Directed Network Topology Inference via Graph Filter Identification

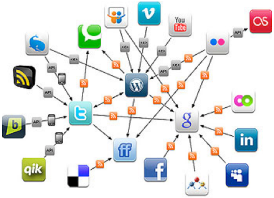
Rasoul Shafipour, **Santiago Segarra**, Antonio G. Marques
and Gonzalo Mateos

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

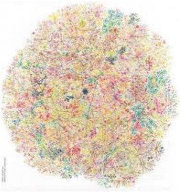
Data Science Workshop, June 5, 2018

Network Science analytics

Online social media



Internet

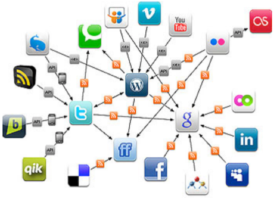


Clean energy and grid analytics

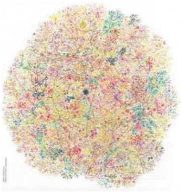
- Desiderata: Process, analyze and learn from network data [Kolaczyk09]

Network Science analytics

Online social media



Internet



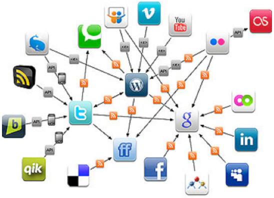
Clean energy and grid analytics

- 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

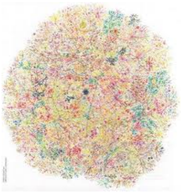
  $\Rightarrow$ Leverage G to process network data $\Rightarrow$ Graph Signal Processing

Network Science analytics

Online social media



Internet

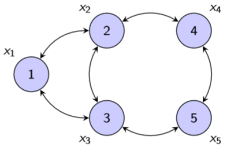


Clean energy and grid analytics

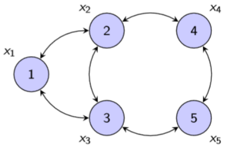
- Desiderata: Process, analyze and learn from network data [Kolaczyk09]
- Network as graph G: encode pairwise relationships
- Sometimes both G and data at the nodes are available
  - $\Rightarrow$ Leverage G to process network data $\Rightarrow$ Graph Signal Processing
- Sometimes we have access to network data but not to G itself
  - $\Rightarrow$ Leverage the relation between them to infer G from the data

Graph signal processing: Notation

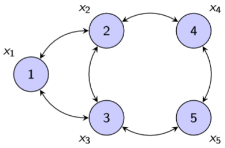
- Graph G with N nodes and adjacency $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i



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



- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ (local structure in G)
  - $\Rightarrow$ Ex: $\mathbf{A}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices

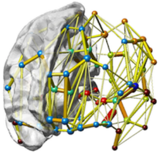


- Graph G with N nodes and adjacency $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i
- Associated with G is the graph-shift operator $\mathbf{S} = \mathbf{V\Lambda V}^{-1} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ (local structure in G)
  - $\Rightarrow$ Ex: $\mathbf{A}$ and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices
- Graph filters $\rightarrow$ Matrix polynomials: $\mathbf{H} = \sum_{l=0}^{N-1} h_l \mathbf{S}^l = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^{-1}$
- Graph Signal Processing $\rightarrow$ Exploit structure encoded in $\mathbf{S}$ to process $\mathbf{x}$
- Take the reverse path. 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

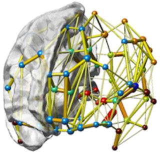


Topology inference: Motivation and context

- 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]

- **Contribution:** Inference for directed networks from diffused signals

Generating structure of a diffusion process

- Signal **y** is the response of a linear network diffusion process to an input **x**

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

$\Rightarrow$ Structure of **y** depends on structure of **x** and **S**

Generating structure of a diffusion process

- Signal **y** is the response of a linear network diffusion process to an input **x**

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

$\Rightarrow$ Structure of **y** depends on structure of **x** and **S**

- Cayley-Hamilton asserts we can write diffusion as

$$\mathbf{y} = \left(\sum_{l=0}^{N-1} h_l \mathbf{S}^l \right) \mathbf{x} := \mathbf{H}\mathbf{x}$$

