NETWORK TOPOLOGY IDENTIFICATION FROM SPECTRAL TEMPLATES

Santiago Segarra*, Antonio G. Marques†, Gonzalo Mateos‡, and Alejandro Ribeiro§ E-mail: segarra@mit.edu

*Massachusetts Institute of Technology, †King Juan Carlos University ‡University of Rochester § University of Pennsylvania

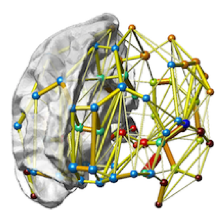
Abstract

We address the problem of identifying a graph from signals defined on it. First, we estimate the eigenvectors or spectral templates of the graph based on the sample covariance and then infer the eigenvalues by imposing desirable properties on the graph to be recovered. We specify theoretical conditions for perfect recovery in the noiseless case and error bounds in the presence of noise.

Motivation and context

- Network topology inference from observations is well-studied
- Some approaches use correlations to construct graphs
- Partial correlations and conditional dependence also used

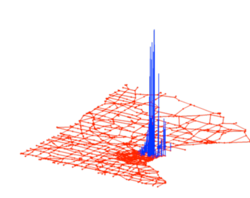
- Paramount importance in neuroscience
  - ⇒ Functional net inferred from activity

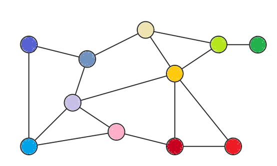


- Most GSP works assume that $\mathbf{S}$ (hence the graph) is known
  - ⇒ Analyze how characteristics of $\mathbf{S}$ affect signals and filters
- We take the reverse path
  - ⇒ How to use GSP to infer the graph topology?

Graph signal processing - 101

- Network as graph $G = (\mathcal{V}, \mathcal{E}, W)$: encode pairwise relationships
- Interest here not in G itself, but in data associated with nodes in $\mathcal{V}$
  - ⇒ The object of study is a graph signal
- Ex: Opinion profile, buffer congestion levels, neural activity



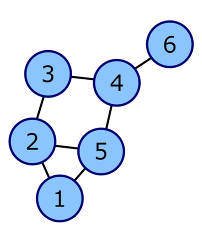


$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{|\mathcal{V}|} \end{bmatrix} = \begin{bmatrix} 0.6 \\ \vdots \\ 0.7 \end{bmatrix}$$

- Graph SP: need to broaden classical SP results to graph signals
  - ⇒ Our view: GSP well suited to study network processes

Graph signals and graph-shift operator

- Graph signals are mappings $x : \mathcal{V} \to \mathbb{R}$
  - ⇒ May be represented as a vector $\mathbf{x} \in \mathbb{R}^N$ (with $|\mathcal{V}| = N$)
- Graph G is endowed with a graph-shift operator $\mathbf{S}$
  - ⇒ Matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ satisfying: $S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$



$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{23} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix}$$

$\mathbf{S}$ captures local structure in G

- Ex: Adjacency $\mathbf{A}$, Laplacian $\mathbf{L}$, normalized Laplacian $\mathcal{L}$

Locality of S and frequency-domain representation

- $\mathbf{S}$ is a local operator ⇒ If $\mathbf{y} = \mathbf{Sx}$, $y_i = \sum_{j\in\mathcal{N}_i} S_{ij}x_j$ ⇒ 1-hop info
- Spectrum of $\mathbf{S}$ useful to analyze $\mathbf{x}$
  - ⇒ Consider the spectral decomposition $\mathbf{S} = \mathbf{V\Lambda V}^H$
- Leverage $\mathbf{S}$ to define graph Fourier transform (GFT) and iGFT

$$\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}, \qquad \mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$$

