{S}$) is obtained by setting this scaling factor so that the first entry of $\tilde{\boldsymbol{\lambda}}$ is -1. ■

Whenever $Q = 1$, Proposition 1 asserts that the objective in (4) is inconsequential since there exists one and only one feasible $\mathbf{S}$. For more general cases, however, the formulation in (4) recovers the sparsest $\mathbf{S}$ among the feasible ones.

Convex ℓ_1-norm relaxation. Non-convexity of the ℓ_0 norm renders the solution of (4) challenging. Hence, we relax the ℓ_0 norm with the ℓ_1 norm (closest convex approximant), i.e.,

$$\min_{\{\mathbf{S},\boldsymbol{\lambda}\}} \|\mathbf{S}\|_1 \quad \text{s. to } \mathbf{S} = \textstyle\sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \mathbf{S} \in \mathcal{S}. \tag{6}$$

Since $\mathcal{S}$ is a convex set [cf. (5)], problem (6) is convex and can be efficiently solved. More importantly, under certain conditions we can ensure that the solution $\mathbf{S}_1^*$ to the relaxed problem (6) coincides with the solution $\mathbf{S}_0^*$ to the original problem (4). To be more specific, denote by $\mathbf{Q}$ an orthogonal matrix whose columns span $\ker(\mathbf{W}_{\mathcal{D}})$ and define $\boldsymbol{\Psi} := \mathbf{WQ}$. Further, define the set $\mathcal{J}$ containing the indices identifying the support of $\mathbf{s}_0^* := \text{vec}(\mathbf{S}_0^*)$ and denote by $\mathcal{J}^c$ its complement. Recalling that matrix subscripts select rows, the following recovery result holds.[2]

Theorem 1 *Problem* (6) *recovers the sparsest solution [cf.* (4)*] if the two following conditions are satisfied:*
1) $\text{rank}(\mathbf{W}_{\mathcal{J}^c}) = N - 1$*; and*
2) There exists a constant $\delta > 0$ *such that*

$$\|\boldsymbol{\Psi}_{\mathcal{J}^c}(\delta^{-2}\boldsymbol{\Psi}_{\mathcal{I}}^T\boldsymbol{\Psi}_{\mathcal{I}} + \boldsymbol{\Psi}_{\mathcal{J}^c}^T\boldsymbol{\Psi}_{\mathcal{J}^c})^{-1}\boldsymbol{\Psi}_{\mathcal{J}}^T\|_\infty < 1. \tag{7}$$

The theorem states *sufficient* conditions under which the relaxation is tight. Simulations in Section IV reveal that the bound in condition 2) is tight by providing examples where the ℓ_∞ norm is exactly one and for which recovery fails. The proof of Theorem 1 builds on recent advances for solving ℓ_1-analysis problems [18]. Condition 1) ensures that the solution to (6) is unique, a necessary requirement to guarantee sparse recovery. Condition 2) is derived from the construction of a dual certificate specially designed to ensure that the unique solution to (6) also has minimum ℓ_0 norm.

Remark 1 (General shifts) The proposed formulation can be broadened to accommodate shifts $\mathbf{S}$ other than the adjacency matrix. To that end, one can modify the set $\mathcal{S}$ in (4) so that it accounts for the properties of the particular GSO. Consider for example a scenario where $\mathbf{S}$ represents a normalized Laplacian [3]. The associated $\mathcal{S}$ would then be

$$\mathcal{S}_{\text{Lap}} := \{\mathbf{S} \,|\, S_{ij} \in [-1, 0] \text{ for } i \neq j, \ \mathbf{S} \in \mathcal{M}_+^N, \ S_{ii} = 1 \text{ for all } i, \ S_{i'j'} = \alpha_{i'j'}, \ i'j' \in \mathcal{I}\}. \tag{8}$$

In (8) we encode the facts that $\mathbf{S}$ is symmetric and positive semi-definite, that its diagonal entries are 1 and its off-diagonal entries are non-positive. Moreover, since $\mathbf{S}$ is a normalized Laplacian we know that the vector $\sqrt{\mathbf{d}}$ containing as entries the square roots of the degrees of the graph is an eigenvector whose eigenvalue is zero, and this can be incorporated as a constraint. The eigenvector $\sqrt{\mathbf{d}}$ can be easily identified since it is the only one whose entries have all the same sign [19]. Finally, properties other than sparsity can be sought by modifying the objective function in the proposed formulations, such as obtaining Laplacians with good mixing conditions by maximizing their second smallest eigenvalue [20].

