Network Topology Inference from Gaussian and Stationary Graph Signals

Andrei Buciulea Vlas

King Juan Carlos University - Madrid, Spain

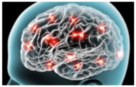
Joint work with A. G. Marques

ICASSP 2023, Rhodes, Greece

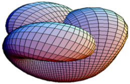
Motivating Examples

- Huge data sets are generated in networks (transportation networks, biological networks, brain networks, computer networks, social networks)
- The data structure carries critical information about the nature of the data
- Modelling the data structure with graphs

Interpolate a brain signal from local observations



Compress a signal in an irregular domain



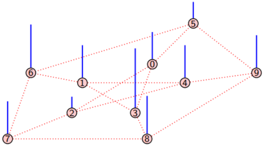
Localize the source of a rumor

Smooth an observed network profile

Predict the evolution of a network process

Infer the topology where the signals reside

Graph Signal Processing (GSP)



- Consider an undirected weighted graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$

  $\Rightarrow \mathcal{V}$, $\mathcal{E}$, $\mathcal{W} \rightarrow$ set of nodes, edges, weights
- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on the top of the graph

  $\Rightarrow x_i =$ value of graph signal (GS) at node i
- Associated with $\mathcal{G}$ is the *Graph-Shift Operator* (GSO)

  $\Rightarrow \mathbf{S} \in \mathbb{R}^{N \times N}$, $S_{ij} \neq 0$ for $i = j$ and $(i, j) \in \mathcal{E}$

  $\Rightarrow$ Ex: Adjacency $\mathbf{A}$, Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, random walk...
- Graph filters (GFi): Linear GS operators $\mathbf{y} = \mathbf{H}\mathbf{x}$ of the form

  $\Rightarrow \mathbf{H} := \sum_{p=0}^{P-1} h_p \mathbf{S}^p$ $\Rightarrow$ I.e., GFi are matrix polynomials of $\mathbf{S}$
- Random GS $\Rightarrow$ Generalizing stationary to GS $\Rightarrow \mathbf{x} = \mathbf{H}\mathbf{w}$ with $\mathbf{w}$ white

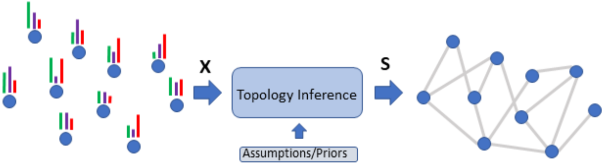
  $\Rightarrow$ Covariance $\mathbf{\Sigma} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbb{E}\left[\mathbf{H}\mathbf{w}(\mathbf{H}\mathbf{w})^{\mathsf{T}}\right] = \mathbf{H}\mathbb{E}\left[\mathbf{w}\mathbf{w}^{\mathsf{T}}\right]\mathbf{H}^{\mathsf{T}} = \mathbf{H}^2$

  $\Rightarrow \mathbf{\Sigma}$ and $\mathbf{H}$ are a polynomials on $\mathbf{S}$ $\Rightarrow \mathbf{x}$ is stationary in $\mathbf{S}$

Network Topology Inference: Motivation and Context

Network topology inference from nodal observations

"Given a collection $\mathbf{X} := [\mathbf{x}_1, ..., \mathbf{x}_R]$ of graph signal observations supported on the unknown graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$ find an optimal $\mathbf{S}$"



- Ill posed problem: optimality, priors, regularizations
  - ⇒ Test Pearson corr., partial corr. and conditional dependence
  - ⇒ Sparsity [Friedman07] and consistency [Meinshausen06]
  - ⇒ Graph Signal Processing (GSP) [Dong17,Mei17,Segarra17]
- **This work**:
  - ⇒ Use GSP to infer the topology
  - ⇒ Assume $\mathbf{x}_r$'s are i.i.d realizations of $\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ and stationary in $\mathbf{S}$

