

GRAPHON-AIDED JOINT ESTIMATION OF MULTIPLE GRAPHS

Madeline Navarro and Santiago Segarra

Electrical and Computer Engineering, Rice University, USA

ABSTRACT

We consider the problem of estimating the topology of multiple networks from nodal observations, where these networks are assumed to be drawn from the same (unknown) random graph model. We adopt a graphon as our random graph model, which is a nonparametric model from which graphs of potentially different sizes can be drawn. The versatility of graphons allows us to tackle the joint inference problem even for the cases where the graphs to be recovered contain different number of nodes and lack precise alignment across the graphs. Our solution is based on combining a maximum likelihood penalty with graphon estimation schemes and can be used to augment existing network inference methods. We validate our proposed approach by comparing its performance against competing methods in synthetic and real-world datasets.

Index Terms— Network topology inference, graph learning, joint inference, graphon.

1. INTRODUCTION

Networks (or graphs) are powerful representations of complex information due to their ability to represent structure via dyadic relationships. Many fields of research utilize network structures for representing and analyzing complex data, such as ecology for predicting animal behavior [1], neuroscience for modeling relationships between neurons [2], and environmental science for discovering and predicting outcomes of climate relationships [3].

While networks are convenient and interpretable tools for tasks on complex data, knowledge of the underlying structure may be unavailable, as is the case for functional connectivity between brain regions [2], or the underlying network may be expensive to obtain, as with structural (anatomical) connectivity between neurons [4]. The inference of network connectivity from nodal observations is a ubiquitous problem that has been well studied in fields such as statistics [5] and signal processing [6]. Data-driven methods for the inference of network structure include graphical models [7, 8], structural equation models [9], and graph signal processing-based approaches [6, 10, 11].

In many of the above mentioned applications, it is often more important to infer the topology of multiple networks. For example, brain functional connectivity is a valuable tool for diagnosis, and the acquisition of multiple functional networks is necessary when considering multiple patients or scenarios [2]. Additionally, a prominent scenario requiring knowledge of multiple networks is when networks vary over time. An ecological example includes the estimation of evolving social networks for a species of interest [12].

In the case of inferring the topologies of multiple networks, separate estimation is a feasible methodology. However, in many scenarios a joint inference method may achieve better performance by

leveraging common structures between the graphs to be inferred. For instance, one would expect certain levels of similarities between the brain networks of different healthy individuals or between the same social network observed at different points in time. Prominent methods for multiple network inference include statistical approaches, primarily consisting of the joint estimation of Gaussian graphical models [13–17]. These methods typically involve modifications on the graphical lasso formulation with additional encouragement of structural similarity. Estimation of time-varying graphs is widely popular, as the relationship between graphs is typically straightforward to implement by considering that graph variation is smooth across time [18, 19]. The above methods for estimating multiple networks typically enforce similar structure, such as promoting similar sparsity patterns [20].

We consider the problem of *estimating the topology of multiple networks sampled from the same (unknown) random graph model*, where graphs have similar global structural characteristics inherited from the model. As our (nonparametric) network model we adopt a graphon [21], but we do not assume the specific graphon model to be known a priori [22]. While estimation of multiple networks is well-studied, to the best of our knowledge no previous method utilizes a shared graphon relationship to jointly estimate graphs of potentially different sizes.

Contributions. The contributions of our paper are threefold:

- 1) We present a methodology to *infer multiple networks that potentially lack node alignment and may have different sizes* by leveraging the assumption that graphs come from the same nonparametric network model.
- 2) We detail how this methodology can be combined with existing network inference methods, effectively providing a whole family of methods to solve the problem of interest.
- 3) Through numerical experiments in synthetic and real-world data we demonstrate the performance of our method in comparison with separate inference and competing joint inference algorithms.

2. PRELIMINARIES

Graph signal processing. We consider undirected, unweighted graphs of the form $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node (vertex) set \mathcal{V} of cardinality N and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The structure of a graph can be represented by its graph shift operator (GSO) [23, 24] as the matrix $\mathbf{S} \in \{0, 1\}^{N \times N}$, where $S_{ij} \neq 0$ if and only if the edge (i, j) exists in the network, and $S_{ij} = 0$ otherwise. We define graph signals as real-valued observations at each of the N nodes, represented by a vector $\mathbf{x} \in \mathbb{R}^N$. We may associate these nodal values with the graph topology via graph signal models. Choices for graph signal models include stationary signals that result from diffusion processes over the graph [23, 25] or as multivariate random numbers, where the graph structure represents statistical dependencies between variables [7, 8].

