Learning Time-Varying Graphs from Data

Alberto Natali

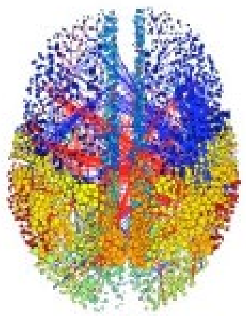
Graphs&Data Seminar // 07 December 2023

In the previous talk …

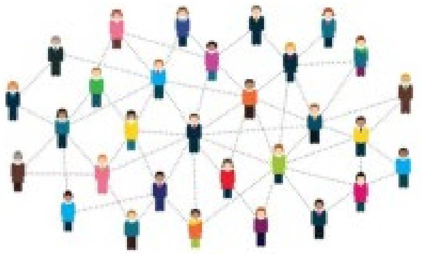
- What is the relationship between classical signal processing and graph signal processing?
- What are graph signals?
- Which architectures can be used to process graph signals?
- Applications?

In this talk...

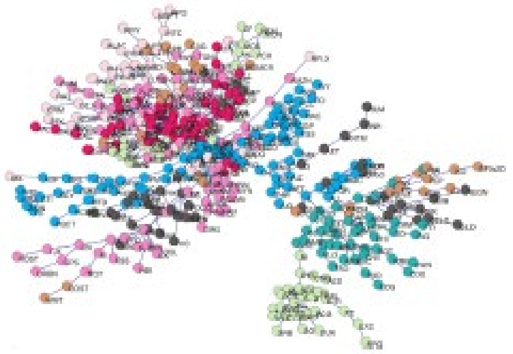
In many scenarios the structure of the network is *unknown*



Brain connections



Social interactions



Stock correlations

However, we can measure data/signals on these networks:

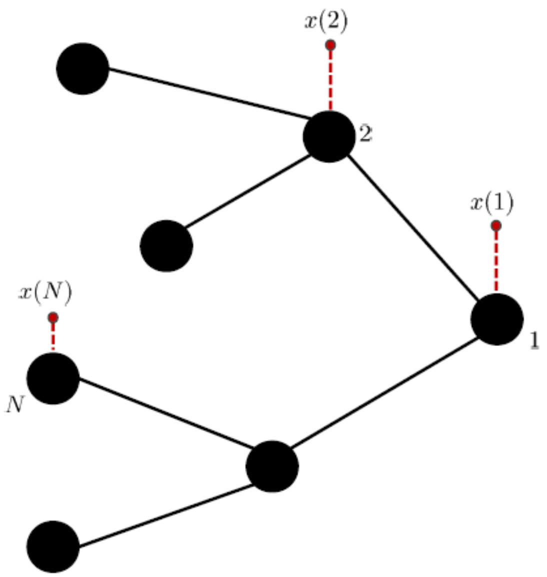
- fMRI, ultrasound, ECoG signals, ...
- User profiling, social networks data, ...
- Daily stock volumes, earnings reports, ...

In this talk…

Learn the topology of the network from the available data

GSP 101: notation

We consider data element $\boldsymbol{x}$ to reside over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{S})$



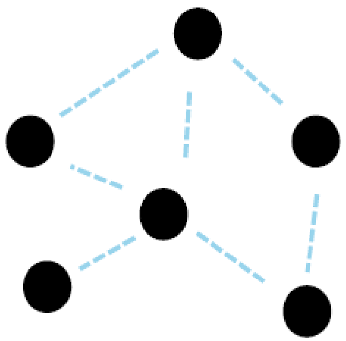
- $\mathcal{V} = \{1, \ldots, N\}$ node set
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ edge set
- $\mathbf{S} \in \mathbb{R}^{N \times N}$ is the shift operator

- $\boldsymbol{x}$ is the **graph signal**
- $x(i)$ is the value at node $i \in \mathcal{V}$
- $\mathbf{S}$ is typically the adjacency matrix or Laplacian $\longleftarrow$ graph structure

Static Topology Inference (graph unknown but fixed over time)

