Robust graph-filter identification with graph denoising regularization

Samuel Rey
King Juan Carlos University - Madrid, Spain

Joint work with:
A. G. Marques

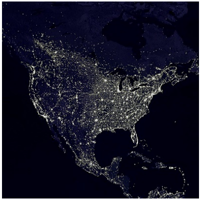
46th IEEE International Conference on acoustic, Speech and Signal Processing - Toronto, Canada - June 6-11, 2021

Network science and GSP

- Data is becoming more heterogeneous and intricate
  - ⇒ Often defined over irregulars domains and networks
  - ⇒ More complex structure demands more complex architectures
- Graph SP: models data structure as a graph [Shuman13], [Sandryhaila13]
  - ⇒ Leverages the graph topology to process the data
  - ⇒ Broadens classical SP to graph signals

Brain network

Social network

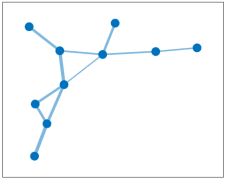


Energy network

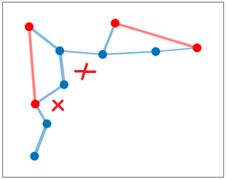
Imperfections in the graph topology

Universidad Rey Juan Carlos

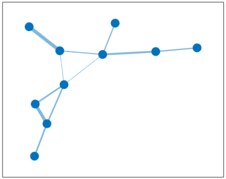
- In GSP it is usually assumed that the graph is perfectly known
- In practical cases the graph contains errors
  - ⇒ Perturbations and observational noise in explicit networks
  - ⇒ Imperfections derived of graphs learned from the data
- Ignored errors will hinder the performance of GSP models
  - ⇒ Filter identification is particularly sensitive to graph errors



Original graph



Added/Removed edges

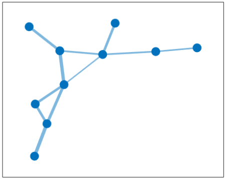


Noisy edges

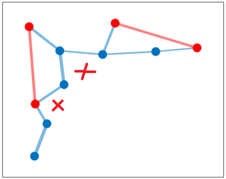
Imperfections in the graph topology

Universidad Rey Juan Carlos

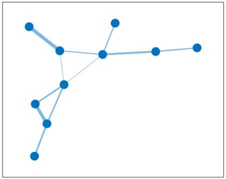
- In GSP it is usually assumed that the graph is perfectly known
- In practical cases the graph contains errors
  - ⇒ Perturbations and observational noise in explicit networks
  - ⇒ Imperfections derived of graphs learned from the data
- Ignored errors will hinder the performance of GSP models
  - ⇒ Filter identification is particularly sensitive to graph errors



Original graph



Added/Removed edges

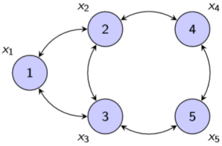


Noisy edges

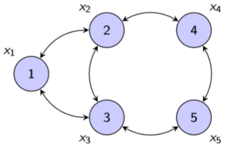
- **This talk**: approach the graph FI accounting for topology imperfections

Universidad Rey Juan Carlos

- Graph $\mathcal{G}$ 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

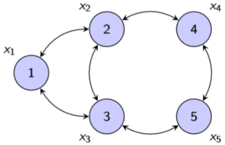


- Graph $\mathcal{G}$ with N nodes and adjacency $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i



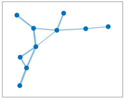
- Associated with $\mathcal{G}$ is the graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$ (e.g. $\mathbf{A}$, $\mathbf{L}$)
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ (local structure in $\mathcal{G}$)
  - $\Rightarrow$ Diagonalized as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$
- Graph filters are defined as $\mathbf{H} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k$
  - $\Rightarrow$ Diagonalized as $\mathbf{H} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T$

- Graph $\mathcal{G}$ 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

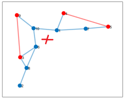


- Associated with $\mathcal{G}$ is the graph-shift operator $\mathbf{S} \in \mathbb{R}^{N\times N}$ (e.g. $\mathbf{A}$, $\mathbf{L}$)
  - $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ (local structure in $\mathcal{G}$)
  - $\Rightarrow$ Diagonalized as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$
- Graph filters are defined as $\mathbf{H} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k$
  - $\Rightarrow$ Diagonalized as $\mathbf{H} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T$
- Graph signal $\mathbf{x}$ is stationary on $\mathcal{G}$ if $\mathbf{C}_x$ is diagonalized by $\mathbf{V}$
  - $\Rightarrow \mathbf{C}_x$ and $\mathbf{S}$ commute $\mathbf{C}_x\mathbf{S} = \mathbf{S}\mathbf{C}_x$

