NETWORK TOPOLOGY INFERENCE WITH GRAPHON SPECTRAL PENALTIES

T. Mitchell Roddenberry, Madeline Navarro, and Santiago Segarra

Rice University, Electrical and Computer Engineering, Houston, TX

ABSTRACT

We consider the problem of inferring the unobserved edges of a graph from data supported on its nodes. In line with existing approaches, we propose a convex program for recovering a graph Laplacian that is approximately diagonalizable by a set of eigenvectors obtained from the second-order moment of the observed data. Unlike existing work, we incorporate prior knowledge about the distribution from where the underlying graph was drawn. In particular, we consider the case where the graph was drawn from a graphon model, and we supplement our convex optimization problem with a provably-valid regularizer on the spectrum of the graph to be recovered. We present the cases where the graphon model is assumed to be known and the more practical setting where the relevant features of the model are inferred from auxiliary network observations. Numerical experiments on synthetic and real-world data illustrate the advantage of leveraging the proposed graphon prior, even when the prior is imperfect.

Index Terms— Network topology inference, spectral graph theory, graph signal processing, graphon.

1. INTRODUCTION

Networks have emerged as an effective tool for analyzing and summarizing complex systems in science and engineering [1–3]. The abstraction of these systems as graphs enables one to examine local structures of influence among agents via direct observation of the edge structure, as well as global behavior via the spectral features of algebraic representations of these graphs. The algebraic properties of networks are used to succinctly describe their behavior, such as the centrality structure [4], community structure [5], or conductance [6]. In some settings, a network is known to the user or can be simply constructed, *e.g.* a rail transportation network, where each station is a node connected by train routes modeled by edges. However, the network structure must often be inferred from the system behavior itself. For instance, the system of functional relationships between regions of the brain may not be known directly, but needs to be inferred from observed neural activity. This is an example of the well-established problem of *network topology inference*.

In this work, we seek to *infer the structure of the graph Laplacian* from so-called *spectral templates* [7], *i.e.*, by first estimating the eigenvectors of the Laplacian matrix from the observed data and then selecting the eigenvalues that yield a graph Laplacian with desirable features. Work in this direction has studied how the structure of a graph shapes data supported on it and how to invert that process for network topology inference, but has neglected the structure endowed by common statistical models of real-world graphs. For instance, wireless sensor networks are well-modeled as geometric

random graphs [8], and social networks often follow power-law degree distributions [9]. In this work, we aim to imbue the network topology inference pipeline with this type of domain knowledge. As a flexible random graph model, we assume the inferred graph is drawn from a graphon, a non-parametric model for large random graphs [10, 11]. Graphons have proven to be an effective model for understanding large graphs, acting as a limit object for the graphs they generate [12–15]. In this work, we leverage convergence results on graphs generated by graphons [16] to regularize the spectral structure of the inferred graph Laplacian.

Related work. The problem of network topology inference is well-studied from a variety of perspectives, each incorporating different assumptions on the data model and imposing characteristics on the graph structure. The statistical perspective represents each node as a random variable, where the graph structure between them captures conditional dependencies [17]. Network inference then seeks to find a sparse inverse covariance (precision) matrix [18, 19]. Indeed, the notion of incorporating constraints on the combinatorial structure of the inferred graph in the context of probabilistic graphical models was studied in [20]. This differs from our work here, where the graph is assumed to be drawn from a distribution exhibiting certain combinatorial properties in a statistical sense, but lacking hard combinatorial constraints.

On the other hand, modeling network processes as the description of a physical system yields another set of methodologies. In these works, topology inference is performed under models of systems related to information diffusion, random walks, and related network processes [21, 22]. These techniques are closely related to those in the graph signal processing literature [23], where the observed data is modeled as a set of graph signals, which are assumed to be smooth on the underlying graph [24, 25], or the output of a graph filter applied to excitation noise [7, 26], perhaps modeling heat diffusion [27] or consensus dynamics [28, 29]. Although our observation model is in line with this body of work, the novelty lies in the incorporation of prior statistical knowledge of the graph to be recovered in the form of a graphon model.