- Key message: the two basic elements of GSP are $\mathbf{x}$ and $\mathbf{S}$

Linear (shift-invariant) graph filter

- A graph filter $H : \mathbb{R}^N \to \mathbb{R}^N$ is a map between graph signals
  - ⇒ Focus on linear filters ⇒ $N \times N$ matrix
- Filter $\mathbf{H}$ is a polynomial in $\mathbf{S}$ with coeffs. $\mathbf{h} = [h_0, \ldots, h_L]^T$

$$\mathbf{H} := h_0\mathbf{S}^0 + h_1\mathbf{S}^1 + \ldots + h_L\mathbf{S}^L = \sum_{l=0}^{L} h_l\mathbf{S}^l$$

- Properties: distributed, only L-hop info, and $\mathbf{H}(\mathbf{Sx}) = \mathbf{S}(\mathbf{Hx})$
- Filter $\mathbf{H}$ is diagonalized by the eigenvectors of the shift operator $\mathbf{S}$

$$\mathbf{H} = \mathbf{V}\,\mathrm{diag}(\tilde{\mathbf{h}})\,\mathbf{V}^H, \qquad \tilde{\mathbf{h}} = \mathrm{diag}\Big(\sum_{l=0}^{L} h_l\mathbf{\Lambda}^l\Big)$$

- We say that $\tilde{\mathbf{h}}$ is the frequency response of $\mathbf{H}$

Stationary graph processes

- Stationarity can be extended from time to graph domains
- Signal $\mathbf{x}$ is the response of linear diffusion applied to a white input

$$\mathbf{x} = \alpha_0\prod_{l=1}^{\infty}(\mathbf{I} - \alpha_l\mathbf{S})\mathbf{w} = \sum_{l=0}^{\infty}\beta_l\mathbf{S}^l\mathbf{w}$$

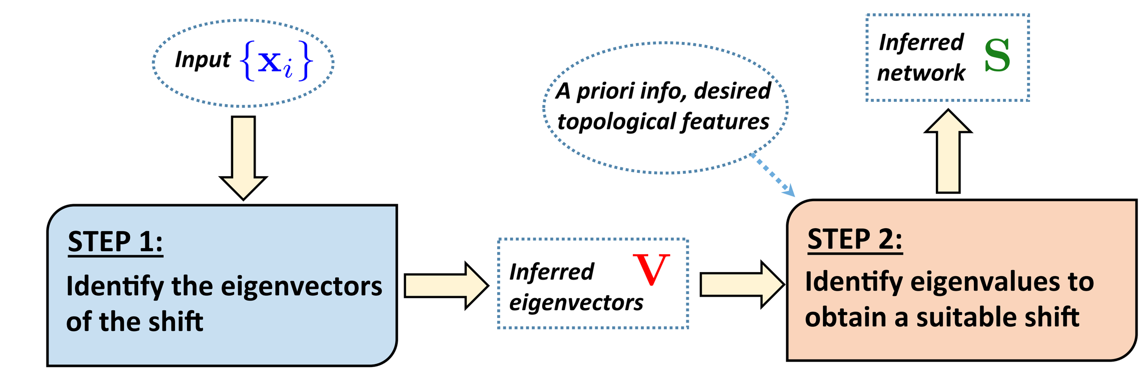
- Common generative model. Heat diffusion if α_l constant
- We say the graph shift $\mathbf{S}$ explains the structure of signal $\mathbf{x}$

- From Cayley Hamilton, diffusion as

$$\mathbf{x} = \Big(\sum_{l=0}^{L-1} h_l\mathbf{S}^l\Big)\mathbf{w} := \mathbf{Hw}$$

Our approach for topology identification

- We propose a two-step approach for graph topology identification



STEP 1: Obtaining the eigenvectors or spectral templates

- The covariance matrix of the signal $\mathbf{x}$ is

$$\mathbf{C}_x = \mathbb{E}\Big(\mathbf{Hw}\big(\mathbf{Hw}\big)^H\Big) = \mathbf{H}\mathbb{E}\Big(\mathbf{ww}^H\Big)\mathbf{H}^H = \mathbf{HH}^H$$