This work was partially supported by NSF under award CCF-2008555. Emails: nav@rice.edu, segarra@rice.edu

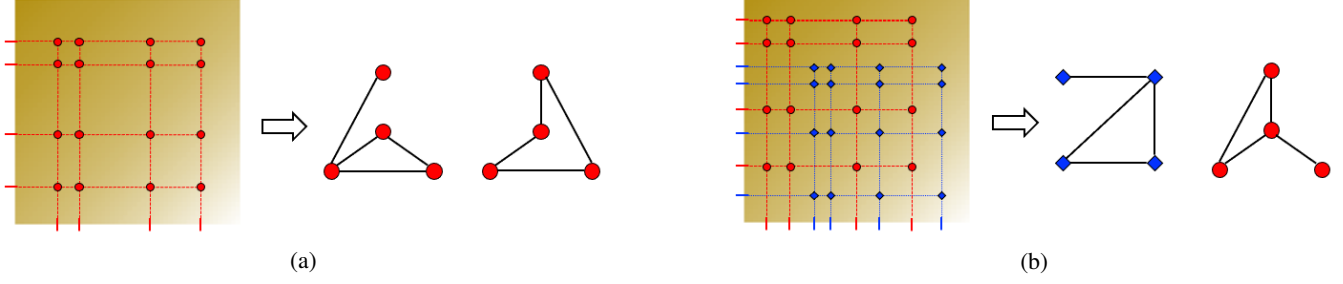


Fig. 1: Schematic depiction of the two problem modalities considered. (a) Multiple graphs sampled from the *same latent point sets* in the *same graphon*. Sampled graphs have not only the same size but also node alignment; see Section 4.1. (b) Multiple graphs sampled from *different latent point sets* in the *same graphon*. Sampled graphs may have different sizes; see Section 4.2.

Graphons. A graphon is a bounded symmetric measurable function $W : [0, 1]^2 \rightarrow [0, 1]$ whose domain can be interpreted as edges in an infinitely large adjacency matrix, while the range of W represents edge probabilities. By this definition, a graphon can be seen as a random graph model from which graphs with similar structural characteristics can be sampled [21, 26, 27]. Generating an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from a graphon W consists of two steps: (1) selecting a random value between 0 and 1 for each node, and (2) assigning an edge between nodes with probability equal to the value of the graphon at their randomly sampled points. Formally, the steps are as follows

$$\zeta_i \sim \text{Uniform}([0, 1]) \quad \forall i \in \mathcal{V}, \quad (1a)$$

$$S_{ij} = S_{ji} \sim \text{Bernoulli}(W(\zeta_i, \zeta_j)) \quad \forall (i, j) \in \mathcal{V} \times \mathcal{V}, \quad (1b)$$

where the latent variables $\zeta_i \in [0, 1]$ are independently drawn for each node i . This notion of graphon encompasses many commonly used exchangeable distributions on networks. Indeed, Erdős-Rényi graph models are represented via constant graphons [28] and stochastic block models via piecewise-constant graphons [29].

In our case, we assume that graphs are sampled from the same graphon, which is also unknown. Therefore, we propose a method to jointly estimate both the graphs and the underlying graphon. Estimating a graphon from observed adjacency matrices is a well-studied task [30, 31], and methods to infer a graphon from binary graphs include estimating the graphon function as a continuous smooth object [32, 33] along with the coarser stochastic blockmodel estimation [34, 35]. Other methods provide only the estimation of the graphon points $W(\zeta_i, \zeta_j)$ where it was sampled, which is equivalent to estimating a probability matrix [36].

3. PROBLEM STATEMENT

Consider a set of K different graphs $\{\mathcal{G}^{(k)}\}_{k=1}^K$ where the k -th graph has $N^{(k)}$ nodes. The set of undirected, unweighted adjacency matrices is represented by the set of GSOs $\mathbf{S} = \{\mathbf{S}^{(k)}\}_{k=1}^K$. Assume also that there is a set of graph signals provided for each graph, represented by $\mathbf{X}^{(k)} := [\mathbf{x}_1^{(k)} \cdots \mathbf{x}_{r_k}^{(k)}] \in \mathbb{R}^{N^{(k)} \times r_k}$, where the r_k columns contain the graph signals corresponding to the k -th graph. We further assume that all graphs are sampled from the same generative model, a graphon W . We present our problem as follows.