$\Rightarrow$ **y** is the output of a GF **H** $\Rightarrow$ Use this and info on $(\mathbf{y}, \mathbf{x})$ to find **S**

$\Rightarrow$ Key property: **H** is diagonalized by the eigenvectors of **S**

- GF $\equiv$ linear maps which are analytic functions of the sparse matrix **S**

Ex.: $\mathbf{S}$, $\mathbf{S}^{-1}$, $(\mathbf{I} - \mathbf{S})^{-1}$, $(\mathbf{I} - \alpha\mathbf{S})^{-2}$, $(\mathbf{I} - \mathbf{S} - \mathbf{S}^2)^{-1}$, $(\mathbf{I} - \beta\mathbf{S})(\mathbf{I} - \alpha\mathbf{S})^{-1}$

Problem formulation

- We have access to M diffusion processes

$$\mathbf{y}_m = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \right) \mathbf{x}_m := \mathbf{H} \mathbf{x}_m$$

- For each process, we gather the realizations $\mathcal{Y}_m := \{\mathbf{y}_m^{(p)}\}_{p=1}^{P_m}$

  $\Rightarrow$ Every realization corresponds to an independent input $\mathbf{x}_m^{(p)}$

- We do not have access to L, h_l, or the inputs

  $\Rightarrow$ We do know that inputs are zero mean with covariance $\mathbf{C}_{\mathbf{x},m}$

Problem formulation

- We have access to M diffusion processes

$$\mathbf{y}_m = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \right) \mathbf{x}_m := \mathbf{H}\mathbf{x}_m$$

- For each process, we gather the realizations $\mathcal{Y}_m := \{\mathbf{y}_m^{(p)}\}_{p=1}^{P_m}$

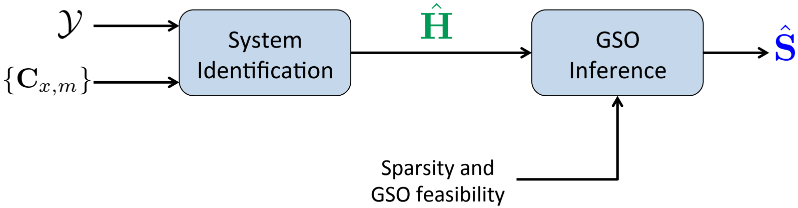
  $\Rightarrow$ Every realization corresponds to an independent input $\mathbf{x}_m^{(p)}$

- We do not have access to L, h_l, or the inputs

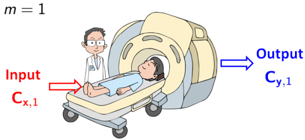
  $\Rightarrow$ We do know that inputs are zero mean with covariance $\mathbf{C}_{\mathbf{x},m}$

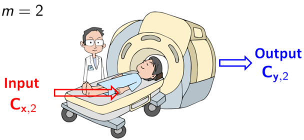
Problem: Given observations $\mathcal{Y} = \bigcup_{m=1}^{M} \mathcal{Y}_m$ and the input covariances $\mathbf{C}_{\mathbf{x},m}$, find sparsest (asymmetric) $\mathbf{S}$ that is consistent with the observations

Blueprint of our approach



A first pass at filter ID



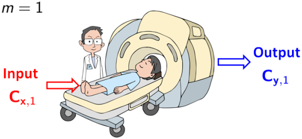


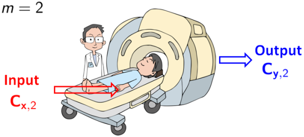
- The covariance matrix of the output process $\mathbf{y}_m$ is

$$\mathbf{C}_{\mathbf{y},m} = \mathbb{E}\left[\mathbf{H}\mathbf{x}_m\left(\mathbf{H}\mathbf{x}_m\right)^T\right] = \mathbf{H}\mathbb{E}\left[\mathbf{x}_m\mathbf{x}_m^T\right]\mathbf{H}^T = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T$$

- Each obs. pair $\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T$ gives rise to a set of potential solutions
  - $\Rightarrow$ Intersection smaller (unique) as $M \uparrow$, try to solve

