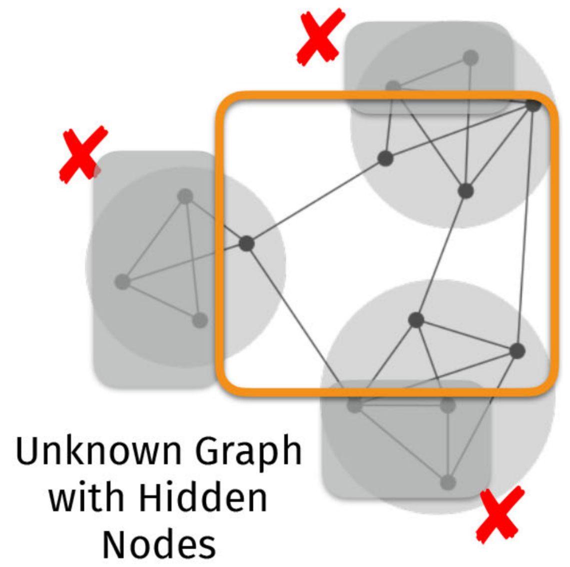
Robustness of Graph Topology Learning with Smooth Signals under Partial Observations

Hoang-Son Nguyen[⋆], Hoi-To Wai[♮]

[⋆]Oregon State University, [♮]The Chinese University of Hong Kong

Motivation



- Graph topology learning via smooth signals is prevalent in GSP [1, 2].
- Modern networks are large – only a **portion** of nodes are observed.

$\Rightarrow$ ***Can we still learn partial graph associated with obs. nodes?***

- Existing works considered sophisticated methods for graph topology learning while accounting for the influence of hidden nodes [4, 5],
- How? Unknown influence signals are low rank if the number of hidden nodes is ***significantly lower*** than the number of observed nodes.

TL;DR

Q: Do we need a new graph topology learning criterion in partial observation settings?

A: No, *if the graph signals are* ***sufficiently smooth***.

Preliminaries: Graph Topology Learning

Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is weighted, undirected, connected, has N nodes, with adjacency matrix $\mathbf{A}$, degree matrix $\mathbf{D}$, and Laplacian matrix

$$\mathbf{L} = \mathrm{Diag}(\mathbf{A1}) - \mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top \in \mathcal{L}_N,$$

where $\mathcal{L}_N$ is the set of N-node Laplacian, i.e.

$$\mathcal{L}_N := \{\hat{\mathbf{L}} : \mathrm{Tr}(\hat{\mathbf{L}}) = N, \hat{\mathbf{L}}\mathbf{1} = \mathbf{0}, \hat{\mathbf{L}} = \hat{\mathbf{L}}^\top\},$$

$\boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, ..., \lambda_N)$ are eigenvalues and $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_N]$ are sorted by ascending eigenvalues $0 = \lambda_1 \leq ... \leq \lambda_N$.

Learning Criterion (full obs.): Assume graph signals $\mathbf{y}_1, ..., \mathbf{y}_M \in \mathbb{R}^N$ are smooth (i.e., $\mathbf{y}_m^\top \mathbf{L} \mathbf{y}_m \approx 0$), [1] and [2] propose to the problem:

$$\min_{\hat{\mathbf{L}}} J_f(\hat{\mathbf{L}}) := \sum_m \mathbf{y}_m^\top \hat{\mathbf{L}} \mathbf{y}_m \text{ s.t. } \hat{\mathbf{L}} \in \mathcal{L}_N \quad \text{(GL-Full)}$$

Let $\mathbf{L}^\star \in \arg\min_{\hat{\mathbf{L}} \in \mathcal{L}_N} J_f(\hat{\mathbf{L}})$ be an optimal solution to GL-Full.

Learning with Partial Observations

Partially Observed Signals: Among N graph signals, only n signals on a certain set of n nodes are observed, i.e.,

$$\mathbf{y}_{o,m} = [\mathbf{I}_{n\times n} \; \mathbf{0}_{n\times(N-n)}]\mathbf{y}_m =: \mathbf{E}_o \mathbf{y}_m,$$

where $\mathbf{E}_o \in \mathbb{R}^{n\times N}$ takes a subset of n nodes from $\mathcal{V}$.

