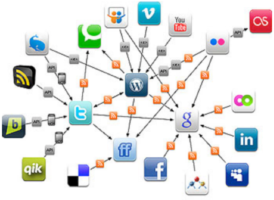
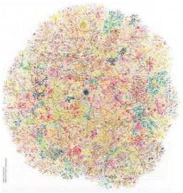
Identifying Undirected Network Structure via Semidefinite Relaxation

Rasoul Shafipour, **Santiago Segarra**, Antonio G. Marques and Gonzalo Mateos

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

ICASSP, April 20, 2018

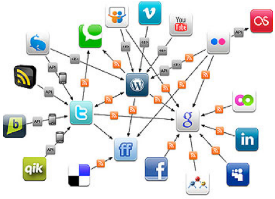
Network Science analytics

Online social media	Internet	Clean energy and grid analytics
		

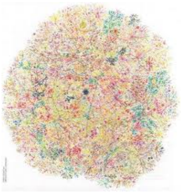
- Desiderata: Process, analyze and learn from network data [Kolaczyk09]

Network Science analytics

Online social media



Internet

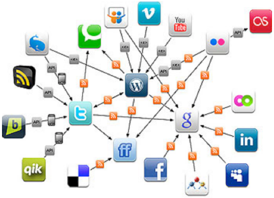


Clean energy and grid analytics

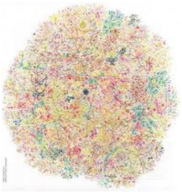
- Desiderata: Process, analyze and learn from network data [Kolaczyk09]
- Network as graph G: encode pairwise relationships
- Sometimes both G and data at the nodes are available

  $\Rightarrow$ Leverage G to process network data $\Rightarrow$ Graph Signal Processing

Online social media



Internet



Clean energy and grid analytics

- Desiderata: Process, analyze and learn from network data [Kolaczyk09]
- Network as graph G: encode pairwise relationships
- Sometimes both G and data at the nodes are available

  $\Rightarrow$ Leverage G to process network data $\Rightarrow$ Graph Signal Processing
- Sometimes we have access to network data but not to G itself

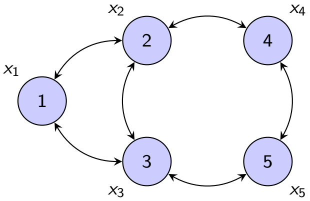
  $\Rightarrow$ Leverage the relation between them to infer G from the data

- Undirected G with adjacency matrix $\mathbf{A}$

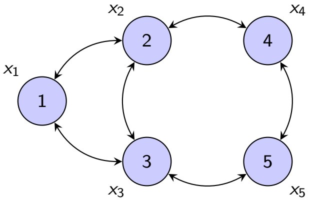
  $\Rightarrow A_{ij}$ = Proximity between i and j

- Define a signal $\mathbf{x}$ on top of the graph

  $\Rightarrow x_i$ = Signal value at node i



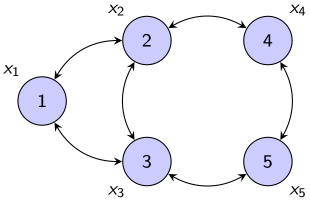
- Undirected G with adjacency matrix $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x}$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i



- 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: adjacency $\mathbf{A}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices

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

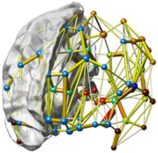


- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \in \mathcal{M}^N$
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ (local structure in G)
  - $\Rightarrow$ Ex: adjacency $\mathbf{A}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices
- Graph filters $H : \mathbb{R}^N \to \mathbb{R}^N$ are maps between graph signals
  - $\Rightarrow$ Polynomial in $\mathbf{S}$ with coefficients $\mathbf{h} \in \mathbb{R}^{L+1} \Rightarrow \mathbf{H} := \sum_{l=0}^{L} h_l \mathbf{S}^l$
- How to use GSP to infer the graph topology?