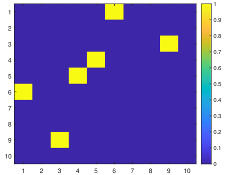
- In graph FI from input/output pairs we estimate $\mathbf{H} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow$ Leveraging that it is a polynomial of the GSO
- We observed the perturbed $\bar{\mathbf{S}} \in \mathbb{R}^{N \times N} \Rightarrow \bar{\mathbf{S}} \neq \mathbf{S}$
  - $\Rightarrow$ The true $\mathbf{S}$ is unknown
- What if we estimate the filter as $\mathbf{H} = \sum_{k=0}^{K} h_k \bar{\mathbf{S}}^k$?
  - $\Rightarrow$ Error between $\mathbf{S}^k$ and $\bar{\mathbf{S}}^k$ grows with k



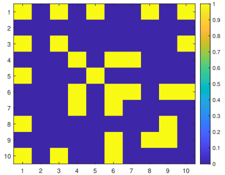
True $\mathcal{G}$



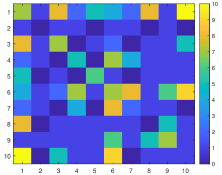
Observed $\mathcal{G}$



$|\mathbf{A} - \bar{\mathbf{A}}|$

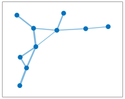


$|\mathbf{A}^2 - \bar{\mathbf{A}}^2|$

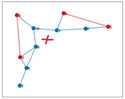


$|\mathbf{A}^4 - \bar{\mathbf{A}}^4|$

- In graph FI from input/output pairs we estimate $\mathbf{H} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow$ Leveraging that it is a polynomial of the GSO
- We observed the perturbed $\bar{\mathbf{S}} \in \mathbb{R}^{N \times N} \Rightarrow \bar{\mathbf{S}} \neq \mathbf{S}$
  - $\Rightarrow$ The true $\mathbf{S}$ is unknown
- What if we estimate the filter as $\mathbf{H} = \sum_{k=0}^{K} h_k \bar{\mathbf{S}}^k$?
  - $\Rightarrow$ Error between $\mathbf{S}^k$ and $\bar{\mathbf{S}}^k$ grows with k

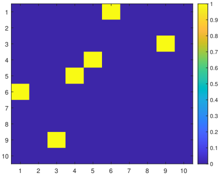


True $\mathcal{G}$

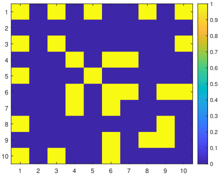


Observed $\mathcal{G}$

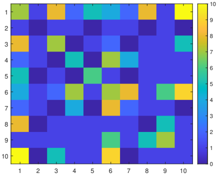
- **Problem:** learn $\mathbf{H}$ as polynomial of $\bar{\mathbf{S}}$ implies a **high estimation error**



$|\mathbf{A} - \bar{\mathbf{A}}|$



$|\mathbf{A}^2 - \bar{\mathbf{A}}^2|$



$|\mathbf{A}^4 - \bar{\mathbf{A}}^4|$

Context, and goal

- Limited number of works dealing with robust GSP
  - ⇒ Graphon based perturbation models [Miettinen19]
  - ⇒ Small perturbation analysis of the spectrum of $\mathbf{L}$ [Ceci20a]
  - ⇒ Combination of TLS with SEMs (TLS-SEM) [Ceci20b]
    - ⇒ TLS-SEM represented as a specific graph filter $\mathbf{H}_{SEM}$