Learning Criterion (partial obs.): Adapting (GL-Full) yields the following *hidden-node agnostic* graph learning criterion (GL-Partial):

$$\min_{\hat{\mathbf{L}}_p} J_p(\hat{\mathbf{L}}_p) := \sum_m \mathbf{y}_{o,m}^\top \hat{\mathbf{L}}_p \mathbf{y}_{o,m} \text{ s.t. } \hat{\mathbf{L}}_p \in \mathcal{L}_n \quad \text{(GL-Partial)}$$

Let $\mathbf{L}_p^\star \in \arg\min_{\hat{\mathbf{L}}_p} J_p(\hat{\mathbf{L}}_p)$ be an optimal solution to GL-Partial.

Main Result

What is the relationship between $\mathbf{L}^\star$ and $\mathbf{L}_p^\star$?

Let us study

$$\widehat{\mathbf{L}} := \frac{N}{n}\mathbf{E}_o^\top \mathbf{L}_p^\star \mathbf{E}_o, \; \widetilde{\mathbf{L}}_p := \mathbf{E}_o \mathbf{L}^\star \mathbf{E}_o^\top + \mathrm{Diag}(\mathbf{L}_{oh}^\star \mathbf{1})$$

where $\widehat{\mathbf{L}}$ is an N-node graph with the only connected part being (scaled) $\mathbf{L}_p^\star$, $\mathbf{E}_o \mathbf{L}^\star \mathbf{E}_o^\top =: \mathbf{L}_{oo}^\star$ is the observed part of $\mathbf{L}^\star$, and $\mathbf{L}_{oh}^\star$ is observed-hidden part.

Assumptions are

- **A1:** $\mathbf{L}_{oh}^\star \mathbf{1} \leq \epsilon \mathbf{1}$ for small ϵ, and
- **A2:** $\mathrm{Tr}(\mathbf{L}_{oo}^\star) - \mathbf{1}^\top \mathbf{L}_{oh}^\star \mathbf{1} \geq cn$ for some $c > 0$, and
- **A3:** $\mathbf{y}_m \in \mathrm{span}(\mathbf{V}_K)$, $\forall m = 1, ..., M$.

$\Rightarrow$ **A3** is plausible in *low-pass graph signals* with bandwidth K [3].

Main Theorem

Consider a random partial observation set $\mathcal{S} = \{s(1), s(2), ..., s(n)\}$, which is sampled with replacements from $\mathcal{V}$. Then, $\forall \delta \in (0, 1)$, $\exists t \in (0, 1)$ such that with probability at least $1 - \delta$,

$$J_p(\mathbf{L}_p^\star) \leq J_p(\widetilde{\mathbf{L}}_p) \leq \left[\frac{1+t}{c}\frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})}\right] J_p(\mathbf{L}_p^\star) + \mathcal{O}\left(\frac{\epsilon}{c}\right)$$

and

$$J_f(\mathbf{L}^\star) \leq J_f(\widehat{\mathbf{L}}) \leq \left[\frac{1+t}{c}\frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})}\right] J_f(\mathbf{L}^\star) - \mathcal{O}\left(\frac{N\epsilon}{cn}\right),$$

provided that the number of observations satisfies

$$\frac{n}{N} \geq \frac{3}{t^2} \max_{1\leq i\leq N} \|\mathbf{V}_K^\top \mathbf{e}_i\|_2^2 \ln\left(\frac{2K}{\delta}\right).$$

Proof sketch: Establish a high-probability one-sided RIP property that

$$\mathbf{y}\mathbf{E}_o^\top \mathbf{E}_o \mathbf{L} \mathbf{E}_o^\top \mathbf{E}_o \mathbf{y} \leq (1+t)\frac{n}{N}\frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})}\mathbf{y}^\top \mathbf{L}\mathbf{y}, \; \forall \mathbf{y} \in \mathrm{span}(\mathbf{V}_K).$$

Interpretations:

- If $\frac{1+t}{c}\frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})} = \Theta(1)$, then the above theorem suggests that $\mathbf{L}_p^\star$ corresponds to a row/column sampled version of $\mathbf{L}^\star$. In this case, such condition can be satisfied for non-modularized graphs.
- Result is ***insensitive*** to the no. of hidden nodes ($N - n \gg 1$).