Consider matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{S})$



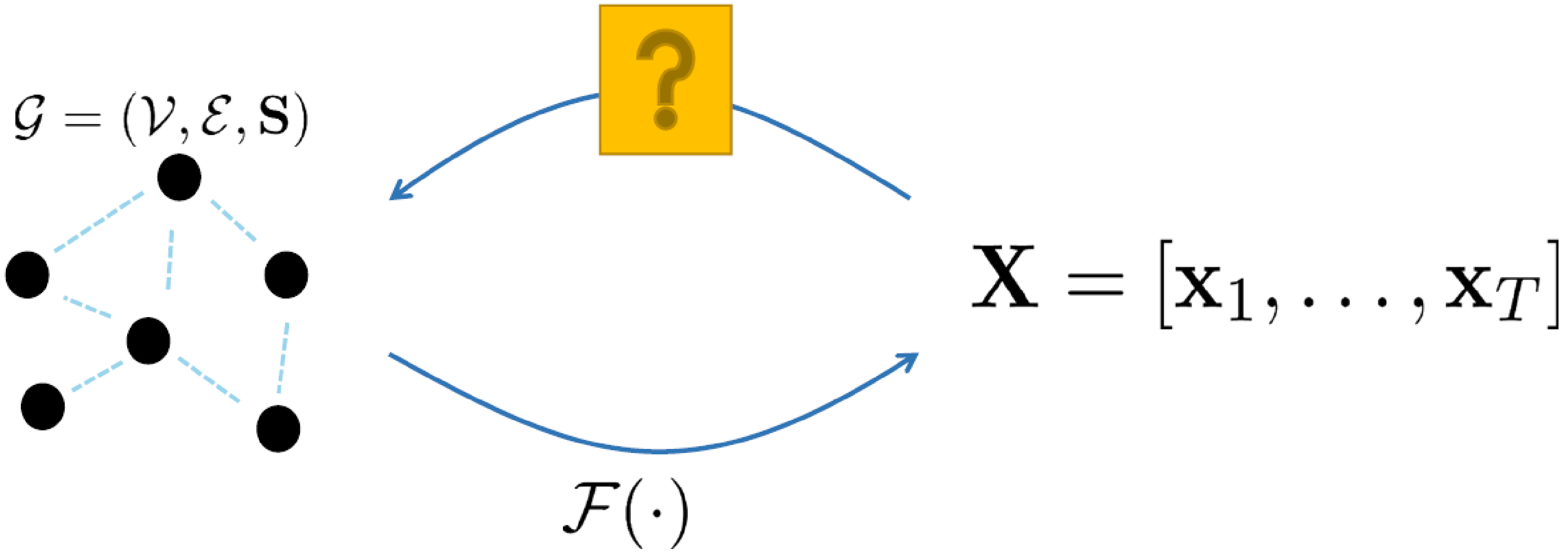
$$\mathcal{F}(\cdot)$$

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$$

TU Delft

Static Topology Inference (graph unknown but fixed over time)

Consider matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$



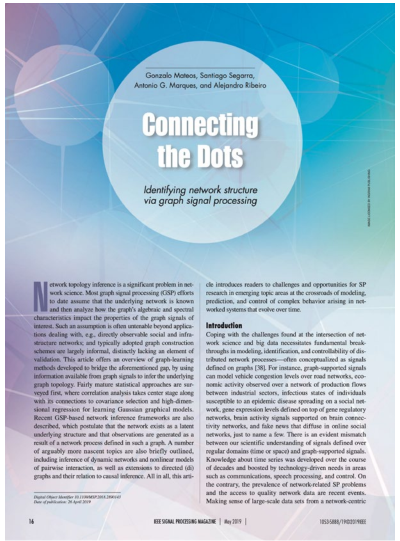
How to infer graph $\mathcal{G}$ from $\mathbf{X}$? (How to infer $\mathbf{S}$?)