Contribution. The contributions of this paper are twofold:
i) We propose an efficient convex optimization problem for network topology inference with a novel spectral shrinkage penalty that leverages the relationship between the *spectrum of the Laplacian* and the *degree distribution of graphs drawn from graphons*.
ii) We then demonstrate that this approach has practical utility beyond the case where the underlying graphon is known, by substituting the limiting (graphon) degree distribution with empirically obtained degree statistics.

2. PRELIMINARIES

Notation. The notation $[n]$ refers to the set of positive integers $\{1, 2, \ldots, n\}$. We refer to matrices using bold uppercase letters,

e.g. $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and to (column) vectors with bold lowercase letters, *e.g.* $\mathbf{v}, \mathbf{w}, \mathbf{x}$. Entries of matrix $\mathbf{A}$ are indicated by A_{ij} while those of vector $\mathbf{x}$ are denoted by x_i. For clarity, we alternatively use the notation $[\mathbf{x}]_i = x_i$. The ℓ_2-norm of a vector is denoted by $\| \cdot \|_2$. The set of symmetric $n \times n$ matrices is denoted by $\mathbb{S}^{n \times n}$. We use the notation $\| \cdot \|_{1,1}$ to denote the element-wise ℓ_1-norm of a matrix, and $\| \cdot \|_F$ for the Frobenius norm. For a linear integral operator W on $L^2[0,1]$, its *operator norm* is specified by $\|\|W\|\|$:

$$\|\|W\|\| = \sup_{\substack{f \in L^2[0,1] \\ \|f\|_2 = 1}} \int_{[0,1]^2} W(x,y) f(x) \, \mathrm{d}x \, \mathrm{d}y. \quad (1)$$

Graphs, graph signals, and graph filters. A *graph* is a finite set of nodes, coupled with a set of edges connecting those nodes. That is, for a set of n nodes denoted by $\mathcal{V}$, the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ forms the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Typically, we endow the set of nodes with arbitrarily ordered integer labels, saying that $\mathcal{V} = [n]$. This representation allows us to represent the graph with the *adjacency matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$, where $A_{ij} = 1$ if $(i,j) \in \mathcal{E}$, taking value 0 otherwise. A graph is *undirected* if for any two nodes i, j, $(i,j) \in \mathcal{E}$ implies that $(j,i) \in \mathcal{E}$: that is, $\mathcal{E}$ is composed of unordered tuples of nodes. Under this condition, the adjacency matrix is symmetric. In this setting, we define the *Laplacian matrix* as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D} = \mathrm{diag}(\mathbf{A1})$ is the diagonal matrix of node degrees.

We model data on the nodes of a graph as a *graph signal*. A graph signal is a real-valued function on the nodes of a graph $x : \mathcal{V} \to \mathbb{R}$. If we endow the set of nodes with a labeling of positive integers so that $\mathcal{V} = [n]$, then a graph signal x has a natural representation $\mathbf{x}$ in $\mathbb{R}^n$, where $[\mathbf{x}]_i = x(i)$ for each $i \in [n]$.

With graph signals represented as vectors in $\mathbb{R}^n$, we now define *graph filters* as linear maps between them. Assuming that $\mathbf{L}$ has the eigenvalue decomposition $\mathbf{L} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, a graph filter is a real polynomial of $\mathbf{L}$, *e.g.* for coefficients α_k,

$$\mathcal{H}(\mathbf{L}) = \sum_{k=0}^{T} \alpha_k \mathbf{L}^k = \sum_{i=1}^{n} h(\lambda_i) \mathbf{v}_i \mathbf{v}_i^\top, \quad (2)$$

where $h : \mathbb{R} \to \mathbb{R}$ is the extension of $\mathcal{H}$ to the real numbers.[1] Notice that graph filters preserve the eigenvectors of the underlying Laplacian, while distorting the eigenvalues.