Synthetic Experiment

Erdos-Renyi graph with connectivity $p = 0.3$, $N = 64$ nodes, and $M = 20$ low-pass signals $\mathbf{y}_m = (\mathbf{I} + \alpha\mathbf{L})^{-1}\mathbf{x}_m$ ($\alpha > 0$ controls low-pass-ness).

$\alpha \uparrow \Rightarrow$ ***more low-pass*** $\Rightarrow$ ***A3 more likely to hold***

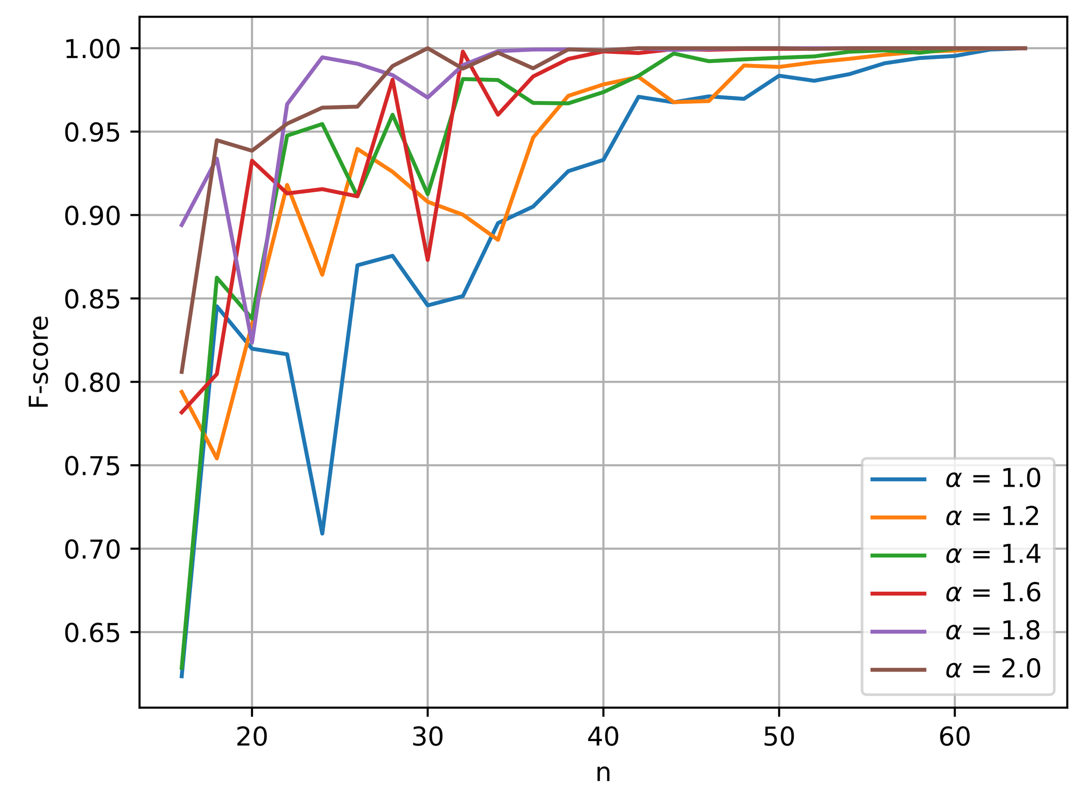


Fig. 1: Median F-score of observed $\mathbf{L}_{oo}^\star$ and learned $\mathbf{L}_p^\star$ vs. no. of hidden nodes n

Two Takeaways

- Even when $N - n \gg 1$, graph learned from (GL-Partial) $\approx$ graph learned from (GL-Full).
- ***More low-pass*** the graph signals $\Rightarrow$ Good F-score is achieved with a smaller number of observations n.

References

[1] Dong et al., "Learning Laplacian Matrix in Smooth Graph Signal Representations", IEEE TSP, 2016.
[2] Kalofolias, "How to Learn a Graph from Smooth Signals", AISTATS, 2016.
[3] Ramakrishna et al. 'A User Guide to Low-Pass Graph Signal Processing and Its Applications: Tools and Applications,' IEEE Signal Processing Magazine, 2020.
[4] Chandrasekaran et al., "Latent Variable Graphical Model Selection via Convex Optimization," The Annals of Statistics, 2012.
[5] Buciulea et al., "Learning Graphs from Smooth and Graph-Stationary Signals with Hidden Variables," IEEE Transactions on Signal and Information Processing over Networks, 2022.