ONLINE NETWORK TOPOLOGY INFERENCE WITH PARTIAL CONNECTIVITY INFORMATION

Rasoul Shafipour and Gonzalo Mateos

Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

ABSTRACT

We develop algorithms for online topology inference from streaming nodal observations and partial connectivity information; i.e., a priori knowledge on the presence or absence of a few edges may be available as in the link prediction problem. The observations are modeled as stationary graph signals generated by local diffusion dynamics on the unknown network. Said stationarity assumption implies the simultaneous diagonalization of the observations' covariance matrix and the so-called graph shift operator (GSO), here the adjacency matrix of the sought graph. When the GSO eigenvectors are perfectly obtained from the ensemble covariance, we examine the structure of the feasible set of adjacency matrices and its dependency on the prior connectivity information available. In practice one can only form an empirical estimate of the covariance matrix, so we develop an alternating algorithm to find a sparse GSO given its imperfectly estimated eigenvectors. Upon sensing new diffused observations in the streaming setting, we efficiently update eigenvectors and perform only one (or a few) online iteration(s) of the proposed algorithm until a new datum is observed. Numerical tests showcase the effectiveness of the novel batch and online algorithms in recovering real-world graphs.

***Index Terms*—** Network topology inference, graph signal processing, diffusion process, link prediction, online algorithm.

1. INTRODUCTION

Network data supported on the vertices of a graph $\mathcal{G}$ are nowadays ubiquitous across disciplines spanning engineering as well as social and the bio-behavioral sciences [1]. Such data can be represented as graph signals, namely vectors indexed by the nodes of $\mathcal{G}$. In this context, the goal of graph signal processing (GSP) is to develop information processing algorithms that fruitfully exploit the relational structure of said network data [2]. However, oftentimes $\mathcal{G}$ is not readily available and a first key step is to use observations of graph signals to learn the underlying network structure (or a meaningful graph model that facilitates signal representations and prediction tasks); see [3, 4] for recent tutorial treatments on graph learning.

To state the problem at hand, consider a weighted undirected graph $\mathcal{G}$, consisting of a node set $\mathcal{N}$ of cardinality N, and symmetric adjacency matrix $\mathbf{A}$ with entry $A_{ij} = A_{ji} \neq 0$ denoting the edge weight between node i and node j. We assume that $\mathcal{G}$ contains no self-loops; i.e., $A_{ii} = 0$. Generally speaking, we can define a generic *graph-shift operator* (GSO) $\mathbf{S} \in \mathbb{R}^{N \times N}$ as any matrix capturing the same sparsity pattern as $\mathbf{A}$ on its off-diagonal entries [5]. Common choices for $\mathbf{S}$ are the adjacency $\mathbf{A}$, the Laplacian $\mathbf{L} := \text{diag}(\mathbf{A1}) - \mathbf{A}$, or their normalized counterparts [2]. Henceforth we focus on $\mathbf{S} = \mathbf{A}$ and aim to recover the adjacency matrix of the unknown graph $\mathcal{G}$. Other GSOs can be accommodated in a similar fashion.

In the sequel, we present an online framework that estimates sparse graphs that explain the structure of streaming random signals. Particularly, in a snapshot, let $\mathbf{y} = [y_1, ..., y_N]^T \in \mathbb{R}^N$ be a zero-mean graph signal in which the ith element y_i denotes the signal value at node i of an *unknown graph* $\mathcal{G}$ with shift operator $\mathbf{S}$. Further consider a zero-mean white signal $\mathbf{x}$ with covariance matrix $\mathbf{C_x} = \mathbf{I}$ (identity matrix). We state 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

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

Under the assumption that $\mathbf{C_x} = \mathbf{I}$, (1) is equivalent to the *stationarity* of $\mathbf{y}$ in $\mathbf{S}$; see e.g., [6, Def. 1], [7], [8]. 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 in (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 *streaming* stationary random signals $\mathcal{Y} := \{\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(p)}, \mathbf{y}^{(p+1)}, \cdots\}$, each of them adhering to linear diffusion dynamics as in (1). Unlike [9] but similar to link prediction problems [1, Ch. 7.2], here we assume a priori knowledge about the presence (or absence) of a few edges; leading to simpler updates and better performance. This is a reasonable assumption, since we may know the status of a few edges (via limited questionnaires or experiments), or might perform edge screening prior to topology inference [10]. Supposing the data acquisition interval is long enough relative to the time required to run a few iterations of the graph learning algorithm, we pursue simple, time-adaptive topology updates that can be used to estimate the GSO in an online fashion. Due to the stationarity assumption, the ensemble covariance matrix of the observations shares the same eigenvectors with the sought GSO; see [11, 12] and Section 2. Leveraging this result, our online algorithm entails two steps, where we: (i) update eigenvectors efficiently using methods described in Section 3.2; and (ii) take one or a few steps of a graph learning algorithm developed in Section 3. In Section 4 we corroborate the effectiveness of the proposed topology inference approaches in both batch and online setups.