Problem 1 Given sets of observations $\mathbf{X} = \{\mathbf{X}^{(k)}\}_{k=1}^K$ for K graphs, find the adjacency matrices $\mathbf{S} = \{\mathbf{S}^{(k)}\}_{k=1}^K$ under the assumptions that (AS1) all graphs are sampled from the same (unknown) graphon W and (AS2) the latent point sets $\zeta^{(k)}$ in (1a) for each graph are known.

The first assumption (AS1) creates a relationship among the graphs, and with it we may improve estimation of graphs by jointly inferring the graph structures given their shared relationship. The second assumption (AS2) eliminates the identifiability problem for graphon estimation, where multiple graphons can lead to the same random graph distribution [26]. When all latent point sets are equivalent, i.e., $\zeta^{(k)} = \zeta$ for all $k \in \{1, 2, \dots, K\}$, (AS2) is equivalent to the assumption in previous joint network inference methods, where node alignment is present and known for all pairs of graphs. However, assuming possibly different known latent point sets is a weaker assumption than that of previous methods, as we do not require node alignment for the graphs.

The assumption (AS2) corresponds intuitively to situations of known sensor placement, such as known locations of electrode placement for neural response data collection or known climate regions to be observed. For example, the brain functional networks of multiple subjects may be measured by considering the same known brain regions or neurons across subjects [2]. Inferred graphs may also correspond to statistical interdependence between pairs of variables in a climate data set, where variables are measured at known spatial regions of earth [3].

4. GRAPHON-AIDED JOINT NETWORK ESTIMATION

In Sections 4.1 and 4.2 we tackle two versions of Problem 1 of increasing difficulty whereas in Section 4.3 we explain how these solutions can be combined with existing network inference methods.

4.1. Graphs and Probability Matrix Estimation

First consider the case where all graphs are sampled as in (1a) from the same points in the graphon space, that is, $\zeta^{(k)} = \zeta$ for all graphs $k \in \{1, 2, \dots, K\}$; see Fig. 1a. In practice, this case arises, e.g., when using the same sensor placement under multiple trials or experiments. Since we only consider edge probabilities in the graphon at points $(x, y) \in \zeta \times \zeta$, we need not consider the whole graphon W but only the probability matrix $\mathbf{T} \in [0, 1]^{N \times N}$ that contains the edge probabilities at the sampled points. The graphs $\mathbf{S}^{(k)}$ are then sampled from the same probability matrix \mathbf{T} , so the graphs all must have the same size, that is, $N^{(k)} = N$ for all $k \in \{1, 2, \dots, K\}$.

Given \mathbf{T} , the log-likelihood of a graph $\mathbf{S}^{(k)}$ is

$$\log \Pr [\mathbf{S}^{(k)} | \mathbf{T}] = \sum_{i < j} S_{ij}^{(k)} \log(T_{ij}) + (1 - S_{ij}^{(k)}) \log(1 - T_{ij}),$$

where we have leveraged the fact that, given \mathbf{T} , edges are drawn independently in our graph model [cf. (1b)]. Furthermore, we can

estimate each edge probability T_{ij} by the sample mean of the graph edges. Thus, we estimate the probability matrix \mathbf{T} as $\frac{1}{K} \sum_{k=1}^K \mathbf{S}^{(k)}$.

Recalling the notation for \mathbf{X} and \mathbf{S} from Problem 1, consider a generic optimization problem to estimate multiple networks that we formalize as

$$\min_{\mathbf{S}} f(\mathbf{S}, \mathbf{X}) + L(\mathbf{S}), \quad (2)$$

where the objective function $f(\mathbf{S}, \mathbf{X})$ estimates graph structures from the observed datasets, and $L(\mathbf{S})$ is an additional graph penalty or regularizer; in Section 4.3 we provide common examples for these functions. To solve our problem at hand, we propose to append the generic formulation in (2) with a negative log-likelihood penalty to obtain