- Network topology inference from nodal observations [Kolaczyk09]
  - Partial correlations and conditional dependence [Dempster74]
  - Sparsity [Friedman07] and consistency [Meinshausen06]
  - [Banerjee08], [Lake10], [Slawski15], [Karanikolas16]

- Key in neuroscience [Sporns10]

  $\Rightarrow$ Functional net inferred from activity

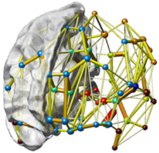


Topology inference: Motivation and context

- Network topology inference from nodal observations [Kolaczyk09]
  - Partial correlations and conditional dependence [Dempster74]
  - Sparsity [Friedman07] and consistency [Meinshausen06]
  - [Banerjee08], [Lake10], [Slawski15], [Karanikolas16]

- Key in neuroscience [Sporns10]

  $\Rightarrow$ Functional net inferred from activity

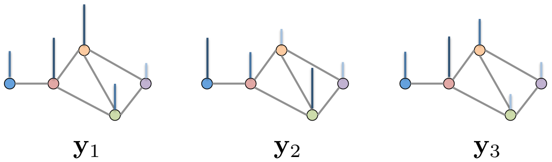


- Noteworthy GSP-based approaches
  - Gaussian graphical models [Egilmez16]
  - Smooth signals [Dong15], [Kalofolias16]
  - Stationary signals [Pasdeloup15], [Segarra16]
  - Directed graphs [Mei15], [Shen16]
  - Low-rank excitation [Wai18]

- **Our contribution:** topology inference from non-stationary graph signals

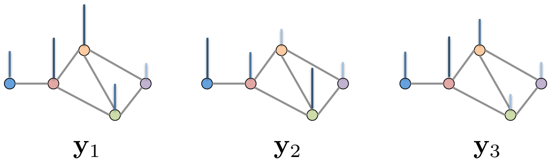
- Underlying graph $\mathcal{G}$ with undirected unknown GSO $\mathbf{S}$
- Observe signals $\{\mathbf{y}_i\}_{i=1}^{K}$ defined on the unknown graph

Setup



- Underlying graph $\mathcal{G}$ with undirected unknown GSO $\mathbf{S}$
- Observe signals $\{\mathbf{y}_i\}_{i=1}^K$ defined on the unknown graph

Setup



Problem statement

Given observations $\{\mathbf{y}_i\}_{i=1}^K$, determine the network $\mathbf{S}$ knowing that: $\{\mathbf{y}_i\}_{i=1}^K$ are outputs of a diffusion process on $\mathbf{S}$.

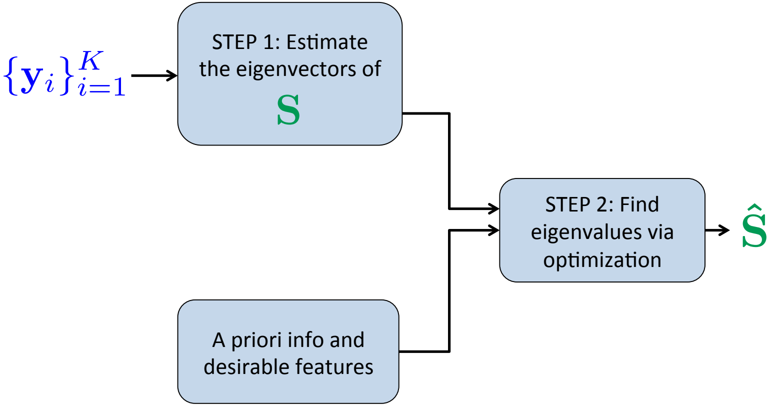
Problem formulation

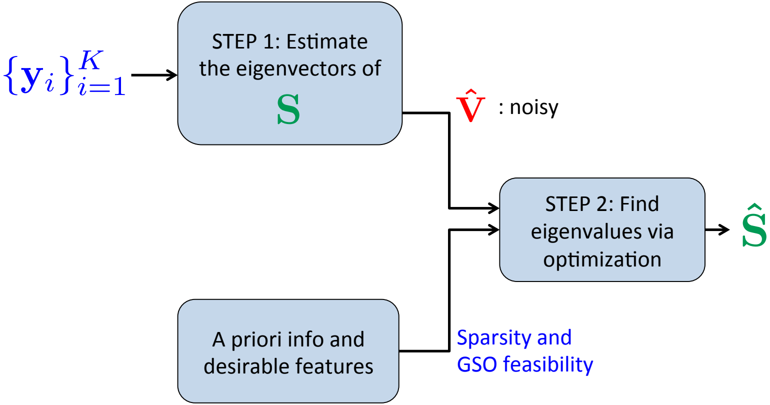
- Consider an arbitrary linear network process on the GSO $\mathbf{S}$
  - $\Rightarrow$ Every realization corresponds to a different input $\mathbf{x}_i$

$$\mathbf{y}_i = \left(\sum_{l=0}^{L} h_l \mathbf{S}^l\right) \mathbf{x}_i = \mathbf{H}\mathbf{x}_i, \quad i = 1, \ldots, K$$