Relation to prior work. Workhorse topology inference approaches construct graphs whose edge weights correspond to nontrivial correlations between signals at incident nodes [1, 13]. Acknowledging that the observed correlations can be due to latent network effects, alternative statistical methods rely on inference of partial correlations [1, Ch. 7.3.2]. Under Gaussianity assumptions, this line of work has well-documented connections with covariance selection [14] and sparse precision matrix estimation [15–18], as well as high-dimensional sparse linear regression [19]. 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 [11, 12, 20–23]. Different from [20, 22, 24, 25] that infer structure from signals assumed to be smooth over the sought undirected graph, here the measurements are assumed related to the graph via filtering. Few works have recently explored this approach by identifying a symmetric GSO given its eigenvectors, either assuming that the input is white [11, 12] – equivalently implying $\mathbf{y}$ is graph stationary [6–8]; or, colored [26, 27]. Building on [9], here

Work in this paper was supported by the NSF award CCF-1750428.

we deal with online graph learning with partial connectivity information. While we assume that the graph signals are stationary, the online scheme in [28] uses observations from a Laplacian-based, continuous-time graph process. Relative to [29] that relies on a single-pole graph filter [30], the filter structure underlying (1) can be arbitrary, but the focus here is on learning undirected graphs.

2. PRELIMINARIES AND PROBLEM STATEMENT

We consider topology inference from stationary signals whereby a small portion of the edges are known a priori. We then examine the size of the feasibility set and finally state the problem of finding a structurally admissible graph with sparse connectivity structure.

2.1. Topology inference under stationarity

To formally state the problem, we consider 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 $\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l \in \mathbb{R}^{N\times N}$ [5], the Cayley-Hamilton theorem asserts that the model in (1) boils down to

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

for some particular $\mathbf{h}$ and $L \leq N$. Note that L specifies the dependency range of the diffusion on the neighhbors.

We first start with the offline setting, where $\mathbf{x}$ is white so that $\mathbf{C_x} = \mathbf{I}$ [11]. The covariance matrix of $\mathbf{y} = \mathbf{H}\mathbf{x}$ is then

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

We used the symmetry of $\mathbf{H}$ to obtain the third equality, as $\mathbf{H}$ is a polynomial in the symmetric GSO $\mathbf{S}$. Using the spectral decomposition of $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ to express the filter as $\mathbf{H} = \sum_{l=0}^{L-1} h_l(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^l = \mathbf{V}(\sum_{l=0}^{L-1} h_l\mathbf{\Lambda}^l)\mathbf{V}^T$, we can diagonalize the covariance matrix as

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

Such a covariance expression is the requirement for a graph signal to be stationary in $\mathbf{S}$ [6, Def. 2.b]. Remarkably, if $\mathbf{y}$ is graph stationary, or equivalently if $\mathbf{x}$ is white (i.e., its covariance is identity matrix), (4) shows that the *eigenvectors* of the shift $\mathbf{S}$, the filter $\mathbf{H}$, and the covariance $\mathbf{C_y}$ are *all the same*. As a result, to estimate $\mathbf{V}$ from the observations $\{\mathbf{y}^{(p)}\}_{p=1}^P$ 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 $\mathbf{S}$ [11, 12]. Note that in estimating $\mathbf{C_y}$, we assume that the observed signals are zero-mean without loss of generality, otherwise we can subtract the mean from the signals.

Furthermore, one can impose constraints to ensure the GSO $\mathbf{S}$ is structurally admissible and incorporate a priori knowledge about $\mathbf{S}$. Namely, 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

$$\mathbf{S} \in \{\mathbf{S} \,|\, S_{ij} \geq 0, \mathbf{S}^T = \mathbf{S}, S_{ii} = 0, \textstyle\sum_j S_{j1} = 1\}. \tag{5}$$

The first condition in (5) 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 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. Other GSOs (e.g., the Laplacian $\mathbf{L}$ and its normalized variants) can be accommodated via minor modifications to (5); see [11].