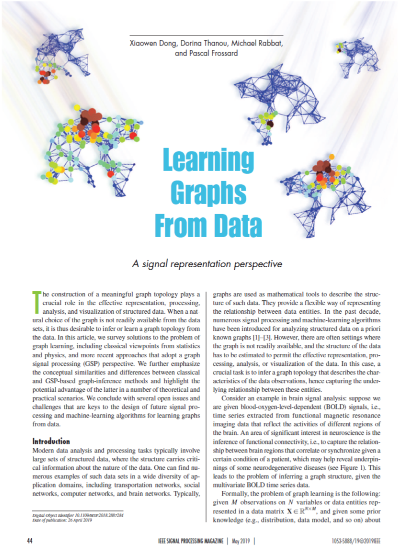
Graph Topology Inference

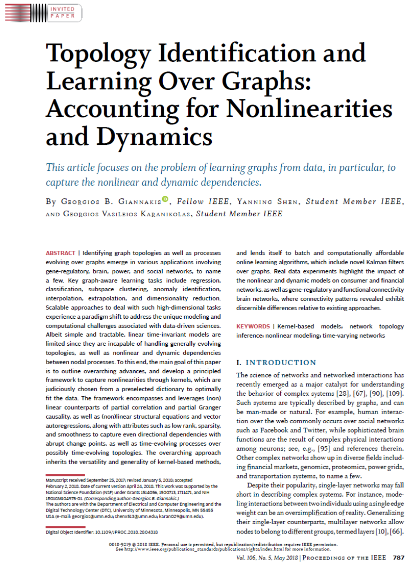
- **Statistical Methods**
  - Covariance Selection (GMRF) [Dempster '72]
  - Neighbourhood-based Regression [Meinshausen '06]
  - Graphical Lasso [Friedman '08]
  - Laplacian-constrained GMRF [Lake '10] [Egilmez '17]

- **Graph Signal Processing**
  - Smoothness-based models [Dong '16] [Kalofolias '16]
  - Edge subset selection [Chepuri '17] [Romero '17]
  - Spectral Templates [Pasdeloup '17] [Segarra '17] [Shafipour '18]
  - Heat-diffusion [Thanou '17]
  - Dynamics and non linearity [Giannakis '18]

- **Machine Learning on Graphs** [Scarselli '08] [Defferard '16]







TU Delft

Graph-Data Model

$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma),\ \Sigma \in \mathbb{S}_{++}^N$

sparsity pattern precision matrix
=
conditional independence

Gaussian Graphical Model (GGM)

$$\underset{\mathbf{S}}{\text{minimize}} - \log\det(\mathbf{S}) + \operatorname{tr}\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}\right)$$
$$\text{s. t.} \quad \mathbf{S} \in \mathbb{S}_{++}^N$$

$x_t(i) = \sum_{j \neq i} S(i, j) x_t(j) + e_t(i),$

Sort of neighbourhood regression

Structural Equation Model (SEM)

$$\underset{\mathbf{S}}{\text{minimize}} \quad \frac{1}{2T}\|\mathbf{X} - \mathbf{S}\mathbf{X}\|_F^2 + g(\mathbf{S}),$$
$$\text{s. t.} \quad \mathbf{S} \in \mathcal{S}$$

$\mathrm{LQ}_{\mathcal{G}}(\mathbf{x}_t) := \mathbf{x}_t^\top \mathbf{L} \mathbf{x}_t = \sum_{i \neq j} W(i, j)(x_t(i) - x_t(j))^2$

Signals are smooth on the graph

Smoothness-Based Model (SBM)

