Gonzalo Mateos, Santiago Segarra,
Antonio G. Marques, and Alejandro Ribeiro

Connecting the Dots

Network topology inference is a significant problem in network science. Most graph signal processing (GSP) efforts to date assume that the underlying network is known and then analyze how the graph's algebraic and spectral characteristics impact the properties of the graph signals of interest. Such an assumption is often untenable beyond applica-

perspective will constitute a crucial step to obtaining new insights in various areas of science and engineering; SP can play a key role in these ventures.

Under the assumption that the signals are related to the topology of the graph where they are supported, the goal of GSP is to develop algorithms that fruitfully leverage this relational structure and can make inferences about these relationships even when they are only partially observed. Most GSP efforts to date assume that the underlying network topology is known and then analyze how the graph's algebraic and spectral characteristics impact the properties of the graph signals of interest. This is feasible in applications involving physical networks or, when the relevant links are tangible and can be directly observed (e.g., when studying flows in transportation networks, monitoring cascading failures in power grids, maximizing influence on social networks, and tracking the dynamic structure of the World Wide Web). However, in many other settings, the network may represent a conceptual model of pairwise relationships among entities. In exploratory studies of, e.g., functional brain connectiv-

the performance and computational complexity of the ensuing algorithms. All of the elements are now in place to state a general network topology identification problem; see "Why Graft Shift?"

Problem

Given a set $\mathbf{X} := \{\mathbf{x}_p\}_{p=1}^{P}$ of graph signal observations supported on the unknown graph $G(V, E, \mathbf{W})$ with $|V| = N$, the goal is to identify the topology encoded in the entries of a graph-shift operator $\mathbf{S}$ that is optimal in some sense. The optimality criterion is usually dictated by the adopted network-dependent model for the signals in $\mathbf{X}$, possibly augmented by priors motivated by physical characteristics of $\mathbf{S}$, to effect statistical regularization, or else to favor more interpretable graphs.

This is admittedly a very general and somewhat loose formulation that will be narrowed down in subsequent sections as we elaborate on various criteria stemming from different models binding the (statistical) signal properties to the graph topology. Indeed, it is clear that one must assume some relation between the signals and the unknown underlying graph, since otherwise, the topology inference exercise would be hopeless. This relation will be henceforth given by statistical generative priors in the "Statistical Methods for Network Topology Inference" section and by properties of the signals with respect to the underlying graph such as smoothness (the "Learning Graphs From Observations of Smooth Signals" section) or stationarity (the "Identifying the Structure of Network Diffusion Processes" section). The observations in $\mathbf{X}$ can be noisy and incomplete, and accordingly the relationship between N, P, and the mechanisms of data errors and missingness will all play a role in the graph recovery performance. Mostly, the focus will be on inference of undirected and static graphs, an active field for which the algorithms and accompanying theory are today better developed. The "Emerging Topic Areas" section will broaden the scope to more challenging directed, dynamic, and multigraphs.

The iGFT formula $\mathbf{x} = \mathbf{V}\mathbf{x} = \sum_{k=1}^{N} x_k \mathbf{v}_k$ allows one to synthesize $\mathbf{x}$ as a sum of orthogonal frequency components $\mathbf{v}_k$. The contribution of $\mathbf{v}_k$ to 97(h)-4M0(i)i(w)1g(e)-12(n)-2(a)-212( )]TJET/PlacedGraphic BMC q8150.005 6890006 6.719 6.937 reW nBT

transform (KLT) [also known as the *principal component analysis* (*PCA*) transform in statistics and data analysis]; see "Encompassing Nature of the GFT." The GFT offers a unifying framework that subsumes all(s)1(s)9(i)-16475Da

an example of a random graph signal that is stationary with respect to any graph shift $\mathbf{S}$. A second, albeit related, observation is that, by definition, any random vector $\mathbf{x}$ is stationary with respect to the shift given by the covariance matrix $\mathbf{S} = \boldsymbol{\Sigma}_{\mathbf{x}}$. The same is true for the precision matrix $\mathbf{S} = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}$. These facts will be leveraged in the "Identifying the Structure of Network Diffusion Processes" section to draw connections between stationary graph signal-based topology inference approaches and some of the classical statistical methods reviewed in the "Statistical Methods for Network Topology Inference" section. Third, notice that the stationarity requirement is tantamount to the covariance of the process being a polynomial in the graph-shift operator. Accordingly, under

$$\text{FDR} := \mathbb{E}\left[\frac{R_f}{R} \middle| R > 0\right] \Pr[R > 0], \tag{10}$$

where *R*d/T18lacedGraphic BMC EMC BT/Cq17 7081 662.968954.15f9 re5 nBT/T15 s 1 scn/T16 1 T110.01 0 Tw -18.3 0 9.5 109.8

Variants of the model penalize only the off-diagonal entries of $\boldsymbol{\Theta}_l$ or incorporate edge-specific penalty parameters $\alpha_{ij} > 0$ to account for structural priors on the graph topology. Estimators of graphs with nonnegative edge weights are of particular interest; see "Learning Gaussian Graphical Models With Laplacian Constraints."