In some pragmatic settings that we consider in this paper, we may know about the existence of a few edges or their corresponding weights as prior information. In that case, we can drop the constraint $\sum_j S_{j1} = 1$ and instead add $S_{ij} = s_{ij}$ for the (i, j) pairs with known weights s_{ij}. This would fix the scale of the GSO which was the reason we had the constraint $\sum_j S_{j1} = 1$ in (5). Accordingly, we can rewrite the set of admissible adjacency matrices as

$$\mathcal{S} := \{\mathbf{S} \,|\, S_{ij} \geq 0, \mathbf{S}^T = \mathbf{S}, S_{ii} = 0, S_{ij} = s_{ij},\ (i,j) \in \Omega\}, \tag{6}$$

where we denote the set of observed edges (i, j) as Ω.

2.2. Size of the feasibility set and sparse recovery

Inspired by [11], here we examine the feasibility set and the degrees of freedom of the GSO $\mathbf{S}$ under the assumption that the perfect spectral templates $\mathbf{V}$ are available and $\mathbf{S} \in \mathcal{S}$ [cf. (6)]. This would shed light on the dependency of the feasibility set's structure and dimensionality (hence the difficulty of recovering $\mathbf{S}$) on the number of observed edges. As we show next, the feasibility set may potentially reduce to a singleton (the graph $\mathbf{S} \in \mathcal{S}$ is completely specified by $\mathbf{V}$), or more generally to a low-dimensional subspace. In the latter (more interesting) case, or more pragmatically when we approximate $\mathbf{V}$ with the eigenvectors $\hat{\mathbf{V}}$ of the observations' sample covariance, we formulate a convex optimization problem as in [11] to search for a sparse and structurally admissible GSO.

Feasibility set. Given the GSO eigenvectors $\mathbf{V}$, consider the mapping $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ between $\mathbf{S}$ and $\mathbf{\Lambda}$. This can precisely be rewritten as $\text{vec}(\mathbf{S}) = (\mathbf{V} \odot \mathbf{V})\boldsymbol{\lambda} = \mathbf{W}\boldsymbol{\lambda}$, where $\odot$ denotes the Khatri-Rao (column-wise Kronecker) product, $\boldsymbol{\lambda} \in \mathbb{R}^N$ collects the diagonal entries of $\mathbf{\Lambda}$, and $\mathbf{W} := \mathbf{V} \odot \mathbf{V} \in \mathbb{R}^{N^2\times N}$. Recall that $\mathbf{S} \in \mathcal{S}$ and accordingly the entries of $\text{vec}(\mathbf{S})$ corresponding to the diagonal entries of $\mathbf{S}$ should be zero. Upon defining the set $\mathcal{T} := \big\{N(i-1)+i \mid i \in \{1, \cdots, N\}\big\}$, we have the mapping $\mathbf{W}_\mathcal{T}\boldsymbol{\lambda} = \mathbf{0}$ to the null diagonal entries of $\mathbf{S}$, where $\mathbf{W}_\mathcal{T} \in \mathbb{R}^{N\times N}$ is a submatrix of $\mathbf{W}$ that contains rows indexed by the set $\mathcal{T}$. Thus, $\mathbf{W}_\mathcal{T}$ is rank-deficient and $\boldsymbol{\lambda}$ belongs to $\ker(\mathbf{W}_\mathcal{T})$, where $\ker(.)$ denotes the null space of its argument. In particular, assume that $\text{rank}(\mathbf{W}_\mathcal{T}) = N-k$, $1 \leq k \leq N$, which implies $\boldsymbol{\lambda}$ lives in a k-dimensional subspace. As some of the entries in $\mathbf{S}$ are known according to $\mathcal{S}$, intuitively we expect that by observing k "*sufficiently different*" edges, the feasible set will boil down to a singleton resulting in a unique feasible $\mathbf{S} \in \mathcal{S}$. To quantify the extra constraints imposed by the partially observed connectivities, let $\mathcal{M} := \{N(j-1)+i \mid (i,j) \in \Omega\}$ correspond to the known entries of $\text{vec}(\mathbf{S})$. Then upon defining $\mathbf{U} \in \mathbb{R}^{N\times k}$ comprising the basis vectors of $\ker(\mathbf{W}_\mathcal{T})$, the condition $\text{rank}(\mathbf{W}_\mathcal{M}\mathbf{U}) = k$ would be sufficient to determine $\mathbf{S}$ uniquely in the k-dimensional null space of $\mathbf{W}_\mathcal{T}$ as summarized in the following proposition.

