IDENTIFYING UNDIRECTED NETWORK STRUCTURE VIA SEMIDEFINITE RELAXATION

Rasoul Shafipour[‡], Santiago Segarra[†], Antonio G. Marques[*] and Gonzalo Mateos[‡]

[‡]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA
[†]Inst. for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA
[*]Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain

ABSTRACT

We address the problem of *inferring* an undirected *graph* from nodal observations, which are modeled as *non-stationary* graph signals generated by local diffusion dynamics on the unknown network. We propose a two-step approach where we first estimate the unknown *diffusion* (graph) *filter*, from which we recover the eigenvectors of the so-called graph-shift operator (a matrix representation of the graph). We then estimate the eigenvalues by imposing desirable properties on the graph to be recovered. To carry out the initial system identification step, we assume that second-order statistics of the inputs are available. While such quadratic filter identification problem boils down to a non-convex fourth order polynomial minimization, we propose a semidefinite relaxation with provable performance guarantees. Finally, numerical tests illustrate the use of the proposed algorithm to unveil urban mobility patterns.

***Index Terms*—** Network topology inference, graph signal processing, diffusion process, semidefinite relaxation.

1. INTRODUCTION

Consider a network represented as a weighted and undirected graph $\mathcal{G}$, consisting of a node set $\mathcal{N}$ of cardinality N, an edge set $\mathcal{E}$ of unordered pairs of elements in $\mathcal{N}$, and edge weights $A_{ij} \in \mathbb{R}$ such that $A_{ij} = A_{ji} \neq 0$ for all $(i,j) \in \mathcal{E}$. The edge weights A_{ij} are collected in the *symmetric* adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$. More broadly, one can define a generic *graph-shift operator* (GSO) $\mathbf{S} \in \mathbb{R}^{N\times N}$ as any matrix having the same sparsity pattern than that of $\mathcal{G}$ [1]. Although the choice of $\mathbf{S}$ can be adapted to the problem at hand, it is often chosen as either $\mathbf{A}$, the Laplacian $\mathbf{L} := \text{diag}(\mathbf{A1}) - \mathbf{A}$, or its normalized counterparts [2].

Our focus in this paper is on identifying graphs that explain the structure of a random signal. Formally, let $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$ be a zero-mean graph signal with covariance matrix $\mathbf{C_x} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]$, in which the ith element x_i denotes the signal value at node i of an *unknown graph* $\mathcal{G}$ with shift operator $\mathbf{S}$. We say that the graph $\mathbf{S}$ represents the structure of the signal $\mathbf{y} \in \mathbb{R}^N$ if there exists a diffusion process in the GSO $\mathbf{S}$ that produces the signal $\mathbf{y}$ from the input signal $\mathbf{x}$, that is

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

When $\mathbf{x}$ is white so that $\mathbf{C_x} = \mathbf{I}$, (1) is equivalent to saying that the graph process $\mathbf{y}$ is *stationary* in $\mathbf{S}$; see e.g., [3, Def. 1], [4], [5] and Section 2.1 for further details. Here though, we deal with more general non-stationary signals $\mathbf{y}$ that adhere to linear diffusion dynamics as in (1), but where the input covariance $\mathbf{C_x}$ can be arbitrary.

Work in this paper was supported by the NSF award CCF-1750428, the Spanish MINECO grants OMICRON (TEC2013-41604-R) and KLINILYCS (TEC2016-75361-R), and MIT IDSS seed fund. Emails: rshafipo@ur.rochester.edu, segarra@mit.edu, antonio.garcia.marques@urjc.es, gmateosb@ece.rochester.edu.

The justification to say that $\mathbf{S}$ represents the structure of $\mathbf{y}$ is that we can think of the edges of $\mathcal{G}$, i.e. the non-zero entries in $\mathbf{S}$, as direct (one-hop) relations between the elements of the signal. The diffusion described by (1) modifies the original correlation by inducing indirect (multi-hop) relations. In this context, our goal is to recover the fundamental relations dictated by $\mathbf{S}$ from a set of independent samples of a non-stationary random signal $\mathbf{y}$, as well as knowledge of $\mathbf{C_x}$. This additional information on the input $\mathbf{x}$ is the price paid to accommodate the more general non-stationary generative models for $\mathbf{y}$, and is not needed when identifying the structure of stationary graph signals [6, 7].

