Robust Network Topology Inference

Santiago Segarra

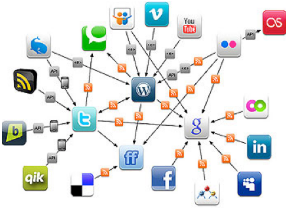
Institute for Data, Systems, and Society
Massachusetts Institute of Technology
segarra@mit.edu
http://www.mit.edu/~segarra/

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

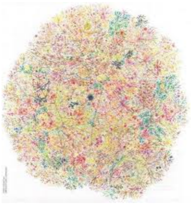
ICASSP, March 9, 2017

Network Science analytics

Online social media



Internet

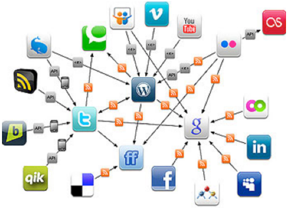


Clean energy and grid analytics

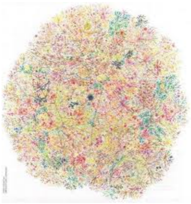
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]

Network Science analytics

Online social media



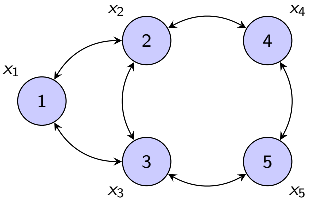
Internet



Clean energy and grid analytics

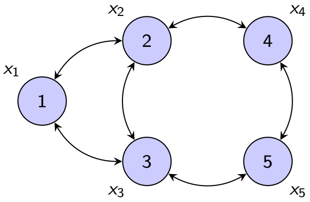
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]
- Network as graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Interest here not in G itself, but in data associated with nodes in $\mathcal{V}$
  - $\Rightarrow$ Object of study is a graph signal
  - $\Rightarrow$ **As.:** Signal properties related to topology of G (e.g., smoothness)

- Undirected G with adjacency matrix $\mathbf{A}$
  $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph
  $\Rightarrow x_i$ = Signal value at node i



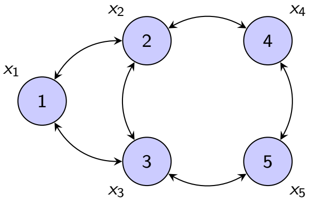
Graph signal processing (GSP)

- Undirected G with adjacency matrix $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i



- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \in \mathcal{M}^N$
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ (local structure in G)
  - $\Rightarrow$ Ex: $\mathbf{A}$, degree $\mathbf{D}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices

- Undirected G with adjacency matrix $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i

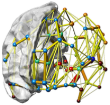


- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \in \mathcal{M}^N$
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ (local structure in G)
  - $\Rightarrow$ Ex: $\mathbf{A}$, degree $\mathbf{D}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices
- Graph Signal Processing $\rightarrow$ Exploit structure encoded in $\mathbf{S}$ to process $\mathbf{x}$
  - $\Rightarrow$ Our view: GSP well suited to study (network) diffusion processes

Motivation and context

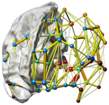
- Network topology inference from nodal observations [Kolaczyk'09]
  - ⇒ Approaches use Pearson correlations to construct graphs [Brovelli04]
  - ⇒ Partial correlations and conditional dependence [Friedman08, Karanikolas16]

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



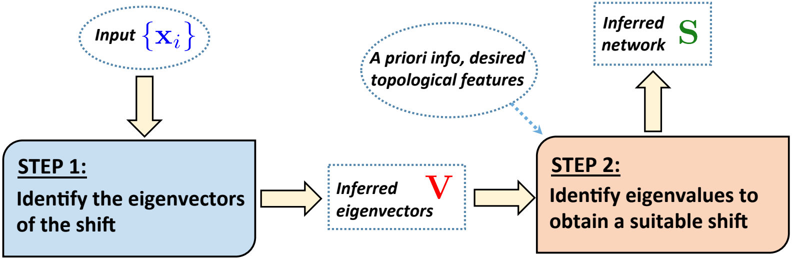
Motivation and context

- Network topology inference from nodal observations [Kolaczyk'09]
  - ⇒ Approaches use Pearson correlations to construct graphs [Brovelli04]
  - ⇒ Partial correlations and conditional dependence [Friedman08, Karanikolas16]
- Key in neuroscience [Sporns'10]
  - ⇒ Functional net inferred from activity