- Since $\mathbf{H}$ and $\mathbf{S}$ share $\mathbf{V}$ ⇒ $\mathbf{C}_x$ and $\mathbf{S}$ also share $\mathbf{V}$

$$\mathbf{C}_x = \mathbf{V}\sum_{l=0}^{L-1} h_l\mathbf{\Lambda}^l\,\mathbf{V}^H\,\mathbf{V}\sum_{l=0}^{L-1} h_l(\mathbf{\Lambda}^l)^H\,\mathbf{V}^H = \mathbf{V}\,\mathrm{diag}(|\tilde{\mathbf{h}}|^2)\,\mathbf{V}^H$$

- Eigenvectors $\mathbf{V}$ are preserved
- Graph and its eigenvalues have been obscured by diffusion

Observations

(a) There are many shifts $\mathbf{S}$ that can explain a signal $\mathbf{x}$
(b) Identifying $\mathbf{S}$ is just a matter of identifying the eigenvalues
(c) In correlation methods the eigenvalues are kept unchanged
(d) In precision methods the eigenvalues are inverted
(e) Our approach valid for any polynomial map. between $\mathbf{S}$ and $\mathbf{C}_x$

STEP 2: Obtaining the eigenvalues

- We can use extra knowledge/assumptions to choose one graph
  - ⇒ Of all graphs, select one that is optimal in some sense

$$\mathbf{S}^* := \underset{\mathbf{S},\boldsymbol{\lambda}}{\mathrm{argmin}}\; f(\mathbf{S},\boldsymbol{\lambda}) \quad \text{s. to} \quad \mathbf{S} = \sum_{k=1}^{N}\lambda_k\mathbf{v}_k\mathbf{v}_k^H, \;\; \mathbf{S}\in\mathcal{S} \qquad (1)$$

- Set $\mathcal{S}$ contains all admissible scaled adjacency matrices

$$\mathcal{S} := \{\mathbf{S}\,|\,S_{ij}\geq 0,\;\; \mathbf{S}\in\mathcal{M}^N,\;\; S_{ii}=0,\;\; \textstyle\sum_j S_{1j}=1\}$$

  - ⇒ Can accommodate Laplacians as well
- Inference problem convex if we select a convex objective $f(\mathbf{S},\boldsymbol{\lambda})$
  - ⇒ Minimum energy ($f(\mathbf{S}) = \|\mathbf{S}\|_F$), Fast mixing ($f(\boldsymbol{\lambda}) = -\lambda_2$)
- The feasibility set in (1) is generally small: Define $\mathbf{W}_{\mathcal{D}} := (\mathbf{V}\odot\mathbf{V})_{\mathcal{D}}$

> Assume that (1) is feasible, then it holds that $\mathrm{rank}(\mathbf{W}_{\mathcal{D}}) \leq N - 1$. Also, if $\mathrm{rank}(\mathbf{W}_{\mathcal{D}}) = N - 1$, then the feasible set of (1) is a singleton.

$\odot$ is the Khatri-Rao product and $\mathcal{D}$ indexes the diagonal

Sparse recovery

- When feasibility set in (1) is non-trivial
  - ⇒ $f(\mathbf{S},\boldsymbol{\lambda})$ determines the features of the recovered graph
- Identify the sparsest shift $\mathbf{S}_0^*$ that explains signals' structure
  - ⇒ Set the cost $f(\mathbf{S},\boldsymbol{\lambda}) = \|\mathbf{S}\|_0$
- Problem is not convex, but can relax to ℓ_1 norm minimization

$$\mathbf{S}_1^* := \underset{\mathbf{S},\boldsymbol{\lambda}}{\mathrm{argmin}}\;\|\mathbf{S}\|_1 \quad \text{s. to} \quad \mathbf{S} = \sum_{k=1}^{N}\lambda_k\mathbf{v}_k\mathbf{v}_k^H, \;\; \mathbf{S}\in\mathcal{S}$$