Relation to prior work. Workhorse topology inference approaches construct graphs whose edge weights correspond to nontrivial correlations between signals at incident nodes [8, 9]. Acknowledging that the observed correlations can be due to latent network effects, alternative statistical methods rely on inference of partial correlations [8, Ch. 7.3.2]. Under Gaussianity assumptions, this line of work has well-documented connections with sparse precision matrix estimation [10–13] as well as high-dimensional sparse linear regression [14]. Extensions to directed graphs include structural equation models (SEMs) [15–17], Granger causality [9,18], or their nonlinear (e.g., kernelized) variants [19, 20]. Recent graph signal processing (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 a graph [6,7,21–24]. Different from [21,23,25,26] that infer structure from signals assumed to be smooth over the sought graph, here the measurements are assumed to be related to the graph via linear filtering as in (1). Two works have recently explored this approach by identifying a symmetric GSO given its eigenvectors [6, 7], but both rely on observations of stationary signals.

Paper outline and contributions. In Section 2 we formulate the problem of identifying a GSO that explains the fundamental structure of a random signal diffused on a graph. To solve this problem, we propose a two-step approach whereby we: i) identify the GSO's eigenbasis from a judicious *graph filter estimate*; and ii) rely on these *spectral templates* to estimate the GSO's eigenvalues such that the inferred graph is sparse. This second step is discussed in Section 2.2. The estimation of the diffusion filter in step i), which is not required when the signals are stationary [6], is analyzed in Section 3 and is one of the main contributions of the paper. More specifically, we focus on scenarios where second-order statistical information of the inputs is available. In this setting, filter identification boils down to a non-convex fourth-order polynomial minimization problem. In Section 4 we develop an efficient algorithm based on a semidefinite relaxation with provable performance guarantees to deal with this non-convex optimization problem. Numerical tests in Section 5 illustrate the application of the proposed method to reveal urban mobility patterns in New York City. Additional comprehensive performance evaluations as well as comparisons to other methods can be found in [27]. Concluding remarks are given in Section 6.

2. PROBLEM STATEMENT

To formally state the topology inference problem, recall the symmetric GSO $\mathbf{S}$ associated with the undirected graph $\mathcal{G}$. Upon defining the vector of coefficients $\mathbf{h} := [h_0, \ldots, h_{L-1}]^T \in \mathbb{R}^L$ and the *symmetric* graph filter $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l \in \mathbb{R}^{N\times N}$ [1], by virtue of the Cayley-Hamilton theorem the model in (1) can be rewritten as

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

for some particular $\mathbf{h}$ and $L \leq N$. Given a set $\mathcal{Y} := \{\mathbf{y}^{(p)}\}_{p=1}^P$ of P independent samples of a non-stationary random signal $\mathbf{y}$ adhering to the network diffusion model (2), the problem is to identify the GSO $\mathbf{S}$ that is optimal in some sense as described in Section 2.2.

Because $\mathbf{S}$ is symmetric it is diagonalizable. Accordingly, define the eigenvector matrix $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ and the diagonal eigenvalue matrix $\mathbf{\Lambda}$ to write $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. Fundamental to the topology inference approach developed here is to note that because $\mathbf{H}$ is a polynomial on $\mathbf{S}$, then: i) all such graph filters (spanned by the unknown coefficients $\mathbf{h}$) have the *same eigenvectors*; and ii) these eigenvectors are the same as those of the shift, namely $\mathbf{V}$. In other words, while the diffusion implicit in $\mathbf{H}$ obscures the eigenvalues of the GSO, the eigenvectors $\mathbf{V}$ are preserved as *spectral templates* of the underlying network topology.

Section 2.1 describes how to leverage (2) to obtain the GSO eigenbasis from a set of nodal observations $\mathcal{Y}$, by first estimating the unknown graph filter $\mathbf{H}$. We show that the information in $\mathcal{Y}$ is in general not enough to uniquely recover $\mathbf{H}$. Hence, we will resort to prior statistical knowledge on (possibly multiple) input signals $\mathbf{x}$ to aid identifiability. Section 2.2 outlines how to use the spectral templates $\mathbf{V}$ to recover the desired GSO by estimating its eigenvalues $\mathbf{\Lambda}$ and, as byproduct, the graph shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ itself.

2.1. Challenges facing non-stationary observations

Consider estimating the eigenbasis $\mathbf{V}$ of the filter $\mathbf{H}$ that governs the diffusion in (2). To gain insights, suppose first that $\mathbf{x}$ is white so that $\mathbf{C}_\mathbf{x} = \mathbf{I}$ and the covariance matrix of $\mathbf{y} = \mathbf{H}\mathbf{x}$ is given by

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

Using the spectral decomposition of $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ to express the filter in the graph frequency domain as $\mathbf{H} = \sum_{l=0}^{L-1} h_l(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l\mathbf{\Lambda}^l)\mathbf{V}^T$, we can diagonalize the covariance matrix as