Proposition 1 *Suppose the GSO eigenvectors $\mathbf{V}$ are given. If $\text{rank}(\mathbf{W}_\mathcal{T}) = N-k$ and $\text{rank}(\mathbf{W}_\mathcal{M}\mathbf{U}) = k$, then $\mathcal{S}$ is a singleton.*

Proof: Since $\boldsymbol{\lambda} \in \ker(\mathbf{W}_\mathcal{T})$, there exists an $\boldsymbol{\alpha} \in \mathbb{R}^k$ such that $\boldsymbol{\lambda} = \mathbf{U}\alpha$. From the known entries of $\text{vec}(\mathbf{S})$ denoted by $\mathbf{w} := [\text{vec}(\mathbf{S})]_\mathcal{M}$ we have $\mathbf{W}_\mathcal{M}\boldsymbol{\lambda} = \mathbf{W}_\mathcal{M}\mathbf{U}\boldsymbol{\alpha} = \mathbf{w}$. Thus, to uniquely identify $\boldsymbol{\alpha}$ and equivalently $\boldsymbol{\lambda}$ (and $\mathbf{S}$), it is sufficient to have $\text{rank}(\mathbf{W}_\mathcal{M}\mathbf{U}) = k$. ■

Proposition 1 further implies that $\text{rank}(\mathbf{W}_\mathcal{M}) \geq k$ under the assumption that $\text{rank}(\mathbf{W}_\mathcal{M}\mathbf{U}) = k$. This is due to the inequality $\text{rank}(\mathbf{W}_\mathcal{M}\mathbf{U}) \leq \min\{\text{rank}(\mathbf{W}_\mathcal{M}), \text{rank}(\mathbf{U})\}$. Observing k "*sufficiently different*" edges for unique recovery of $\mathbf{S}$ is the intuition behind the rank constraint on $\mathbf{W}_\mathcal{M}$. In real-world graphs, we have observed that k is typically much smaller than N; see also Section 4 and [11, Section 3]. This would make it feasible to uniquely identify the graph, given only its eigenvectors and k sufficiently different pre-observed edges. However, in practice we may not know about the status of those many edges, or, the graph eigenvectors may only be imperfectly estimated via eigendecomposition of the sample covariance matrix $\hat{\mathbf{C}}_\mathbf{y}$. This motivates searching for an optimal graph while accounting for the (finite sample size) approximation errors and the prescribed structural constraints, the subject dealt with next.

Sparse recovery. Given estimates $\hat{\mathbf{V}}$ of the covariance eigenvectors, recovery of $\mathbf{S}$ amounts to selecting its eigenvalues $\mathbf{\Lambda}$ and to that end

we assume that the shift of interest is sparse. At the same time, we should account for the discrepancies between $\hat{\mathbf{V}}$ and the underlying eigenvectors of $\mathbf{S}$, due to finite sample size constraints and unavoidable errors in estimating the filter. Accordingly, we build on [11] and seek for the shift operator $\mathbf{S}$ that: (a) is sparse, meaning that few edge weights are non-zero; (b) belongs to the convex set $\mathcal{S}$ that specifies structural constraints and edge status information; and (c) is close to $\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T$ in the Frobenius-norm sense. One can thus solve

$$\hat{\mathbf{S}} := \underset{\mathbf{\Lambda}\in\mathcal{D},\mathbf{S}\in\mathcal{S}}{\operatorname{argmin}} \ \|\mathbf{S}\|_1, \quad \text{subject to: } \|\mathbf{S}-\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T\|_F \leq \epsilon, \quad (7)$$

where $\mathcal{D}$ is the set of $N\times N$ diagonal matrices. Problem (7) 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.

Going back to the online setting where we acquire streaming stationary signals $\mathcal{Y}=\{\mathbf{y}^{(1)},\cdots,\mathbf{y}^{(p)},\mathbf{y}^{(p+1)},\cdots\}$, the goal is to update the GSO in a computationally efficient way at each time step. We assume that the data acquisition period (interval between two samples) is much shorter than the time required to update the GSO by eigendecomposition of an updated $\hat{\mathbf{C}}_{\mathbf{y}}$ and by solving (7) using off-the-shelf algorithms. Accordingly, the idea is to only take one (or a few) simple step(s) of an iterative algorithm for solving (7) upon sensing a new graph signal. Beyond the time constraint, it is not prudent to devote major efforts in fully solving (7) if we are going to refine $\mathbf{S}$ soon thereafter, once a new datum arrives. Specifically, upon arrival of a new observation we first update the covariance $\hat{\mathbf{C}}_{\mathbf{y}}$ using a weighted re-averaging, or, by resorting to a cheaper approach that directly updates the covariance eigenvectors $\hat{\mathbf{V}}$ using rank-one perturbations of $\hat{\mathbf{C}}_{\mathbf{y}}$; see Section 3.2. Given the updated eigenvectors $\hat{\mathbf{V}}$, we then run few iterations of the alternating minimzation algorithm in Section 3.1 to refine $\hat{\mathbf{S}}$ before we acquire a new signal.

