Network Topology Identification from Imperfect Spectral Templates

Santiago Segarra

Institute for Data, Systems, and Society
Massachusetts Institute of Technology
segarra@mit.edu
http://www.mit.edu/~segarra/

Co-authors: Antonio G. Marques, Gonzalo Mateos, and Alejandro Ribeiro

Asilomar Conference, November 8, 2016
Desiderata: Process, analyze and learn from network data [Kolaczyk’09]
Network Science analytics

Online social media  Internet  Clean energy and grid analytics

Desiderata: Process, analyze and learn from network data [Kolaczyk’09]

Network as graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships

Interest here not in $G$ itself, but in data associated with nodes in $\mathcal{V}$

⇒ Object of study is a graph signal $\mathbf{x} \in \mathbb{R}^N$ ($|\mathcal{V}| = N$)

⇒ As.: Signal properties related to topology of $G$ (e.g., smoothness)
Graph signal processing (GSP)

▶ Graph $G$ with adjacency matrix $A$
  $\Rightarrow A_{ij} = \text{Proximity between } i \text{ and } j$

▶ Define a signal $x$ on top of the graph
  $\Rightarrow x_i = \text{Signal value at node } i$

▶ Graph Signal Processing $\Rightarrow$ Exploit structure encoded in $G$ to process $x$
  $\Rightarrow$ Our view: GSP well suited to study (network) diffusion processes
Graph G with adjacency matrix $A$

$A_{ij} = \text{Proximity between } i \text{ and } j$

Define a signal $x$ on top of the graph

$x_i = \text{Signal value at node } i$

Graph Signal Processing $\rightarrow$ Exploit structure encoded in $G$ to process $x$

$\Rightarrow$ Our view: GSP well suited to study (network) diffusion processes

Associated with $G$ is the graph-shift operator $S = \mathbf{V}\Lambda\mathbf{V}^H \in \mathbb{R}^{N \times N}$

$S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ (captures local structure in $G$)

Ex: Adjacency $A$, degree $D$, and Laplacian $L = D - A$ matrices
Network topology inference from nodal observations [Kolaczyk’09]
  ⇒ Test Pearson correlations to construct graphs
  ⇒ Partial correlations and conditional dependence

Key in neuroscience [Sporns’10]
  ⇒ Functional net inferred from activity
Network topology inference from nodal observations [Kolaczyk’09]
  ⇒ Test Pearson correlations to construct graphs
  ⇒ Partial correlations and conditional dependence

Key in neuroscience [Sporns’10]
  ⇒ Functional net inferred from activity

Most GSP works assume that $S$ (i.e., $G$) is known [Shuman et al’13]
  ⇒ Analyze how the characteristics of $S$ affect signals and filters

We take the reverse path
  ⇒ How to use GSP to infer the graph topology?
  ⇒ Other approaches: [Dong15, Mei15, Pavez16, Pasdeloup16]
Our approach for topology inference

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

1. **STEP 1:** Identify the eigenvectors of the shift
2. **STEP 2:** Identify eigenvalues to obtain a suitable shift

- Alternative sources for **spectral templates**
  - Design of graph filters [Segarra et al’15]
  - Graph sparsification
  - Network deconvolution [Feizi et al’13]
Step 1: Obtaining the eigenvectors

- \( \mathbf{x} \) is a **stationary process** on the unknown graph \( S \)
- \( \Rightarrow \) Observed \( \{x_i\} \) are random realizations of \( \mathbf{x} \)
- \( \Rightarrow \) Eigenvectors \( \mathbf{V} \) can be recovered from covariance \( \mathbf{C}_x \)

- Signal \( \mathbf{x} \) is the response of a linear diffusion process to a white input

\[
\mathbf{x} = \alpha_0 \prod_{l=1}^{\infty} (\mathbf{I} - \alpha_l \mathbf{S}) \mathbf{w} = \sum_{l=0}^{\infty} \beta_l \mathbf{S}^l \mathbf{w} = \left( \sum_{l=0}^{N-1} \mathbf{h}_l \mathbf{S}^l \right) \mathbf{w} := \mathbf{H} \mathbf{w}
\]

- Common generative model. Heat diffusion if \( \alpha_k \) constant
- \( \mathbf{H} \) is a **graph filter** on the unknown graph
- \( \mathbf{H} \) diagonalized by the eigenvectors \( \mathbf{V} \) of the shift operator \( \mathbf{S} \)
Step 1: Obtaining the eigenvectors