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

Such a covariance expression is precisely the requirement for a graph signal to be stationary in $\mathbf{S}$ [3, Def. 2.b)]. In this setting, (4) shows that the *eigenvectors* of the shift $\mathbf{S}$, the filter $\mathbf{H}$, and the covariance $\mathbf{C}_\mathbf{y}$ coincide. As a result, to estimate $\mathbf{V}$ from the observations $\mathcal{Y} = \{\mathbf{y}^{(p)}\}_{p=1}^P$ it suffices to form the sample covariance $\hat{\mathbf{C}}_\mathbf{y} = \frac{1}{P}\sum_{p=1}^P \mathbf{y}^{(p)}(\mathbf{y}^{(p)})^T$ and use its eigenvectors as spectral templates to recover $\mathbf{S}$ [6, 7]; see also Section 2.2.

The broader focus of the present paper is on identifying the GSO $\mathbf{S}$ that is considered to be the best possible description of the structure of a *non-stationary* signal $\mathbf{y} = \mathbf{H}\mathbf{x}$ [cf. (2), where $\mathbf{x}$ is not white]. For generic (non-identity) input covariance matrix $\mathbf{C}_\mathbf{x}$, we face the challenge that the signal covariance [cf. (3)]

$$\mathbf{C}_\mathbf{y} = \mathbf{H}\mathbf{C}_\mathbf{x}\mathbf{H} \tag{5}$$

is no longer simultaneously diagonalizable with $\mathbf{S}$. This rules out using the eigenvectors of the sample covariance $\hat{\mathbf{C}}_\mathbf{y}$ as spectral templates of $\mathbf{S}$. Still, as argued following (2) the eigenvectors of the GSO coincide with those of $\mathbf{H}$. This motivates using realizations of observed signals together with prior statistical information on the excitation inputs $\mathbf{x}$ to *identify the filter* $\mathbf{H}$, with the ultimate goal of estimating its eigenvectors $\mathbf{V}$. This system identification task in the graph setting is the subject of Section 3, but before we close the loop showing how to recover a sparse $\mathbf{S}$ given its estimated eigenbasis $\hat{\mathbf{V}}$.

2.2. Using the spectral templates to recover a sparse shift

Given estimates $\hat{\mathbf{V}}$ of the filter eigenvectors, recovery of $\mathbf{S}$ amounts to selecting its eigenvalues $\mathbf{\Lambda}$ and to that end we assume that the shift of interest is optimal in some sense. At the same time, we should account for the discrepancies between $\hat{\mathbf{V}}$ and the actual eigenvectors of $\mathbf{S}$, due to finite sample size constraints and unavoidable errors in estimating the filter. Accordingly, we build on [6] and seek for the shift operator $\mathbf{S}$ that: (a) is sparse, meaning that few edge weights are non-zero; (b) belongs to a convex set $\mathcal{S}$ that specifies the desired type of shift operator (e.g., the adjacency $\mathbf{A}$ or Laplacian $\mathbf{L}$); and (c) is close to $\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T$ in the Frobenius-norm sense. One can thus solve

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

which is a convex optimization problem for the choice of a sparsity-promoting ℓ_1-norm criterion, and ϵ is a tuning parameter chosen based on a priori information on the imperfections.

The constraint $\mathbf{S} \in \mathcal{S}$ in (6) incorporates a priori knowledge about $\mathbf{S}$. If we let $\mathbf{S} = \mathbf{A}$ represent the adjacency matrix of an undirected graph with non-negative weights and no self-loops, we can explicitly write $\mathcal{S} = \mathcal{S}_\text{A}$ as follows

$$\mathcal{S}_\text{A} := \{\mathbf{S} \,|\, S_{ij} \geq 0,\ \mathbf{S}\in\mathcal{M}_N,\ S_{ii} = 0,\ \textstyle\sum_j S_{j1} = 1\}. \tag{7}$$

The first condition in $\mathcal{S}_\text{A}$ encodes the non-negativity of the weights whereas the second condition incorporates that $\mathcal{G}$ is undirected, hence, $\mathbf{S}$ must belong to the set $\mathcal{M}_N$ of real and symmetric $N\times N$ matrices. The third condition encodes the absence of self-loops, thus, each diagonal entry of $\mathbf{S}$ must be null. Finally, the last condition fixes the scale of the admissible graphs by setting the weighted degree of the first node to 1, and rules out the trivial solution $\mathbf{S} = \mathbf{0}$. Other GSOs (e.g., the Laplacian $\mathbf{L}$ and its normalized variants) can be accommodated via minor modifications to $\mathcal{S}$; see [6].

