Enhanced graph-learning schemes driven by similar distributions of motifs

S. Rey[*], T. M. Roddenberry[†], S. Segarra[†], A. G. Marques[*]

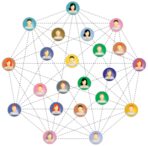
[*]Rey Juan Carlos University - [†]Rice University

Graph Signal Processing Workshop 2024 (GSP 2024) - Delft, The Netherlands - June 24-26, 2024

Introduction

- Graph-based methods are a popular option to process contemporary data
  - ⇒ Graph topology useful to model underlying irregular structure
  - ⇒ The graph topology is unknown in many relevant applications

Road network



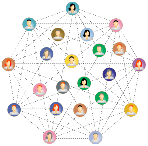
Social network

Home automation network

Introduction

- Graph-based methods are a popular option to process contemporary data
  - ⇒ Graph topology useful to model underlying irregular structure
  - ⇒ The graph topology is unknown in many relevant applications

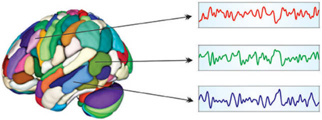
Road network

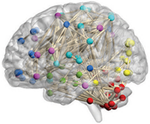


Social network

Home automation network

- Graph learning: estimates the graph topology from nodal observations
  - ⇒ Properties of the signals depend on the topology





Context and goal

- Graph learning is a well-studied problem with many different approaches
  - ⇒ Assuming different relation between the graph and the data
  - ⇒ Smoothness, Gaussianity, stationarity [Dong16][Friedman08][Segarra17]

Context and goal

- Graph learning is a well-studied problem with many different approaches
  - ⇒ Assuming different relation between the graph and the data
  - ⇒ Smoothness, Gaussianity, stationarity [Dong16][Friedman08][Segarra17]
- **Limitation**: focus placed on the model relating graphs and signals
  - ⇒ Prior graph structural information is rarely considered
- Works starting to pay attention to this problem
  - ⇒ Joint network inference [Navarro24]
  - ⇒ Spectral Laplacian constraints [Kumar19]
  - ⇒ Graphon-based method [Roddenberry21]

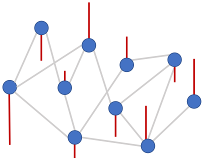
Context and goal

Universidad Rey Juan Carlos

- Graph learning is a well-studied problem with many different approaches
  - ⇒ Assuming different relation between the graph and the data
  - ⇒ Smoothness, Gaussianity, stationarity [Dong16][Friedman08][Segarra17]
- **Limitation**: focus placed on the model relating graphs and signals
  - ⇒ Prior graph structural information is rarely considered
- Works starting to pay attention to this problem
  - ⇒ Joint network inference [Navarro24]
  - ⇒ Spectral Laplacian constraints [Kumar19]
  - ⇒ Graphon-based method [Roddenberry21]
- **Contribution**: exploit motif density information when learning a graph
  - ⇒ Assuming observed signals are Gaussian

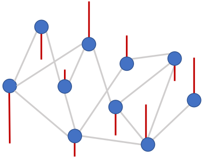
Notation and preliminaries

- Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 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

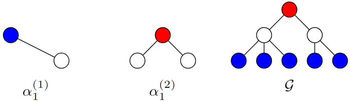


- Associated with $\mathcal{G}$ is the graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$ (e.g. $\mathbf{A}$, $\mathbf{L}$)
  - $\Rightarrow S_{ij} \neq 0$ if $i = j$ or $(i, j) \in \mathcal{E}$ (local structure in $\mathcal{G}$)
  - $\Rightarrow$ Diagonalized as $\mathbf{S} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$

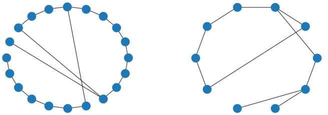
Notation and preliminaries

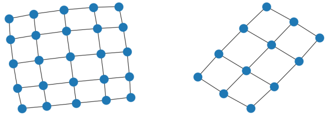


- Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes and adjacency $\mathbf{A}$
  - $\Rightarrow A_{ij}$ = Proximity between i and j
- Define a signal $\mathbf{x} \in \mathbb{R}^N$ on top of the graph
  - $\Rightarrow x_i$ = Signal value at node i
- Associated with $\mathcal{G}$ is the graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$ (e.g. $\mathbf{A}$, $\mathbf{L}$)
  - $\Rightarrow S_{ij} \neq 0$ if $i = j$ or $(i, j) \in \mathcal{E}$ (local structure in $\mathcal{G}$)
  - $\Rightarrow$ Diagonalized as $\mathbf{S} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$
- **Motifs** are subgraphs with a specific pattern (e.g., star graph, triangle)
  - $\Rightarrow$ Density of motifs measures the frequency with which a motif appears



