Learning Graphs and Simplicial Complexes from Data

Andrei Buciulea

Joint work with E. Isufi, G. Leus and A. G. Marques

April 17, ICASSP 2024

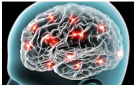
Universidad Rey Juan Carlos

ICASSP 2024 KOREA

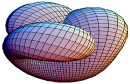
Motivating Examples: Networked Data

- Huge data sets are generated in networks (transportation, biological, brain, computer, social networks)
- Data structure carries critical information about the nature of the data
- Modelling the data structure using graphs

Interpolate a brain signal from local observations



Compress a signal in an irregular domain



Localize the source of a rumor

Smooth an observed network profile

Predict the evolution of a network process

Infer the topology where the signals reside

Graph Signal Processing (GSP)

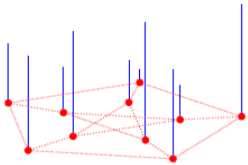
- Consider an undirected weighted graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$

  $\Rightarrow \mathcal{V}$, $\mathcal{E}$, $\mathcal{W} \rightarrow$ set of nodes, edges, weights

- Associated with $\mathcal{G} \rightarrow$ *Graph-Shift Operator* (GSO)

  $\Rightarrow$ **S** $\in \mathbb{R}^{N \times N}$, $S_{ij} \neq 0$ for $i = j$ and $(i, j) \in \mathcal{E}$

  $\Rightarrow$ Ex: Adjacency **A**, Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$...



- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on top of the graph

  $\Rightarrow x_i =$ value of graph signal (GS) at node i

- Sometimes the graph is not enough to explain the data structure

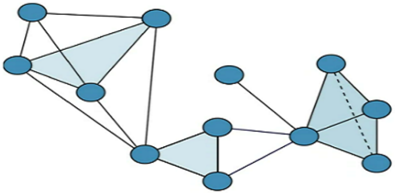
  $\Rightarrow$ Need for structures more complex than a graph

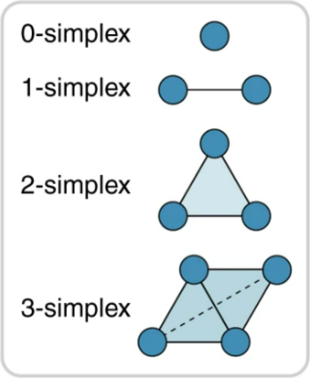
  $\Rightarrow$ Use Simplicial Complexes (SCs)

What is a Simplicial Complex?

- Mathematical structure that generalizes the concept of a graph to higher dimensions
- Building blocks

  $\Rightarrow$ Vertices, edges, triangles, tetrahedra, etc
- Graphs as 1-dimensional simplicial complexes
- Social structure, simplicial complex

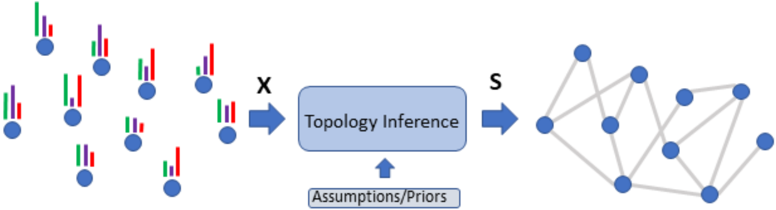




Graph Learning: Motivation and Context

Network topology inference from nodal observations

"Given a collection $\mathbf{X} := [\mathbf{x}_1, ..., \mathbf{x}_R]$ of graph signal observations supported on the unknown graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$ find an optimal $\mathbf{S}$"



- **This work**:
  - ⇒ Use data to learn both, the graph and the higher-order interactions
  - ⇒ Modelling data and graph using Autoregressive Graph Volterra Models

Related work (I): Graph Learning

- Goal: use $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_R] \in \mathbb{R}^{N \times R}$ to learn $\mathbf{S}$ with $\hat{\mathbf{\Sigma}} = \frac{1}{R}\mathbf{X}\mathbf{X}^{\mathsf{T}}$
  - **Correlation networks** $\Rightarrow$ $\mathbf{X}$ supported on $\mathcal{G}$

    $$\hat{\mathbf{S}} \approx \hat{\mathbf{\Sigma}} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^{\mathsf{T}}\right] \;(\hat{\mathbf{S}} \text{ is a thresholded version of } \hat{\mathbf{\Sigma}})$$