B. Shift inference from noisy spectral templates

Oftentimes the spectral templates $\mathbf{V}$ are not known perfectly but, rather, only a noisy version $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N]$ of them is available. This occurs if, e.g., the templates are obtained as the eigenvectors of the *sample* covariance of a diffusion process.

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

The question then is how to update the formulation in (4) – or its convex relaxation in (6) – to accommodate for the discrepancies between the estimated spectral templates and the actual eigenvectors of $\mathbf{S}$. A possible reformulation is to include $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ as decision variables and postulate the following optimization problem

$$\min_{\{\mathbf{S},\boldsymbol{\lambda},\mathbf{V}\}} \ \|\mathbf{S}\|_1 \tag{9}$$
$$\text{s. to } \ \mathbf{S} = \textstyle\sum_{k=1}^N \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \ \mathbf{S} \in \mathcal{S}, \ \ \max_k d(\mathbf{v}_k, \hat{\mathbf{v}}_k) \leq \epsilon,$$

where we directly stated the relaxation with the ℓ_1 norm in the objective function and $d(\cdot, \cdot)$ is a *convex* vector distance function, such as the ℓ_p norm of the vector difference for $p \geq 1$. The idea in (9) is to find a sparse $\mathbf{S}$ that satisfies the desired properties in $\mathcal{S}$ while its eigenvectors $\mathbf{v}_k$ are each of them close to the observed ones $\hat{\mathbf{v}}_k$. This problem is more challenging than its noiseless counterpart (6) since the first constraint is non-convex given that both λ_k and $\mathbf{v}_k$ are optimization variables. A more tractable alternative is to form $\hat{\mathbf{S}} := \sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T$ and look for shifts $\mathbf{S}$ that possess the desired properties while being close to $\hat{\mathbf{S}}$. Formally,

$$\min_{\{\mathbf{S},\boldsymbol{\lambda},\hat{\mathbf{S}}\}} \ \|\mathbf{S}\|_1 \tag{10}$$
$$\text{s. to } \ \hat{\mathbf{S}} = \textstyle\sum_{k=1}^N \lambda_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k^T, \ \ \mathbf{S} \in \mathcal{S}, \ \ d(\mathbf{S}, \hat{\mathbf{S}}) \leq \epsilon,$$

where $d(\cdot, \cdot)$ is a *convex* matrix distance function.

C. Shift inference from incomplete spectral templates

Our last contribution is to study scenarios where only some of the eigenvectors are available, e.g., if $\mathbf{V}$ is found from the sample covariance of *bandlimited* graph signals. Supposing that the K first eigenvectors $\mathbf{V}_K = [\mathbf{v}_1, ..., \mathbf{v}_K]$ are those which are known, the goal in this case is to solve [cf. (4)]

$$\min_{\{\mathbf{S},\mathbf{S}_{\bar{K}},\boldsymbol{\lambda}\}} \ \|\mathbf{S}\|_0, \quad \text{s. to} \quad \mathbf{S} \in \mathcal{S}, \tag{11}$$
$$\mathbf{S} = \mathbf{S}_{\bar{K}} + \textstyle\sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \ \mathbf{S}_{\bar{K}} \in \mathcal{M}_K^N, \ \ \mathbf{S}_{\bar{K}} \mathbf{v}_k = \mathbf{0},$$

where $\mathcal{M}_K^N$ is the set of symmetric $N \times N$ matrices with rank at most $N - K$. In addition to requiring $\mathbf{S}_{\bar{K}} \in \mathcal{M}_K^N$, the last constraint in (11) enforces the given spectral templates to lie in the null space of $\mathbf{S}_{\bar{K}}$. Our approach to deal with the constraint $\mathbf{S}_{\bar{K}} \in \mathcal{M}_K^N$ is to add to the objective a nuclear-norm regularizer $\eta \|\mathbf{S}_{\bar{K}}\|_*$, with higher values of η promoting solutions with lower rank. This leads to the convex problem

$$\min_{\{\mathbf{S},\mathbf{S}_{\bar{K}},\boldsymbol{\lambda}\}} \ \eta \|\mathbf{S}_{\bar{K}}\|_* + \|\mathbf{S}\|_1 \quad \text{s. to} \quad \mathbf{S} \in \mathcal{S}, \tag{12}$$
$$\mathbf{S} = \mathbf{S}_{\bar{K}} + \textstyle\sum_{k=1}^K \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \ \mathbf{S}_{\bar{K}} \mathbf{v}_k = \mathbf{0}.$$

