école normale supérieure paris–saclay

Offline detection of change-points for stationary graph signals

Alejandro de la Concha, Nicolas Vayatis, Argyris Kalogeratos

Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, France

CENTRE BORELLI

Background

1. Motivations

Previous work in change-point detection

- *Offline detection* Truong et al. [2020]
- *Online detection*: Aminikhanghahi et al. [2016] Tartakovsky [2020]
- Application examples: speech recognition, climate change evaluation, medical condition monitoring.

Main contribution

An offline change-point detector using tools coming from Graph Signal Processing to automatically infer the number of change-points, their location in the graph, and a sparse representation of the mean vector based on the input graph structure.

2. Problem statement

A stream of graph signals (SGS) Y is observed over the same graph G of p vertices. Let $Y = \{y_t\}_{t=0}^{T}$, where $\forall t \colon y_t \in \mathbb{R}^p$, and also $\mu_t = \mathbb{E}[y_t]$ is its unknown mean value. These expected values are the rows of matrix $\mu \in \mathbb{R}^{T\times p}$. Let an unknown ordered set $\tau = \{\tau_0 = 0, \ldots, \tau_d = T\} \subset \{0, \ldots, T\}$ indicating $d+1$ *change-points*.

Hypothesis: the expected values in each of the segments induced by τ remain constant.

Goal: Infer the set of change-points τ and the μ that will be an element of the space:

$$F_\tau = \{\mu \in \mathbb{R}^{T\times p} \mid \mu_{\tau_{l-1}+1} = \ldots = \mu_{\tau_l}, \forall \tau_l \in \tau \backslash \{0\}\}.$$

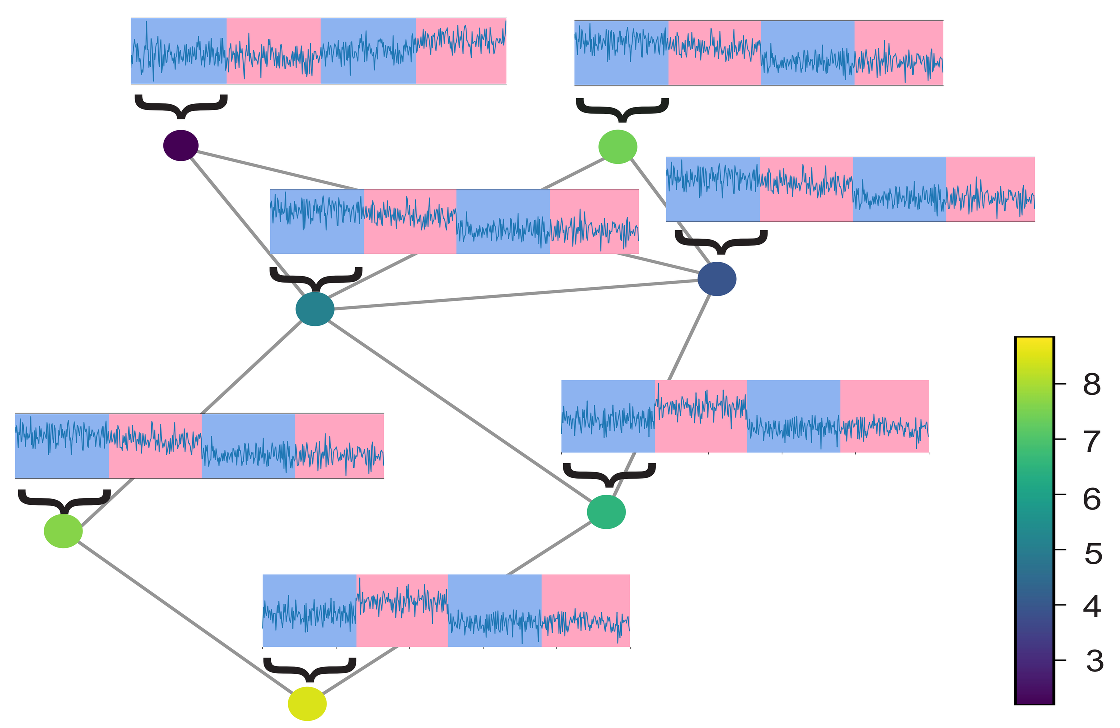

Figure: Example stream of graph signals (SGS) with five change-points in the mean.

Methods

3. Basic concepts

1. A **graph shift operator** (GSO) S associated with a graph $G = (V, E)$, is a $p \times p$ matrix whose entry $S_{i,j} \neq 0$ iff $i = j$ or $(i,j) \in E$, and it admits an eigenvector decomposition $S = U\Theta U^*$.
2. The **Graph Fourier Transform** (GFT) of a graph signal $y : V \to \mathbb{R}$ is defined by $\tilde{y} = U^* y$.
3. A zero-mean graph signal $y : V \to R$ with covariance matrix Σ_y is **stationary with respect to the vertex domain encoded by** S, iff Σ_y and S are simultaneously diagonalizable: $\Sigma_y = U \operatorname{diag}(P_y) U^*$. The vector $P_y \in \mathbb{R}^p$ is **graph power spectral density** (PSD).