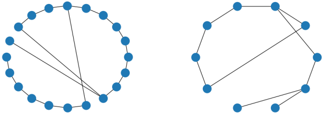
- Knowing the true motifs distribution of $\mathcal{G}$ is useful but unrealistic
- Assume access to a reference graph $\tilde{\mathcal{G}}$ with a similar motif distribution
  - $\Rightarrow$ Learn $\mathcal{G}$ with a similar distribution to that of $\tilde{\mathcal{G}}$
  - $\Rightarrow$ Similar assumption to joint graph learning methods

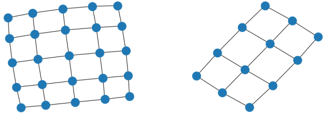




Motif density from a reference graph

- Knowing the true motifs distribution of $\mathcal{G}$ is useful but unrealistic
- Assume access to a reference graph $\tilde{\mathcal{G}}$ with a similar motif distribution
  - $\Rightarrow$ Learn $\mathcal{G}$ with a similar distribution to that of $\tilde{\mathcal{G}}$
  - $\Rightarrow$ Similar assumption to joint graph learning methods





Why density of motifs?

- The density of motifs can be computed **locally**
  - $\Rightarrow$ Allows comparing graphs of different sizes
  - $\Rightarrow$ It can be approximated from a subgraph of $\tilde{\mathcal{G}}$
- Requiring similar motifs is laxer assumption than similar support

Universidad Rey Juan Carlos

Problem statement

- Let $\mathbf{S}$ be the GSO of the unknown graph $\mathcal{G}$
- Given M graph signals $\mathbf{X} := [\mathbf{x}^{(1)}, ..., \mathbf{x}^{(M)}] \in \mathbb{R}^{N \times M}$
  $\Rightarrow$ Sampled from a Gaussian distribution $\mathbf{x}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^{-1})$
- Given a reference graph $\tilde{\mathcal{G}}$ with eigenvalues $\tilde{\boldsymbol{\lambda}}$
  $\Rightarrow$ With a similar density of motifs to that of the sought graph
- **Goal**: find the maximum likelihood estimator $\hat{\mathbf{S}}$

Problem statement

- Let $\mathbf{S}$ be the GSO of the unknown graph $\mathcal{G}$
- Given M graph signals $\mathbf{X} := [\mathbf{x}^{(1)}, ..., \mathbf{x}^{(M)}] \in \mathbb{R}^{N \times M}$
  - $\Rightarrow$ Sampled from a Gaussian distribution $\mathbf{x}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^{-1})$
- Given a reference graph $\tilde{\mathcal{G}}$ with eigenvalues $\tilde{\boldsymbol{\lambda}}$
  - $\Rightarrow$ With a similar density of motifs to that of the sought graph
- **Goal**: find the maximum likelihood estimator $\hat{\mathbf{S}}$

Challenges

- How to tractably compare densities of motifs of two graphs?
- How to incorporate motif information into the MLE?
  - $\Rightarrow$ Optimization problem will most likely be non-convex

From motif density to spectral test functions

- Computing the density of motifs of a graph is a combinatorial task
- Consider evaluating a test function $g(\lambda)$ over the spectral distribution

$$c_g(\boldsymbol{\lambda}) = \int g(\lambda)\, d\mu_{\boldsymbol{\lambda}}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} g(\lambda_i)$$

$\Rightarrow$ With empirical spectral density function $\mu_{\boldsymbol{\lambda}}$

From motif density to spectral test functions

Universidad Rey Juan Carlos

- Computing the density of motifs of a graph is a combinatorial task
- Consider evaluating a test function $g(\lambda)$ over the spectral distribution

$$c_g(\boldsymbol{\lambda}) = \int g(\lambda)\, d\mu_{\boldsymbol{\lambda}}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} g(\lambda_i)$$

$\Rightarrow$ With empirical spectral density function $\mu_{\boldsymbol{\lambda}}$