$$\begin{aligned} \min_{\mathbf{S}, \mathbf{T}} \quad & f(\mathbf{S}, \mathbf{X}) + L(\mathbf{S}) - \sum_{k=1}^K \log \Pr [\mathbf{S}^{(k)} | \mathbf{T}] \\ \text{s.to} \quad & \mathbf{S}^{(k)} \in \mathcal{S}_A, \quad \mathbf{T} = \frac{1}{K} \sum_{k=1}^K \mathbf{S}^{(k)}, \end{aligned} \quad (3)$$

where we jointly estimate the graphs and their shared generative probability matrix \mathbf{T} . The estimation of the probability matrix entries is included as the sample mean of the edges in the graphs. The set \mathcal{S}_A enforces valid binary adjacency matrices, that is,

$$\mathcal{S}_A = \left\{ \mathbf{S} : \mathbf{S} = \mathbf{S}^\top, S_{ii} = 0, S_{ij} \in \{0, 1\} \right\},$$

where we consider undirected graphs without self-loops and edges that are unweighted.

As mentioned in the problem statement, the assumption $\zeta^{(k)} = \zeta$ for all $k \in \{1, 2, \dots, K\}$ is equivalent to node alignment for all graphs. We relate (3) to the task of estimating functional networks among the same brain regions of one subject under a set of discrete stimuli, or observing climate variables among the same geographical regions over several time instances.

4.2. Graphs and Graphon Estimation

We now consider the case where each graph is sampled from different latent point sets $\zeta^{(k)}$, and graphs may possibly have different sizes $N^{(k)}$; see Fig. 1b. Therefore, each graph is sampled from a potentially different probability matrix $\mathbf{T}^{(k)}$, which is the value of the graphon at the points $(x, y) \in \zeta^{(k)} \times \zeta^{(k)}$. The probability matrices provide estimates of the graphon at the known latent point pairs, and each graph provides information about the value of its respective probability matrix. We present an optimization problem to jointly estimate the graphs, the probability matrices, and the graphon as

$$\begin{aligned} \min_{\mathbf{S}, \mathbf{T}, W} \quad & f(\mathbf{S}, \mathbf{X}) + L(\mathbf{S}) - \sum_{k=1}^K \log \Pr [\mathbf{S}^{(k)} | \mathbf{T}^{(k)}] + g(W) \\ \text{s.to} \quad & \mathbf{S}^{(k)} \in \mathcal{S}_A, \\ & \mathbf{T}^{(k)} = h(\mathbf{S}^{(k)}), \quad T_{ij}^{(k)} = W(\zeta_i^{(k)}, \zeta_j^{(k)}), \\ & W : [0, 1]^2 \rightarrow [0, 1], \quad W(x, y) = W(y, x), \end{aligned} \quad (4)$$

where we include the same negative log-likelihood term as in (3), but each graph is associated with a different probability matrix $\mathbf{T}^{(k)}$. The function $h(\cdot)$ is a probability matrix estimation method that takes an adjacency matrix as input, such as network histogram or

stochastic block model approximations [33, 34]. The third constraint fits the graphon W at the known latent point pairs to the values of the probability matrices $\mathbf{T}^{(k)}$, and the regularization term $g(\cdot)$ in the objective imposes a prior on the overall graphon structure. For example, we may apply a thin-plate spline term [37] to estimate a smooth graphon assuming that points $\zeta^{(k)}$ are relatively evenly spaced throughout the interval $[0, 1]$.

The assumptions in (4) are weaker than those in (3), thus a wider range of applications are available. In the example of estimating brain functional networks, functional connectivity of the same subject may be inferred for different sets of brain regions. Additionally, climate network inference is often based on correlation or mutual inference measures, which decreases with geographic distance [3]. Thus, separating inference of climate networks into multiple networks of subregions and applying (4) may be more practical than estimating a single climate network for a large region, as connectivity is expected to be very low for far apart geographical locations.

4.3. Examples for Network Inference Methods

Up to this point, we have been considering a generic network inference problem in (2). Both the formulations in (3) and (4) are applicable to existing network inference methods through specific choices of functions $f(\mathbf{S}, \mathbf{X})$ and $L(\mathbf{S})$.