- Does the solution $\mathbf{S}_1^*$ coincide with the ℓ_0 solution $\mathbf{S}_0^*$?
  - ⇒ Construct $\mathbf{R} := [(\mathbf{I} - \mathbf{WW}^\dagger)_{\mathcal{D}^c},\; \mathbf{e}_1\otimes\mathbf{1}_{N-1}]$
  - ⇒ Denote by $\mathcal{K}$ the indices of the support of $\mathbf{s}_0^* = \mathrm{vec}(\mathbf{S}_0^*)$

> $\mathbf{S}_1^*$ and $\mathbf{S}_0^*$ coincide if the two following conditions are satisfied:
> 1) $\mathrm{rank}(\mathbf{R}_{\mathcal{K}}) = |\mathcal{K}|$; and
> 2) There exists a constant $\delta > 0$ such that
> $$\psi_{\mathbf{R}} := \|\mathbf{I}_{\mathcal{K}^c}(\delta^{-2}\mathbf{RR}^T + \mathbf{I}_{\mathcal{K}^c}^T\mathbf{I}_{\mathcal{K}^c})^{-1}\mathbf{I}_{\mathcal{K}}^T\|_\infty < 1.$$

- Cond. 1) guarantees uniqueness of solution $\mathbf{S}_1^*$
- Cond. 2) ensures existence of a dual certificate for ℓ_0 optimality

Recovery from noisy spectral templates

- When approximating $\mathbf{C}_x$ with the sample covariance $\hat{\mathbf{C}}_x$
  - ⇒ We have access to $\hat{\mathbf{V}}$, a noisy version of the eigenvectors
- With $d(\cdot,\cdot)$ denoting a (convex) distance between matrices

$$\hat{\mathbf{S}}_1^* := \underset{\{\mathbf{S},\boldsymbol{\lambda},\mathbf{S}'\}}{\mathrm{argmin}}\;\|\mathbf{S}\|_1 \quad \text{s. to} \quad \mathbf{S}' = \sum_{k=1}^{N}\lambda_k\hat{\mathbf{v}}_k\hat{\mathbf{v}}_k^H, \;\; \mathbf{S}\in\mathcal{S}, \;\; d(\mathbf{S},\mathbf{S}')\leq\epsilon$$

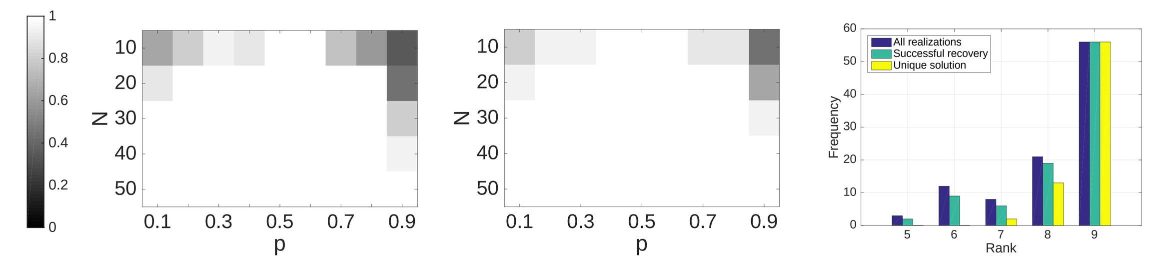
- How does the recovery depend on the noise level ϵ?
- Assume that $d(\mathbf{S},\mathbf{S}') = \|\mathbf{S}-\mathbf{S}'\|_F$ and $d(\mathbf{S}_0^*,\mathbf{S}')\leq\epsilon$