Graphons as random graph models. A *graphon* is a symmetric, measurable function from the unit square $[0,1]^2$ to the unit interval $[0,1]$, with the set of all such functions denoted by $\mathcal{W}$ [10]. For a graphon $W \in \mathcal{W}$, we sample a graph of size n by first drawing *i.i.d.* uniform random variables $\{\zeta_i\}_{i=1}^n \sim \mathcal{U}[0,1]$. Then, the adjacency matrix of our drawn graph is determined by setting $A_{ij} = A_{ji} = 1$ with probability $W(\zeta_i, \zeta_j)$, and $A_{ij} = A_{ji} = 0$ otherwise. A graphon determines a probability distribution over graphs of fixed size in this way. For instance, the constant graphon $W(x,y) = p$ for all $x, y \in [0,1]$ yields Erdős-Rényi graphs of size n and edge density p. From this sampling method, one can see that any measure-preserving bijection $\tau : [0,1] \to [0,1]$ yields an equivalence relation between graphons. That is, if for all $x, y \in [0,1]$ we say that $W^\tau(x,y) = W(\tau(x), \tau(y))$, then $W^\tau \sim W$. Additionally, if two graphons W^0, W^1 are equal almost everywhere, we also say that $W^0 \sim W^1$.

---

[1]Graph filters can be defined in terms of generic graph shift operators [30], but here we focus on those defined based on $\mathbf{L}$.

3. NETWORK TOPOLOGY INFERENCE

In order to formally define our problem, we need to assume a model on how the underlying graph structure shapes the data on its nodes. Without such a relationship, inferring the graph from the observed graph signals would be a hopeless endeavor. To achieve this, we leverage the versatility of graph filters in representing network processes [30] and model the observed signals as the outputs of graph filters. For simplicity we focus on a class of network processes corresponding to discrete-time consensus dynamics, a well-studied model for network topology inference [28, 29]. More precisely, we model our observed graph signals $\mathbf{y} \in \mathbb{R}^n$ as being generated by

$$\mathbf{y} = \mathcal{H}(\mathbf{L}) \mathbf{w} = \prod_{k=1}^{T} (\mathbf{I} - \alpha_k \mathbf{L}) \mathbf{w}, \quad (3)$$

with unknown coefficients $0 \le \alpha_k \le \lambda_{max}(\mathbf{L})$ and $\mathbf{w}$ being an unknown random vector drawn from a zero-mean distribution with identity covariance. With this notation in place, we can formally state our problem.

Problem 1. *Consider an unknown graph $\mathcal{G}$ drawn from a known graphon W. Given a collection of graph signals $\{\mathbf{y}^\ell\}_{\ell=1}^m$ defined in $\mathcal{G}$ following the model in* (3)*, estimate the Laplacian $\mathbf{L}$ of $\mathcal{G}$.*

Given perfect knowledge of a graph filter $\mathcal{H}(\mathbf{L})$, it follows from (2) that the eigenvectors of $\mathbf{L}$ can be readily obtained from an eigendecomposition of $\mathcal{H}(\mathbf{L})$. Moreover, even if the filter is unknown, we can estimate $\{\mathbf{v}_i\}_{i=1}^n$ from the observations in $\{\mathbf{y}^\ell\}_{\ell=1}^m$ as in previous works [7, 23]. In a nutshell, a simple computation of the covariance $\mathbf{C_y}$ of the signals in (3) reveals that

$$\mathbf{C_y} = \mathcal{H}^2(\mathbf{L}) = \sum_{i=1}^{n} h^2(\lambda_i) \mathbf{v}_i \mathbf{v}_i^\top, \quad (4)$$

so that the covariance matrix of $\mathbf{y}$ shares a set of eigenvectors with $\mathbf{L}$. Thus, from the observed signals $\{\mathbf{y}^\ell\}_{\ell=1}^m$ we may estimate $\mathbf{C_y}$ via a sample covariance estimator $\widehat{\mathbf{C}}_\mathbf{y}$ and obtain approximate eigenvectors (or spectral templates) $\{\widehat{\mathbf{v}}_i\}_{i=1}^n$. Note that in obtaining these estimated spectral templates we have not used the knowledge of the specific form of the network process in (3) nor any knowledge of the graphon W. In Section 3.1, we present a convex optimization problem that takes as inputs $\{\widehat{\mathbf{v}}_i\}_{i=1}^n$ and these additional sources of information to provide an estimate of $\mathbf{L}$, as desired in Problem 1.

