Network Topology Identification from Imperfect Spectral Templates

Santiago Segarra

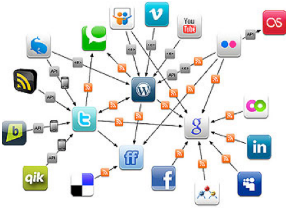
Institute for Data, Systems, and Society
Massachusetts Institute of Technology
segarra@mit.edu
http://www.mit.edu/~segarra/

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

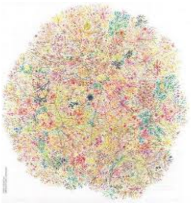
Asilomar Conference, November 8, 2016

Network Science analytics

Online social media



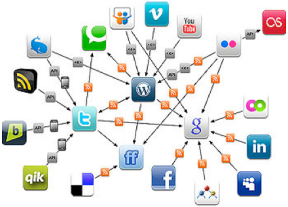
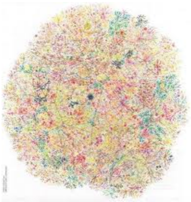
Internet



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
		

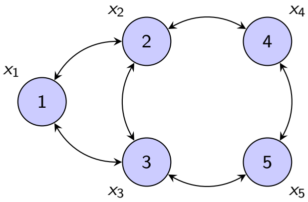
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]
- Network as graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Interest here not in G itself, but in data associated with nodes in $\mathcal{V}$
  - $\Rightarrow$ Object of study is a graph signal $\mathbf{x}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

- Graph 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

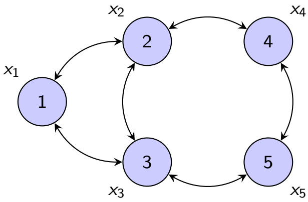


- Graph Signal Processing $\rightarrow$ Exploit structure encoded in G to process $\mathbf{x}$

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

Graph signal processing (GSP)

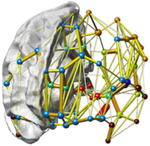
- Graph G with adjacency matrix $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i



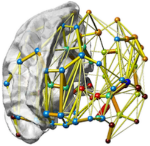
- Graph Signal Processing $\rightarrow$ Exploit structure encoded in G to process $\mathbf{x}$
  - $\Rightarrow$ Our view: GSP well suited to study (network) diffusion processes
- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \in \mathbb{R}^{N \times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

Ex: Adjacency $\mathbf{A}$, degree $\mathbf{D}$, and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ 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

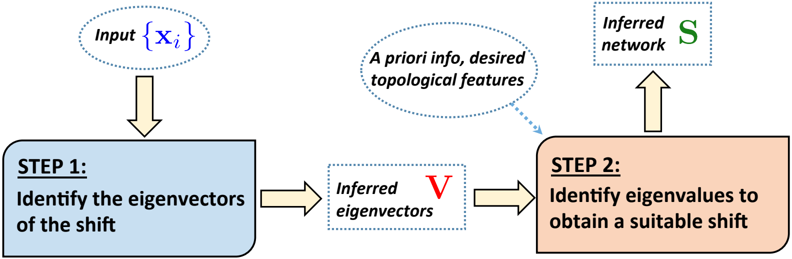


- Network topology inference from nodal observations [Kolaczyk'09]
  - ⇒ Test Pearson correlations to construct graphs
  - ⇒ Partial correlations and conditional dependence
- Key in neuroscience [Sporns'10]
  - ⇒ Functional net inferred from activity



- Most GSP works assume that $\mathbf{S}$ (i.e., G) is known [Shuman et al'13]
  - ⇒ Analyze how the characteristics of $\mathbf{S}$ affect signals and filters
- We take the reverse path
  - ⇒ How to use GSP to infer the graph topology?
  - ⇒ Other approaches: [Dong15, Mei15, Pavez16, Pasdeloup16]

- 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
  - Network deconvolution [Feizi et al'13]

- $\mathbf{x}$ 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 $\mathbf{C}_x$

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

- 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

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

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

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

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

  - $\Rightarrow$ Can accommodate Laplacian matrices as well

- Problem is convex if we select a convex objective $f(\mathbf{S},\boldsymbol{\lambda})$

Ex: Minimum energy ($f(\mathbf{S}) = \|\mathbf{S}\|_F$), fast mixing ($f(\boldsymbol{\lambda}) = -\lambda_2$)

- Whenever the feasibility set of (1) is non-trivial

  $\Rightarrow$ $f(\mathbf{S}, \boldsymbol{\lambda})$ determines the features of the recovered graph

  Ex: Identify sparsest shift $\mathbf{S}_0^*$ that explains observed signal structure

  $\Rightarrow$ Set the cost $f(\mathbf{S}, \boldsymbol{\lambda}) = \|\mathbf{S}\|_0$

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

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

- Does the solution $\mathbf{S}_1^*$ coincide with the ℓ_0 solution $\mathbf{S}_0^*$?