Theorem

Let $\mathcal{G}$ and $\tilde{\mathcal{G}}$ be two graphs such that the distance between their densities of motifs is upper-bounded by ϵ. Then

$$|c_g(\boldsymbol{\lambda}) - c_g(\tilde{\boldsymbol{\lambda}})| \leq \delta_\epsilon + \delta_r$$

- Similar densities of motifs render similar values when evaluating $c_g(\boldsymbol{\lambda})$

  $\Rightarrow$ Enables easy integration into an optimization problem

  $\Rightarrow$ Less expressive than motive densities in describing the structure

Motif graph learning formulation

- We have a tractable approach to harnessing information about motif density
  - $\Rightarrow$ Evaluating a test function over the spectrum of $\mathbf{S}$ and $\tilde{\mathbf{S}}$
- The constrained maximum likelihood estimator is given by

$$\min_{\mathbf{S},\mathbf{V},\boldsymbol{\lambda}} \; \mathrm{tr}(\hat{\mathbf{C}}\mathbf{S}) - \log\det(\mathrm{diag}(\boldsymbol{\lambda})) + \alpha\|\mathbf{S}\|_1 + \frac{\beta}{2}\|\mathbf{S} - \mathbf{V}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{V}^\top\|_F^2$$

$$\text{s.t}: \quad |c_g(\boldsymbol{\lambda}) - c_g(\tilde{\boldsymbol{\lambda}})| \leq \delta, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{V}^\top\mathbf{V} = \mathbf{I}.$$

  - $\Rightarrow$ Similarity constraint stems from similar motifs of $\mathcal{G}$ and $\tilde{\mathcal{G}}$
  - $\Rightarrow$ $c_g(\tilde{\boldsymbol{\lambda}})$ is a constant obtained from $\tilde{\mathbf{S}}$
  - $\Rightarrow$ Spectrum of $\mathbf{S}$ is close to the one satisfying similarity constraint

Motif graph learning formulation

- We have a tractable approach to harnessing information about motif density
  - $\Rightarrow$ Evaluating a test function over the spectrum of $\mathbf{S}$ and $\tilde{\mathbf{S}}$
- The constrained maximum likelihood estimator is given by

$$\min_{\mathbf{S},\mathbf{V},\boldsymbol{\lambda}} \operatorname{tr}(\hat{\mathbf{C}}\mathbf{S}) - \log\det(\operatorname{diag}(\boldsymbol{\lambda})) + \alpha\|\mathbf{S}\|_1 + \frac{\beta}{2}\|\mathbf{S} - \mathbf{V}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{V}^\top\|_F^2$$

$$\text{s.t}: \quad |c_g(\boldsymbol{\lambda}) - c_g(\tilde{\boldsymbol{\lambda}})| \leq \delta, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{V}^\top\mathbf{V} = \mathbf{I}.$$

  - $\Rightarrow$ Similarity constraint stems from similar motifs of $\mathcal{G}$ and $\tilde{\mathcal{G}}$
  - $\Rightarrow$ $c_g(\tilde{\boldsymbol{\lambda}})$ is a constant obtained from $\tilde{\mathbf{S}}$
  - $\Rightarrow$ Spectrum of $\mathbf{S}$ is close to the one satisfying similarity constraint
- Solving the resulting non-convex optimization problem is non-trivial
  - $\Rightarrow$ Similarity constraint will be non-convex for most test functions
  - $\Rightarrow$ Orthogonality of $\mathbf{V}$ and bilinear terms involving $\mathbf{V}$ and $\boldsymbol{\lambda}$ and

Relaxed similarity constraints

- Focus on **concave test functions** g
- Split constraint into concave/convex terms avoiding function composition

$$|c_g(\boldsymbol{\lambda}) - c_g(\tilde{\boldsymbol{\lambda}})| \leq \delta \quad \Longrightarrow \quad c_g(\boldsymbol{\lambda}) \leq c_g(\tilde{\boldsymbol{\lambda}}) + \delta, \quad c_g(\boldsymbol{\lambda}) \geq c_g(\tilde{\boldsymbol{\lambda}}) - \delta$$

- Reformulate the optimization problem as

$$\min_{\mathbf{S},\mathbf{V},\boldsymbol{\lambda}} \operatorname{tr}(\hat{\mathbf{C}}\mathbf{S}) - \log\det(\operatorname{diag}(\boldsymbol{\lambda})) + \alpha\|\mathbf{S}\|_1 + \frac{\beta}{2}\|\mathbf{S} - \mathbf{V}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{V}^\top\|_F^2 + \gamma c_g(\boldsymbol{\lambda})$$

$$\text{s.t}: \quad c_g(\boldsymbol{\lambda}) \geq c_g(\tilde{\boldsymbol{\lambda}}) - \delta, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{V}^\top\mathbf{V} = \mathbf{I}.$$