Consider examples for the function $f(\mathbf{S}, \mathbf{X})$ that relate observed graph signals to the structure of the graphs. Graph signals may be assumed to be smooth over their respective graphs [10], and we apply the penalty

$$f(\mathbf{S}, \mathbf{X}) = \sum_{k=1}^K \|\mathbf{S}^{(k)} \circ \mathbf{Z}^{(k)}\|_1, \quad (5)$$

where $\mathbf{Z}_{ij}^{(k)} = \|\mathbf{X}_i^{(k)} - \mathbf{X}_j^{(k)}\|^2$ as in [10]. Alternatively, we may have graph signals that are the diffusion of noise through a graph filter [6, 11]. In this case, we have stationary graph signals, where the signal covariance \mathbf{C} commutes with the adjacency matrix \mathbf{S} . Defining sample covariance matrices as $\mathbf{C}^{(k)} = \frac{1}{R^{(k)}} \mathbf{X}^{(k)} (\mathbf{X}^{(k)})^\top$, we can write

$$f(\mathbf{S}, \mathbf{X}) = \sum_{k=1}^K \|\mathbf{S}^{(k)} \mathbf{C}^{(k)} - \mathbf{C}^{(k)} \mathbf{S}^{(k)}\|_F^2. \quad (6)$$

In many applications, the graphs of interest are sparse, so it is common to apply a sparsity constraint for each graph [7, 11]. We may apply this with the penalty function $L(\mathbf{S})$ as

$$L(\mathbf{S}) = \sum_{k=1}^K \|\text{vec}(\mathbf{S}^{(k)})\|_1. \quad (7)$$

If, instead of separately inferring each graph, we wish to promote similar sparsity patterns, we may encourage edge similarity between graphs [17, 20] as

$$L(\mathbf{S}) = \sum_{k < k'} \|\text{vec}(\mathbf{S}^{(k)} - \mathbf{S}^{(k')})\|_1, \quad (8)$$

which requires graphs that are not only the same size, but are also on the same node set. Thus, the regularizer in (8) is applicable to our formulation in (3) but not to the one in (4).

Combinations of the described examples for $f(\mathbf{S}, \mathbf{X})$ and $L(\mathbf{S})$ are common in existing works. For instance, graph signal stationarity in (6) and sparsity penalties for each graph via (7) are applied in [11]. Moreover, joint inference is performed in [20] by combining (6) and (8).

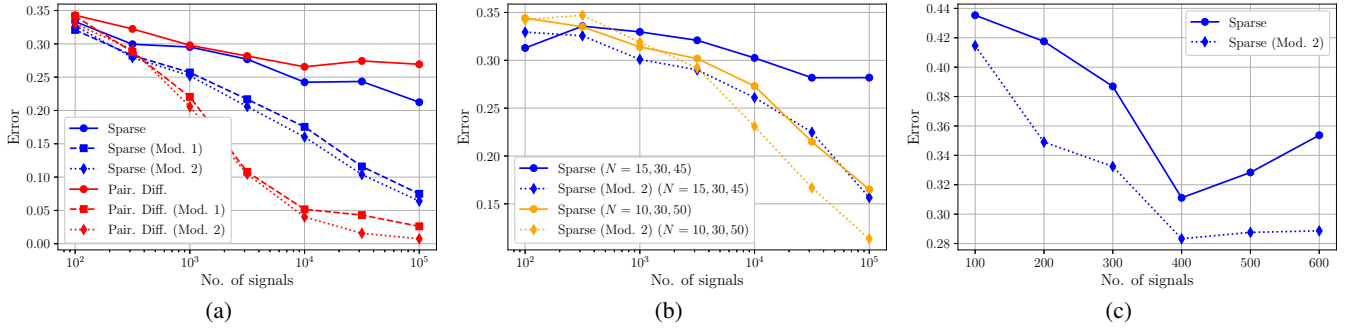


Fig. 2: (a) Recovery error for $K = 3$ graphs sampled from the *same latent point sets* in the same graphon as a function of the number of observed signals. Incorporating the joint estimation of the probability matrix or the graphon both improve estimation performance. (b) Recovery error for $K = 3$ graphs sampled from *latent point sets of different sizes* in the same graphon as a function of the number of observed signals. Separate inference of graphs is outperformed by including joint estimation of the underlying graphon. (c) Recovery error of *induced subgraphs* of three senate networks of sizes $N = 15, 30, 45$ as a function of the number of observed signals. Joint network and graphon inference outperforms separate network inference for all sets of observed signals.