Context, and goal

- Limited number of works dealing with robust GSP
  - $\Rightarrow$ Graphon based perturbation models [Miettinen19]
  - $\Rightarrow$ Small perturbation analysis of the spectrum of $\mathbf{L}$ [Ceci20a]
  - $\Rightarrow$ Combination of TLS with SEMs (TLS-SEM) [Ceci20b]
    - $\Rightarrow$ TLS-SEM represented as a specific graph filter $\mathbf{H}_{SEM}$
- **Challenges**
  - $\Rightarrow$ Translate errors in $\mathcal{G}$ to its spectrum is hard
  - $\Rightarrow$ Numerical issues in vertex-domain due to high-order polynomials

- Limited number of works dealing with robust GSP
  - ⇒ Graphon based perturbation models [Miettinen19]
  - ⇒ Small perturbation analysis of the spectrum of $\mathbf{L}$ [Ceci20a]
  - ⇒ Combination of TLS with SEMs (TLS-SEM) [Ceci20b]
    - ⇒ TLS-SEM represented as a specific graph filter $\mathbf{H}_{SEM}$
- **Challenges**
  - ⇒ Translate errors in $\mathcal{G}$ to its spectrum is hard
  - ⇒ Numerical issues in vertex-domain due to high-order polynomials
- **Our Goal**: identify a graph filter $\mathbf{H}$ from M inputs/outputs pairs
  - ⇒ Assuming imperfect knowledge of the GSO
  - ⇒ Obtaining a denoised version of the GSO
- **Key**: Introduce a graph denoising regularization term

Robust graph-filter identification

- Let $\mathbf{X}/\mathbf{Y}$ be the observed $N \times M$ input/output
- Assume that $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{E}$
- The goal is to estimate $\mathbf{H}$, a graph filter of $\mathbf{S}$, but only $\bar{\mathbf{S}}$ is known

- Let $\mathbf{X}/\mathbf{Y}$ be the observed $N \times M$ input/output
- Assume that $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{E}$
- The goal is to estimate $\mathbf{H}$, a graph filter of $\mathbf{S}$, but only $\bar{\mathbf{S}}$ is known
- The usual FI formulation transforms

$$\min_{\tilde{\mathbf{h}}} \|\mathbf{Y} - \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T\mathbf{X}\|_F^2$$

- Let $\mathbf{X}/\mathbf{Y}$ be the observed $N \times M$ input/output
- Assume that $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{E}$
- The goal is to estimate $\mathbf{H}$, a graph filter of $\mathbf{S}$, but only $\bar{\mathbf{S}}$ is known
- The usual FI formulation transforms

$$\min_{\tilde{\mathbf{h}},\mathbf{V}} \|\mathbf{Y} - \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T\mathbf{X}\|_F^2 \quad \text{s. t.} \;\; \mathbf{V}\mathbf{V}^T = \mathbf{I}$$

Robust graph-filter identification

- Let $\mathbf{X}/\mathbf{Y}$ be the observed $N \times M$ input/output
- Assume that $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{E}$
- The goal is to estimate $\mathbf{H}$, a graph filter of $\mathbf{S}$, but only $\bar{\mathbf{S}}$ is known
- The usual FI formulation transforms

$$\min_{\tilde{\mathbf{h}},\mathbf{V}} \|\mathbf{Y} - \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T\mathbf{X}\|_F^2 \quad \text{s. t.} \;\; \mathbf{V}\mathbf{V}^T = \mathbf{I}$$

Proposed robust filter identification (RFI) formulation

$$\min_{\mathbf{S}\in\mathcal{S},\mathbf{H}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2 + \lambda d(\mathbf{S}, \bar{\mathbf{S}}) + \beta\|\mathbf{S}\|_0 \quad \text{s. t.} \;\; \mathbf{S}\mathbf{H} = \mathbf{H}\mathbf{S}$$

- Perform joint estimation of $\mathbf{H}$ and $\mathbf{S}$ in vertex domain
- The constraint captures the fact that $\mathbf{H}$ is a polynomial of $\mathbf{S}$
- Second term is a distance measure between $\bar{\mathbf{S}}$ and $\mathbf{S}$

Graph perturbation model

- Several types of graph perturbations
- We assume that $\mathcal{G}$ is perturbed by adding/deleting edges

  $\Rightarrow$ Edges are pertubed independently with a given probability
- We set $d(\mathbf{S}, \bar{\mathbf{S}}) = \|\mathbf{S} - \bar{\mathbf{S}}\|_0$

Graph perturbation model

- Several types of graph perturbations
- We assume that $\mathcal{G}$ is perturbed by adding/deleting edges