3. QUADRATIC GRAPH FILTER IDENTIFICATION

Consider $m = 1, \ldots, M$ diffusion processes on $\mathcal{G}$, and assume that the observed non-stationary signal $\mathbf{y}_m$ corresponds to an input $\mathbf{x}_m$ diffused by an unknown graph filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l\mathbf{S}^l$. While realizations of the excitation input process $\mathbf{x}_m$ may be challenging to acquire, information about the *statistical* description of $\mathbf{x}_m$ could still be available and used to estimate $\mathbf{H}$. Specifically, assume that the excitation input processes are zero mean and their covariance $\mathbf{C}_{\mathbf{x},m} = \mathbb{E}[\mathbf{x}_m\mathbf{x}_m^T]$ is known. Further suppose that for each input $\mathbf{x}_m$, we have access to a set of output observations $\mathcal{Y}_m := \{\mathbf{y}_m^{(p)}\}_{p=1}^{P_m}$, which are then used to estimate the output covariance $\hat{\mathbf{C}}_{\mathbf{y},m}$ via sample averaging. Since under (2) the *ensemble covariance* is $\mathbf{C}_{\mathbf{y},m} = \mathbb{E}[\mathbf{y}_m\mathbf{y}_m^T] = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}$ [cf. (5)], the aim is to identify a filter $\mathbf{H}$ such that matrices $\hat{\mathbf{C}}_{\mathbf{y},m}$ and $\mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}$ are close.

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

$$\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}, \quad m = 1, \ldots, M. \tag{8}$$

It is insightful to consider first the case where $M = 1$ and henceforth we drop the subindex m so that we can write (8) as (5). Given

the eigendecomposition of the symmetric and positive semidefinite (PSD) covariance matrix $\mathbf{C_y} = \mathbf{V_y}\mathbf{\Lambda_y}\mathbf{V_y}^T$, the *principal square root* of $\mathbf{C_y}$ is the unique symmetric and PSD matrix $\mathbf{C_y}^{1/2}$ which satisfies $\mathbf{C_y} = \mathbf{C_y}^{1/2}\mathbf{C_y}^{1/2}$. It is given by $\mathbf{C_y}^{1/2} = \mathbf{V_y}\mathbf{\Lambda_y}^{1/2}\mathbf{V_y}^T$, where $\mathbf{\Lambda_y}^{1/2}$ stands for a diagonal matrix with the nonnegative square roots of the eigenvalues of $\mathbf{C_y}$. With this notation in place, the solutions of the matrix quadratic equation $\mathbf{C_y} = \mathbf{H}\mathbf{C_x}\mathbf{H}$ in (5) are specified in the following proposition (the proof is omitted due to lack of space, but the details can be found in [27]).

Proposition 1 *Introduce the symmetric and PSD matrix* $\mathbf{C_{xyx}} := \mathbf{C_x}^{1/2}\mathbf{C_y}\mathbf{C_x}^{1/2}$*, whose eigenvectors are denoted by* $\mathbf{V_{xyx}}$*. If the input covariance matrix* $\mathbf{C_x}$ *is nonsingular, all symmetric solutions* $\mathbf{H} \in \mathcal{M}_N$ *of* (5) *are given by*

$$\mathbf{H} = \mathbf{C_x}^{-1/2}\mathbf{C_{xyx}}^{1/2}\mathbf{V_{xyx}}\text{diag}(\mathbf{b})\mathbf{V_{xyx}}^T\mathbf{C_x}^{-1/2}, \tag{9}$$

where $\mathbf{b} \in \{-1,1\}^N$ *is a binary (signed) vector.*

Inspection of (9) shows that the filter is non-identifiable for $M = 1$. Indeed, there are 2^N possible solutions to the quadratic equation (5), which are parametrized by the binary vector $\mathbf{b}$.

For $M > 1$, the set of feasible solutions to the system of equations (8) is naturally given by

$$\mathcal{H}_M = \bigcap_{m=1}^{M} \Big\{\mathbf{H} \in \mathcal{M}_N \mid \mathbf{b}_m \in \{-1,1\}^N \text{ and } \mathbf{H} = \mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{C}_{\mathbf{xyx},m}^{1/2}\mathbf{V}_{\mathbf{xyx},m}\text{diag}(\mathbf{b}_m)\mathbf{V}_{\mathbf{xyx},m}^T\mathbf{C}_{\mathbf{x},m}^{-1/2}\Big\}. \tag{10}$$

It is thus conceivable that as M grows and, therefore, the number of equations increases, the cardinality of $\mathcal{H}_M$ shrinks and the problem is rendered identifiable (up to an unavoidable sign ambiguity because if $\mathbf{H} \in \mathcal{M}_N$ is a solution of (8), so is $-\mathbf{H}$). Next, we show that with as few as $M = 2$ excitation inputs having covariances $\mathbf{C}_{\mathbf{x},1}$ and $\mathbf{C}_{\mathbf{x},2}$ with *identical* eigenvectors, uniqueness can be attained as long as their eigenvalues are sufficiently different (see [27] for a proof).