A first pass at filter ID





- The covariance matrix of the output process $\mathbf{y}_m$ is

$$\mathbf{C}_{\mathbf{y},m} = \mathbb{E}\left[\mathbf{H}\mathbf{x}_m(\mathbf{H}\mathbf{x}_m)^T\right] = \mathbf{H}\mathbb{E}\left[\mathbf{x}_m\mathbf{x}_m^T\right]\mathbf{H}^T = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T$$

- Each obs. pair $\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T$ gives rise to a set of potential solutions
  - ⇒ Intersection smaller (unique) as $M \uparrow$, try to solve

$$\underset{\mathbf{H_L},\mathbf{H_R} \in \mathcal{M}_N}{\text{argmin}} \sum_{m=1}^{M} ||\mathbf{C}_{\mathbf{y},m} - \mathbf{H_L}\mathbf{C}_{\mathbf{x},m}\mathbf{H_R}^T||_F^2 \quad \text{s. to } \mathbf{H_L} = \mathbf{H_R}$$

  - ⇒ Variables of size N^2, smarter way to formulate the recovery?
  - ⇒ Parametrize the set of feasible solutions

Filter ID for directed networks

- For each m, we have a matrix equation of the form

$$\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T \tag{1}$$

Filter ID for directed networks

- For each m, we have a matrix equation of the form

$$\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T \tag{1}$$

If $\mathbf{C}_{\mathbf{x},m}$ and $\mathbf{C}_{\mathbf{y},m}$ are full rank, the set $\mathcal{H}_m$ containing all the (possibly asymmetric) matrices $\mathbf{H}$ that solve (1) for a particular m is given by

$$\mathcal{H}_m = \{\mathbf{H} \,|\mathbf{H} = \mathbf{C}_{\mathbf{y},m}^{1/2}\mathbf{U}\mathbf{C}_{\mathbf{x},m}^{-1/2} \text{ and } \mathbf{U}\mathbf{U}^T = \mathbf{I}\}.$$

Filter ID for directed networks

- For each m, we have a matrix equation of the form

$$\mathbf{C}_{\mathbf{y},m} = \mathbf{H}\mathbf{C}_{\mathbf{x},m}\mathbf{H}^T \tag{1}$$

If $\mathbf{C}_{\mathbf{x},m}$ and $\mathbf{C}_{\mathbf{y},m}$ are full rank, the set $\mathcal{H}_m$ containing all the (possibly asymmetric) matrices $\mathbf{H}$ that solve (1) for a particular m is given by

$$\mathcal{H}_m = \{\mathbf{H} \,|\mathbf{H} = \mathbf{C}_{\mathbf{y},m}^{1/2}\mathbf{U}\mathbf{C}_{\mathbf{x},m}^{-1/2} \text{ and } \mathbf{U}\mathbf{U}^T = \mathbf{I}\}.$$

- Optimization over unitary matrices
  - $\Rightarrow$ $N(N-1)/2$ degrees of freedom in lieu of N^2
  - $\Rightarrow$ Each m kills $N(N+1)/2$ degrees of freedom $\Rightarrow$ $M = 2$ may suffice
  - $\Rightarrow$ Non convex, but tailored algorithms are available
- Solving system id (non-symm. square roots) by leveraging structure

Manopt for non-symmetric graph-filter id

- Original formulation:

$$\underset{\mathbf{H_L},\mathbf{H_R}}{\text{argmin}} \sum_{m=1}^{M} ||\hat{\mathbf{C}}_{\mathbf{y},m} - \mathbf{H_L}\mathbf{C}_{\mathbf{x},m}\mathbf{H_R}^T||_F^2 \quad \text{s. to } \mathbf{H_L} = \mathbf{H_R} \quad (\mathcal{P}1)$$

- **Leveraging structure:** optimize over $\mathbf{U}_m \in \mathcal{U}_N$

$$\underset{\mathbf{H},\{\mathbf{U}_m\}_{m=1}^M}{\text{argmin}} \sum_{m=1}^{M} \|\mathbf{H} - \hat{\mathbf{C}}_{\mathbf{y},m}\mathbf{U}_m\mathbf{C}_{\mathbf{x},m}^{-1/2}\|_F^2 \quad (\mathcal{P}2)$$