3. ALGORITHMS

In this section, we first derive an iterative algorithm to solve (7) in the (offline) batch setting, and then we adapt it for online processing.

3.1. Alternating minimization scheme

Here we develop an iterative algorithm to find a sparse and structurally admissible $\mathbf{S}$ that solves (7). To make the optimization problem amenable to closed form iterations, we reformulate (7) as

$$\begin{aligned} \underset{\mathbf{\Lambda},\mathbf{S}}{\text{minimize}} \quad & \delta(\mathbf{\Lambda},\mathbf{S}) := \mu\|\mathbf{S}\|_1 + \frac{1}{2}\|\mathbf{S}-\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^T\|_F^2, \\ \text{subject to} \quad & \mathbf{\Lambda}\in\mathcal{D}, \quad \mathbf{S}^T=\mathbf{S}, \quad S_{ij}\geq 0, \quad S_{ii}=0, \\ & S_{ij}=s_{ij},\ \forall (i,j)\in\Omega. \end{aligned} \quad (\mathcal{P})$$

Problem $(\mathcal{P})$ is still convex and can be solved using a block coordinate-descent method. In particular, we solve $(\mathcal{P})$ cyclically over each variable $\mathbf{S}$ and $\mathbf{\Lambda}$ while fixing the other variable to its most up to date value. This procedure provably converges to the global optimum of $(\mathcal{P})$ [31]. The closed-form updates for the blocks $\mathbf{\Lambda}$ and $\mathbf{S}$ [initialized randomly at $\mathbf{S}(0)$] are outlined next.

$\mathbf{\Lambda}$-update. At each iteration $k=0,1,2,\ldots$ we fix $\mathbf{S}=\mathbf{S}(k)$ and update $\mathbf{\Lambda}(k)$ by solving

$$\begin{aligned} \mathbf{\Lambda}(k) &:= \underset{\mathbf{\Lambda}\in\mathcal{D}}{\operatorname{argmin}} \ \delta(\mathbf{\Lambda},\mathbf{S}(k)) \\ &= \underset{\mathbf{\Lambda}\in\mathcal{D}}{\operatorname{argmin}} \ \|\mathbf{S}(k)-\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^\top\|_F^2. \end{aligned} \quad (8)$$

Problem (8) has a closed-form solution given by

$$\mathbf{\Lambda}(k) := \operatorname{diag}(\hat{\mathbf{V}}^T\mathbf{S}(k)\hat{\mathbf{V}}), \quad (9)$$

because one can equivalently rewrite the cost in (8) as $\|\hat{\mathbf{V}}^T\mathbf{S}(k)\hat{\mathbf{V}}-\mathbf{\Lambda}\|_F^2$ due to the orthonormality of $\hat{\mathbf{V}}$.

S-update. After updating $\mathbf{\Lambda}(k)$ at iteration k, we keep the latest $\mathbf{\Lambda}$ fixed and solve $(\mathcal{P})$ with respect to $\mathbf{S}$. The optimization variable $\mathbf{S}$ in $\delta(\mathbf{\Lambda}(k),\mathbf{S})$ is component-wise separable. In particular, $\mathbf{S}(k+1)$ can be obtained in a simple closed form in terms of non-negative soft-thresholding operations and suitable projections. Defining the matrix $\mathbf{B}(k):=\hat{\mathbf{V}}\mathbf{\Lambda}(k)\hat{\mathbf{V}}^T$, the entries of $\mathbf{S}(k+1)$ are given by

$$S_{ij}(k+1)=\begin{cases} 0, & i=j \\ s_{ij}, & (i,j)\in\Omega \\ \max(0,B_{ij}(k)-\mu), & \text{otherwise,} \end{cases}. \quad (10)$$

The overall algorithm is referred to as AltTopoId, which entails alternating updates for $\mathbf{\Lambda}$ and $\mathbf{S}$ until a certain termination condition is met; e.g., $\|\mathbf{S}(k)-\mathbf{S}(k-1)\|_F/\|\mathbf{S}(k-1)\|_F\leq\varepsilon$ for a parameter ε. Next, we build on the simple AltTopoId iterations to develop an online topology inference algorithm which is capable of processing streaming signals and learn the graph structure adaptively.