Graph Topology Inference: Related Work

- Goal: use $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_R] \in \mathbb{R}^{N\times R}$ to infer $\mathbf{S}$ with sample cov. $\hat{\boldsymbol{\Sigma}} = \frac{1}{R}\mathbf{X}\mathbf{X}^{\mathsf{T}}$
  - Let $\mathbf{X}$ be R samples supported on the graph $\mathcal{G} \Rightarrow$ {Correlation networks}

    $$\hat{\mathbf{S}} \approx \hat{\boldsymbol{\Sigma}} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^{\mathsf{T}}\right] \text{ (}\hat{\mathbf{S}} \text{ is a thresholded version of } \hat{\boldsymbol{\Sigma}}\text{)}$$

  - Let $\mathbf{X}$ be R i.i.d samples of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \Rightarrow$ {Partial correlation networks} GL

    $$\hat{\mathbf{S}} = \underset{\mathbf{S}\succeq 0, \mathbf{S}\in\mathcal{S}_{\boldsymbol{\Theta}}}{\text{argmin}} \; -\log(\det(\mathbf{S})) + \text{tr}(\hat{\boldsymbol{\Sigma}}\mathbf{S}) + \rho \text{h}(\mathbf{S})$$

    $\Rightarrow$ Good performance in low-sample scenarios
    $\Rightarrow$ Specific covariance model $\boldsymbol{\Sigma}_{MRF} = (\sigma\mathbf{I} + \delta\mathbf{S})^{-1}$

  - Let $\mathbf{X}$ be stationary w.r.t $\mathbf{S} \Rightarrow$ {Graph-stationary diffusion processes} GSR

    $$\hat{\mathbf{S}} = \underset{\mathbf{S}\in\mathcal{S}}{\text{argmin}} \; \|\mathbf{S}\|_0 \quad \text{s. to} \quad \hat{\boldsymbol{\Sigma}}\mathbf{S} = \mathbf{S}\hat{\boldsymbol{\Sigma}} \quad \text{[Segarra17]}$$

    $\Rightarrow$ More general covariance model $\boldsymbol{\Sigma}_{poly} = poly(\mathbf{S}) \begin{cases} Corr.\ netw. \rightarrow \hat{\boldsymbol{\Sigma}} = \mathbf{S} \\ Part.\ corr. \rightarrow \hat{\boldsymbol{\Sigma}} = \mathbf{S}^{-1} \end{cases}$
    $\Rightarrow$ Higher number of samples are needed for an accurate estimation

  - Other approaches: Smoothness [Dong17], Sparse SEM [Bazerque13]

Graphical Models with St. Signals: Problem Statement

- Given the sample covariance matrix $\hat{\mathbf{\Sigma}}$ estimate $\mathbf{S}$ under assumptions
  - $\Rightarrow$ (AS1): $\{\mathbf{x}_r\}$ are i.i.d realizations of $\mathcal{N}(\mathbf{0}, \mathbf{\Theta}^{-1})$
  - $\Rightarrow$ (AS2): $\{\mathbf{x}_r\}$ are stationary in $\mathbf{S}$
  - ML approach leveraging (AS1) and the associated log-likelihood function

$$L(\mathbf{X}|\mathbf{\Theta}) := \prod_{r=1}^{R} f_{\mathbf{\Theta}}(\mathbf{x}_r), \quad \mathcal{L}(\mathbf{X}|\mathbf{\Theta}) := \sum_{r=1}^{R} \log\left((2\pi)^{-N/2} \cdot det^{\frac{1}{2}}(\mathbf{\Theta}) \cdot e^{-\frac{1}{2}\mathbf{x}^T\mathbf{\Theta}\mathbf{x}}\right)$$

  - Maximizing the log-likelihood function under (AS2)

Problem I: Minimize $-\mathcal{L}(\mathbf{X}|\mathbf{\Theta})$ under (AS2)

$$\hat{\mathbf{\Theta}}, \hat{\mathbf{S}} = \underset{\mathbf{\Theta}\succeq 0, \mathbf{S}\in\mathcal{S}}{\operatorname{argmin}} \; -\log(\det(\mathbf{\Theta})) + \operatorname{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Theta}),$$

$$\text{s. to} \qquad \|\mathbf{S}\|_0 \leq \kappa \;\text{ and }\; \mathbf{\Theta}\mathbf{S} = \mathbf{S}\mathbf{\Theta}$$