- **Goal**: Recover $\mathbf{S}$ from the observation of K signals $\{\mathbf{y}_i\}_{i=1}^{K}$
- **Additional unknowns**
  - $\Rightarrow$ The degree of the filter L
  - $\Rightarrow$ The filter coefficients $\{h_l\}_{l=0}^{L}$
  - $\Rightarrow$ The specific inputs $\mathbf{x}_i$; but we know that $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{x}})$

Blueprint of our solution




$\{\mathbf{y}_i\}_{i=1}^K$
STEP 1: Estimate the eigenvectors of $\mathbf{S}$
$\hat{\mathbf{V}}$: noisy
STEP 2: Find eigenvalues via optimization
$\hat{\mathbf{S}}$
A priori info and desirable features
Sparsity and GSO feasibility

Step 1: Estimating the eigenvectors of $\mathbf{S}$

- $\mathbf{y}$ is the output of a local diffusion process on the graph

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

- Whenever the input $\mathbf{x}$ is white

  $\Rightarrow$ graph stationary process on $\mathbf{S}$ [Marques17, Girault15, Perraudin17]

- $\mathbf{y}$ is the output of a local diffusion process on the graph

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

- Whenever the input $\mathbf{x}$ is white

  $\Rightarrow$ graph stationary process on $\mathbf{S}$ [Marques17, Girault15, Perraudin17]

Stationary case

- The covariance $\mathbf{C_y}$ of $\mathbf{y}$ shares $\mathbf{V}$ with $\mathbf{S}$

$$\mathbf{C_y} = \mathbf{H}^2 = h_0^2 \mathbf{I} + 2 h_0 h_1 \mathbf{S} + h_1^2 \mathbf{S}^2 + \ldots$$

- Estimate covariance from $\{\mathbf{y}_i\}_{i=1}^{K}$ as $\hat{\mathbf{C}}_\mathbf{y}$ $\Rightarrow$ Diagonalize $\Rightarrow$ Obtain $\hat{\mathbf{V}}$

- **Q:** What if the signal $\mathbf{y} = \mathbf{H}\mathbf{x}$ is not stationary (i.e., $\mathbf{x}$ colored)?

  $\Rightarrow$ Matrices $\mathbf{S}$ and $\mathbf{C_y}$ no longer simultaneously diagonalizable since

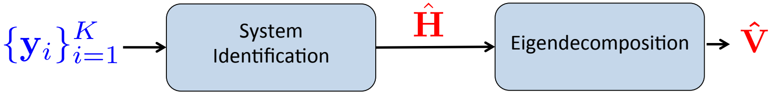
$$\mathbf{C_y} = \mathbf{H}\mathbf{C_x}\mathbf{H}$$

Non-stationary graph signals

- **Q:** What if the signal $\mathbf{y} = \mathbf{H}\mathbf{x}$ is not stationary (i.e., $\mathbf{x}$ colored)?
  - $\Rightarrow$ Matrices $\mathbf{S}$ and $\mathbf{C_y}$ no longer simultaneously diagonalizable since

$$\mathbf{C_y} = \mathbf{H}\mathbf{C_x}\mathbf{H}$$

- Key: still $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ diagonalized by the eigenvectors $\mathbf{V}$ of $\mathbf{S}$
  - $\Rightarrow$ Infer $\mathbf{V}$ by estimating the unknown diffusion (graph) filter $\mathbf{H}$
  - $\Rightarrow$ Step 1 boils down to system identification + eigendecomposition



Define $\mathbf{C}_{\mathbf{xyx}} := \mathbf{C}_{\mathbf{x}}^{1/2}\mathbf{C}_{\mathbf{y}}\mathbf{C}_{\mathbf{x}}^{1/2}$, with eigenvectors $\mathbf{V}_{\mathbf{xyx}}$. If $\mathbf{C}_{\mathbf{x}}$ is non-singular then all admissible symmetric filters $\mathbf{H}$ are of the form

$$\mathbf{H} = \mathbf{C}_{\mathbf{x}}^{-1/2}\mathbf{C}_{\mathbf{xyx}}^{1/2}\mathbf{V}_{\mathbf{xyx}}\mathrm{diag}(\mathbf{b})\mathbf{V}_{\mathbf{xyx}}^{T}\mathbf{C}_{\mathbf{x}}^{-1/2},$$

where $\mathbf{b} \in \{-1, 1\}^N$ is a binary (signed) vector.