- The covariance matrix of the signal $x$ is

$$C_x = E \left[ \left( Hw \left( Hw^H \right) \right) \right] = H E \left[ \left( ww^H \right) \right] H^H = HH^H$$

- Since $H$ is diagonalized by $V$, so is the covariance $C_x$

$$C_x = V \left| \sum_{l=0}^{L-1} h_l \Lambda_l \right|^2 V^H$$

- Any shift with eigenvectors $V$ can explain $x$

  $\Rightarrow$ $G$ and its specific eigenvalues have been obscured by diffusion

Observations

(a) Identifying $S$ $\rightarrow$ Identifying the eigenvalues

(b) Correlation methods $\rightarrow$ Eigenvalues are kept unchanged

(c) Precision methods $\rightarrow$ Eigenvalues are inverted
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

\[ S^* := \underset{S, \lambda}{\arg\min} f(S, \lambda) \quad \text{s. to} \quad S = \sum_{k=1}^{N} \lambda_k v_k v_k^H, \quad S \in S \quad (1) \]

▶ Set \( S \) contains all admissible scaled adjacency matrices

\[ S := \{S \mid S_{ij} \geq 0, \ S \in \mathcal{M}^N, \ S_{ii} = 0, \ \sum_j S_{1j} = 1\} \]

⇒ Can accommodate Laplacian matrices as well

▶ Problem is convex if we select a convex objective \( f(S, \lambda) \)

\textbf{Ex:} Minimum energy \((f(S) = \|S\|_F)\), fast mixing \((f(\lambda) = -\lambda_2)\)
Whenever the feasibility set of (1) is non-trivial
\[ f(S, \lambda) \] determines the features of the recovered graph

**Ex:** Identify sparsest shift \( S_0^* \) that explains observed signal structure
\[ \Rightarrow \text{Set the cost } f(S, \lambda) = \|S\|_0 \]

Non-convex problem, relax to \( \ell_1 \)-norm minimization, e.g., [Tropp'06]

\[ S_1^* := \arg\min_{S,\lambda} \|S\|_1 \quad \text{s. to } \quad S = \sum_{k=1}^{N} \lambda_k v_k v_k^H, \quad S \in S \]

Does the solution \( S_1^* \) coincide with the \( \ell_0 \) solution \( S_0^* \)?
Recovery guarantee

- Define $\mathbf{W} := \mathbf{V} \odot \mathbf{V}$, where $\odot$ is the Khatri-Rao product
  - Denote by $\mathcal{D}$ the index set such that $\text{vec}(\mathbf{S})_{\mathcal{D}} = \text{diag}(\mathbf{S})$
- Build $\mathbf{M} := (\mathbf{I} - \mathbf{WW}^\dagger)_{\mathcal{D}^c}$ the orthogonal projector onto $\text{range}(\mathbf{W})$
  - Construct $\mathbf{R} := [\mathbf{M}, \mathbf{e}_1 \otimes 1_{N-1}]$
  - Denote by $\mathcal{K}$ the indices of the support of $\mathbf{s}_0^* = \text{vec}(\mathbf{S}_0^*)$

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

- Cond. 1) ensures uniqueness of solution $\mathbf{S}_1^*$
- Cond. 2) guarantees existence of a dual certificate for $\ell_0$ optimality
We might have access to \( \hat{V} \), a noisy version of the spectral templates. With \( d(\cdot, \cdot) \) denoting a (convex) distance between matrices,

\[
\min_{\{S, \lambda, \hat{S}\}} \|S\|_1 \quad \text{s. to} \quad \hat{S} = \sum_{k=1}^{N} \lambda_k \hat{v}_k \hat{v}_k^H, \quad S \in S, \quad d(S, \hat{S}) \leq \epsilon
\]

How does the noise in \( \hat{V} \) affect the recovery?
We might have access to $\hat{V}$, a noisy version of the spectral templates.

With $d(\cdot, \cdot)$ denoting a (convex) distance between matrices,

$$\min_{\{S, \lambda, \hat{S}\}} \|S\|_1 \quad \text{s. to} \quad \hat{S} = \sum_{k=1}^{N} \lambda_k \hat{v}_k \hat{v}_k^H, \quad S \in S, \quad d(S, \hat{S}) \leq \epsilon$$

How does the noise in $\hat{V}$ affect the recovery?

Stable (robust) recovery can be established.