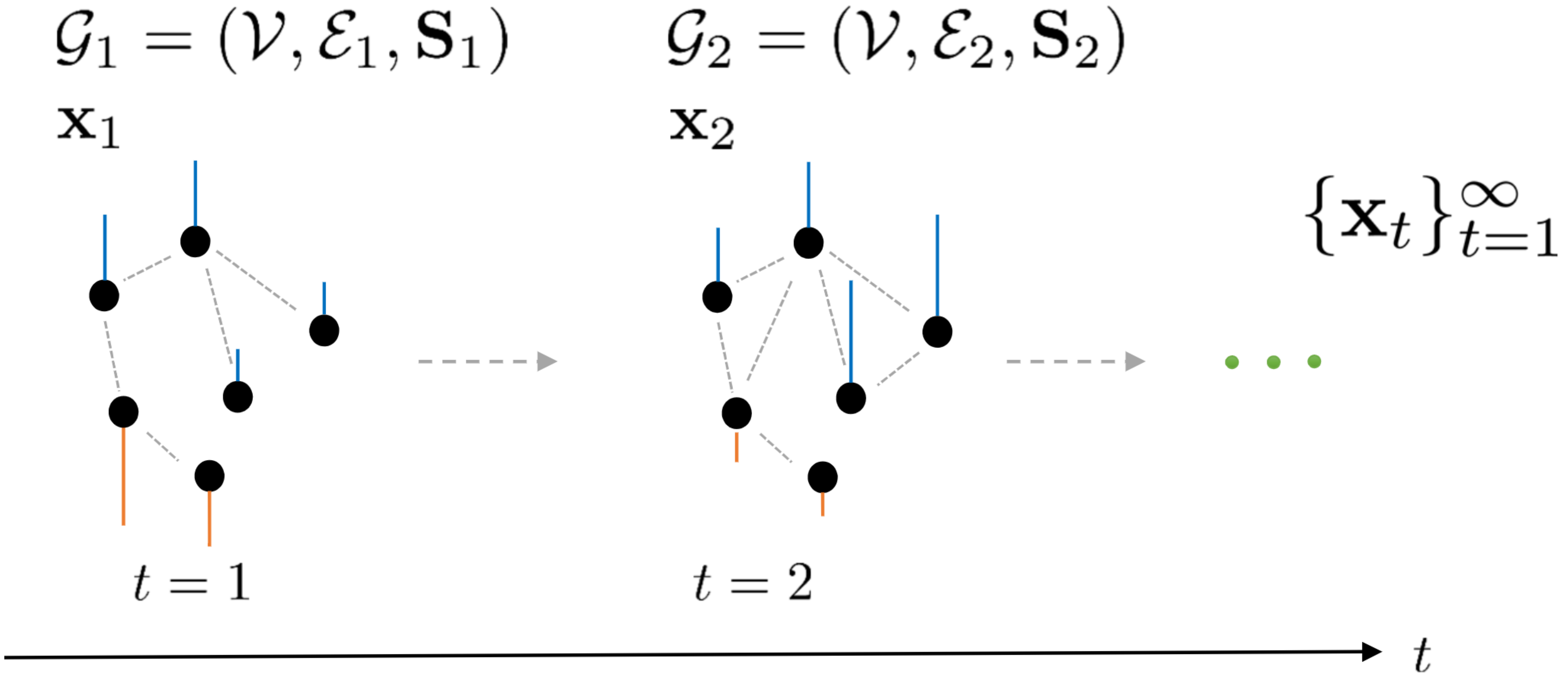
$$\underset{\mathbf{S}}{\text{minimize}} \quad \frac{1}{T} \operatorname{tr}\left(\mathbf{X}^\top \mathbf{S} \mathbf{X}\right) + g(\mathbf{S}),$$
$$\text{s. t.} \quad \mathbf{S} \in \mathcal{S}$$