  $\Rightarrow$ Edges are pertubed independently with a given probability
- We set $d(\mathbf{S},\bar{\mathbf{S}}) = \|\mathbf{S}-\bar{\mathbf{S}}\|_0$

Convex relaxations

- ℓ_1 as convex surrogate for ℓ_0
- Constraint $\mathbf{SH}=\mathbf{HS}$ rewritten as a regularizer

$$\min_{\mathbf{S}\in\mathcal{S},\mathbf{H}} \|\mathbf{Y}-\mathbf{HX}\|_F^2 + \lambda\|\mathbf{S}-\bar{\mathbf{S}}\|_1 + \beta\|\mathbf{S}\|_1 + \gamma\|\mathbf{SH}-\mathbf{HS}\|_F^2$$

- Still non-convex, but amenable to alternating optimization

Alternating optimization algorithm

Universidad Rey Juan Carlos

Step 1: Filter Identification

- Assume $\hat{\mathbf{S}}$ is known, estimate $\hat{\mathbf{H}}$

$$\hat{\mathbf{H}} = \arg\min_{\mathbf{H}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2 + \gamma\|\hat{\mathbf{S}}\mathbf{H} - \mathbf{H}\hat{\mathbf{S}}\|_F^2$$

$\Rightarrow$ LS problem with closed-form solution

Step 2: Graph Denoising

- Assume $\hat{\mathbf{H}}$ is known, estimate $\hat{\mathbf{S}}$

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}\in\mathcal{S}} \lambda\|\mathbf{S} - \bar{\mathbf{S}}\|_1 + \beta\|\mathbf{S}\|_1 + \gamma\|\mathbf{S}\hat{\mathbf{H}} - \hat{\mathbf{H}}\mathbf{S}\|_F^2$$

- Convergence is sensitive to the value of γ

  $\Rightarrow$ If γ is close to 0 the problems decouple

  $\Rightarrow$ If γ is too large convergence to non-robust solution

  $\Rightarrow$ Start with small γ and increase it progressively

- $\mathbf{X}$ and $\mathbf{Y}$ can exhibit properties depending on $\mathcal{G}$
  - $\Rightarrow$ Bandlimited, diffused, stationary
  - $\Rightarrow$ This information can be leveraged to enhance the performance

Leveraging additional structure

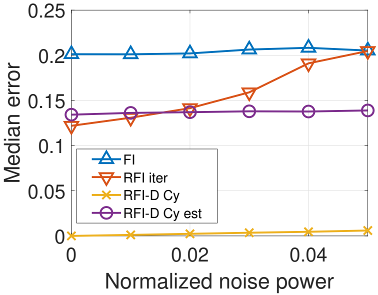
- $\mathbf{X}$ and $\mathbf{Y}$ can exhibit properties depending on $\mathcal{G}$
  - $\Rightarrow$ Bandlimited, diffused, stationary
  - $\Rightarrow$ This information can be leveraged to enhance the performance
- Focus on stationary $\mathbf{X}$ and/or $\mathbf{Y}$

$$\min_{\mathbf{S}\in\mathcal{S},\mathbf{H}} \|\mathbf{Y}-\mathbf{H}\mathbf{X}\|_F^2 + \lambda\|\mathbf{S}-\bar{\mathbf{S}}\|_1 + \beta\|\mathbf{S}\|_1 + \gamma\|\mathbf{S}\mathbf{H}-\mathbf{H}\mathbf{S}\|_F^2$$

$$\text{s.t.} \quad \|\mathbf{C}_Y\mathbf{S}-\mathbf{S}\mathbf{C}_Y\|_F \leq \epsilon_Y, \ \|\mathbf{C}_X\mathbf{S}-\mathbf{S}\mathbf{C}_X\|_F \leq \epsilon_X,$$

  - $\Rightarrow$ Constraints considered in the graph denoising step
- Filter identification step can be augmented with
  - $\Rightarrow$ $\mathbf{C}_Y\mathbf{H} = \mathbf{H}\mathbf{C}_Y$ and $\mathbf{C}_X\mathbf{H} = \mathbf{H}\mathbf{C}_X$

- Erdős Rényi (ER) graphs with $N = 20$ and $p = 0.25$
- Edge perturbation iid with probability $\delta = 0.1$
- $M = 10$ input/output observations