Although (14) is convex, the objective is nonsmooth and has an unbounded constraint set. As shown in [2], the resulting complexity for off-the-shelf interior point methods adopted in [64] is $O(N^6)$. Additionally, interior point meth-

Learning graphs from observations of smooth signals

In various GSP applications, it is desirable to construct a graph on which network data admit certain regularity. Accordingly, in this section, we survey a family of topology identification approaches that deal with the following general problem. Given a set $\mathbf{X} := \{\mathbf{x}_p\}_{p=1}^{P}$ of possibly noisy graph signal observations, the goal is to learn a graph $G(V, E, \mathbf{W})$ with $|V| = N$ nodes such that the observations in $\mathbf{X}$ are smooth on $\mathbf{G}$. Recall that a graph signal is said to be smooth if the values associated with vertices incident to edges with large weights in the graph tend to be similar. As discussed in the "Graph Fourier Transform and Signal Smoothness" section, the so-defined smoothness of a signal can be quantified by means of a TV measure given by the Laplacian quadratic form in (1). Such a measure offers a natural criterion to search for the best topology (encoded in the entries of the Laplacian), which endows the signals in $\mathbf{X}$ with the desired smoothness property.

There are several reasons that motivate this graph-learning paradigm. First, smooth signals admit low-pass, band-limited (i.e., sparse) representations using the GFT basis [cf. the discussion following (1)]. From this vantage point, the graph-learning problem can be equivalently viewed as one of finding efficient information-processing transforms for graph signals. Second, smoothness is a cornerstone property at the heart of several graph-based statisticSt644c1phsi in99(e)3(a)-0.030(l -1.u)-2(d3(o)12()3(a)-19()17)2(n9n)aphitinr63(f)]TJ-0.015 Tc a4moo-21(6(i)-157

the other hand, small eigenvalues associated with low frequencies are translated to high-power factor loadings—a manifestation of the model imposing a smoothness prior on $\mathbf{x}$.

Given the observed signal $\mathbf{x}$, the maximum a posteriori (MAP) estimator of the latent variables is given by (^{2} is subsequently assumed known and absorbed into the parameter $\alpha > 0$)

$$\boldsymbol{\chi}_{\text{MAP}} = \arg\min_{\boldsymbol{\chi}} \left\{ \| \mathbf{x} - \mathbf{V}\boldsymbol{\chi} \|^2 + \alpha \boldsymbol{\chi}^T \quad \boldsymbol{\chi} \right\}, \tag{19}$$

which is, of course, parameterized by the unknown eigenvectors and eigenvalues of the Laplacian. With $\mathbf{y} := \mathbf{V}\boldsymbol{\chi}$ denoting the predicted graph signal (or error-free representation of $\mathbf{x}$), it follows that [cf. (1)]

$$\boldsymbol{\chi}^T \quad \boldsymbol{\chi} = \mathbf{y}^T \mathbf{V} \quad \mathbf{V}^T \mathbf{y} = \mathbf{y}^T \mathbf{L}\mathbf{y} = \text{TV}(\mathbf{y}). \tag{20}$$

Consequently, one can interpret the MAP estimator (19) as a Laplacian-based TV denoiser of $\mathbf{x}$, which effectively imposes a smoothness prior on the recovered signal $\mathbf{y} = \mathbf{V}\boldsymbol{\chi}$. One can also view (19) as a kernel ridge-regression estimator with (unknown) Laplacian kernel $\mathbf{K} := \mathbf{L}$ [29, Sec. 8.4.1].

Building on (19) and making the graph topology an explicit variable in the optimization, the idea is to jointly search for the graph Laplacian $\mathbf{L}$ and a denoised representation $\mathbf{y} = \mathbf{V}\boldsymbol{\chi}$ of $\mathbf{x}$, thus solving

$$\min_{,} \quad \mathbf{x} \quad \mathbf{y} \qquad \mathbf{y} \ \mathbf{L}\mathbf{y} \ .$$

in the "Graph Filters as Models of Network Diffusion" section. As we will see, this is a more general model where we require

Diffused nonstationary graph signals

We now deal with more general nonstationary signals $\mathbf{x}$ that adhere to linear diffusion dynamics $\mathbf{x} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l \mathbf{w} = \mathbf{H}\mathbf{w}$, but where the input covariance $\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbb{E}[\mathbf{w}\mathbf{w}^T]$ can be arbitrary. In other words, we relax the assumption of $\mathbf{w}$ being white, which led to the stationary signal model dealt with so far [cf. Definition 2 and (30)]. Such a model is, e.g., relevant to (geographically) correlated sensor network data or to models of opinion dynamics, where (even before engaging in discussion) the network agents can be partitioned into communities according to their standing on the subject matter.