> If 1) and 2) are fulfilled for $\hat{\mathbf{R}}$, the solution $\hat{\mathbf{s}}_1^* := \mathrm{vec}(\hat{\mathbf{S}}_1^*)$ satisfies
> $$\|\hat{\mathbf{s}}_1^* - \mathbf{s}_0^*\|_1 \leq C\epsilon, \quad \text{with} \quad C = 2C_1 + 2C_2C_3,$$
> where the constants C_1, C_2, and C_3 are given by
> $$C_1 = \frac{\sqrt{|\mathcal{K}|}}{\sigma_{\min}(\hat{\mathbf{R}}_{\mathcal{K}}^T)}, \quad C_2 = \frac{1 + \|\hat{\mathbf{R}}^T\|_2C_1}{1 - \psi_{\hat{\mathbf{R}}}}, \quad C_3 = \|\hat{\mathbf{R}}^\dagger\|_2 N.$$

- $\hat{\mathbf{S}}_1^*$ is a consistent estimator of $\mathbf{S}_0^*$ under conditions 1) and 2)

Topology inference in random graphs

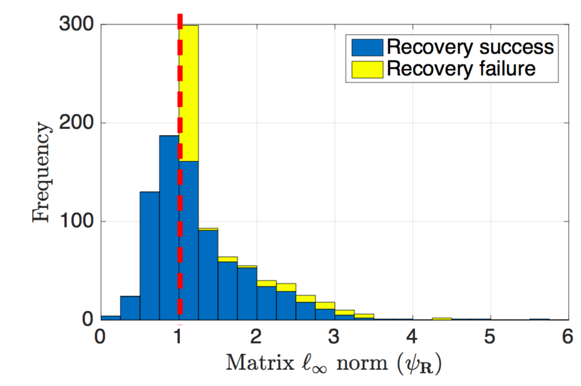
- Erdős-Rényi (ER) graphs of varying size $N\in\{10, 20, \ldots, 50\}$
  - ⇒ Edge probabilities $p\in\{0.1, 0.2, \ldots, 0.9\}$
- Recovery rates for adjacency (left) and normalized Laplacian (mid)



- Recovery is easier for intermediate values of p
- Rate of recovery related to the rank of $\mathbf{W}_{\mathcal{D}}$
  - ⇒ As rank decreases, there is a detrimental effect on recovery

Sparse recovery guarantees

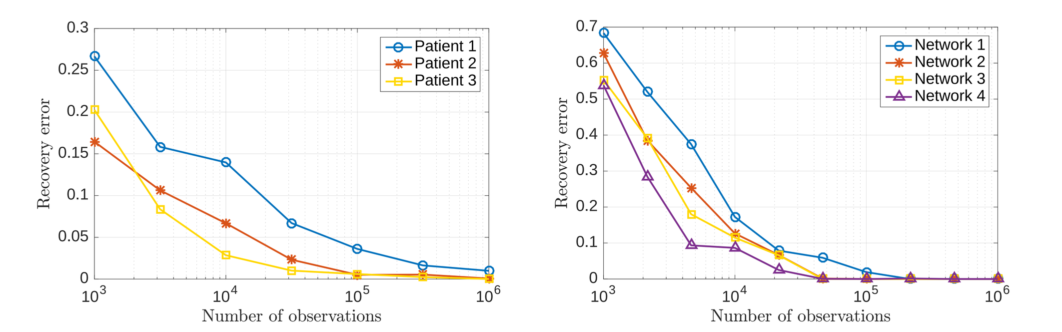
- Generate 1000 ER random graphs ($N = 20$, $p = 0.25$) such that
  - ⇒ Feasible set is not a singleton and Cond. 1) is satisfied



- ℓ_1 norm recovery success as a function of $\psi_{\mathbf{R}}$
- Condition 2) is sufficient but not necessary
  - ⇒ Tightest bound on $\psi_{\mathbf{R}}$

Inference from noisy spectral templates

- Identification of brain graphs (left) and social networks (right)
- Test recovery for noisy spectral templates $\hat{\mathbf{V}}$
  - ⇒ Obtained from sample covariances of diffused signals