Define $\mathbf{C}_{\mathsf{xyx}} := \mathbf{C}_{\mathsf{x}}^{1/2}\mathbf{C}_{\mathsf{y}}\mathbf{C}_{\mathsf{x}}^{1/2}$, with eigenvectors $\mathbf{V}_{\mathsf{xyx}}$. If $\mathbf{C}_{\mathsf{x}}$ is non-singular then all admissible symmetric filters $\mathbf{H}$ are of the form

$$\mathbf{H} = \mathbf{C}_{\mathsf{x}}^{-1/2}\mathbf{C}_{\mathsf{xyx}}^{1/2}\mathbf{V}_{\mathsf{xyx}}\mathrm{diag}(\mathbf{b})\mathbf{V}_{\mathsf{xyx}}^{T}\mathbf{C}_{\mathsf{x}}^{-1/2},$$

where $\mathbf{b} \in \{-1, 1\}^N$ is a binary (signed) vector.

- Even if we get $\mathbf{C}_{\mathsf{y}}$ exactly, $\mathbf{H}$ is not identifiable
  - $\Rightarrow$ Not surprising since we only have second moment info
- Consider having access to multiple input distributions $\{\mathbf{C}_{\mathsf{x},m}\}_{m=1}^{M}$

Multiple input processes

- Define $\mathbf{A}_m := (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{V}_{\mathbf{xyx},m}) \odot (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{C}_{\mathbf{xyx},m}^{1/2}\mathbf{V}_{\mathbf{xyx},m})$

$$\boldsymbol{\Psi} := \begin{bmatrix} \mathbf{A}_1 & -\mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & -\mathbf{A}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{M-1} & -\mathbf{A}_M \end{bmatrix}$$

- $\mathbf{b}_m \in \{-1, 1\}^N$ and $\mathbf{b} = [\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_M^T]^T$, then $\boldsymbol{\Psi}\mathbf{b}^* = 0$

- Define $\mathbf{A}_m := (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{V}_{\mathbf{xyx},m}) \odot (\mathbf{C}_{\mathbf{x},m}^{-1/2}\mathbf{C}_{\mathbf{xyx},m}^{1/2}\mathbf{V}_{\mathbf{xyx},m})$

$$\mathbf{\Psi} := \begin{bmatrix} \mathbf{A}_1 & -\mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & -\mathbf{A}_3 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{M-1} & -\mathbf{A}_M \end{bmatrix}$$

- $\mathbf{b}_m \in \{-1, 1\}^N$ and $\mathbf{b} = [\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_M^T]^T$, then $\mathbf{\Psi}\mathbf{b}^* = 0$

Whenever only estimates $\hat{\mathbf{C}}_{\mathbf{y},m}$ are available, we can estimate $\mathbf{b}^*$ as

$$\hat{\mathbf{b}}^* = \underset{\mathbf{b}\in\{-1,1\}^{NM}}{\operatorname{argmin}} \ \mathbf{b}^T \hat{\mathbf{\Psi}}^T \hat{\mathbf{\Psi}} \mathbf{b},$$

obtaining our estimate for the filter $\mathbf{H}$ as

$$\hat{\mathbf{H}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{C}_{\mathbf{x},m}^{-1/2} \hat{\mathbf{C}}_{\mathbf{xyx},m}^{1/2} \hat{\mathbf{V}}_{\mathbf{xyx},m} \mathrm{diag}(\hat{\mathbf{b}}_m^*) \hat{\mathbf{V}}_{\mathbf{xyx},m}^T \mathbf{C}_{\mathbf{x},m}^{-1/2}$$