For generic (nonidentity) $\boldsymbol{\Sigma}_{\mathbf{w}}$ we face the challenge that the signal covariance [cf. (6)]

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbf{H}\,\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{H}^T \tag{35}$$

is no longer simultaneously diagonalizable with $\mathbf{S}$. This rules out using the eigenvectors of the sample covariance $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ as eigenbasis of $\mathbf{S}$, as proposed in Step 1 for the stationary case. Still, observe that the eigenvectors of the shift coincide with those of the graph filter $\mathbf{H}$ that governs the underlying diffusion dynamics. This motivates adapting Step 1 in the "Step 1: Inferring the Eigenvectors" section when given observations of nonstationary graph processes. Simply put, the approach in [56] is to use snapshot observations $\mathbf{X}$ together with additional (statistical) information on the excitation input $\mathbf{w}$ to identify the filter $\mathbf{H}$, with the ultimate goal of estimating its eigenvectors $\mathbf{V}$. These estimated eigenvectors $\mathbf{V}$ are then used as inputs to the shift identification problem (32), exactly as in the robust version of Step 2 in the "Step 2: Inferring the Eigenvalues" section. Accordingly, the focus is placed on the graph filter (i.e., system) identification task; see Figure 2(b).

Identification of the graph filter $\mathbf{H}$ from nonstationary signal observations is studied in detail in [56], for various scenarios that differ on what is known about the input process $\mathbf{w}$. Of particular interest is the setting where realizations of the excitation input are challenging to acquire, but information about the statistical description of $\mathbf{w}$ is still available. Concretely, consider M different excitation processes that are zero mean and their covariance $\boldsymbol{\Sigma}_{\mathbf{w},m} = \mathbb{E}[\mathbf{w}_m \mathbf{w}_m^T]$ is known for all $m = 1, \ldots, M$. Further suppose that for each input process $\mathbf{w}_m$ we have access to a set of independent realizations $\mathcal{X}_m = \{\mathbf{x}_m^{(p)}\}_{p=1}^{P_m}$

with (37). For $M > 1$, the set of feasible solutions to the system of (36) is naturally given by $\mathsf{H}_{1:M}^{\text{sym}} = \bigcap_{m=1}^{M} \mathsf{H}_{m}^{\text{sym}}$.

If only empirical covariances $\{\hat{\mathbf{x}}_{,m}\}_{m=1}^{M}$ are available, (39) can be leveraged to define the matrices $\mathbf{A}_m := \hat{\mathbf{w}}_{,m}^{-1/2} \mathbf{V}_{\mathbf{wxw},m} \odot$

$\boldsymbol{\Theta} = \mathbf{L}$ [cf. (1)]. This connection notwithstanding, the optimization problems (24) and (29) are motivated by smooth signal priors, but there is no explicit generative model for the observations, unlike the graphical models in the "Statistical Methods for Network Topology Inference" section, which have a clear interpretation.

signal in the graph spectral domain [6]. For image restoration tasks such as denoising and deblurring, a major challenge is how to design appropriate signal priors to regularize otherwise ill-posed inverse problems. Learning graph Laplacians that endow the signal representations with desired sparsity or smoothness properties is thus well motivated and an active area of research.

Increasingly, applications call for learning graph representations of dynamic, multiaspect data, possibly accounting for nonlinear and directional (causal) effects among nodal signals. While a thorough treatment is beyond the scope of this article, for completeness, we offer a brief account in the next section. For a comprehensive survey of these emerging topics, the reader is referred to [15].

Emerging topic areas

Thus far, the focus has been on learning static and undirected graphs from data. In this section, we first consider the identification of digraphs given nodal time series, which is intimately related to the problem of causal inference. We then cross the boundary of linear time-invariant network models and outline recent advances for tracking topologies of dynamic graphs as well as mechanisms to account for nonlinear pairwise interactions among vertex processes.

Digraphs and causality

Undirected graphs, like correlation networks, can inform proximity between nodal signals but cannot inform causality. Here we will lift the assumption that graph-shift operators are symmetric and consider estimation of digraphs with the intent of inferring causality from snapshot observations.

To that end, structural equation modeling encapsulates a family of statistical methods that model causal relationships between interacting variables in a complex system. This is pursued by estimating linear relationships among endogenous as well as exogenous traits, and symmetric structural equation models (SEMs) have been extensively adopted in economics, psychometrics, social sciences, and genetics, among others; see, e.g., [27]. The appeal of SEMs can be attributed to simplicity and the inherent ability to capture edge directionality in graphs, represented through a (generally) asymmetric adjacency matrix

central to popular digraph topology identification approaches

also touches upon prediction of (nonlinear and dynamic) processes supported on graphs.