$\Rightarrow$ Convex feasible set yet non-convex objective

Relaxed similarity constraints

Universidad Rey Juan Carlos

- Focus on **concave test functions** g
- Split constraint into concave/convex terms avoiding function composition

$$|c_g(\boldsymbol{\lambda}) - c_g(\tilde{\boldsymbol{\lambda}})| \leq \delta \quad \Longrightarrow \quad c_g(\boldsymbol{\lambda}) \leq c_g(\tilde{\boldsymbol{\lambda}}) + \delta, \quad c_g(\boldsymbol{\lambda}) \geq c_g(\tilde{\boldsymbol{\lambda}}) - \delta$$

- Reformulate the optimization problem as

$$\min_{\mathbf{S},\mathbf{V},\boldsymbol{\lambda}} \operatorname{tr}(\hat{\mathbf{C}}\mathbf{S}) - \log\det(\operatorname{diag}(\boldsymbol{\lambda})) + \alpha\|\mathbf{S}\|_1 + \frac{\beta}{2}\|\mathbf{S} - \mathbf{V}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{V}^\top\|_F^2 + \gamma c_g(\boldsymbol{\lambda})$$

$$\text{s.t}: \quad c_g(\boldsymbol{\lambda}) \geq c_g(\tilde{\boldsymbol{\lambda}}) - \delta, \;\; \mathbf{S} \in \mathcal{S}, \;\; \mathbf{V}^\top\mathbf{V} = \mathbf{I}.$$

$\Rightarrow$ Convex feasible set yet non-convex objective

- Approximate concave term minimizing an upper bound as in MM

$$u(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(t-1)}) = \nabla c_g(\boldsymbol{\lambda}^{(t-1)})^\top \boldsymbol{\lambda}$$

- Iterative alternating minimization algorithm to avoid biconvexity

  $\Rightarrow$ Similar to the approach introduced in [Kumar2019]

Step 1: Solved approximately in closed-form

$$\mathbf{S}^{(t+1)} = \underset{\mathbf{S}\in\mathcal{S}}{\operatorname{argmin}} \operatorname{tr}(\hat{\mathbf{C}}\mathbf{S}) + \alpha\|\mathbf{S}\|_1 + \frac{\beta}{2}\|\mathbf{S} - \mathbf{V}^{(t)}\boldsymbol{\Lambda}^{(t)}\mathbf{V}^{(t)\top}\|_F^2$$

Step 2. Solution given by eigendecomposition of $\mathbf{S}^{(t+1)}$

$$\mathbf{V}^{(t+1)} = \underset{\mathbf{V}}{\operatorname{argmin}} \frac{\beta}{2}\|\mathbf{S}^{(t+1)} - \mathbf{V}\boldsymbol{\Lambda}^{(t)}\mathbf{V}^{\top}\|_F^2 \;\; \text{s.t}: \;\; \mathbf{V}^{\top}\mathbf{V} = \mathbf{I}$$

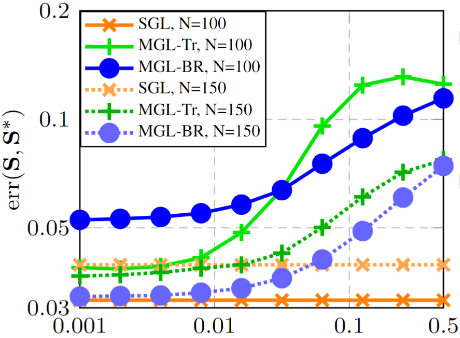
Step 3. Solved with over-the-shelf cvx solver

$$\boldsymbol{\lambda}^{(t+1)} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} -\sum_{j=1}^{N}\log(\boldsymbol{\lambda}_j) + \frac{\beta}{2}\|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|_2^2 + \gamma u(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(t)}) \;\; \text{s.t}: \;\; c_g(\boldsymbol{\lambda}) \geq c_g(\tilde{\boldsymbol{\lambda}}) - \delta$$

- The algorithm is guaranteed to converge to a stationary point
- Computational complexity of $\mathcal{O}(N^3)$