Remark 1 (Alternative problem formulations). Although we focus on Problem 1 as stated, several variations and relaxations can be tackled by incorporating the methodology proposed here with existing approaches. For instance, knowledge of the coefficients α_k in (3) can be further exploited as in [29], and the identity covariance of $\mathbf{w}$ can be relaxed as in [31, 32]. Moreover, since the signals $\{\mathbf{y}^\ell\}_{\ell=1}^m$ are used to obtain the estimated eigenvectors $\{\widehat{\mathbf{v}}_i\}_{i=1}^n$, the optimization method proposed in the next section can be directly applied given any other source of approximated spectral templates [7]. Lastly, we might consider the case where the graphon W is also unknown. This empirically-relevant variant is discussed toward the end of Section 3.1 and illustrated in Section 4.

3.1. Inferring the Laplacian with graphon penalties

We now consider how knowledge of the graphon W from which a graph was drawn can inform our inferred Laplacian spectrum. The view of graphons as generative models for dense, simple graphs

yields probabilistic bounds relating properties of graphs drawn from a graphon and the analogous property of the graphon itself. For instance, it is stated in [11, Corollary 10.4] that the density of a motif in a sufficiently large graph concentrates around the density of that motif in the graphon from which it was drawn, *e.g.* the homomorphism density of triangles in a graph:

$$\frac{1}{n^3}\operatorname{tr}\left(\mathbf{A}^3\right) \xrightarrow{n} \int_{[0,1]^3} W(x,y)\,W(y,z)\,W(z,x)\ \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \quad (5)$$

In particular, the following result characterizes the *spectrum of the Laplacian matrix* of large graphs drawn from a graphon.

Theorem 1 (From [16])**.** *Let W be a piecewise-Lipschitz graphon such that $W(x_1,y) \leq W(x_2,y)$ when $x_1 \leq x_2$, and the infimum of W over $[0,1]^2$ is bounded away from 0. Then, for a graph $\mathcal{G}$ of size n sampled from W, where $0 = \lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of the Laplacian matrix of $\mathcal{G}$, with probability at least $1 - 3\nu$,*

$$\|\mu(x) - d(x)\|_2 \leq C_0 + \sqrt[4]{\frac{2}{n}}\sqrt{\|W\| + C_1} + C_1, \quad (6)$$

where $\mu(x) = \lambda_{\lfloor nx \rfloor + 1}/n$ for all $x \in [0,1)$, $\mu(1) = \lambda_n/n$, $d(x) = \int_0^1 W(x,y)\mathrm{d}y$, and C_0, C_1 are dependent on ν, the Lipschitz smoothness of W, and the infimum of W.

Notice that $\mu(x)$ is a piecewise constant function given by the eigenvalues of the Laplacian and the degree function $d(x)$ is the analogous of a degree sequence defined directly on the continuous graphon [15]. Informally, Theorem 1 states that the ordered Laplacian eigenvalues concentrate near the ordered degree function as n grows, especially for smooth graphons whose values are bounded far away from zero. It is key to notice that this result does not depend on knowing the particular degrees of the nodes in the graph nor their latent positions $\zeta_i \in [0,1]$ in the sampling procedure (cf. Section 2).

With Theorem 1 in hand, we propose a method for the recovery of the graph Laplacian, with tolerance for imperfections in the spectral templates brought about by noise or finite sampling effects. We combine an ℓ_1-penalty on the inferred Laplacian matrix to promote edge-sparsity [7, 29], as well as a spectral shrinkage penalty towards a prior degree function. One key property of the class of filters in (3) is that the extension of $\mathcal{H}$ to the real numbers $h : \mathbb{R} \to \mathbb{R}$ is *monotonically decreasing* and *non-negative* in the spectrum of $\mathbf{L}$, so that it preserves, in reverse, the ordering of the eigenvalues of $\mathbf{L}$, even after being squared as in (4). This is captured in the following convex program, where the spectral templates $\{\widehat{\mathbf{v}}_i\}_{i=1}^n$ are used as an approximate ordered eigenbasis for the inferred matrix,