Applications

This section presents numerical tests conducted on real data to demonstrate the effectiveness of selected graph-learning methods, ranging from ad hoc thresholding-based network constructions all the way to algorithms for identification of di, time-varying graphs. Through test cases, we show impact to diverse application domains including the economy, computational biology, neuroscience, and online social media.

Efficient representation of signals supported on a network of U.S. economic sectors

The Bureau of Economic Analysis of the U.S. Department of Commerce publishes a yearly table of inputs and outputs organized by economic sectors. More precisely, we have a set of 62 industrial sectors as defined by the North American Industry Classification System and a similarity function $\mathbf{W} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}_{+}$, where W_{ij} represents how much of the production of sector i, expressed in trillions of U.S. dollars per year, was used as an input of sector j on average from 2008 to 2010. Moreover, for each sector we are given two economic markers: the added value (VA) generated and the level of pro-
16(-6-1(r2w)23(e a)-29(r)-6(e g)-13(iJ0.075 T)-13() g)11 1 1)-1)4(o)5(m)Qnoma
duction 0(s)23(:6:C91(f)-24(i)-16(n)-1(e3 Td[((n)3(d t)-20(a8(n)3[3c)1-24(i)-16(n)-1)4(o)5(m)QBT/T11 1 T(cT..(e)13f)-24f p)-4(r)-4(o)]TJ

agnostic to the form of the filter [53]. This naturally leads back to formulation (31), with $\mathbf{V}$ given by the eigenvectors of $\mathbf{T}$ and $f(\mathbf{S})$ chosen as an edge sparsity-promoting criterion. Different from the problem dealt with in the "Identifying the Structure of Network Diffusion Processes" section, note that matrix $\mathbf{T}$ is not necessarily a covariance matrix.

Consider identifying the structural properties of proteins from a mutual information graph of the covariation between the constitutional amino acids [13], [53]. Pictorially, for a particular protein we want to recover the structural graph in the top left of Figure 4(a) when given the graph of mutual information in the top right corner. The graph recovered by network deconvolution [13] is illustrated in the bottom left corner of Figure 4(a), whereas the one recovered using the approach in (31) (with the sparsity-promoting $f(\mathbf{S}) = \|\mathbf{S}\|_1$) is depicted in the bottom right corner. The comparison of the recovered graphs demonstrates that using a general filter model translates to a sparser graph that captures more accurately the desired structure. To quantify this latter assertion, Figure 4(b) depicts the fraction of the real contact edges recovered for each method as a function of the number of edges considered. For example, if for a given method the 100 edges with largest weight in the recovered graph contain 40% of the edges in the ground-truth graph, we say that the 100 top edge predictions achieve a fraction of recovered edges equal to 0.4. From Figure 4(b) it follows that the method in the "Learning Graphs From Obser-

values $u_i^{(c)}$ were uniformly sampled over the interval [0, 0.01]; see [1].

The algorithm in [1] was run on the data set, and Figure 6(a) and (b) shows visualizations of the inferred network at $t = 10$ and $t = 40$ weeks. Speculation about the possible successor of the dying North Korean ruler, Kim Jong-il, rose until his death on 17 December 2011 (week 38). He was succeeded by Kim Jong-un on 30 December 2011 (week 40). The visualizations show an increasing number of edges over the 45 weeks, illustrating the growing interest of international news websites and blogs in the new ruler, about whom little was known in the first 10 weeks. Unfortunately, the observation horizon does not go beyond $T = 45$ weeks. A longer span of data would have been useful to investigate the rate at which global news coverage on the topic eventually subsided. Figure 6(c) depicts the time evolution of the total number of edges in the inferred dynamic network. Of particular interest are the weeks during which 1) Kim Jong-un was appointed as the vice chair of the North Korean military commission; 2) Kim Jong-il died; and 3) Kim Jong-un became the ruler of North Korea. These events were the topics of many online news articles and political blogs, an observation that is reinforced by the experimental results shown in the plot.

Concluding remarks and research outlook

Contending that GSP prl8eTw i5t is rmgChu letae rer3 Tm[(C)(i)10(c)3(s)1315( )]TJ0.0nd ic21(u)-1(rd(c21d2(r3 t)-19(a)-22(16

nonlinear interactions for network data such as those given by median or other nonlinear graph filters.

In terms of computational complexity, there is room for improving scalability of some of the algorithms described via parallelization and decentralized implementations. Moreover, adaptive algorithms that can track the (possibly) time-varying structure of the network and achieve both memory and computational savings by processing the signals on the fly are naturally desirable, but so far largely unexplored.

Finally, one can explore the links between network deconvolution—as described in the "Identifying Protein Structure via Network Deconvolution" section—and graph sparsification approaches. The latter consists on approximating a given