$\Rightarrow$ Approach: projected gradient descent (manopt)

$\Rightarrow$ Sensitive to initialization

Manopt for non-symmetric graph-filter id

- Original formulation:

$$\underset{\mathbf{H_L},\mathbf{H_R}}{\text{argmin}} \sum_{m=1}^{M} ||\hat{\mathbf{C}}_{\mathbf{y},m} - \mathbf{H_L}\mathbf{C}_{\mathbf{x},m}\mathbf{H_R}^T||_F^2 \quad \text{s. to } \mathbf{H_L} = \mathbf{H_R} \quad (\mathcal{P}1)$$

- **Leveraging structure:** optimize over $\mathbf{U}_m \in \mathcal{U}_N$

$$\underset{\mathbf{H},\{\mathbf{U}_m\}_{m=1}^M}{\text{argmin}} \sum_{m=1}^{M} \|\mathbf{H} - \hat{\mathbf{C}}_{\mathbf{y},m}\mathbf{U}_m\mathbf{C}_{\mathbf{x},m}^{-1/2}\|_F^2 \quad (\mathcal{P}2)$$

$\Rightarrow$ Approach: projected gradient descent (manopt)

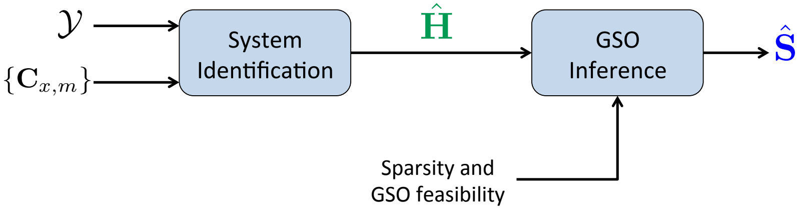
$\Rightarrow$ Sensitive to initialization

Enhanced algorithm: Smart initialization + Projected gradient

(s1) Use $(\mathcal{P}1)$ to find $\hat{\mathbf{H}}_L^{(0)}$ initial estimate of $\mathbf{H}$

(s2) Use the $\hat{\mathbf{H}}_L^{(0)}$ generated by (s1) as input for $(\mathcal{P}2)$ to obtain $\hat{\mathbf{H}}$

We completed our first step



GSO inference

- Finding $\mathbf{S}$ from $\mathbf{H} = h_0\mathbf{I} + h_1\mathbf{S} + h_2\mathbf{S}^2$ non-convex but...
  - $\Rightarrow$ $\mathbf{S}$ and $\mathbf{H}$ are simultaneously diagonalizable

GSO inference

- Finding $\mathbf{S}$ from $\mathbf{H} = h_0\mathbf{I} + h_1\mathbf{S} + h_2\mathbf{S}^2$ non-convex but...
  - $\Rightarrow$ $\mathbf{S}$ and $\mathbf{H}$ are simultaneously diagonalizable
- We can use extra knowledge/assumptions to choose one graph
  - $\Rightarrow$ Of all graphs, select one that sparsest one

$$\mathbf{S}^* := \underset{\mathbf{S}}{\text{argmin}} \ \|\mathbf{S}\|_0 \quad \text{s. to} \quad \mathbf{HS} = \mathbf{SH}, \ \ \mathbf{S} \in \mathcal{S}$$

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

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

GSO inference

- Finding $\mathbf{S}$ from $\mathbf{H} = h_0\mathbf{I} + h_1\mathbf{S} + h_2\mathbf{S}^2$ non-convex but...
  - $\Rightarrow$ $\mathbf{S}$ and $\mathbf{H}$ are simultaneously diagonalizable
- We can use extra knowledge/assumptions to choose one graph
  - $\Rightarrow$ Of all graphs, select one that sparsest one

$$\mathbf{S}^* := \operatorname*{argmin}_{\mathbf{S}} \ \|\mathbf{S}\|_0 \quad \text{s. to} \quad \mathbf{HS} = \mathbf{SH}, \ \ \mathbf{S} \in \mathcal{S}$$