3.2. Online topology inference

Towards online topology inference from stationary signals, an important first step is to update the eigenvectors of $\hat{\mathbf{C}}_{\mathbf{y}}$ efficiently and circumvent repeated eigendecompositions. This can be achieved in $\mathcal{O}(N^2)$ complexity using results on the eigendecomposition of rank-one modifications of a symmetric matrix [32]. In particular, let $\hat{\mathbf{C}}_{\mathbf{y}}^{(P)}$ denote the signal's sample covariance obtained from P observations. The updated sample covariance upon sensing $\mathbf{y}^{(P+1)}$ is given by

$$\hat{\mathbf{C}}_{\mathbf{y}}^{(P+1)}=\frac{1}{P+1}\left(P\hat{\mathbf{C}}_{\mathbf{y}}^{(P)}+\mathbf{y}^{(P+1)}\mathbf{y}^{(P+1)^T}\right), \quad (11)$$

which is a rank-one modification of $\hat{\mathbf{C}}_{\mathbf{y}}^{(P)}$. It is shown in [32] that the eigenvalues of the rank-one modification of a symmetric matrix with eigenvalues $\{d_j\}_{j=1}^N$ are the roots γ of the characteristic equation

$$1+\sum_{j=1}^{N}\frac{\mathbf{z}_j^2}{Pd_j-\gamma}=0, \quad (12)$$

where $\mathbf{z}=\hat{\mathbf{V}}^{(P)^T}\mathbf{y}^{(P+1)}$ and $\hat{\mathbf{V}}^{(P)}$ are the eigenvectors of $\hat{\mathbf{C}}_{\mathbf{y}}^{(P)}$. The so-termed secular equation (12) can be solved using, e.g., the Newton method in $\mathcal{O}(N^2)$ complexity. After finding the eigenvalues $\{\gamma_j\}_{j=1}^N$ efficiently, the corresponding eigenvectors can be updated via

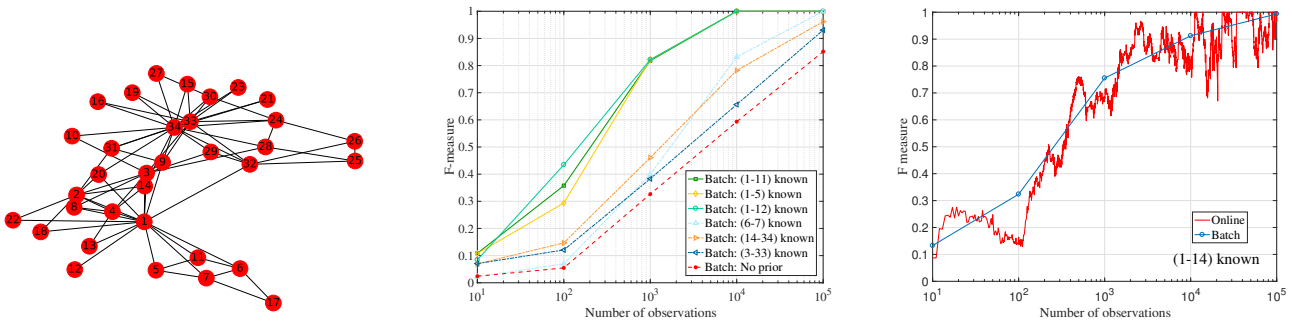
$$\mathbf{v}_j^{(P+1)}=q_j\mathbf{y}^{(P+1)}\circ\mathbf{q}_j, \quad j=1,\ldots,N \quad (13)$$

where q_j is a normalizing factor ensuring that $\|\mathbf{v}_j\|_2=1$, $\mathbf{q}_j=[1/(Pd_1-\gamma_j),\ldots,1/(Pd_N-\gamma_j)]^T$, and $\circ$ denotes the Hadamard (entrywise) product. So, upon receiving a new signal $\mathbf{y}^{(P+1)}$, the updated eigenvectors $\hat{\mathbf{V}}^{(P+1)}$ can be adaptively computed using $\hat{\mathbf{V}}^{(P)}$ and $\mathbf{y}^{(P+1)}$ in $\mathcal{O}(N^2)$ time. This effectively avoids the $\mathcal{O}(N^3)$ cost of directly computing the eigendecomposition of $\hat{\mathbf{C}}^{(P+1)}$.