Proposition 2 *Consider the system of equations* (8) *for* $M = 2$ *and suppose that* $\mathbf{C}_{\mathbf{x},1} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}_1)\mathbf{U}^T$ *and* $\mathbf{C}_{\mathbf{x},2} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}_2)\mathbf{U}^T$*. Then* (8) *has a unique symmetric solution, i.e.,* $\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ *is identifiable up to a sign ambiguity if the following conditions hold:*
C-1) All eigenvalues in $\boldsymbol{\lambda}_1$ *are distinct;*
C-2) $\lambda_{1,i}\,\lambda_{2,j} \neq \lambda_{1,j}\,\lambda_{2,i}$ *for all* i, j*;*
C-3) $\mathbf{V}$ *and* $\mathbf{U}$ *do not share any eigenvector; and*
C-4) $\text{rank}(\mathbf{H}) = N$.

Conditions C-1) and C-2) encode a notion of richness on the input signals. In fact, condition C-2) is the specialization for $M = 2$ of a requirement on the Kruskal rank of a matrix related to the eigenvalues of the excitation processes. From this vantage point, it follows that larger values of M facilitate the fulfillment of this requirement, corroborating our initial intuition regarding uniqueness of $\mathbf{H}$. Moreover, from the proof arguments it follows that symmetry of $\mathbf{H}$ is essential [27]. Actually, if one lifts the symmetry assumption and all input covariances have the same eigenvectors, the problem remains non-identifiable for any M.

4. SEMIDEFINITE RELAXATION

Here we show that the graph filter identification task can be tackled using semidefinite relaxation (SDR) [28], a convexification technique which has been successfully applied to a wide variety of nonconvex quadratically-constrained quadratic programs (QCQP) in applications such as MIMO detection [29] and transmit beamforming [30]. To that end, we first cast the problem of identifying $\mathbf{H}$ – given by the simultaneous fulfillment of the M equalities in (10) – as a Boolean quadratic program (BQP). With $\odot$ denoting the Khatri-Rao product, the ensuing result follows.

Proposition 3 *For* $m = 1, \ldots, M$ *form* $\mathbf{A}_m := (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{V}_{\mathbf{xyx},m}) \odot (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{C}_{\mathbf{xyx},m}^{1/2}\mathbf{V}_{\mathbf{xyx},m}) \in \mathbb{R}^{N^2 \times N}$ *and consider the unknown binary vectors* $\mathbf{b}_m \in \{-1,1\}^N$*. Define* $\mathbf{\Psi} \in \mathbb{R}^{N^2(M-1)\times NM}$ *as*

$$\mathbf{\Psi} := \begin{bmatrix} \mathbf{A}_1 & -\mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & -\mathbf{A}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{M-1} & -\mathbf{A}_M \end{bmatrix} \tag{11}$$

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

$$\mathbf{b}^* = \underset{\mathbf{b}\in\{-1,1\}^{NM}}{\text{argmin}}\ \mathbf{b}^T\mathbf{\Psi}^T\mathbf{\Psi}\mathbf{b}. \tag{12}$$

Problem (12) offers a natural formulation for the setting whereby $\{\mathbf{C}_{\mathbf{y},m}\}_{m=1}^M$ are replaced by sample estimates, and one aims at minimizing the residuals $\sum_{m=1}^{M-1}\|\hat{\mathbf{A}}_m\mathbf{b}_m - \hat{\mathbf{A}}_{m+1}\mathbf{b}_{m+1}\|^2 = \mathbf{b}^T\hat{\mathbf{\Psi}}^T\hat{\mathbf{\Psi}}\mathbf{b}$ in the least-squares sense. Given a solution of (12) with $\hat{\mathbf{\Psi}}$ replacing $\mathbf{\Psi}$, $\mathbf{H} \in \mathcal{M}_N$ can be estimated as [cf. (9)]

$$\hat{\mathbf{H}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{C}_{\mathbf{x},m}^{-1/2}\hat{\mathbf{C}}_{\mathbf{xyx},m}^{1/2}\hat{\mathbf{V}}_{\mathbf{xyx},m}\text{diag}(\mathbf{b}_m^*)\hat{\mathbf{V}}_{\mathbf{xyx},m}^T\mathbf{C}_{\mathbf{x},m}^{-1/2}. \tag{13}$$