- Most GSP works: How known graph **S** affects signals and filters
- Here, reverse path: How to use GSP to infer the graph topology?
  - ⇒ Gaussian graphical models [Egilmez16]
  - ⇒ Smooth signals [Dong15], [Kalofolias16]
  - ⇒ Stationary signals [Segarra16], [Pasdeloup16]
  - ⇒ Directed graphs [Mei-Moura15], [Shen16]
- **Today's talk:** Guarantees of robustness in topology inference

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



- Alternative sources for spectral templates $\mathbf{V}$
  - Design of graph filters [Segarra et al'15]
  - Graph sparsification and Network deconvolution [Feizi et al'13]
- Small number of $\{\mathbf{x}_i\}$ or specific signal features

  $\Rightarrow$ May lead to noisy or incomplete eigenvectors $\hat{\mathbf{V}}$
- How good is the recovery of $\mathbf{S}$ when $\hat{\mathbf{V}}$ (instead of $\mathbf{V}$) is available?

Step 1: Obtaining the eigenvectors

- $\mathbf{x}$ is a stationary process on the unknown graph $\mathbf{S}$
  - $\Rightarrow$ Observed $\{\mathbf{x}_i\}$ 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

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

- Common generative model. Heat diffusion if α_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 $\mathbf{x}$ is

$$\mathbf{C}_x = \mathbb{E}\left[\left(\mathbf{Hw}\left(\mathbf{Hw}\right)^H\right)\right] = \mathbf{H}\mathbb{E}\left[\left(\mathbf{ww}^H\right)\right]\mathbf{H}^H = \mathbf{HH}^H$$

- Since $\mathbf{H}$ is diagonalized by $\mathbf{V}$, so is the covariance $\mathbf{C}_x$

$$\mathbf{C}_x \;=\; \mathbf{V}\left|\sum_{l=0}^{L-1} h_l \boldsymbol{\Lambda}^l\right|^2 \mathbf{V}^H$$

- Any shift with eigenvectors $\mathbf{V}$ can explain $\mathbf{x}$

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

Observations

(a) Identifying $\mathbf{S}$ $\rightarrow$ 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
  - $\Rightarrow$ Of all graphs, select one that is optimal in some sense

$$\mathbf{S}_0^* := \underset{\mathbf{S},\boldsymbol{\lambda}}{\text{argmin}} \ \|\mathbf{S}\|_0 \quad \text{s. to} \quad \mathbf{S} = \sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^H, \ \ \mathbf{S} \in \mathcal{S}$$

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

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

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

$$\mathbf{S}_0^* := \underset{\mathbf{S},\boldsymbol{\lambda}}{\text{argmin}} \ \|\mathbf{S}\|_0 \quad \text{s. to} \quad \mathbf{S} = \sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^H, \ \ \mathbf{S} \in \mathcal{S}$$

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

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

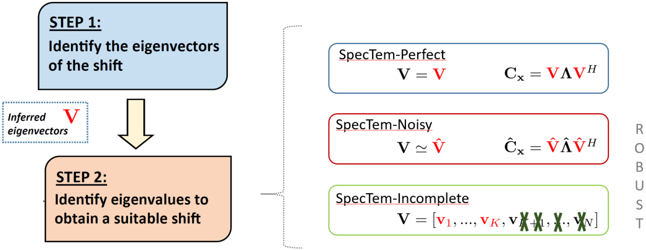
- Non-convex problem, relax to ℓ_1-norm minimization, e.g., [Tropp'06]

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

- What if $\mathbf{V}$ is not available? $\Rightarrow$ Noisy and/or incomplete $\hat{\mathbf{V}}$

Robust shift identification

- Two-step algorithm based on perfect spectral templates
  - ⇒ However, perfect knowledge of **V** may not be available
  - ⇒ Robust designs?