TUDelft

However…

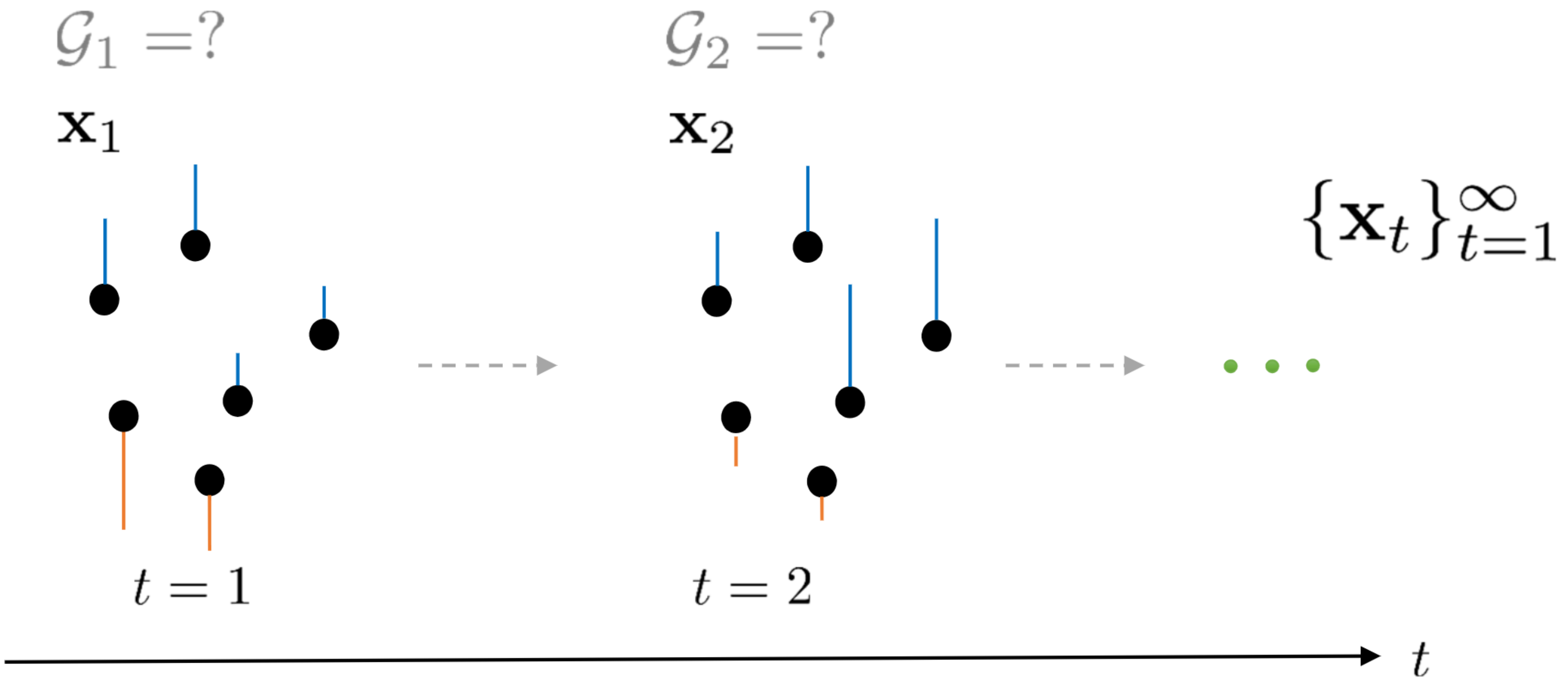
Often the underlying network changes over time, and data arrive in a streaming fashion

Scenario



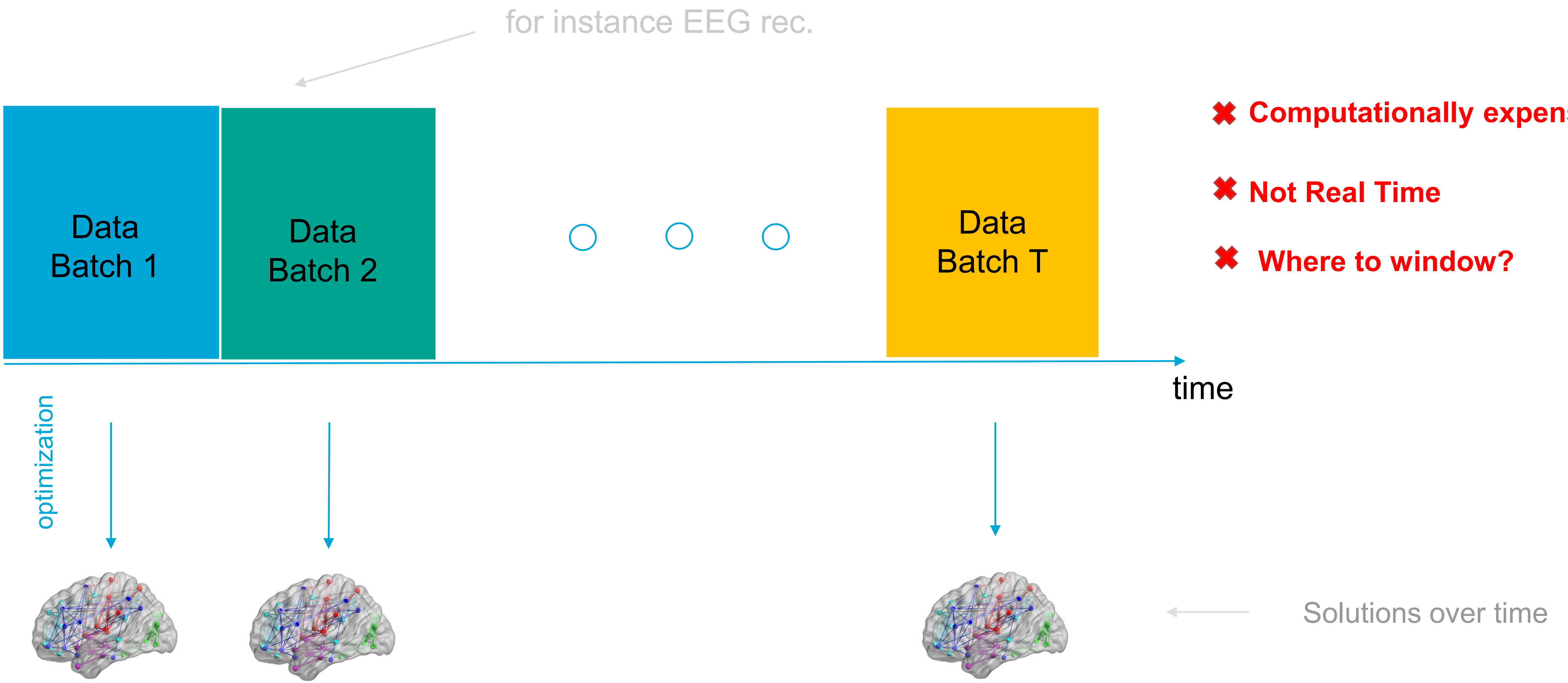
$\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1, \mathbf{S}_1)$
$\mathbf{x}_1$
$\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2, \mathbf{S}_2)$
$\mathbf{x}_2$
$\{\mathbf{x}_t\}_{t=1}^{\infty}$
$t = 1$
$t = 2$
t