Boolean quadratic program

- Our problem then reduces to solving the BQP

$$\hat{\mathbf{b}}^* = \operatorname*{argmin}_{\mathbf{b} \in \{-1,1\}^{NM}} \mathbf{b}^T \hat{\mathbf{\Psi}}^T \hat{\mathbf{\Psi}} \mathbf{b}$$

- Define $\hat{\mathbf{W}} = \hat{\mathbf{\Psi}}^T \hat{\mathbf{\Psi}}$ and $\mathbf{B} = \mathbf{b}\mathbf{b}^T$

$$\min_{\mathbf{B} \succeq \mathbf{0}} \operatorname{tr}(\hat{\mathbf{W}}\mathbf{B}) \quad \text{s. to} \;\; \operatorname{rank}(\mathbf{B}) = 1, \; B_{ii} = 1, \; i = 1, \ldots, NM$$

- Drop source of non-convexity to obtain the semi-definite relaxation

$$\mathbf{B}^* = \operatorname*{argmin}_{\mathbf{B} \succeq \mathbf{0}} \operatorname{tr}(\hat{\mathbf{W}}\mathbf{B}) \quad \text{s. to} \;\; B_{ii} = 1, \; i = 1, \ldots, NM$$

Performance guarantee

- For $l = 1, \ldots, L$, draw $\mathbf{z}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^*)$, round $\tilde{\mathbf{b}}_l = \mathrm{sign}(\mathbf{z}_l)$, to obtain

$$l^* = \operatorname*{argmin}_{l=1,\ldots,L} \tilde{\mathbf{b}}_l^T \hat{\mathbf{W}} \tilde{\mathbf{b}}_l$$

Performance guarantee

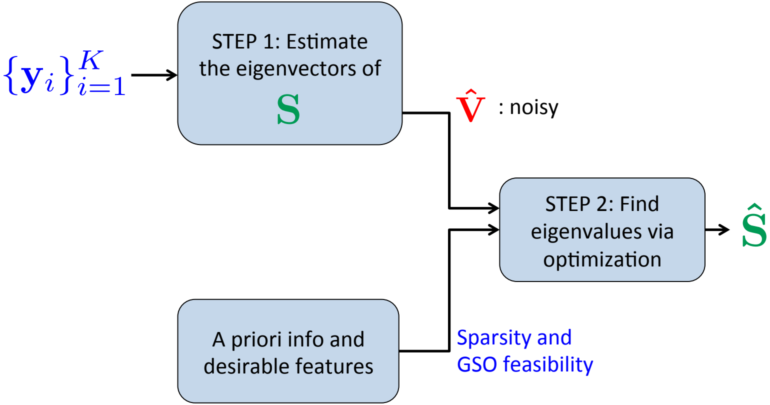
- For $l = 1, \ldots, L$, draw $\mathbf{z}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^*)$, round $\tilde{\mathbf{b}}_l = \mathrm{sign}(\mathbf{z}_l)$, to obtain

$$l^* = \operatorname*{argmin}_{l=1,\ldots,L} \tilde{\mathbf{b}}_l^T \hat{\mathbf{W}} \tilde{\mathbf{b}}_l$$

Let $\hat{\mathbf{b}}^*$ be the true solution of the BQP and let $\tilde{\mathbf{b}}_{l^*}$ be the output of our method. Then,

$$(\hat{\mathbf{b}}^*)^T \hat{\mathbf{W}} \hat{\mathbf{b}}^* \leq \mathbb{E}\left[(\tilde{\mathbf{b}}_{l^*})^T \hat{\mathbf{W}} \tilde{\mathbf{b}}_{l^*} \right] \leq \frac{2}{\pi} (\hat{\mathbf{b}}^*)^T \hat{\mathbf{W}} \hat{\mathbf{b}}^* + \gamma,$$

where $\gamma = \left(1 - \frac{2}{\pi}\right) \lambda_{\mathrm{max}} N M$.