Putting all the pieces together, our online topology inference algorithm entails two steps every time a new observation becomes available: (i) update eigenvectors $\hat{\mathbf{V}}$ using the aforementioned rank-one modification technique; and (ii) run one (or a few iterations) of the AltTopoId algorithm to update $\hat{\mathbf{S}}$ before a new datum is received. Note that if the signals arrive faster, one can create a buffer and perform each iteration of the AltTopoId algorithm on a $\hat{\mathbf{V}}$ updated with a sliding window of newly observed signals. On the other hand, for a slower rate of arrivals, additional algorithmic iterations would likely favor recovery performance.

4. NUMERICAL RESULTS

In this section we assess the performance of the proposed algorithm in recovering sparse real-world graphs. To that end, we illustrate the effect of the partial connectivity information on the behavior of the AltTopoId algorithm in an offline setting. We also evaluate the



(a) (b) (c)

Fig. 1: (a) Zachary's karate club graph with $N=34$ nodes. (b) F-measusre of the recovered graph in an offline batch setting using the AltTopo algorithm versus number of observations for the karate club and different a priori information on the connectivities. (c) Evolution of the F-measure in inferring the karate club by performing one step of the proposed online algorithm upon sensing each new signal, superimposed with the offline batch counterpart. This is plotted assuming that we observe the edge between node 1 and node 14 as a priori knowledge.

performance of the proposed online scheme in recovering the graph adaptively using streaming signals.

Throughout this section, we infer unweighted real-world networks from the observation of diffusion processes that are synthetically generated via graph filtering as in (2). For the graph shift $\mathbf{S}=\mathbf{A}$, the adjacency matrix of the sought network, we consider a second-order filter $\mathbf{H}=\sum_{l=0}^{2} h_l \mathbf{S}^l$, where the coefficients $\{h_l\}$ are drawn uniformly from $[0,1]$. To measure the edge-support recovery, we compute the F-measure defined as the harmonic mean of edge precision and recall (precision is the percentage of correct edges in $\hat{\mathbf{S}}$, and recall is the fraction of edges in $\mathbf{S}$ that are retrieved).

Zachary's karate club: Offline. We consider the social network of Zachary's karate club [33] represented by a graph $\mathcal{G}$ consisting of $N=34$ nodes or members of the club and 78 undirected edges symbolizing friendships among them; see Fig. 1-a. Note that the rank of $\mathbf{W}_\mathcal{T}$ (cf. Proposition 1) for this graph is 32. This implies that the knowledge of the perfect spectral templates $\mathbf{V}$ leaves the GSO $\mathbf{S}$ in a 2-dimensional subspace which can lead to a singleton feasibility set by observing only 2 different edges. However, here we apply the AltTopoId algorithm on *noisy* eigenvectors $\hat{\mathbf{V}}$ of the sample covariance and assume that we know one of the 78 edges as a priori information and aim to infer the rest of the edges. Synthetic signals $\{\mathbf{y}^{(p)}\}_{p=1}^{P}$ are generated through diffusion process $\mathbf{H}$ where the entries of the inputs $\{\mathbf{x}^{(p)}\}_{p=1}^{P}$ are drawn independently from the normal Gaussian distribution to make the observations stationary. Fig. 1-b plots the F-measure averaged over 1000 experiments as a function of the number of observed signals P and the a priori knowledge – top and bottom 3 in terms of performance. We also superimpose Fig. 1-b with the recovery performance of the case when we have no a priori knowledge on the edges; i.e., approach in [11]. First we notice that as the number of observations increase, the estimate $\hat{\mathbf{V}}$ becomes more reliable which leads to a better performance (i.e., larger F-measure) of the underlying GSO. As expected, having a priori information on the edges would help the performance. Finally, we calculate the edge betweenness centralities for all the edges as in [34]. Interestingly, we notice that the worst 3 edges to know as a priori knowledge in the topology inference are among the top links in terms of the edge betweenness centrality. From a feasibility standpoint, we conjecture that the edges that are more different from others would result in an easier subspace to search for an optimal GSO; e.g., see $(1-12)$ a priori connectivity information in Fig. 1.