Conditions 1) and 2) but based on $\hat{R}$, guaranteed $d(S^*, S_0^*) \leq C\epsilon$

$\Rightarrow \epsilon$ large enough to guarantee feasibility of $S_0^*$

$\Rightarrow$ Constant $C$ depends on $\hat{V}$ and the support $K$
Incomplete spectral templates

- Partial access to $\mathbf{V} \Rightarrow$ Only $K$ known eigenvectors $[v_1, \ldots, v_K]$
  $\Rightarrow$ E.g., if (sample) covariance is rank-deficient

$$\min_{\{S, S_{\tilde{K}}, \lambda\}} \|S\|_1 \text{ s. to } S = S_{\tilde{K}} + \sum_{k=1}^{K} \lambda_k v_k v_k^H, \ S \in S, \ S_{\tilde{K}} v_k = 0$$

- How does the (partial) knowledge of $\mathbf{V}_K$ affect the recovery?
Incomplete spectral templates

- Partial access to $\mathbf{V} \Rightarrow$ Only $K$ known eigenvectors $[v_1, \ldots, v_K]$
  $\Rightarrow$ E.g., if (sample) covariance is rank-deficient

\[
\min_{\{S, S_{\overline{K}}, \lambda\}} \|S\|_1 \; \text{s. to} \; S = S_{\overline{K}} + \sum_{k=1}^{K} \lambda_k v_k v_k^H, \; S \in S, \; S_{\overline{K}} v_k = 0
\]

- How does the (partial) knowledge of $\mathbf{V}_K$ affect the recovery?

- Define $\mathbf{P} := [P_1, P_2]$ in terms of $\mathbf{V}_K$, and $\Upsilon := [I_{N^2}, 0_{N^2 \times N^2}]$

$\mathbf{S}^*$ and $\mathbf{S}_0^*$ coincide if the two following conditions are satisfied:
1) $\text{rank}([P_1^T, P_2^T]) = |\mathcal{K}| + N^2$; and
2) There exists a constant $\delta > 0$ such that

\[
\eta_P := \|\Upsilon_{\mathcal{K}^c} (\delta^{-2} \mathbf{P} \mathbf{P}^T + \Upsilon_{\mathcal{K}^c} \Upsilon_{\mathcal{K}^c}^{-1} \Upsilon_{\mathcal{K}^c}^T)\|_\infty < 1.
\]

- For $K = N$, guarantees boil down to the noiseless case
Social graphs from imperfect templates

- Identification of multiple social networks $N = 32$
  - Defined on the same node set of students from Ljubljana
  - Synthetic signals from diffusion processes in the graphs
- Recovery for noisy (left) and incomplete (right) spectral templates

- Error (left) decreases with increasing number of observed signals
- Error (right) decreases with increasing nr. of spectral templates
Performance comparisons

- Comparison with graphical lasso and sparse correlation methods
  - Evaluated on 100 realizations of ER graphs with $N = 20$ and $p = 0.2$

- Graphical lasso implicitly assumes a filter $H_1 = (\rho I + S)^{-1/2}$
  - For this filter spectral templates work, but not as well

- For general diffusion filters $H_2$ spectral templates still work fine
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 eigenvalues \( \hat{\mathbf{V}} \) of given network as noisy templates

**Ex:** Infer contact between amino-acid residues in BPT1 BOVIN
  - Use mutual information of amino-acid covariation as input

- Network deconvolution assumes a specific filter model [Feizi et al’13]
  - We achieve better performance by being agnostic to this
Network topology inference cornerstone problem in Network Science
- Most GSP works analyze how $S$ affect signals and filters
- Here, reverse path: How to use GSP to infer the graph topology?

Our GSP approach to network topology inference
⇒ Two step approach: i) Obtain $V$; ii) Estimate $S$ given $V$
Network topology inference cornerstone problem in Network Science

Most GSP works analyze how $S$ affect signals and filters

Here, reverse path: How to use GSP to infer the graph topology?

Our GSP approach to network topology inference

Two step approach: i) Obtain $V$; ii) Estimate $S$ given $V$

How to obtain the spectral templates $V$

Based on covariance of diffused signals

Other sources: network operators, network deconvolution
Network topology inference cornerstone problem in Network Science

- Most GSP works analyze how $S$ affect signals and filters
- Here, reverse path: How to use GSP to infer the graph topology?

Our GSP approach to network topology inference

- Two step approach: i) Obtain $V$; ii) Estimate $S$ given $V$

How to obtain the spectral templates $V$

- Based on covariance of diffused signals
- Other sources: network operators, network deconvolution

Infer $S$ via convex optimization

- Objectives promotes desirable properties
- Constraints encode structure a priori info and structure
- Formulations for perfect and imperfect templates
- Sparse recovery results for adjacency and normalized Laplacian