  - **Partial correlation networks** $\Rightarrow$ $\mathbf{X}$ i.i.d. $\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ GL

    $$\hat{\mathbf{S}} = \underset{\mathbf{S} \succeq 0, \mathbf{S} \in \mathcal{S}_{\Theta}}{\operatorname{argmin}} \; -\log(\det(\mathbf{S})) + \operatorname{tr}(\hat{\mathbf{\Sigma}}\mathbf{S}) + \rho \mathrm{h}(\mathbf{S})$$ [Fr.08]

  - **Graph-stationary diffusion processes** $\Rightarrow$ $\mathbf{X}$ st. w.r.t $\mathbf{S}$ GSR

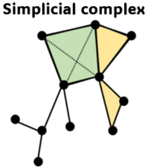
    $$\hat{\mathbf{S}} = \underset{\mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \; \|\mathbf{S}\|_0 \quad \text{s. to} \quad \hat{\mathbf{\Sigma}}\mathbf{S} = \mathbf{S}\hat{\mathbf{\Sigma}}$$ [Segarra17]

- Related to graphical Lasso:

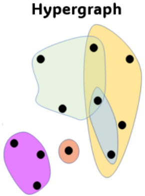
Sparse SEM: $\hat{\mathbf{S}} = \underset{\mathbf{S} \in \mathcal{S}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{S}\mathbf{X}\|_F^2 + \mathrm{g}(\mathbf{S})$ [Bazerque13]

Related work (II): Learning higher-order interactions

- Goal: use $\mathbf{X}$ and $\mathbf{S}$ to learn higher-order interactions
- **Vietoris–Rips complex approach** [Zomorodian10] RC
  - $\Rightarrow$ Form topological space from distances between points
  - $\Rightarrow$ Learn SCs from the data (i.e. $\hat{\boldsymbol{\Sigma}} = \mathbb{E}\left[\mathbf{X}\mathbf{X}^{\mathsf{T}}\right]$)



- **Learning SCs from data** [Barbarossa20] MTV-SC
  - $\Rightarrow$ Specific nature edge data $\mathbf{x}_1 = \mathbf{B}_1^\top s_0 + \mathbf{s}_H + \mathbf{B}_2\mathbf{s}_2 + \mathbf{w}$
  - $\Rightarrow$ Learn SCs ($\mathbf{B}_2$) from edge data ($\mathbf{X}_1$) and graph ($\mathbf{B}_1$)



- **Learning hypergraphs from data** [Tang23] HGSL
  - $\Rightarrow$ Graph structure is learned from node data
  - $\Rightarrow$ Hyperedges are obtained from the learned graph

Problem Formulation: Data Modelling

- Data Modelling: Autoregressive Graph Volterra Model of order 2

$$\mathbf{X} = \mathbf{H}_1\mathbf{X} + \mathbf{H}_2\mathbf{Y} + \mathbf{V} + \mathbf{E}, \text{ with } \mathbf{Y} = \mathbf{X} \odot \mathbf{X} \in \mathbb{R}^{N^2 \times R}$$

$\mathbf{H}_1 \in \mathbb{R}^{N\times N}$ pairwise interactions, $\mathbf{H}_2 \in \mathbb{R}^{N\times N^2}$ node-pair interactions
$\mathbf{V} \in \mathbb{R}^{N\times R}$ exogenous variable, $\mathbf{E} \in \mathbb{R}^{N\times R}$ zero-mean white noise

- $\mathbf{H}_1\mathbf{X}$ is a linear combination of the signals in the other nodes
- $\mathbf{H}_2\mathbf{Y}$ is a product of the signals in the other tuples of nodes

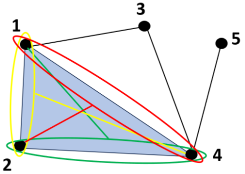
- Example of signal representation in terms of $\mathbf{H}_1$ and $\mathbf{H}_2$

$$x_2 = \mathbf{H}_1[2,1]x_1 + \mathbf{H}_1[2,4]x_4 + \mathbf{H}_2[2,(1,4)]x_1x_4 + \mathbf{H}_2[2,(4,1)]x_1x_4 + v_2 + e_2.$$