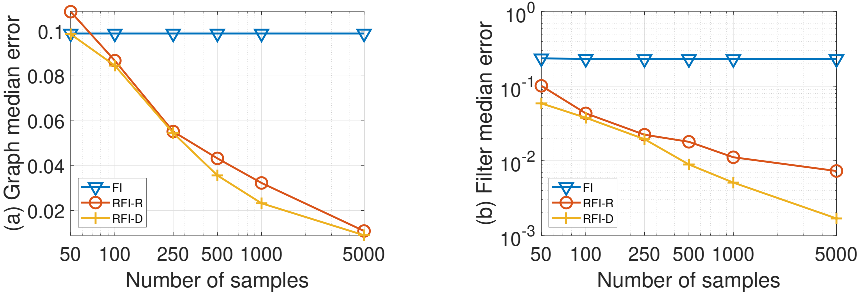


- "FI": ignores perturbations
- "RFI iter": iterative RFI algorithm
- "RFI-D": leverages stationarity
  - $\Rightarrow$ First $\hat{\mathbf{S}}$ with $\gamma = 0$
  - $\Rightarrow$ Then $\hat{\mathbf{H}}$ with large γ

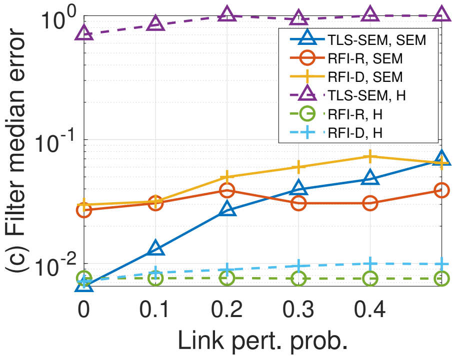
- Clear advantage of robust formulations
- "RFI iter" ignores stationarity and is more sensitive to noise

Numerical results - Test case 2

- Zachary's Karate graph used as original GSO
- "RFI-R": low complexity variant replacing $\mathbf{SH} = \mathbf{HS}$ with $\hat{\mathbf{C}}_Y \mathbf{H} = \mathbf{H}\hat{\mathbf{C}}_Y$



- As M grows, robust models recover a more accurate $\hat{\mathbf{S}}$
- Better estimates $\hat{\mathbf{S}}$ and increasing M improve estimate $\mathbf{H}$
  - $\Rightarrow$ "RFI-R" performs worse than "RFI-D" since it not using $\hat{\mathbf{S}}$



- ER graphs as in Test case 1
- $M = 200$ observations
- "SEM": $\mathbf{Y} = \mathbf{H}_{SEM}\mathbf{X}$
- "H": $\mathbf{Y} = \mathbf{H}\mathbf{X}$

- TLS-SEM outperform RFI algorithms under "SEM" model
  - $\Rightarrow$ Only for small perturbation probability
- RFI always outperform TLS-SEM under $\mathbf{H}$ model
- Good performance of RFI algorithms for both signal models
  - $\Rightarrow$ Benefits of more general assumptions

Conclusions

- Proposed general robust graph filter identification model
  - $\Rightarrow$ Joint graph denoising and filter identification performed
  - $\Rightarrow$ Approached in the vertex-domain
- Formulated as a non-convex optimization problem
  - $\Rightarrow$ Convex relaxation are leveraged
  - $\Rightarrow$ Proposed an alternated minimization algorithm
  - $\Rightarrow$ Stationarity assumptions are incorporated
- Numerical evaluation over synthetic and real-world graphs
  - $\Rightarrow$ Comparison with TLS-SEM

Conclusions

Universidad Rey Juan Carlos

- Proposed general robust graph filter identification model
  - ⇒ Joint graph denoising and filter identification performed
  - ⇒ Approached in the vertex-domain
- Formulated as a non-convex optimization problem
  - ⇒ Convex relaxation are leveraged
  - ⇒ Proposed an alternated minimization algorithm
  - ⇒ Stationarity assumptions are incorporated
- Numerical evaluation over synthetic and real-world graphs
  - ⇒ Comparison with TLS-SEM
- **Future research directions** include
  - ⇒ Establishing theoretical guarantees
  - ⇒ Extension to other GSP tasks