Universidad Rey Juan Carlos

- Assess the influence of discrepancies between $\mathcal{G}$ and $\tilde{\mathcal{G}}$
  - $\Rightarrow$ $\mathcal{G}$: SBM graph with 100/150 nodes and 5 disconnected communities
  - $\Rightarrow$ Setting tailored to GSL method

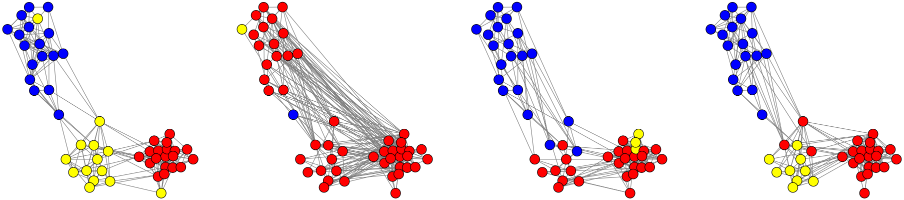


(c) Intercluster probability

- MGL robust to discrepancies in the intercluster connectivity
  - $\Rightarrow$ Information in $\tilde{\mathcal{G}}$ beyond conn. comp.
- Error of MGL improves with N
  - $\Rightarrow$ $N = 150$ outperforms SGL
  - $\Rightarrow$ Similarity constraint more informative for large graphs

Numerical evaluation - Real world data

- Dataset contains votes of the US Senate
  - ⇒ Political graph is inferred from senators votes
  - ⇒ $\mathcal{G}$ is the 115th congress and $\tilde{\mathcal{G}}$ is the 114th
  - ⇒ Labels represent ideological representation (Rep, Dem, Mix)
- First we learn the graph topology using 100 observations
  - ⇒ Then use spectral clustering to infer node labels



Ground truth   Unc.   SGL   MGL-BR

Conclusions and future work

Universidad Rey Juan Carlos

- We proposed a tractable approach to leverage graph structural information
  - ⇒ Assume access to a reference graph with a similar motif density
  - ⇒ Evaluate a test function over the graph spectrum
  - ⇒ Relation motif density/test function applicable to other tasks
- The density of motifs and the similarity constraint can be computed locally
  - ⇒ Allows to compare graphs of different sizes
  - ⇒ Is less restrictive than other assumptions
- The MLE with the similarity constraint was a non-convex problem
  - ⇒ Proposed convex iterative algorithm with guaranteed convergence

Conclusions and future work

- We proposed a tractable approach to leverage graph structural information
  - ⇒ Assume access to a reference graph with a similar motif density
  - ⇒ Evaluate a test function over the graph spectrum
  - ⇒ Relation motif density/test function applicable to other tasks
- The density of motifs and the similarity constraint can be computed locally
  - ⇒ Allows to compare graphs of different sizes
  - ⇒ Is less restrictive than other assumptions
- The MLE with the similarity constraint was a non-convex problem
  - ⇒ Proposed convex iterative algorithm with guaranteed convergence

Future research direction

- Exploit information about motifs as prior information in other GSP tasks
  - ⇒ Settings with hidden variables or imperfect topology knowledge
- Characterize and design good test functions

Questions?

- Paper [Rey23] available at QR

Questions at: **samuel.rey.escudero@urjc.es**

References

Dong16 X. Dong, D. Thanou, P. Frossard, P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations", IEEE Trans. Signal Process., 2016.

Friedman08 J. Friedman, T. Hastie, R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso", Biostatistics, 2008.

Segarra17 S. Segarra, A. G. Marques, G. Mateos, A. Ribeiro. "Network topology inference from spectral templates", IEEE Trans. Signal Inf. Process. Netw., 2017.

Navarro24 M. Navarro, S. Rey, A. Buciulea, A. G. Marques, S. Segarra, "Joint Network Topology Inference in the Presence of Hidden Nodes", IEEE Trans. Signal Process., 2024.

Kumar19 S. Kumar, J. Ying, J. V. de Miranda Cardoso, D. Palomar, "Structured graph learning via Laplacian spectral constraints", NEURIPS 2019.

Roddenberry21 T. M. Roddenberry, M. Navarro, S. Segarra, "Network Topology Inference with Graphon Spectral Penalties", ICASSP 2021.

Rey23 S. Rey, T. M. Roddenberry, S. Segarra, A. G. Marques, "Enhanced graph-learning schemes driven by similar distributions of motifs", IEEE Trans. Signal Process., 2023.