Part of x_2 is described by:

$\Rightarrow$ **node-to-node** interactions ($\mathbf{H}_1$)

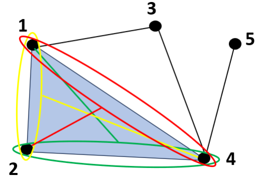
$\Rightarrow$ **node-to-pair** interactions ($\mathbf{H}_2$)



Problem Formulation: Graph & SC Modelling

- Recalling the signal modelling

$$\mathbf{X} = \mathbf{H}_1\mathbf{X} + \mathbf{H}_2\mathbf{Y} + \mathbf{V} + \mathbf{E}, \text{ with } \mathbf{Y} = \mathbf{X} \odot \mathbf{X}.$$



- **Graph Modelling:** pairwise interactions $\mathbf{H}_1$.

$\Rightarrow \mathcal{H}_1 = \{\mathbf{H}_1 \geq \mathbf{0}, \mathbf{B}_1 \circ \mathbf{H}_1 = \mathbf{0}, \ \mathbf{H}_1 = \mathbf{H}_1^\top\}$

$\Rightarrow$ Pos. weights, no self-loops ($\mathbf{B}_1 = \mathbf{I}$), symmetry.

H1	1	2	3	4	5
1					
2					
3					
4					
5					

- **SC Modelling:** node-to-pair interactions $\mathbf{H}_2$.

$\Rightarrow \mathcal{H}_2 = \{\mathbf{H}_2 \geq \mathbf{0}, \mathbf{B}_2 \circ \mathbf{H}_2 = \mathbf{0}\}$

$\Rightarrow$ Positive weights, no self-loops

H2	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)
1																									
2																									
3																									
4																									
5																									

Proposed Approach

Proposed formulation for learning graphs and simplicial complexes

$$(\hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2) = \underset{\mathbf{H}_1 \in \mathcal{H}_1, \mathbf{H}_2 \in \mathcal{H}_2}{\operatorname{argmin}} \quad \|\mathbf{X} - \mathbf{H}_1\mathbf{X} - \mathbf{H}_2\mathbf{Y} - \mathbf{V}\|_F^2 + \alpha\|\mathbf{H}_1\|_1 + \beta\|\mathbf{H}_2\|_1$$

$$\text{s. t.} \quad \mathbf{H}_2[k,(i,j)] \leq \theta \mathbb{1}(\mathbf{H}_1[k,i]\mathbf{H}_1[k,j]\mathbf{H}_1[i,j]);$$

$\Rightarrow \|\mathbf{X} - \mathbf{H}_1\mathbf{X} - \mathbf{H}_2\mathbf{Y} - \mathbf{V}\|_F^2 \rightarrow$ Fitting $\mathbf{X}$ to the considered model

$\Rightarrow \|\mathbf{H}_1\|_1 \rightarrow$ Controlling the number of node-to-node interactions with α

$\Rightarrow \|\mathbf{H}_2\|_1 \rightarrow$ Controlling the number of node-to-pair interactions with β

$\Rightarrow \mathbf{H}_2[k,(i,j)] \leq \theta \mathbb{1}(\mathbf{H}_1[k,i]\mathbf{H}_1[k,j]\mathbf{H}_1[i,j])$

$\rightarrow$ Filled triangle can exist if nodes i, j, and k are interconnected

- Non-convex formulation because of the trilinear constraint
  - $\Rightarrow$ Next $\rightarrow$ convex formulation to address non-convexities

Proposed Convex Approach

Convex formulation for learning graphs and simplicial complexes

$$(\hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2) = \underset{\mathbf{H}_1 \in \mathcal{H}_1, \mathbf{H}_2 \in \mathcal{H}_2}{\text{argmin}} \|\mathbf{X} - \mathbf{H}_1\mathbf{X} - \mathbf{H}_2\mathbf{Y} - \mathbf{V}\|_F^2 + \alpha\|\mathbf{H}_1\|_1 + \beta\|\mathbf{H}_2\|_1 + \gamma \sum_{i,j,k=1}^{N} \|\mathbf{Q}^{(i,j,k)} \circ [\mathbf{H}_1, \mathbf{H}_2]\|_F$$