Even though the BQP is a classical NP-hard combinatorial optimization problem [28], via SDR one can obtain near-optimal solutions with provable approximation guarantees. To derive the SDR of (12), first introduce the $NM \times NM$ symmetric PSD matrices $\mathbf{W} := \mathbf{\Psi}^T\mathbf{\Psi}$ and $\mathbf{B} := \mathbf{b}\mathbf{b}^T$. By construction, the binary matrix $\mathbf{B}$ has rank one and its diagonal entries are $B_{ii} = b_i^2 = 1$. Conversely, any matrix $\mathbf{B} \in \mathbb{R}^{NM\times NM}$ that is PSD (henceforth denoted $\mathbf{B} \succeq \mathbf{0}$), satisfies $B_{ii} = 1$, and $\text{rank}(\mathbf{B}) = 1$ necessarily has the form $\mathbf{B} = \mathbf{b}\mathbf{b}^T$, for some $\mathbf{b} \in \{-1,1\}^{NM}$. Using these definitions, one has $\mathbf{b}^T\mathbf{W}\mathbf{b} = \text{tr}(\mathbf{W}\mathbf{B})$ and (12) is equivalent to

$$\min_{\mathbf{B}\succeq\mathbf{0}} \text{tr}(\mathbf{W}\mathbf{B}) \quad \text{s. to } \text{rank}(\mathbf{B}) = 1,\ B_{ii} = 1,\ i = 1, \ldots, NM.$$

The only source of non-convexity in the above formulation is the rank constraint, and dropping it yields the convex SDR

$$\mathbf{B}^* = \underset{\mathbf{B}\succeq\mathbf{0}}{\text{argmin}}\ \text{tr}(\mathbf{W}\mathbf{B}), \quad \text{s. to } B_{ii} = 1,\ i = 1, \ldots, NM, \tag{14}$$

which coincides with the bidual (dual of the dual) problem of (12). Problem (14) is a semidefinite program (SDP) and can be solved using an off-the-shelf interior-point method [31]. It is immediate that a rank-one optimal solution $\mathbf{B}^* = \mathbf{b}^*(\mathbf{b}^*)^T$ of (14) solves the original BQP as well; however, in general $\text{rank}(\mathbf{B}^*) \neq 1$. To generate a feasible solution of (12) from $\mathbf{B}^*$, we adopt the so-termed Gaussian randomization procedure [28, 32]. The overall scheme is tabulated under Algorithm 1.

All in all, the recipe to estimate the graph filter via the SDR approach entails the following steps. First we calculate $\{\hat{\mathbf{A}}_m\}_{m=1}^M$ from $\{\hat{\mathbf{C}}_{\mathbf{y},m}, \mathbf{C}_{\mathbf{x},m}\}_{m=1}^M$ using the expression in the statement of Proposition 3, and form $\hat{\mathbf{\Psi}}$ as in (11) to finally obtain $\hat{\mathbf{W}} = \hat{\mathbf{\Psi}}^T\hat{\mathbf{\Psi}}$. Next, a feasible solution $\hat{\mathbf{b}}_{l^*}$ to the BQP is obtained after running Algorithm 1 with $\hat{\mathbf{W}}$ and an appropriate choice of L as inputs. Finally, $\hat{\mathbf{H}}$ is estimated using (13).

Algorithm 1 Graph filter identification using SDR

Input: $\mathbf{W} = \mathbf{\Psi}^T\mathbf{\Psi} \in \mathcal{M}_{NM}$ and $L \in \mathbb{N}$.
Solve the SDP in (14) to obtain $\mathbf{B}^*$.
for $l = 1, \ldots, L$ **do**
  Draw $\mathbf{z}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^*)$.
  Round $\hat{\mathbf{b}}_l = \text{sgn}(\mathbf{z}_l)$.
end for
Determine $l^* = \underset{l=1,\ldots,L}{\text{argmin}}\, \hat{\mathbf{b}}_l^T \mathbf{W} \hat{\mathbf{b}}_l$.
Return $\hat{\mathbf{b}}_{l^*}$

Computational complexity. The SDR entails dropping a rank constraint after "lifting" a (binary) vector-valued problem with NM variables to a matrix-valued one with $NM(NM+1)/2$ variables. Accordingly, the complexity of a general-purpose interior point method to solve the resulting SDP is $\mathcal{O}(N^7M^7\log(1/\epsilon))$, for a prescribed solution accuracy $\epsilon > 0$ [28]. Such a cost could hinder applicability of the SDR approach in Algorithm 1 to problems involving very large networks. For those scenarios, lightweight proximal-gradient iterations in [27] can still find stationary solutions with lower memory requirements and $\mathcal{O}(MN^3)$ complexity per iteration. However, nothing can be said a priori on the quality of the aforementioned stationary points, while the SDR-based solutions of this paper offer quantifiable approximation guarantees.