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

$$\psi_{\mathbf{R}} := \| \mathbf{I}_{\mathcal{K}^c} (\delta^{-2} \mathbf{R}\mathbf{R}^T + \mathbf{I}_{\mathcal{K}^c}^T \mathbf{I}_{\mathcal{K}^c})^{-1} \mathbf{I}_{\mathcal{K}}^T \|_{\infty} < 1.$$

- Cond. 1) ensures uniqueness of solution $\mathbf{S}_1^*$
- 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

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

- 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 (robust) recovery can be established
- Conditions 1) and 2) but based on $\hat{\mathbf{R}}$, 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 $\mathcal{K}$

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

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

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

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

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

- How does the (partial) knowledge of $\mathbf{V}_K$ affect the recovery?
- Define $\mathbf{P} := [\mathbf{P}_1, \;\mathbf{P}_2]$ in terms of $\mathbf{V}_K$, and $\boldsymbol{\Upsilon} := [\mathbf{I}_{N^2}, \mathbf{0}_{N^2 \times N^2}]$

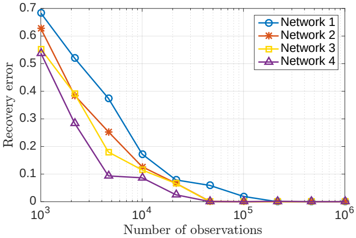
$\mathbf{S}^*$ and $\mathbf{S}_0^*$ coincide if the two following conditions are satisfied:
1) $\text{rank}([\mathbf{P}_1{}_{\mathcal{K}}^T, \mathbf{P}_2{}^T]) = |\mathcal{K}| + N^2$; and
2) There exists a constant $\delta > 0$ such that

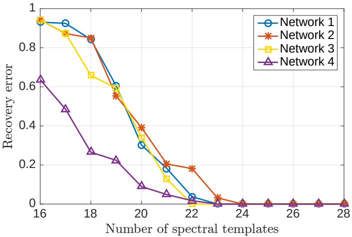
$$\eta_{\mathbf{P}} := \|\boldsymbol{\Upsilon}_{\mathcal{K}^c}(\delta^{-2}\mathbf{P}\mathbf{P}^T + \boldsymbol{\Upsilon}_{\mathcal{K}^c}^T \boldsymbol{\Upsilon}_{\mathcal{K}^c})^{-1} \boldsymbol{\Upsilon}_{\mathcal{K}}^T\|_\infty < 1.$$

- For $K = N$, guarantees boil down to the noiseless case

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 noisy (left) and incomplete (right) spectral templates

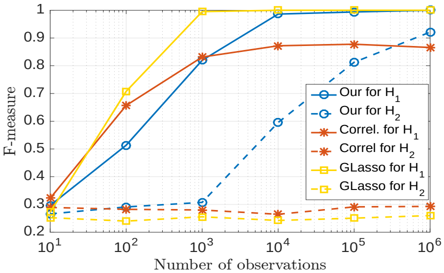




- Error (left) decreases with increasing number of observed signals
- Error (right) decreases with increasing nr. of spectral templates

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

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

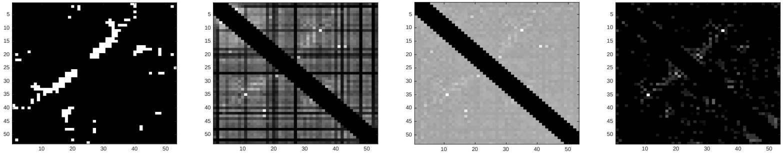
- For general diffusion filters $\mathbf{H}_2$ spectral templates still work fine

Inferring direct relations

- Our method can be used to sparsify a given network
  - $\Rightarrow$ Keep direct and important edges or relations
  - $\Rightarrow$ 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

$\Rightarrow$ 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]
  - $\Rightarrow$ 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 diffused 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 diffused 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 perfect and imperfect templates

  $\Rightarrow$ Sparse recovery results for adjacency and normalized Laplacian