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

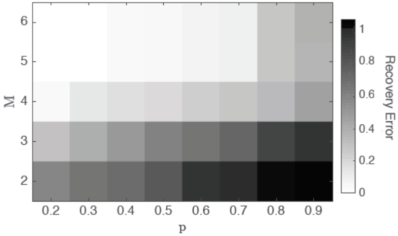
$$\mathcal{S} := \{\mathbf{S} \mid S_{ij} \geq 0, \ \ \mathbf{S} \in \mathcal{M}^N, \ S_{ii} = 0, \ \ \sum_j S_{1j} = 1\}$$

- In practice we solve the robust convex relaxation

$$\mathbf{S}^* := \operatorname*{argmin}_{\mathbf{S}} \ \|\mathbf{S}\|_1 \quad \text{s. to} \quad \|\hat{\mathbf{H}}\mathbf{S} - \mathbf{S}\hat{\mathbf{H}}\|_{\mathrm{F}} \leq \epsilon, \ \ \mathbf{S} \in \mathcal{S}$$

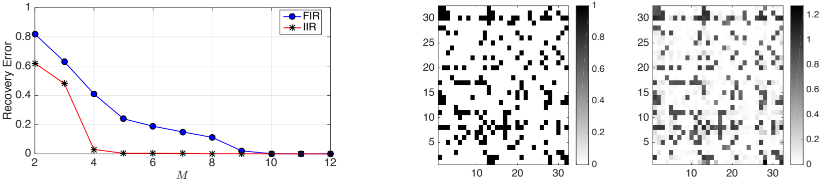
Simulations: Perfect covariances

- Consider Erdős-Rényi digraphs with 20 nodes and link probability p
  - $\Rightarrow$ Generate covariances as $\mathbf{C}_{\mathbf{x},m} = \mathbf{B}_m \mathbf{B}_m^T$, with $\mathbf{B}_m$ normal
  - $\Rightarrow$ Diffusing filters: FIR with $L = 3$



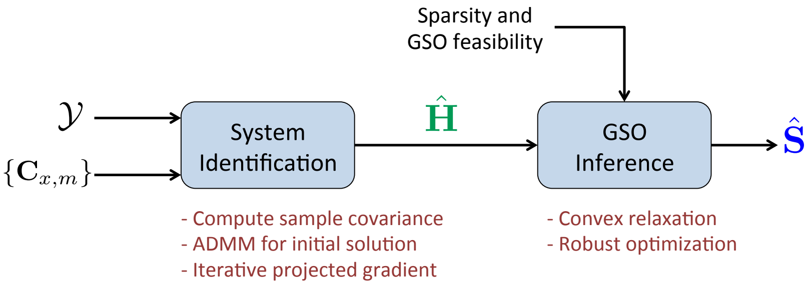
- Recovery for different M and p averaged for 10 graphs
  - $\Rightarrow$ Recovery increases with M
  - $\Rightarrow$ For high p fails oftentimes due to sparse recovery (Step 2)

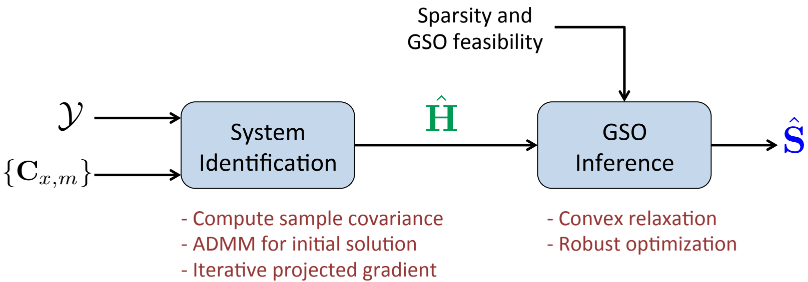
- Social net. $\mathcal{G}$ of $N = 32$ students in a class at the University of Ljubljana
  - $\Rightarrow$ Directed edges between students represent perceived friendships
  - $\Rightarrow$ Signals generated synthetically: FIR and IIR



- Recov. over 10 realizations as a function of M and $P = 10^6$ (left)
- $M = 5$, $P = 10^4$ filtered by FIR (right)

Conclusions





- Guarantees for system ID (manifold optimization)
- Guarantees for GSO inference
- Incorporation of priors on the filter and the GSO

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)