Performance guarantee. To obtain approximation bounds we will leverage a result in [32] stated as Theorem 1 below, which offers guarantees for an extension to the maximum cut problem

$$\max_{\mathbf{b}\in\{-1,1\}^{NM}} \mathbf{b}^T\mathbf{W}'\mathbf{b}, \tag{15}$$

where the weight matrix $\mathbf{W}' \succeq \mathbf{0}$ may have negative entries. As we show next, (15) can be rendered equivalent to the BQP (12). Define $\mathbf{W}' := \lambda_{\max}\mathbf{I}_{NM} - \mathbf{W} \succeq \mathbf{0}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{W}$. For $\mathbf{b} \in \{-1,1\}^{NM}$, then $\mathbf{b}^T\mathbf{W}'\mathbf{b} = NM\lambda_{\max} - \mathbf{b}^T\mathbf{W}\mathbf{b}$ and the problems are equivalent.

Theorem 1 ([32]) *Let $\mathbf{b}^*$ be the solution of problem* (12) *or equivalently* (15) *for $\mathbf{W}' := \lambda_{\max}\mathbf{I} - \mathbf{W} \succeq \mathbf{0}$. Let $\hat{\mathbf{b}}_{l^*}$ be the output of Algorithm 1. Then,*

$$\frac{2}{\pi}(\mathbf{b}^*)^T\mathbf{W}'\mathbf{b}^* \leq \mathbb{E}\left[(\hat{\mathbf{b}}_{l^*})^T\mathbf{W}'\hat{\mathbf{b}}_{l^*}\right] \leq (\mathbf{b}^*)^T\mathbf{W}'\mathbf{b}^*.$$

A guarantee for the BQP (12) thus follows immediately.

Corollary 1 *Let $\mathbf{b}^*$ be the solution of problem* (12) *and $\hat{\mathbf{b}}_{l^*}$ be the output of Algorithm 1. For $\gamma = \left(1 - \frac{2}{\pi}\right)\lambda_{\max}NM > 0$, then*

$$\gamma + \frac{2}{\pi}(\mathbf{b}^*)^T\mathbf{W}\mathbf{b}^* \geq \mathbb{E}\left[(\hat{\mathbf{b}}_{l^*})^T\mathbf{W}\hat{\mathbf{b}}_{l^*}\right] \geq (\mathbf{b}^*)^T\mathbf{W}\mathbf{b}^*. \tag{16}$$

Notice that although the bounds in (16) offer guarantees in terms of the expected objective value, particular realizations $\hat{\mathbf{b}}_{l^*}$ tend to fall within those bounds with high probability if the number L of random draws in Algorithm 1 is chosen sufficiently large.

5. UNVEILING URBAN MOBILITY PATTERNS

We implement our SDR graph topology inference method (Algorithm 1) in order to detect mobility patterns in New York City from Uber pickups data[1]. We have access to times and locations of pickups from January 1st to June 29th 2015 for 263 known location IDs. For simplicity, we cluster the locations into $N = 30$ zones based on their geographical proximity; shown as red pins in Fig. 1. These zones represent the nodes of the graph to be recovered. The total number of pickups aggregated by zone during a specific time horizon can be regarded as graph signals defined on the unknown graph. More specifically, we consider $M = 2$ graph processes: weekday ($m = 1$) and weekend ($m = 2$) pickups. Moreover, we consider that the pickups from 6am to 11am constitute the inputs of our process whereas the pickups from 3pm to 8pm comprise the outputs. To be more precise, for a specific day we aggregate all the pickups within 6-11am to form an input signal $\mathbf{x}$ and similarly we group all the pickups within 3-8pm to generate the associated output signal $\mathbf{y}$. If the day considered is a weekday, we think of this pair as being generated from process $m = 1$, and if it is a weekend we consider the pair coming from process $m = 2$. We repeat this procedure for all the days included in the period of study, and estimate input-output covariance pairs $\{\hat{\mathbf{C}}_{\mathbf{x},m}, \hat{\mathbf{C}}_{\mathbf{y},m}\}_{m=1}^2$. We then run Algorithm 1 to infer an underlying graph filter $\hat{\mathbf{H}}$ and solve (6) given the estimated eigenbasis of $\hat{\mathbf{H}}$ to find a sparse mobility pattern.