Noisy and incomplete spectral templates can be handled by combining the formulations in (10) and (12). In practice, the recovery performance of the convex relaxations presented in (6), (10), and (12) can be improved by replacing the ℓ_1 and nuclear norms with their iteratively reweighted versions [21].

IV. Numerical experiments

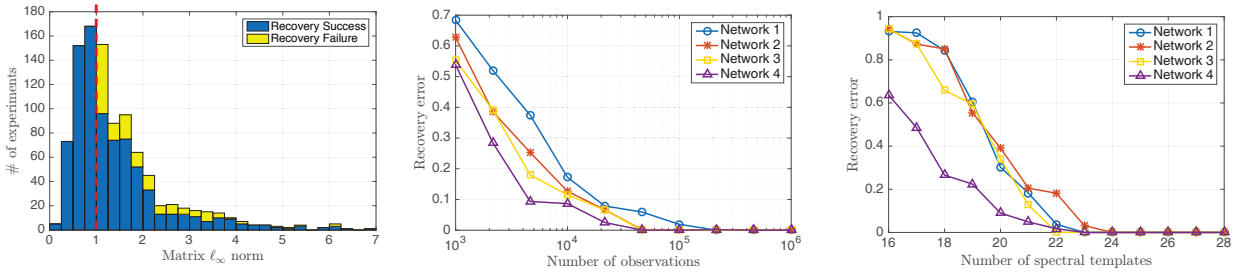
Through synthetic and real-world graphs, we validate our theoretical claims and illustrate the performance when inferring topologies given noisy and incomplete spectral templates.

Conditions for noiseless recovery. We draw Erdős-Rényi random graphs [22] of size $N = 20$ and edge-formation probability $p = 0.1$, and try to recover their adjacency matrices $\mathbf{A}$ from the corresponding spectral templates $\mathbf{V}$ by solving (6). We assume that the position of *one* random edge is known a priori, i.e., $|\mathcal{I}| = 1$. For each graph, we make sure that the associated $\mathbf{W}$ matrix (cf. Proposition 1) has a nullspace of dimension greater than 1 – to rule out the cases where the feasible set is a singleton –, and that the first condition in Theorem 1 is satisfied. In Fig. 1(a) we plot the number of successes and failures in recovering the adjacency as a function of the ℓ_∞ norm in (7). The number of realizations is 1000, and for each of them the constant δ in (7) is chosen to minimize the ℓ_∞ norm. Fig. 1(a) clearly depicts the result of Theorem 1 in that, for all cases in which the norm of interest is smaller than one, relaxation (6) achieves perfect recovery. Equally important, from the figure it is clear that the bound stated in (7) is tight since a large proportion of the realization with norms exactly equal to 1 or just above this value led to failed recoveries.

Inference of social graphs. Consider four social networks defined on a common set of nodes representing 32 students from the University of Ljubljana in Slovenia. Each network represents different types and levels of interactions among the students, and were built by asking each student to select a group of preferred college mates for different situations, e.g., to discuss a personal issue or to invite to a birthday party[3]. The considered graphs are unweighted and symmetric, and the edge between i and j exists if either student i picked j in the questionnaire or vice versa. We test the recovery performance of formulation (10) for noisy spectral templates $\hat{\mathbf{V}}$ obtained from sample covariances of signals generated through diffusion processes (cf. Section II). Fig. 1(b) plots the reconstruction error (averaged over 50 realizations) as a function of the number of observed signals for the different networks studied. We quantify the error as the proportion of edges misidentified, i.e. $\|\mathbf{A}_i - \hat{\mathbf{A}}_i\|_0 / \|\mathbf{A}_i\|_0$, where $\mathbf{A}_i$ and $\hat{\mathbf{A}}_i$ are the real and estimated adjacency matrices of graph i. Notice that for an increasing number of observed signals we see a monotonous decrease in recovery error. For instance, when going from 10^3 to 10^4 observations the error averaged across the networks is (approximately) divided by five. This is expected since a larger number of observations entails a more reliable estimate of the covariance matrix leading to less noisy versions of the spectral templates.