5. NUMERICAL EXPERIMENTS

We compare the performance of network topology inference methods with and without the augmentations in (3) and (4), denoted by "Mod. 1" and "Mod. 2", respectively. For all experiments, we apply the same signal model assumption $f(\mathbf{S}, \mathbf{X})$ as (6), and we compare separate network inference via sparsity penalties (7) and joint network inference via pairwise difference penalties (8). For synthetic experiments, we sample from the graphon $W(x, y) = \frac{1}{2}(x^2 + y^2)$. The error of estimator $\hat{\mathbf{S}}$ is calculated as $\|\mathbf{S} - \hat{\mathbf{S}}\|_F / \|\mathbf{S}\|_F$, where the true GSO is given by \mathbf{S} . Optimization problems are solved via the alternating direction method of multipliers (ADMM) [38].¹

Same node sets. We consider the case where all graphs are sampled from the same points within the graphon space, $\zeta^{(k)} = \zeta$. We estimate $K = 3$ graphs with $N = 30$ nodes as we observe an increasing number of signals for sample covariance computation. We present in Fig. 2a the comparison of separate and joint network inference methods with the augmentations in (3) and (4), and without either. In both methods, the augmented formulations improve estimation performance significantly. The pairwise joint penalty (8) enjoys the greatest improvement, as graphs not only possess node alignment required by (8), but they also follow our graph model assumption.

Node sets of different sizes. We consider the challenging case where the graphs have different latent point sets of different sizes $N^{(k)} \neq N^{(k')}$. Unlike the previous experiment, we cannot apply (3) or (8), so we consider only (6) and (7) with and without the joint graphon estimation from (4). We consider $K = 3$ graphs for node sets of $N = 10, 30, 50$ and $N = 15, 30, 45$. For both cases, application of joint graphon inference results in consistent improvement, with increasing performance gap for larger number of observed signals.

Senate networks. Finally, we performed different-sized graph estimation with real-world data of U.S. congress roll-call votes [39], and we set up senate vote signals as in [20]. We observe the 724, 919, and 612 votes of congresses 103, 104, and 105, respectively, and

we let the underlying true networks be obtained by separate estimation of each network using all available votes. The number of nodes $N = 101$ corresponds to the number of voters (100 senators and 1 President). We estimate induced subgraphs of size $N = 15, 30, 45$, where we only observe votes of senators corresponding to these subsets of nodes. In Fig. 2c we observe that joint network and graphon estimation consistently outperforms separate inference, even though the true networks were estimated separately. This demonstrates that the versatile nonparametric nature of graphons can aid the recovery of real-world graphs, which have not been explicitly drawn from a graphon model in the first place.

6. CONCLUSION AND FUTURE WORK

We demonstrated a method to jointly estimate multiple networks under the assumption that they are sampled from the same graphon. To the best of our knowledge, this is the first method that leverages graphons to solve the challenging problem of inferring graphs of different sizes. We demonstrated that our proposed maximum-likelihood-based method improves network estimation in synthetic and real-world experiments. In terms of future directions, we plan to consider: i) The more challenging case where only noisy or partial information about the latent variables $\zeta^{(k)}$ is available, and ii) Other random graph models (beyond graphons) that can also generate graphs of different size while promoting different graph structural characteristics.

7. REFERENCES

- [1] D. R. Farine and H. Whitehead, "Constructing, conducting and interpreting animal social network analysis," *Journal of Animal Ecology*, vol. 84, no. 5, pp. 1144–1163, 2015.
- [2] M. Narayan and G. I. Allen, "Mixed effects models for resampled network statistics improves statistical power to find differences in multi-subject functional connectivity," *Frontiers in Neuroscience*, vol. 10, p. 108, 2016.
- [3] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics," *The European Physical Journal Special Topics*, vol. 174, no. 1, pp. 157–179, 2009.

¹Implementations of our method are available at https://github.com/mn51/jointinf_graphs_graphon.