$$\mathbf{L}^*, \boldsymbol{\lambda}^* = \underset{\mathbf{L},\boldsymbol{\lambda}}{\operatorname{argmin}} \|\mathbf{L}\|_{1,1} + \beta\|\mu(x) - d(x)\|_2^2 \quad (7)$$

$$\text{s.t.}\quad \rho\left(\mathbf{L}, \sum_{i=1}^n \lambda_i \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i^\top\right) \leq \epsilon, \ \mathbf{L} \in \mathcal{L},$$

$$\mu(x) = \lambda_{\lfloor nx \rfloor + 1}/n \quad \text{for all } x \in [0,1),$$

$$\mu(1) = \lambda_n/n,$$

$$\lambda_i \leq \lambda_{i+1+\eta} \quad \text{for all } i \in [n-1-\eta],$$

where $\mathcal{L}$ indicates the set of valid graph Laplacians, the function $\rho : \mathbb{S}^{n \times n} \times \mathbb{S}^{n \times n} \to \mathbb{R}^+$ is a distance function on symmetric matrices, and ϵ dictates how far the eigenvectors of the inferred graph Laplacian can be from the approximated spectral templates. The parameter η allows for slack in the ordering of the eigenvalues, due to spectral perturbation in the estimation of the spectral templates. Finally, $\mu(x)$ and $d(x)$ are the same as in Theorem 1.

The program (7) jointly recovers a spectrum $\boldsymbol{\lambda}^*$ and a valid Laplacian that is close to $\boldsymbol{\lambda}^*$ in the eigenbasis $\{\widehat{\mathbf{v}}_i\}_{i=1}^n$. The spectrum is encouraged towards the degree function of the graphon based on Theorem 1, and sparsity is promoted in the Laplacian. We emphasize that shrinking the Laplacian spectrum towards the graphon degree function *does not* require knowledge of the degrees of particular nodes. That is, since the spectral shrinkage penalty operates only on the eigenvalues of $\mathbf{L}$, it is blind to permutations of the nodes.

Inference when the graphon is unknown. In practice, it may be unrealistic to assume that the graphon W is known. However, the rich statistical structures of this modeling assumption can still be leveraged for the application of (7). Indeed, via Theorem 1 we have distilled the signature of the graphon W down to its degree function $d(x)$. Therefore, if we can obtain an estimate of $d(x)$ from auxiliary information, then (7) can still be applied. If a sufficiently large graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is drawn from an unknown graphon W, then its degree distribution will be similar to that of W. In particular, suppose that a uniformly-sampled random subgraph of $\mathcal{G}$ is observed, where a subset of the nodes $\mathcal{V}_0 \subset \mathcal{V}$ induces the graph $\mathcal{G}_0 = (\mathcal{V}_0, \mathcal{E} \cap \mathcal{V}_0 \times \mathcal{V}_0)$. It is clear that $\mathcal{G}_0$ is also distributed according to the sampling procedure described in Section 2 applied to W. Therefore, the expected normalized degree distribution of $\mathcal{G}_0$ is approximately the graphon degree function $d(x)$, following naturally by considering the homomorphism density of a star graph as a way to query the degree distribution of a graph [33, Example 2.2].

With this in mind, we propose a modification to (7), where a uniformly-sampled induced subgraph $\mathcal{G}_0$ on n_0 nodes of $\mathcal{G}$ is observed, with ordered vector of degrees $\mathbf{d}_0 = \operatorname{sort}(\mathbf{A}_0 \mathbf{1})$. In this case, the natural approximate degree function is $\widehat{d}(x) = [\mathbf{d}_0]_{\lfloor n_0 x \rfloor + 1}/n_0$ for $0 \leq x < 1$, and $\widehat{d}(1) = [\mathbf{d}_0]_{n_0}$, which we use to then solve (7) with $d(x)$ replaced by $\widehat{d}(x)$. Finally, even if the observation of a subgraph $\mathcal{G}_0$ is infeasible, the assumption of a graphon distribution can be used to incorporate information from similar graphs potentially supported on a different set of nodes. For instance, when inferring the topology of a particular brain network, using the degree distribution of another, better-mapped brain would be a reasonable approximation of the assumed graphon degree function W for use in (7).