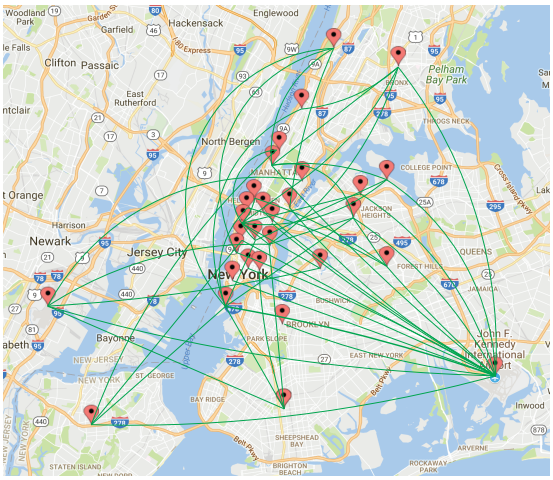

Fig. 1: New York's mobility pattern inferred from 2015 Uber pickups data. Most edges connect Manhattan with the other boroughs indicating that Uber is widely used to commute to/from the suburbs.

The recovered graph is depicted in Fig. 1, where the weights of the recovered edges are represented by the edge widths. Given the nature of the input and output processes considered, the graph is a sparse description of the mobility pattern of people throughout the day. Most connections occur between Manhattan and the other boroughs (Queens, Bronx, Staten Island, Brooklyn, and Newark), while only a few edges connect zones within Manhattan. This indicates that people use Uber to commute from their homes in the suburbs to their work (or leisure activities in the weekends) in the city. These findings are consistent with exploratory research of this same dataset [33] as well as a recent New York Times article: *"Uber has deployed thousands of black cars across Manhattan . . . but the ride-hail app is booming in the other boroughs, with half of all Uber rides now starting outside Manhattan. . . "* [34]. Lastly, observe that the JFK, Newark, and LaGuardia airports are strongly connected with Manhattan and the other boroughs, as expected.

6. CONCLUSIONS

We studied the inference of an undirected network from observations of *non-stationary* signals diffused on the unknown graph. Relative to the stationary setting, the main challenge was that the GSO eigenvectors differ from those of the output signal covariance matrix. Thus, we leveraged that the sought eigenbasis is preserved by the diffusing graph filter. However, the identification of this filter from the covariance led to a non-convex optimization, which was handled by a semi-definite relaxation with provable recovery guarantees.

[1]Dataset from https://github.com/fivethirtyeight/uber-tlc-foil-response

7. REFERENCES

[1] 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.

[2] 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, May 2013.

[3] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Aug. 2017.

[4] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, July 2017.

[5] B. Girault, "Stationary graph signals using an isometric graph translation," in *European Signal Process. Conf. (EUSIPCO)*, Aug 2015, pp. 1516–1520.

[6] S. Segarra, A.G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Aug. 2017.

[7] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. PP, no. 99, pp. 1–1, 2017.

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

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

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

[11] B. M. Lake and J. B. Tenenbaum, "Discovering structure by learning sparse graph," in *Annual Cognitive Sc. Conf.*, 2010, pp. 778 – 783.

[12] 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.

[13] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under Laplacian and structural constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 825–841, Sept. 2017.

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

[15] 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.

[16] B. Baingana, G. Mateos, and G. B. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 563–575, Aug. 2014.

[17] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3675–3687, Jul. 2017.

[18] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality," *PNAS*, vol. 101, pp. 9849–9854, 2004.

[19] G. V. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Shanghai, China, Mar. 20-25, 2016.

[20] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2503–2516, Feb. 2017.

[21] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, Aug. 2016.

[22] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.

[23] V. Kalofolias, "How to learn a graph from smooth signals," in *Intl. Conf. Artif. Intel. Stat. (AISTATS)*. J Mach. Learn. Res., 2016, pp. 920–929.

[24] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 484–499, Sept 2017.

[25] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, "Learning sparse graphs under smoothness prior," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, New Orleans, LA, Mar. 5-9, 2017, pp. 6508–6512.

[26] M. G. Rabbat, "Inferring sparse graphs from smooth signals with theoretical guarantees," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, New Orleans, LA, Mar. 5-9, 2017, pp. 6533–6537.

[27] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *IEEE J. Sel. Topics Signal Process.*, 2017, (submitted).

[28] Z. Q. Luo, W. K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[29] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1442–1449, 2001.

[30] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, 2010.

[31] Y. Ye, *Interior Point Algorithms: Theory and Analysis.*, Wiley-Interscience series in discrete mathematics and optimization, 1997.

[32] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optimization methods and software*, vol. 9, no. 1-3, pp. 141–160, 1998.

[33] C. Bialik, A. Flowers, R. Fischer-Baum, and D. Mehta, "Uber is Serving New York's Outer Boroughs More Than Taxis Are," Aug. 10 2015.

[34] W. Hu, "Uber, Surging Outside Manhattan, Tops Taxis in New York City," Oct. 12 2017.