4. Our method

Algorithm: Variable Selection-based GS change-point detector (VSGS)

Input : $Y \in \mathbb{R}^{T\times p}$ representing the stream of the graph signals
$d_{\max}$: Maximum number of change-points
w: length of the warming period
U: eigenvectors of the GSO
Λ: penalization constants associated the sparsity of the GFT of μ

Output : $\hat{d}, \hat{\tau}(\hat{d})$: number of change-points, set of change-points
$\hat{\tilde{\mu}}_{\hat{\tau}}(\hat{d}) \in \mathbb{R}^{\hat{d}\times p}$ with rows being the GFT of the mean in each segment

Estimate the GFT of the dataset $\tilde{Y} = YU$

Compute an estimation of P_y using w observations Perraudin [2017].

for $\lambda \in \Lambda$ **do**

  Solve the Lasso problem: $\tilde{\mu}_{\text{Lasso}} := \underset{\tilde{\mu} \in R^{T\times p}}{\arg\min} \sum_{t=1}^{T} \sum_{i=1}^{p} \frac{\left\|\tilde{y}_t^{(i)} - \tilde{\mu}^i\right\|^2}{T(P_y^{(i)})} + \lambda \|\tilde{\mu}\|_1$

  Define $D_{m_\lambda} := \|\tilde{\mu}_{\text{Lasso}}\|_0$ and $S_{(D_m,\tau)} := \{\tilde{\mu} \in F_\tau \mid \tilde{\mu}_{\tau_l} \in S_{D_m}, \forall l \in \{1, \ldots, d\}\}$, where S_{D_m} the space generated by m specific elements of the standard basis of $\mathbb{R}^p$.

  for $d \in \{1, \ldots, d_{\max}\}$ **do**

    Solve the change-point detection problem via dynamic programming

$$\hat{\tau}(d, D_{m_\lambda}), \hat{\tilde{\mu}}_{\hat{\tau}}(d, D_{m_\lambda}) := \underset{(\tilde{\mu},\tau) \in S_{(D_{m_\lambda}, \tau(d))}}{\arg\min} C^{\text{LSE}}(\tilde{\mu}(d, D_{m_\lambda}), \tau(d, D_{m_\lambda}))$$

$$:= \underset{(\tilde{\mu},\tau) \in S_{(D_{m_\lambda}, \tau(d))}}{\arg\min} \sum_{l=1}^{d} \sum_{t=\tau_{l-1}+1}^{\tau_l} \sum_{i=1}^{p} \frac{(\tilde{y}_t^{(i)} - \tilde{\mu}_{\tau_l}^{(i)})^2}{T(P_y^{(i)})}$$

  end

end

Find K_1, K_2, K_3 using the slope heuristic (Arlot et al. [2019]) and solve:

$$(\hat{\lambda}, \hat{d}) := \underset{\substack{\lambda \in \Lambda, \\ d = \{1, \ldots, d_{\max}\}}}{\operatorname{argmin}} C^{\text{LSE}}(\hat{\tau}(d, D_{m_\lambda}), \hat{\tilde{\mu}}_{\hat{\tau}}^{\text{LSE}}(d, D_{m_\lambda})) + K_1 \frac{D_{m_\lambda}}{T} + \frac{d}{T}\left(K_2 + K_3 \log \frac{T}{d}\right)$$

Keeping the segmentation $\hat{\tau}(\hat{d}, \hat{D}_{m_{\hat{\lambda}}})$ and $\hat{\lambda}$ fixed, recover $\hat{\tilde{\mu}}_{\hat{\tau}}(\hat{d})$ via

$$\hat{\tilde{\mu}}_{\hat{\tau}_l}^{(i)}(\hat{d}) = \operatorname{sign}\left(\bar{\tilde{y}}_{\hat{\tau}_l}^{(i)}\right) * \max\left(\left|\bar{\tilde{y}}_{\hat{\tau}_l}^{(i)}\right| - \frac{\hat{\lambda} P_y^{(i)}}{2}, 0\right),$$

where $\bar{\tilde{y}}_{\tau_l}^{(i)} = \frac{1}{I_l} \sum_{t=\tau_{l-1}+1}^{\tau_l} \tilde{y}_t^{(i)}$

Return $\hat{\tau}(\hat{d}, \hat{D}_{m_{\hat{\lambda}}})$ and $\hat{\tilde{\mu}}_{\hat{\tau}}(\hat{d})$