$\{\mathbf{y}_i\}_{i=1}^K$
STEP 1: Estimate the eigenvectors of $\mathbf{S}$
$\hat{\mathbf{V}}$: noisy
STEP 2: Find eigenvalues via optimization
$\hat{\mathbf{S}}$
A priori info and desirable features
Sparsity and GSO feasibility

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

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

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

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

  - $\Rightarrow$ Can accommodate Laplacian matrices as well

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

$$\hat{\mathbf{S}} := \underset{\mathbf{S},\boldsymbol{\lambda}}{\operatorname{argmin}} \ f(\mathbf{S},\boldsymbol{\lambda}) \quad \text{s. to} \quad \mathbf{S} = \sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^T, \ \ \mathbf{S} \in \mathcal{S}$$

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

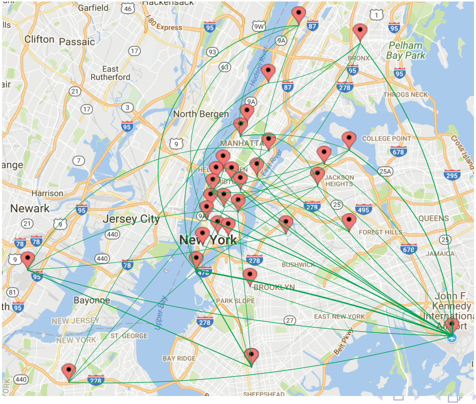
$$\mathcal{S} := \{\mathbf{S} \mid S_{ij} \geq 0, \ \ \mathbf{S} \in \mathcal{M}^N, \ S_{ii} = 0, \ \ \textstyle\sum_j S_{1j} = 1\}$$

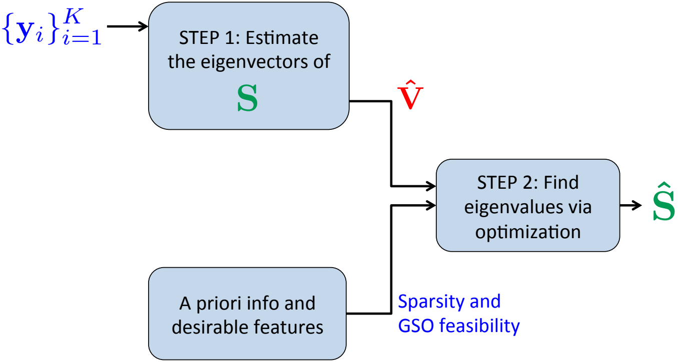
  $\Rightarrow$ Can accommodate Laplacian matrices as well

- Problem is convex if we select a convex objective $f(\mathbf{S},\boldsymbol{\lambda})$
  Ex: Sparsity ($f(\mathbf{S}) = \|\mathbf{S}\|_1$), min. energy ($f(\mathbf{S}) = \|\mathbf{S}\|_F$), mixing ($f(\boldsymbol{\lambda}) = -\lambda_2$)
- Robust recovery from imperfect or incomplete $\hat{\mathbf{V}}$ [Segarra16]