- Q1: How to modify the optimization in step 2?
  - ⇒ Distance for noise, orthogonal subspace for incomplete
- Q2: Recovery guarantees?

Incomplete spectral templates

- Partial access to $\mathbf{V}$ $\Rightarrow$ Only K known eigenvectors $[v_1, \ldots, v_K]$

$$\min_{\{\mathbf{S}, \mathbf{S}_{\bar{K}}, \boldsymbol{\lambda}\}} \|\mathbf{S}\|_1 \;\; \text{s. to} \;\; \mathbf{S} = \mathbf{S}_{\bar{K}} + \textstyle\sum_{k=1}^{K} \lambda_k \mathbf{v}_k \mathbf{v}_k{}^H, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{S}_{\bar{K}} \mathbf{v}_k = \mathbf{0}$$

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

- Partial access to $\mathbf{V}$ $\Rightarrow$ Only K known eigenvectors $[v_1, \ldots, v_K]$

$$\min_{\{\mathbf{S}, \mathbf{S}_{\bar{K}}, \boldsymbol{\lambda}\}} \|\mathbf{S}\|_1 \text{ s. to } \mathbf{S} = \mathbf{S}_{\bar{K}} + \sum_{k=1}^{K} \lambda_k \mathbf{v}_k \mathbf{v}_k{}^H, \ \mathbf{S} \in \mathcal{S}, \ \mathbf{S}_{\bar{K}} \mathbf{v}_k = \mathbf{0}$$

- How does the (partial) knowledge of $\mathbf{V}_K$ affect the recovery?
- Define $\mathbf{P} := [\mathbf{P}_1, \ \mathbf{P}_2]$ in terms of $\mathbf{V}_K$, and $\boldsymbol{\Upsilon} := [\mathbf{I}_{N^2}, \mathbf{0}_{N^2 \times N^2}]$
  $\Rightarrow$ Goal is to reformulate problem as $\min_{\mathbf{t}} \|\boldsymbol{\Upsilon}\mathbf{t}\|_1$ s.to $\mathbf{P}^T\mathbf{t} = \mathbf{b}$

$\mathbf{S}^*$ and $\mathbf{S}_0^*$ coincide if the two following conditions are satisfied:
1) $\text{rank}([\mathbf{P}_1{}_{\mathcal{K}}^T, \mathbf{P}_2{}^T]) = |\mathcal{K}| + N^2$; and
2) There exists a constant $\delta > 0$ such that

$$\eta_{\mathbf{P}} := \|\boldsymbol{\Upsilon}_{\mathcal{K}^c}(\delta^{-2}\mathbf{P}\mathbf{P}^T + \boldsymbol{\Upsilon}_{\mathcal{K}^c}^T \boldsymbol{\Upsilon}_{\mathcal{K}^c})^{-1} \boldsymbol{\Upsilon}_{\mathcal{K}}^T\|_\infty < 1.$$

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

Noisy spectral templates

- We might have access to $\hat{\mathbf{V}}$, a noisy version of the spectral templates

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

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

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

Noisy spectral templates

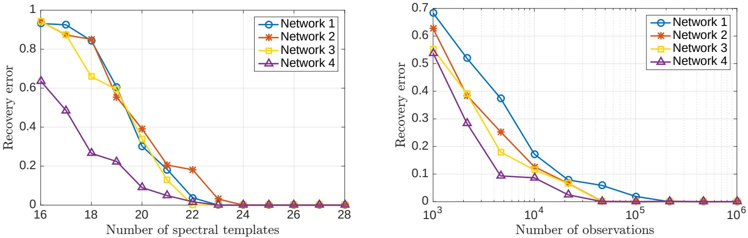
- We might have access to $\hat{\mathbf{V}}$, a noisy version of the spectral templates
  - $\Rightarrow$ With $d(\cdot,\cdot)$ denoting a (convex) distance between matrices