- Entries of binary matrix $\mathbf{Q}^{(i,j,k)} \in \mathbb{R}^{N \times (N+N^2)}$ involving three nodes

$\Rightarrow$ Node-node interactions

$$\mathbf{Q}^{(i,j,k)}[i, j] = 1$$
$$\mathbf{Q}^{(i,j,k)}[i, k] = 1$$
$$\mathbf{Q}^{(i,j,k)}[j, k] = 1$$

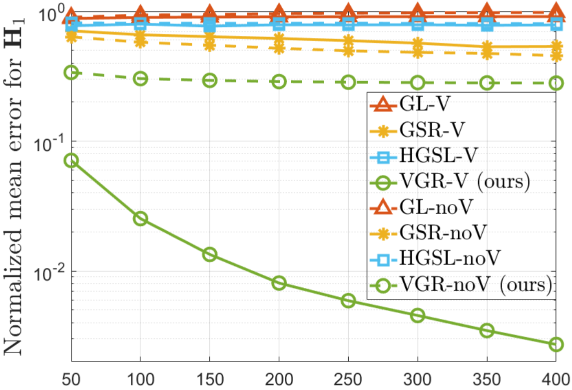
$\Rightarrow$ Node-pair interactions

$$\mathbf{Q}^{(i,j,k)}[i, Nj + k] = 1$$
$$\mathbf{Q}^{(i,j,k)}[j, Ni + k] = 1$$
$$\mathbf{Q}^{(i,j,k)}[k, Ni + j] = 1$$

- Group entries of $\mathbf{H}_1$ and $\mathbf{H}_2$ that participate in a triangle using $\mathbf{Q}^{(i,j,k)}$
- Controlling the number of filled triangles ($\mathbf{H}_2$) with β

Synthetic Data Results

- Estimation performance ($\mathrm{err}(\mathbf{H}_1)$) of different algorithms as R increases



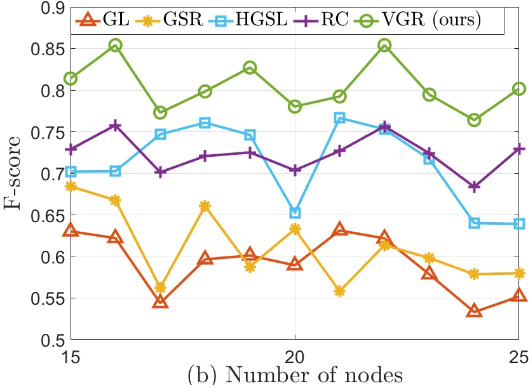
(a) Number of samples

- Normalized error when estimating filled triangles ($\mathrm{err}(\mathbf{H}_2)$)

Alg. \ R	50	100	200	300	400	500
MTV-SC	1.505	1.496	1.497	1.493	1.494	1.490
RC	0.790	0.767	0.761	0.753	0.748	0.751
VGR	0.559	0.428	0.294	0.214	0.165	0.133

Real Data Results

- Estimation performance (F-score) of different algorithms as N increases



- F-score and $\mathrm{err}(\mathbf{H}_2)$ when estimating filled triangles

	F-score			Error		
Alg. \ N	15	20	25	15	20	25
MTV-SC	0.093	0.058	0.056	7.418	7.536	7.530
RC	0.667	0.650	0.585	1.350	2.101	2.837
VGR	0.718	0.676	0.625	0.548	0.558	0.649

Conclusions

- New scheme that jointly learns graphs and simplicial complexes
- Key **assumptions**:
  - ⇒ Model data using autoregressive graph Volterra models
  - ⇒ Model network as graph ($\mathbf{H}_1$) and simplicial complexes ($\mathbf{H}_2$)
- Jointly learn from data node-pair interactions and filled triangles
- Challenge: non-convex approach due to filled triangle modelling
  - ⇒ Convex approach using group sparsity term
- Encouraging results in both synthetic and real data sets

- **THANKS!**
  - ⇒ Feel free to contact me for questions and code andrei.buciulea@urjc.es