Scenario



$\mathcal{G}_1 =?$
$\mathbf{x}_1$
$\mathcal{G}_2 =?$
$\mathbf{x}_2$
$\{\mathbf{x}_t\}_{t=1}^{\infty}$
$t = 1$
$t = 2$
t

Standard Approach



- ✖ **Computationally expensive**
- ✖ **Not Real Time**
- ✖ **Where to window?**

Our work

- A model-agnostic framework for learning time-varying graphs from streaming data
    - An umbrella for all the methods we have seen so far
    - Theoretical analysis on the abstract formulation
    - Operates in non-stationary environments
    - Accelerated through a prediction-correction scheme

TU Delft

Natali, A., Isufi, E., Coutino, M. and Leus, G., 2022. Learning time-varying graphs from online data. *IEEE Open Journal of Signal Processing*, *3*, pp.212-228.

Our work

Data fidelity Regularization and constraints

For every new graph signal $\mathbf{x}_t$ we have a new time-varying function $F(\mathbf{S};t) := \underbrace{f(\mathbf{S};t)}_{\text{Data fidelity}} + \underbrace{\lambda g(\mathbf{S};t)}_{\text{Regularization and constraints}}$

$$\mathbf{S}_t^\star := \operatorname{argmin}_{\mathbf{S}} f(\mathbf{S};t) + \lambda g(\mathbf{S};t) \quad \text{for } t = 1, 2, \ldots$$

TV-GGM

$$f(\mathbf{S};t) = -\log\det(\mathbf{S}) + \operatorname{tr}(\mathbf{S}\hat{\boldsymbol{\Sigma}}_t)$$
$$g(\mathbf{S};t) = \iota_{\mathcal{S}}(\mathbf{S})$$

TV-SEM

$$f(\mathbf{S};t) = \frac{1}{2}[\operatorname{tr}(\mathbf{S}^2\hat{\boldsymbol{\Sigma}}_t) - 2\operatorname{tr}(\mathbf{S}\hat{\boldsymbol{\Sigma}}_t) + \operatorname{tr}(\hat{\boldsymbol{\Sigma}}_t)]$$
$$g(\mathbf{S};t) = \|\mathbf{S}\|_1 + \iota_{\mathcal{S}}(\mathbf{S})$$

TV-SBM

$$f(\mathbf{S};t) = \operatorname{tr}(\operatorname{Diag}(\mathbf{S}\mathbf{1})\hat{\boldsymbol{\Sigma}}_t) - \operatorname{tr}(\mathbf{S}\hat{\boldsymbol{\Sigma}}_t)$$
$$g(\mathbf{S};t) = \frac{\lambda_1}{4}\|\mathbf{S}\|_F^2 - \lambda_2\mathbf{1}^\top\log(\mathbf{S}\mathbf{1}) + \iota_{\mathcal{S}}(\mathbf{S})$$

Dependence of the data through covariance matrix $\hat{\boldsymbol{\Sigma}}_t$ $\longrightarrow$ $\hat{\boldsymbol{\Sigma}}_t = \gamma\hat{\boldsymbol{\Sigma}}_{t-1} + (1-\gamma)\mathbf{x}_t\mathbf{x}_t^\top$ EWMA

TU Delft

Natali, A., Isufi, E., Coutino, M. and Leus, G., 2022. Learning time-varying graphs from online data. *IEEE Open Journal of Signal Processing*, *3*, pp.212-228.