- Sparsity constraint $\|\mathbf{S}\|_0 \leq \kappa \rightarrow$ Non-convex
- Key novelty: stationarity also implies $\mathbf{\Theta}\mathbf{S} = \mathbf{S}\mathbf{\Theta} \rightarrow$ Bi-linear term

- Step I. Reformulate Problem I adding convex relaxation ℓ_1 and auxiliary $\boldsymbol{\Theta}_2$

$$\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2, \hat{\mathbf{S}} = \underset{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \succeq 0; \mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \operatorname{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_1) - \log\det(\boldsymbol{\Theta}_2) + \rho\|\mathbf{S}\|_1$$

$$\text{s. to} \quad \boldsymbol{\Theta}_1\mathbf{S} = \mathbf{S}\boldsymbol{\Theta}_1 \ \text{ and } \ \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2$$

- Step II. Define augmented Lagrangian with multipliers $\Rightarrow$ $\mathbf{Y}$ and $\mathbf{Z}$

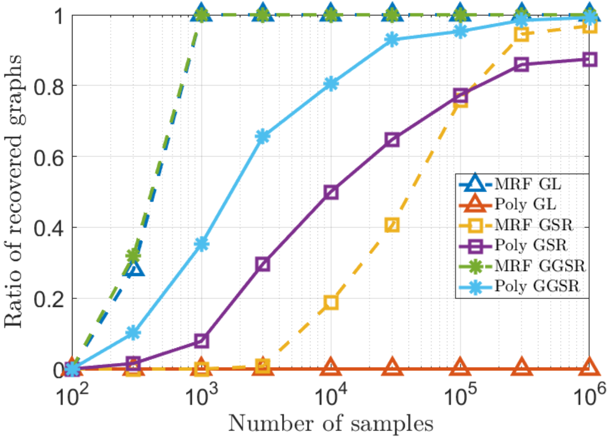
$$\hat{\boldsymbol{\Theta}}_1, \hat{\boldsymbol{\Theta}}_2, \hat{\mathbf{S}} = \underset{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \succeq 0; \mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \operatorname{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Theta}_1) - \log\det(\boldsymbol{\Theta}_2) + \rho\|\mathbf{S}\|_1 + \langle \mathbf{Z}, \boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2 \rangle + \frac{\lambda}{2}\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\|_F^2 + \langle \mathbf{Y}, \boldsymbol{\Theta}_1\mathbf{S} - \mathbf{S}\boldsymbol{\Theta}_1 \rangle + \frac{\lambda}{2}\|\boldsymbol{\Theta}_1\mathbf{S} - \mathbf{S}\boldsymbol{\Theta}_1\|_F^2$$

- Step III. Solving the problem via alternating algorithm

1. $\hat{\boldsymbol{\Theta}}_1^{(t+1)} = \underset{\boldsymbol{\Theta}_1 \succeq 0}{\operatorname{argmin}} \ L_f(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2^{(t)}, \mathbf{S}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{Z}^{(t)})$
2. $\hat{\boldsymbol{\Theta}}_2^{(t+1)} = \underset{\boldsymbol{\Theta}_2 \succeq 0}{\operatorname{argmin}} \ L_f(\boldsymbol{\Theta}_1^{(t+1)}, \boldsymbol{\Theta}_2, \mathbf{S}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{Z}^{(t)})$
3. $\hat{\mathbf{S}}^{(t+1)} = \underset{\mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \ L_f(\boldsymbol{\Theta}_1^{(t+1)}, \boldsymbol{\Theta}_2^{(t+1)}, \mathbf{S}, \mathbf{Y}^{(t)}, \mathbf{Z}^{(t)})$
4. $\mathbf{Y}^{(t+1)} = \mathbf{Y}^{(t)} + \lambda\left(\boldsymbol{\Theta}_2^{(t+1)}\mathbf{S}^{(t+1)} - \mathbf{S}^{(t+1)}\boldsymbol{\Theta}_2^{(t+1)}\right)$
5. $\mathbf{Z}^{(t+1)} = \mathbf{Z}^{(t)} + \lambda\left(\boldsymbol{\Theta}_2^{(t+1)} - \boldsymbol{\Theta}_1^{(t+1)}\right)$