4. EXPERIMENTS

We demonstrate the utility of incorporating prior information on the graph model when inferring both real and synthetic networks.[2] Since we are interested in recovering the network structure and not the scale of $\mathbf{L}$, we normalize all graph matrices to have unit Frobenius norm before computing the error. That is, for a ground-truth Laplacian $\mathbf{L}$ and an estimated Laplacian $\mathbf{L}^*$, our error function is

$$\operatorname{err}(\mathbf{L}^*, \mathbf{L}) = \left\| \frac{\mathbf{L}^*}{\|\mathbf{L}^*\|_F} - \frac{\mathbf{L}}{\|\mathbf{L}\|_F} \right\|_F. \quad (8)$$

To minimize confounders when comparing methods, in estimating $\mathbf{L}^*$ we do not include any thresholding or rounding operation. In solving (7), we use the distance $\rho(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_F$, and set $\eta = 0$ for the first experiment where the eigenvectors are perfectly known and $\eta = 2$ for all other experiments.

Convergence with graph size. As stated in Theorem 1, the Laplacian spectrum concentrates near the graphon degree distribution as the sampled graph size grows. To illustrate this, we estimate synthetic graphs of increasing sizes $n \in \{20, 40, 60, 80, 100\}$ drawn

---

[2]Code available at git.roddenberry.xyz/graphon-nti

5392
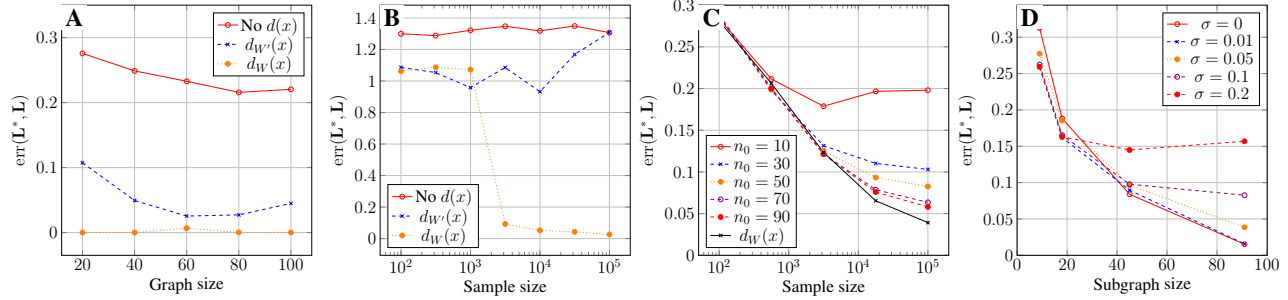
Fig. 1. **(A)** Convergence of the inferred Laplacian for increasing graph size. 'No $d(x)$' indicates the absence of a spectral shrinkage penalty, as in [29, Algorithm 3]. **(B)** Laplacian inference using noisy spectral templates and a graphon degree prior. **(C)** Inference of a synthetic graph Laplacian from noisy spectral templates using subgraph degree statistics. **(D)** Denoising of the Rhesus Macaque cerebral cortex network using subgraph degree statistics.

from the graphon $W(x,y) = \frac{1}{2}\gamma(x^2 + y^2) + (1 - \gamma)$, where $\gamma = 0.7$. Three optimization programs are compared for graph estimation: i) Solving (7) for $d(x) = d_W(x)$ computed based on the true graphon W, ii) Solving (7) for $d(x) = d_{W'}(x)$ computed based on the alternate graphon $W'(x,y) = \frac{1}{2}(1 - \gamma)(x^2 + y^2) + \gamma$, and iii) Solving (7) for $\beta = 0$, as proposed in [29]. For the first two methods, the weight β is chosen for optimal estimation for each graph size. As shown in Fig. 1A for all graph sizes, including a spectral prior achieves a consistently higher accuracy than using no spectral information, even for the imperfect case when W' is considered. Moreover, Laplacian inference with the true generating graphon degree distribution uniformly attains minimal error.