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

- How does the noise in $\hat{\mathbf{V}}$ affect the recovery?
- Stable recovery can be established $\Rightarrow$ depends on noise level
  - $\Rightarrow$ Reformulate problem as $\min_{\mathbf{t}} \|\mathbf{t}\|_1$ s. to $\|\mathbf{R}^T\mathbf{t} - \mathbf{b}\|_2 \leq \epsilon$
- Conditions 1) and 2) but based on $\mathbf{R}$, guaranteed $d(\mathbf{S}^*, \mathbf{S}_0^*) \leq C\epsilon$
  - $\Rightarrow$ ϵ large enough to guarantee feasibility of $\mathbf{S}_0^*$
  - $\Rightarrow$ Constant C depends on $\hat{\mathbf{V}}$ and the support $\mathcal{K}$

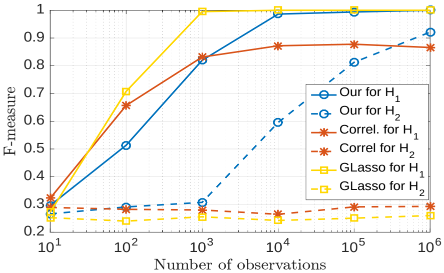
Social graphs from imperfect templates

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



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

- 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 $\mathbf{H}_1 = (\rho\mathbf{I} + \mathbf{S})^{-1/2}$

  $\Rightarrow$ For this filter spectral templates work, but not as well

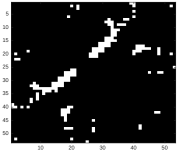
- For general diffusion filters $\mathbf{H}_2$ spectral templates still work fine

Inferring direct relations

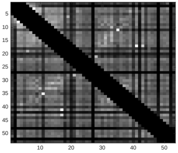
- Our method can be used to sparsify a given network
  - ⇒ Keep direct and important edges or relations
  - ⇒ Discard indirect relations that can be explained by direct ones
- Use eigenvectors $\hat{\mathbf{V}}$ of given network as noisy templates

Ex: Infer contact between amino-acid residues in BPT1 BOVIN

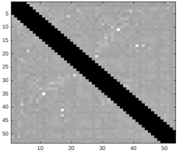
⇒ Use mutual information of amino-acid covariation as input



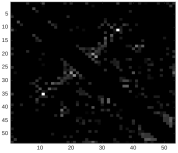
Ground truth



Mutual info.



Network deconv.



Our approach

- Network deconvolution assumes a specific filter model [Feizi et al'13]
  - ⇒ We achieve better performance by being agnostic to this

Closing remarks

- Network topology inference cornerstone problem in Network Science
  - Most GSP works analyze how $\mathbf{S}$ affect signals and filters
  - Here, reverse path: How to use GSP to infer the graph topology?
- Our GSP approach to network topology inference

  $\Rightarrow$ Two step approach: i) Obtain $\mathbf{V}$; ii) Estimate $\mathbf{S}$ given $\mathbf{V}$

Closing remarks

- Network topology inference cornerstone problem in Network Science
  - Most GSP works analyze how $\mathbf{S}$ affect signals and filters
  - Here, reverse path: How to use GSP to infer the graph topology?
- Our GSP approach to network topology inference

  $\Rightarrow$ Two step approach: i) Obtain $\mathbf{V}$; ii) Estimate $\mathbf{S}$ given $\mathbf{V}$
- How to obtain the spectral templates $\mathbf{V}$

  $\Rightarrow$ Based on covariance of stationary signals

  $\Rightarrow$ Other sources: network operators, network deconvolution

Closing remarks

- Network topology inference cornerstone problem in Network Science
  - Most GSP works analyze how $\mathbf{S}$ affect signals and filters
  - Here, reverse path: How to use GSP to infer the graph topology?
- Our GSP approach to network topology inference
  - $\Rightarrow$ Two step approach: i) Obtain $\mathbf{V}$; ii) Estimate $\mathbf{S}$ given $\mathbf{V}$
- How to obtain the spectral templates $\mathbf{V}$
  - $\Rightarrow$ Based on covariance of stationary signals
  - $\Rightarrow$ Other sources: network operators, network deconvolution
- Infer $\mathbf{S}$ via convex optimization
  - $\Rightarrow$ Objectives promotes desirable properties
  - $\Rightarrow$ Constraints encode structure a priori info and structure
  - $\Rightarrow$ Formulations for noisy and incomplete templates