- [4] O. Sporns, "The human connectome: origins and challenges," *NeuroImage*, vol. 80, pp. 53–61, 2013.
- [5] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.
- [6] 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.
- [7] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.
- [8] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.
- [9] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Sparse structural equation modeling for inference of gene regulatory networks exploiting genetic perturbations," *PLoS, Computational Biology*, vol. 9, no. 5, 2013.
- [10] V. Kalofolias, "How to learn a graph from smooth signals," in *Artificial Intelligence and Statistics*, pp. 920–929, PMLR, 2016.
- [11] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 3, no. 3, pp. 467–483, 2017.
- [12] S. de Silva, A. D. Ranjeewa, and S. Kryazhimskiy, "The dynamics of social networks among female Asian elephants," *BMC Ecology*, vol. 11, no. 1, pp. 1–16, 2011.
- [13] Y. Wang, S. Segarra, and C. Uhler, "High-dimensional joint estimation of multiple directed Gaussian graphical models," *Electronic Journal of Statistics*, vol. 14, no. 1, pp. 2439–2483, 2020.
- [14] T. T. Cai, H. Li, W. Liu, and J. Xie, "Joint estimation of multiple high-dimensional precision matrices," *Statistica Sinica*, vol. 26, no. 2, p. 445, 2016.
- [15] L. Gan, X. Yang, N. Narisetty, and F. Liang, "Bayesian joint estimation of multiple graphical models," *Advances in Neural Information Processing Systems*, pp. 9802–9812, 2019.
- [16] J. Honorio and D. Samaras, "Multi-task learning of Gaussian graphical models," in *Intl. Conf. on Machine Learning (ICML)*, 2010.
- [17] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 2, p. 373, 2014.
- [18] V. Kalofolias, A. Loukas, D. Thanou, and P. Frossard, "Learning time varying graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, pp. 2826–2830, 2017.
- [19] K. Yamada, Y. Tanaka, and A. Ortega, "Time-varying graph learning based on sparseness of temporal variation," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, pp. 5411–5415, 2019.
- [20] M. Navarro, Y. Wang, A. G. Marques, C. Uhler, and S. Segarra, "Joint inference of multiple graphs from matrix polynomials," *arXiv preprint arXiv:2010.08120*, 2020.
- [21] L. Lovász, *Large networks and graph limits*, vol. 60. American Mathematical Soc., 2012.
- [22] T. M. Roddenberry, M. Navarro, and S. Segarra, "Network topology inference with graphon spectral penalties," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, pp. 5390–5394, 2021.
- [23] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [24] 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.
- [25] 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, 2017.
- [26] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," *arXiv preprint arXiv:0712.2749*, 2007.
- [27] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Trans. Network Science and Eng.*, vol. 7, no. 1, pp. 520–537, 2020.
- [28] P. Erdős and A. Rényi, "On random graphs I," *Publicationes Mathematicae*, 1959.
- [29] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [30] C. Gao, Y. Lu, and H. H. Zhou, "Rate-optimal graphon estimation," *The Annals of Statistics*, vol. 43, no. 6, pp. 2624–2652, 2015.
- [31] O. Klopp and N. Verzelen, "Optimal graphon estimation in cut distance," *Probability Theory and Related Fields*, vol. 174, no. 3, pp. 1033–1090, 2019.
- [32] B. Sischka and G. Kauer mann, "EM-based smooth graphon estimation using MCMC and spline-based approaches," *Social Networks*, vol. 68, pp. 279–295, 2022.
- [33] S. Chan and E. Airoldi, "A consistent histogram estimator for exchangeable graph models," in *Intl. Conf. on Machine Learning (ICML)*, pp. 208–216, PMLR, 2014.
- [34] S. C. Olhede and P. J. Wolfe, "Network histograms and universality of blockmodel approximation," *Proceedings of the Nat. Acad. of Sciences (PNAS)*, vol. 111, no. 41, pp. 14722–14727, 2014.
- [35] E. M. Airoldi, T. B. Costa, and S. H. Chan, "Stochastic block-model approximation of a graphon: Theory and consistent estimation," *arXiv preprint arXiv:1311.1731*, 2013.
- [36] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *The Annals of Statistics*, vol. 43, no. 1, pp. 177–214, 2015.
- [37] J. Duchon, "Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces," *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, vol. 10, no. R3, pp. 5–12, 1976.
- [38] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [39] J. B. Lewis, K. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet, "Voteview: Congressional roll-call votes database." <https://voteview.com/>, 2021.