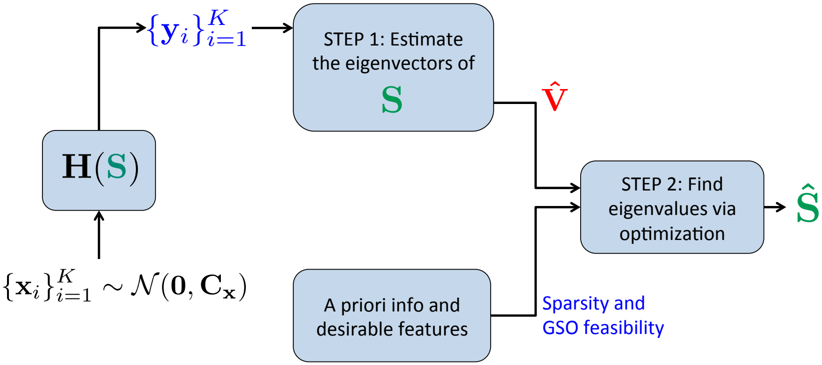
Unveiling urban mobility patterns

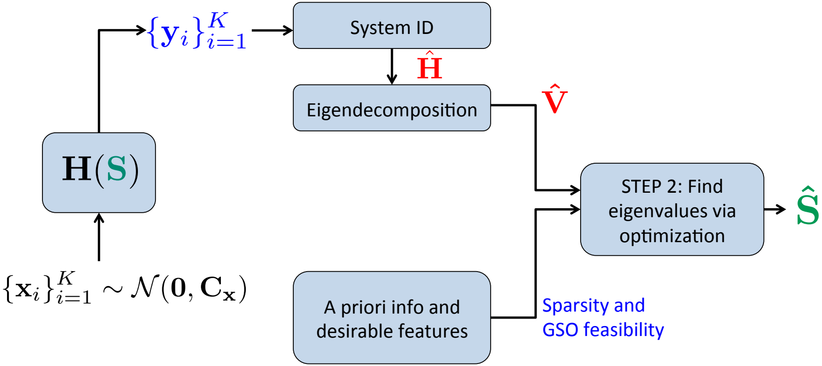
- Detect mobility patterns in New York City from Uber pickup data
- Times and locations ($N = 30$) from January 1st to June 29th 2015
- $M = 2$ graph processes: weekday ($m = 1$) and weekend ($m = 2$) pickups
- Pickups within 6-11am as input signal $\mathbf{x}$ and 3-8pm as output $\mathbf{y}$



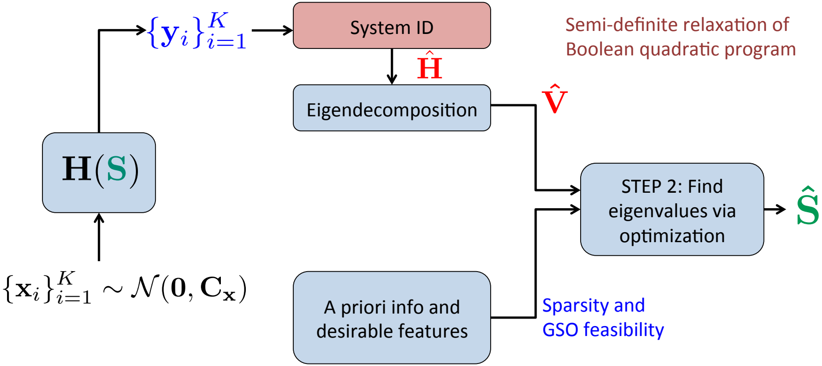


$\{\mathbf{y}_i\}_{i=1}^K$
STEP 1: Estimate the eigenvectors of $\mathbf{S}$
$\hat{\mathbf{V}}$
STEP 2: Find eigenvalues via optimization
$\hat{\mathbf{S}}$
A priori info and desirable features
Sparsity and GSO feasibility


$\{\mathbf{y}_i\}_{i=1}^K$
STEP 1: Estimate the eigenvectors of $\mathbf{S}$
$\hat{\mathbf{V}}$
$\mathbf{H}(\mathbf{S})$
STEP 2: Find eigenvalues via optimization
$\hat{\mathbf{S}}$
$\{\mathbf{x}_i\}_{i=1}^K \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{x})$
A priori info and desirable features
Sparsity and GSO feasibility



$\{\mathbf{y}_i\}_{i=1}^K$
System ID
$\hat{\mathbf{H}}$
Eigendecomposition
$\hat{\mathbf{V}}$
$\mathbf{H}(\mathbf{S})$
STEP 2: Find eigenvalues via optimization
$\hat{\mathbf{S}}$
$\{\mathbf{x}_i\}_{i=1}^K \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{x})$
A priori info and desirable features
Sparsity and GSO feasibility


$\{\mathbf{y}_i\}_{i=1}^K$
System ID
$\hat{\mathbf{H}}$
Semi-definite relaxation of
Boolean quadratic program
Eigendecomposition
$\hat{\mathbf{V}}$
$\mathbf{H}(\mathbf{S})$
STEP 2: Find
eigenvalues via
optimization
$\hat{\mathbf{S}}$
$\{\mathbf{x}_i\}_{i=1}^K \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{x})$
A priori info and
desirable features
Sparsity and
GSO feasibility

GlobalSIP'18 Symposium on GSP

Symposium on Graph Signal Processing

Topics of interest

- Graph-signal transforms and filters
- Distributed and non-linear graph SP
- Statistical graph SP
- Prediction and learning for graphs
- Network topology inference
- Recovery of sampled graph signals
- Control of network processes
- Signals in high-order and multiplex graphs
- Neural networks for graph data
- Topological data analysis
- Graph-based image and video processing
- Communications, sensor and power networks
- Neuroscience and other medical fields
- Web, economic and social networks

Paper submission due: **June 17, 2018**

Organizers:

Gonzalo Mateos (Univ. of Rochester)

Santiago Segarra (MIT)

Sundeep Chepuri (TU Delft)