Results

5. Theoretical results

Let $\hat{\tau}$ and $\hat{\tilde{\mu}}_{\hat{\tau}}$ be solutions to the optimization problem:

$$\underset{\substack{d \in 1,..,T \\ \tau \in \mathcal{T}^d, \\ \tilde{\mu} \in S_{(D_m,\tau)}}}{\operatorname{argmin}} C_T^{\text{LSE}}(\tau, \hat{\mu}, \hat{Y}) + pen(m,\tau) := \underset{\substack{d \in 1,..,T \\ \tau \in \mathcal{T}^d, \\ \tilde{\mu} \in S_{(D_m,\tau)}}}{\operatorname{argmin}} \sum_{l=1}^{d} \sum_{t=\tau_{l-1}+1}^{\tau_l} \sum_{i=1}^{p} \frac{(\tilde{y}_t^{(i)} - \tilde{\mu}_{\tau_l}^{(i)})^2}{T(P_y^{(i)})} + K_1 \frac{D_m}{T} + \frac{d}{T}\left(K_2 + K_3 \log \frac{T}{d}\right)$$

where $\mathcal{T}^d$ is the set of all possible segmentations of length d and D_m is the number non-zero entries of the GFT. Then, there exist constants K_1, K_2, K_3 defining the penalty term for all $(m,\tau) \in M$, where $M \subset \{1, \ldots, p\} \times \mathcal{T}$, and there exists a positive constant $C(K)$, and $K > 1$ a given constant, such that:

$$\mathbb{E}\left[\frac{\left\|\hat{\tilde{\mu}}_{\hat{\tau}} - \tilde{\mu}^*\right\|_F^2}{T}\right] \leq C(K)\left[\inf_{(m,\tau) \in M} \inf_{\tilde{\mu} \in S(D_m,\tau)} \frac{\|\tilde{\mu} - \tilde{\mu}^*\|_F^2}{T} + pen(m,\tau) + \left(1 + \left(\frac{1}{(e^\gamma - 1)(e-1)}\right)\right)\epsilon^2\right]$$

where $\mathcal{T}$ is the set of all possible segmentations of the SGS, $\gamma = \frac{1}{K}(\sqrt{\log p + L} - \sqrt{\log p + \log 2})$.

6. Experiments

We tested our method using synthetic graph signals on two random and one real graph. The signals are generated differently (wrt noise distribution and the nature of the change) in each of the three scenarios. More details and results are provided in main document.

Scenario	# Nodes, Segments	Noise distribution	Graph	Hausdorff (↓)	Rand (↑)	Recall (↑)	Precision (↑)	F1 (↑)
I	100,5	Uniform with unit variance	Erdos–Rényi	1.73 (70.88)	0.99 (0.02)	1.00 (0.05)	0.88 (0.13)	0.93 (0.08)
II	100,3	Standard Normal	Barabasi-Albert	1.57 (07.28)	0.99 (0.01)	1.00 (0.00)	0.98 (0.07)	0.99 (0.04)
I	500,3	Uniform with unit variance	Erdos–Rényi	6.29 (17.13)	0.98 (0.04)	0.97 (0.11)	1.00 (0.05)	0.98 (0.09)
II	500,5	Standard Normal	Barabasi-Albert	12.48 (16.54)	0.96 (0.06)	0.89 (0.16)	1.00 (0.00)	0.93 (0.09)
I	1000,5	Uniform with unit variance	Erdos–Rényi	7.36 (23.46)	0.98 (0.07)	0.96 (0.12)	1.00 (0.00)	0.98 (0.09)
II	1000,3	Standard Normal	Barabasi-Albert	33.81 (17.05)	0.89 (0.05)	0.71 (0.12)	1.00 (0.00)	0.83 (0.08)
III - 5 rand. regions 10 rand. nodes	2642,3	Student-t	Minnesota Road Network	120.81 (63.38)	0.82 (0.09)	0.60 (0.20)	1.00 (0.00)	0.73 (0.13)
III - 10 rand. regions 20 rand. nodes	2642,3	Student-t	Minnesota Road Network	8.85 (32.28)	0.99 (0.04)	0.97 (0.12)	1.00 (0.00)	0.98 (0.08)
III - 20 rand. regions 40 rand. nodes	2642,3	Student-t	Minnesota Road Network	0.72 (00.45)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

7. Conclusions

- Our method exploits the interplay between the graph structure and multivariate time series thanks to the concept of *graph stationarity*.
- Our model-selection framework allows the automatic inference of i) the relevant parameters for the recovery of the number of change-points and ii) a sparse representation of the graph signals.
- By combining techniques coming from Graph Signal Processing and model-selection, we defined a tractable formulation of the problem and obtain theoretical guarantees for the performance of our method.

References

S. Arlot, A. Celisse, and Z. Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 2019.

N. Perraudin. Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, 2017.

C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 2020.