Traditional methods like graphical lasso [9] fail to recover $\mathbf{S}$ from the sample covariance of filtered white signals. This happens because the filter $\mathbf{H}$ introduces conditional dependence between signal values *more than one hop* apart. For instance, based on 10^5 observations, the recovery error of

[3]Access to the data and additional details are available in http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:pajek:students



(a) (b) (c)

Fig. 1. (a) Experimental validation of Theorem 1. For every realization where the ℓ_∞ norm in (7) is strictly less than 1, perfect recovery is achieved. (b) Recovery error for four social networks as a function of the number of signals observed in the estimation of the spectral templates. (c) Recovery error for four social networks (with $N = 32$ nodes) as a function of K, the number of spectral templates that are known.

graphical lasso with optimal tuning parameters and averaged over 50 realizations is 0.548, 0.348, 0.313, and 0.429 for the four networks studied.

Finally, we illustrate the recovery performance in the presence of incomplete spectral templates by solving (12) for the aforementioned four networks. In Fig. 1(c) we plot the recovery error as a function of the number K of eigenvectors of $\mathbf{S}$ that are given. Each point in the plot is the average across 50 realizations in which different K eigenvectors were chosen as given from the $N = 32$ possible ones. As expected, performance for all networks increases with the number of spectral templates known. The performance improvement is sharp and precipitous going from a large error of over 0.85 for three of the networks when 17 spectral templates are known to a perfect recovery for all the networks when 24 eigenvectors are given. Moreover, notice that network 4 is consistently the easiest to identify both for noisy [cf. Fig. 1(b)] and incomplete [cf. Fig. 1(c)] spectral templates. For example, when given 19 spectral templates the error associated with network 4 is 0.224 whereas the average across the other three networks is 0.584. This hints towards the fact that some graphs are inherently more robust for identification when given noisy or incomplete spectral templates. A formal analysis of this phenomenon is left as future work.

V. Conclusions

The problem of identifying a graph-shift operator $\mathbf{S}$ – encoding the topology of a graph $\mathcal{G}$ of interest – given its eigenbasis $\mathbf{V}$ was studied. Three setups were investigated corresponding to varying knowledge about $\mathbf{V}$, namely: 1) perfect knowledge of $\mathbf{V}$; 2) access to a noisy version of $\mathbf{V}$; and 3) access to a subset of the columns of $\mathbf{V}$. We formulated optimization problems to recover $\mathbf{S}$, presented convex relaxations, and derived theoretical results characterizing the recovery. Finally, we illustrated the performance of the proposed approach via the identification of real-world social graphs.

References

[1] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamic Processes on Complex Networks*, Cambridge University Press, Cambridge, UK, 2012.

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

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

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

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

[6] J. Mei and J. M. F. Moura, "Signal processing on graphs: Modeling (causal) relations in big data," *arXiv preprint arXiv:1503.00173v1 [cs.IT]*, 2015.

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

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

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

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

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

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

[13] C. Godsil and G. Royle, *Algebraic Graph Theory*, vol. 207, Springer-Verlag, Graduate Texts in Mathematics, 2001.

[14] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *arXiv preprint arXiv:1507.08364*, 2015.

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

[16] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters," *arXiv preprint arXiv:1604.07234*, 2016.

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

[18] H. Zhang, M. Yan, and W. Yin, "One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimizations," *arXiv preprint arXiv:1304.5038*, 2013.

[19] T. Biyikoğu, J. Leydold, and P. Stadler, *Laplacian Eigenvectors of Graphs: Perron-Frobenius and Faber-Krahn Type Theorems*, Springer-Verlag Berlin Heidelberg, 2007.

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

[21] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, Dec. 2008.

[22] B. Bollobás, *Random Graphs*, Springer, 1998.