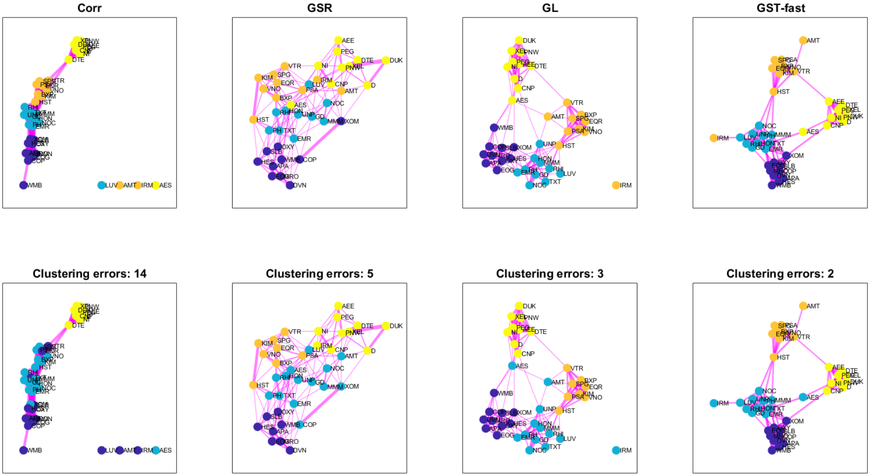
Synthetic Data Results

- Recovery performance for different algorithms from sample covariance considering 2 setups: $\mathbf{\Sigma} = (\sigma \mathbf{I} + \delta \mathbf{S})^{-1}$ (MRF) and $\mathbf{\Sigma} = \sum_{p=0}^{2P-2} c_p \mathbf{S}^p$ (Poly)



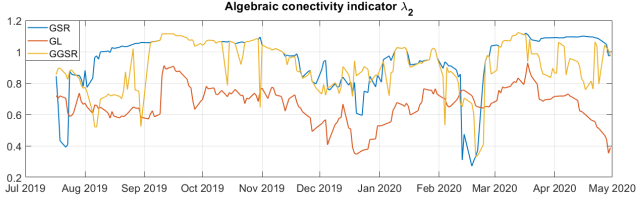
- Similar performance results for GGSR and GL with MRF setup
- Significant improvement for GGSR compared to GSR with Poly setup

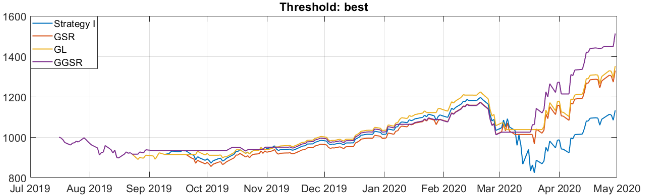
- Estimation of the connections between 40 companies from 4 different sectors of the SP500 index using daily stock closing price during 2010-2016



Financial Data Experiment II

- Time-varying graph learning for investment strategies. [Cardoso20]
  - ⇒ How good the graph estimates are? ⇒ No real ground truth
  - ⇒ Approach: Using the graph to design an investment strategy





⇒ GGSR performs best ⇒ Implicit validation

Conclusions

- New graph learning scheme that subsumes GL
- Key **assumptions**: graph sparse, signals Gaussian, and relation between those two (stationarity)
- Requires way less signals than non-Gaussian stationarity-based approaches
- Challenge: ML estimation non-convex

  $\Rightarrow$ Relaxations and alternating minimization algorithm
- Encouraging results in both synthetic and real data sets

- **THANKS!**

  $\Rightarrow$ Additional details can be found in the paper

  $\Rightarrow$ Feel free to contact me for questions and code andrei.buciulea@urjc.es