Noisy spectral templates. We demonstrate the estimation of a graph of size $n = 40$ drawn from the graphon $W(x,y) = \frac{1}{2}\gamma(x^2 + y^2) + (1 - \gamma)$ when noisy spectral templates are available. These noisy eigenvectors are obtained from sample covariance matrices computed based on a varying number of observed filtered graph signals (cf. discussion before Remark 1), where the filter is given by $\mathcal{H}(\mathbf{L}) = \mathbf{I} - \alpha\mathbf{L}$. The same three methods from the previous experiment are compared in Fig. 1B. Spectral shrinkage toward the true graphon degree distribution greatly improves estimation compared to the alternate graphon degree distribution or no spectral shrinkage. This effect becomes more conspicuous for larger number of observed signals, i.e., for the setting where the Laplacian eigenvectors are better estimated.

Incorporating subgraph degree statistics. While the previous two experiments assume knowledge of the graphon model, this might not be readily available in practice. However, if the graph was indeed drawn from a graphon, its subgraphs are distributed according to the graphon as well. In particular, we use the degree distribution of an observed subgraph for (7), as discussed at the end of Section 3.1. To understand the performance of using empirical degree distributions in (7), we do not constrain the recovered graph to match the observed subgraph, although this would be done in practice. This can also be interpreted as a valid illustration of the case where the degree distribution is estimated from a *different* graph (as opposed to a subgraph) drawn from the same (unknown) graphon.

Synthetic example. The results of the described procedure for graphs drawn from the graphon $W(x,y) = 1 - 0.8\max(x,y)$ and of size $n = 100$ are plotted in Fig. 1C. We obtain a set of noisy spectral templates using the same graph filter and increasing number of graph signals as in the previous experiment. Then, degree functions are extracted for subgraph sizes $n_0 \in \{10, 30, 50, 70, 90\}$, as well as the underlying graphon for comparison. Clearly, when the size of the observed subgraph increases, the estimated degree profile improves,

yielding better performance. Moreover, as the number of sampled graph signals m increases, the inferred network improves in quality as well, due to the more accurately estimated spectral templates. The experiment also reveals that, in the presence of noisy eigenvectors, approximate knowledge of the degree distribution can be almost as valuable as complete access to $d_W(x)$. The marginal value of the latter becomes apparent for large number of available signals.

Sparsifying a noisy brain network. We consider the connectome of the Rhesus Macaque's cerebral cortex [34, 35], where each node ($|\mathcal{V}| = 91$) is a region in the cortex and each edge ($|\mathcal{E}| = 1401$) corresponds to an interareal pathway. For $n_0 \in \{9, 18, 45, 91\}$ nodes, we randomly select n_0 nodes without replacement, and then compute the empirical degree function $\widehat{d}(x)$ of the subgraph as before. Then, we inject Gaussian noise of increasing variance into the adjacency matrix, and use the eigenvectors of the corresponding Laplacian as noisy spectral templates in (7).

As shown in Fig. 1D, as the subgraph size n_0 increases, the relative error between the recovered and true Laplacians decreases. This indicates the utility of the degree function for the shrinkage penalty, since a more accurate estimate of the degree distribution leads to better recovery performance. Additionally, as the power of the injected noise increases, the recovery performance degrades. This is intuitive, as a corrupted set of spectral templates will bound the performance of the recovery algorithm.

5. CONCLUSION AND FUTURE WORK

We considered the network topology inference problem through the lens of a graphon modeling assumption. We demonstrated how the spectral properties of graph Laplacians drawn from a graphon relate to the degree function of that graphon, and how this relationship can be used to recover the graph topology from so-called spectral templates. The proposed method based on this property was then extended to the case where the graphon is not known, but degree information is available from either a subgraph or an auxiliary graph.

Future research avenues will incorporate more of the rich structures and statistical properties afforded by graphon-based models. Specifically, rather than relying on a shrinkage penalty for the Laplacian spectrum, one can relate the spectral moments of the adjacency matrix to the homomorphism densities of cycles in a graphon [11]. Additionally, we will seek to generalize the considered penalties based on conventional graphons to the empirically relevant class of sparse exchangeable graph models [10]. Finally, we are currently investigating the use of spectral priors for the inference of higher-order network models, where edges are not strictly pairwise, but can relate arbitrarily large sets of nodes.