Zachary's karate club: Online. We use the same graph to study the online behavior of the algorithm. We assume that we know the existence of the edge between nodes 1 and 14 as an a priori edge which leads to a moderate performance among all other connectivity a priori information. We generate streaming signals $\{\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(p)}, \mathbf{y}^{(p+1)}, \cdots\}$ by diffusing inputs $\{\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(p)}, \mathbf{x}^{(p+1)}, \cdots\}$ through filter $\mathbf{H}$, where the inputs are formed similarly to the ones for previous experiment. Upon sensing an observation $\mathbf{y}^{(p)}$, we first update sample covariance eigenvectors $\hat{\mathbf{V}}$ using the procedure described in Section 3.2 and then run the AltTopoId algorithm only for one iteration. This is under the assumption that differences between the arrival times of signals are longer than one step of our pipeline. In Fig. 1-c, we depict the evolution of F-measure averaged over 10 instances. We further plot the (average) offline behavior similar to the previous experiment in order to gauge the loss of online estimation. We notice that the adaptive online scheme can successfully track the performance of the offline counterpart on expectation.

Facebook friendship graph: Offline. Finally, we consider a directed network of $N=2888$ Facebook users, where the 2981 edges represent friendships among the users [35, 36]. More precisely, an edge from node i to node j exists if user i is a friend of the user j. To make the graph amenable to our framework, we assume that the friendships are bilateral and ignore the directions. First, we notice that the rank of $\mathbf{W}_\mathcal{T}$ is 2882. This means that knowing GSO's spectral templates and $k=6$ different edges as a priori information would lead to a singleton feasibility set (cf. Proposition 1). To assess the AltTopoId algorithm for this large scale graph, we perform 10 experiments wherein we assume that we know the existence of 5 random edges in each experiment as a priori knowledge. We then generate 10^6 synthetic random signals similar to the first offline experiment and estimate $\hat{\mathbf{V}}$ via eigenvectors of the outputs sample covariance. In this case, estimated $\hat{\mathbf{S}}$ from the AltTopoId algorithm would result in an average F-measure of 0.862 which is promising. Testing this graph for different number of observations and in the online setting would be a valuable extension, and results will be reported elsewhere.

5. CONCLUSION

We studied the inference of an undirected network from streaming observations of *stationary* signals diffused on the graph with partially known connectivity information. We first examined the size of the feasibility set in the ideal scenario whereby the the GSO eigenvectors are perfectly known, as a function of the a priori information available on the status of a few edges. For the pragmatic setup of finite sample size (hence the eigenvectors can only be estimated with error) or limited a priori information, we developed an iterative algorithm to find a sparse adjacency matrix explaining the structure of the diffused signals. This algorithm was then used in an online scheme wherein we adaptively updated the graph as new signals became available sequentially in time. The overall procedure was validated in offline and online settings when recovering real-world graphs from synthetically generated signals.

6. REFERENCES

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

[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] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 16–43, 2019.

[4] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, 2019.

[5] 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.

[6] 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.

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

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

[9] R. Shafipour, A. Hashemi, G. Mateos, and H. Vikalo, "Online topology inference from streaming stationary graph signals," in *IEEE Data Science Wrkshp. (DSW)*, June 2019, pp. 140–144.

[10] T. Ahmed and W. U. Bajwa, "Correlation-based ultrahigh-dimensional variable screening," in *IEEE Intl. Wrksp. Computat. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Dec 2017, pp. 1–5.

[11] 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.

[12] B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat, "Characterization and inference of graph diffusion processes from observations of stationary signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 3, pp. 481–496, 2018.

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

[14] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.

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

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

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

[26] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *IEEE Trans. Signal Process.*, 2018, (submitted; see also arXiv:1801.03862 [eess.SP]).

[27] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, New Orleans, LA, Mar. 5-9, 2017.

[28] S. Vlaski, H. P. Maretić, R. Nassif, P. Frossard, and A. H. Sayed, "Online graph learning from sequential data," in *IEEE Data Science Wrkshp. (DSW)*, 2018, pp. 190–194.

[29] 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.

[30] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, Jan. 2017.

[31] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[32] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen, "Rank-one modification of the symmetric eigenproblem," *Numerische Mathematik*, vol. 31, no. 1, pp. 31–48, 1978.

[33] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[34] U. Brandes, "A faster algorithm for betweenness centrality," *The Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[35] J. Kunegis, "KONECT – The Koblenz Network Collection," in *Proc. Int. Conf. on World Wide Web Companion*, 2013, pp. 1343–1350.

[36] J. Leskovec and J.J. Mcauley, "Learning to discover social circles in ego networks," in *Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 539–547.