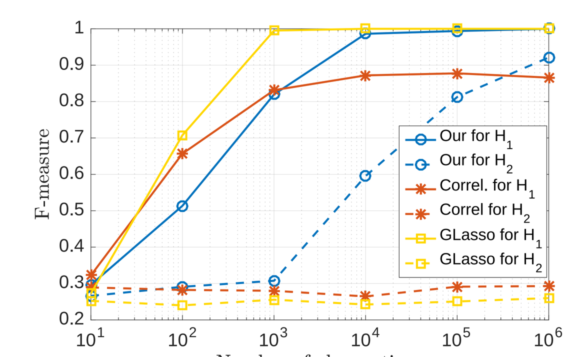


- Recovery error decreases with more observed signals
  - ⇒ More reliable estimate of the covariance ⇒ Less noisy $\hat{\mathbf{V}}$
- Traditional methods like graphical lasso fail to recover $\mathbf{S}$

Performance comparison

- Comparison with other GSP methods and established methods
  - ⇒ 100 ER graphs with $N = 20$ and $p = 0.2$

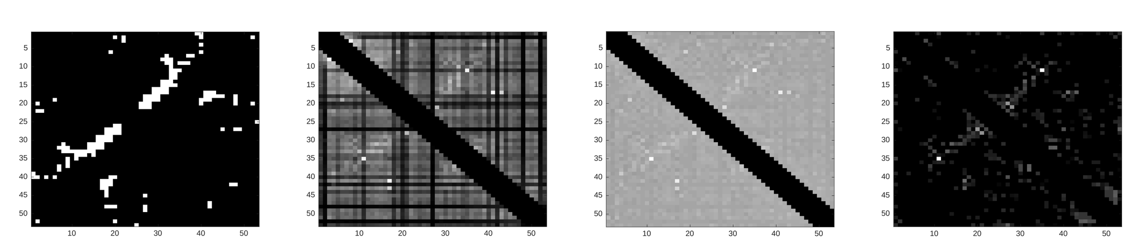
	Our	Kalof.	Dong
F-measure	**0.896**	0.791	0.818
edge error	**0.108**	0.152	0.168
degree error	**0.058**	0.071	0.105



- Recovery of a Laplacian from smooth graph signals (left)
  - ⇒ We achieve better F-measure and smaller errors
- Comparison with graphical lasso and correlation (right)
- Comparable when the model adheres exactly to graphical lasso
  - ⇒ Particular filter given by $\mathbf{H} = (\rho\mathbf{I} + \mathbf{S})^{-1/2}$
  - ⇒ For general diffusion filters $\mathbf{H}$ we outperform both methods

Inferring direct relations

- Our method can be used to sparsify a given network
- Keep direct and important edges or relations
  - ⇒ Discard indirect relations that can be explained by direct ones
- Use eigenvectors $\hat{\mathbf{V}}$ of given network as noisy templates
- Infer contact between amino-acid residues in BPT1 BOVIN
  - ⇒ Use mutual information of amino-acid covariation as input



Ground truth Mutual info. Network deconv. Our approach

- Network deconvolution assumes a specific filter model
  - ⇒ We achieve better performance by being agnostic to this

References

- V. Kalofolias, "How to learn a graph from smooth signals," AISTATS, 2016
- V. Dong, et al., "Learning Laplacian matrix in smooth graph signal representations," IEEE TSP, 2016
- B. Pasdeloup, et al., "Characterization and Inference of Graph Diffusion Processes from Observations of Stationary Signals," *arXiv preprint*, 2016
- S. Feizi, et al., "Network deconvolution as a general method to distinguish direct dependencies in networks," *Nat Biotech*, vol. 31, no. 8, pp. 726-733, 2013.
- S. Segarra, et al.,S "Network Topology Inference from Spectral Templates," *arXiv preprint*, 2016