6. REFERENCES

[1] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.

[2] M. O. Jackson, *Social and Economic Networks*, Princeton university press, 2010.

[3] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, USA, Mar. 2010.

[4] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," Technical Report 1999-66, Stanford InfoLab, Nov. 1999.

[5] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[6] D. A. Levin and Y. Peres, *Markov chains and mixing times*, vol. 107, American Mathematical Soc., 2017.

[7] 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, 2017.

[8] H. Kenniche and V. Ravelomananana, "Random geometric graphs as model of wireless sensor networks," in *Intl. Conf. on Comp. and Auto. Eng. (ICCAE)*. IEEE, 2010, vol. 4, pp. 103–107.

[9] L. Muchnik, S. Pei, L. C. Parra, S. D. Reis, J. S. Andrade Jr, S. Havlin, and H. A. Makse, "Origins of power-law degree distribution in the heterogeneity of human activity in social networks," *Scientific Reports*, vol. 3, no. 1, pp. 1–8, 2013.

[10] C. Borgs and J. Chayes, "Graphons: A nonparametric method to model, estimate, and design algorithms for massive networks," in *Proc. of the 2017 ACM Conf. on Econ. and Comp.*, 2017, pp. 665–672.

[11] L. Lovász, *Large networks and graph limits*, vol. 60, American Mathematical Soc., 2012.

[12] M. W. Morency and G. Leus, "Signal processing on kernel-based random graphs," in *European Signal Process. Conf. (EUSIPCO)*. IEEE, 2017, pp. 365–369.

[13] C. E. Lee and D. Shah, "Unifying framework for crowd-sourcing via graphon estimation," *arXiv preprint arXiv:1703.08085*, 2017.

[14] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *IEEE Trans. Auto. Control*, vol. 65, no. 10, pp. 4090–4105, 2020.

[15] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 520–537, 2020.

[16] R. Vizuete, F. Garin, and P. Frasca, "The Laplacian spectrum of large graphs sampled from graphons," *arXiv preprint arXiv:2004.09177*, 2020.

[17] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.

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

[19] B. M. Lake and J. B. Tenenbaum, "Discovering structure by learning sparse graphs," in *Proc. of the 32nd Annual Cog. Sci. Conf.*, 2010, pp. 778–783.

[20] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar, "A unified framework for structured graph learning via spectral constraints.," *J. Mach. Learn. Res*, vol. 21, no. 22, pp. 1–60, 2020.

[21] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Trans. Knowl. Disc. Data (TKDD)*, vol. 5, no. 4, pp. 1–37, 2012.

[22] S. Shahrampour and V. M. Preciado, "Topology identification of directed dynamical networks via power spectral analysis," *IEEE Trans. Auto. Control*, vol. 60, no. 8, pp. 2260–2265, 2014.

[23] 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.

[24] 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, 2016.

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

[26] 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, 2017.

[27] 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, 2017.

[28] S. Shahrampour and V. M. Preciado, "Reconstruction of directed networks from consensus dynamics," in *American Control Conf. (ACC)*. IEEE, 2013, pp. 1685–1690.

[29] Y. Zhu, M. T. Schaub, A. Jadbabaie, and S. Segarra, "Network inference from consensus dynamics with unknown parameters," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 300–315, 2020.

[30] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, 2017.

[31] 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)*. IEEE, 2017, pp. 5870–5874.

[32] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Identifying the topology of undirected networks from diffused non-stationary graph signals," *arXiv preprint arXiv:1801.03862*, 2018.

[33] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, "Counting graph homomorphisms," in *Topics in Discrete Mathematics*, pp. 315–371. Springer, 2006.

[34] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Conf. Artif. Intell. (AAAI)*, 2015, pp. 4292–4293.

[35] N. T. Markov, M. M. Ercsey-Ravasz, A. R. Ribeiro Gomes, C. Lamy, et al., "A weighted and directed interareal connectivity matrix for macaque cerebral cortex," *Cerebral Cortex*, vol. 24, no. 1, pp. 17–36, 2014.