Our work

For every new graph signal $\mathbf{x}_t$ we have a new time-varying function $F(\mathbf{S};t) := \overbrace{f(\mathbf{S};t)}^{\text{Data fidelity}} + \overbrace{\lambda g(\mathbf{S};t)}^{\text{Regularization and constraints}}$

$$\mathbf{S}_t^\star := \mathrm{argmin}_{\mathbf{S}}\, f(\mathbf{S};t) + \lambda g(\mathbf{S};t) \quad \text{for } t = 1, 2, \ldots$$

Solved through an iterative routine relying on a **prediction-correction scheme** (onto the half-vec space)

(at time t)
Prediction

next time-step function prediction

$$\mathrm{s}_{t+1|t}^\star := \underset{\mathrm{s}}{\mathrm{argmin}}\, \hat{F}(\mathrm{s}; t+1). \longrightarrow \hat{\mathbf{s}}^{p+1} = \hat{\mathcal{T}}\hat{\mathbf{s}}^p, \quad p = 0, 1, \ldots, P-1 \longrightarrow \hat{\mathbf{s}}_{t+1|t} = \hat{\mathbf{s}}^P$$

(at time t+1)
Correction

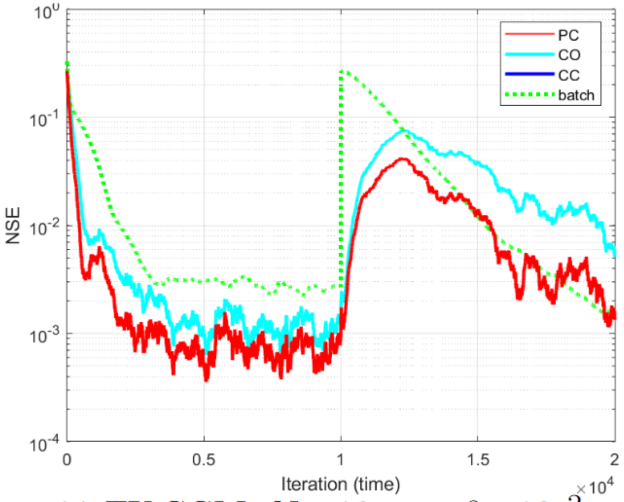
$$\mathrm{s}_{t+1}^\star := \underset{\mathrm{s}}{\mathrm{argmin}}\, F(\mathrm{s}; t+1). \longrightarrow \hat{\mathbf{s}}^{c+1} = \mathcal{T}\hat{\mathbf{s}}^c, \quad c = 0, 1, \ldots, C-1 \longrightarrow \hat{\mathbf{s}}_{t+1} = \hat{\mathbf{s}}^C$$

TU Delft

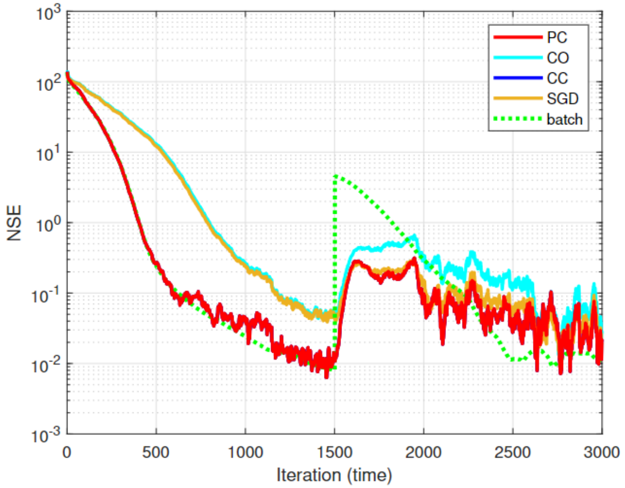
✓ Light computation ✓ Non-stationarity ✓ Accelerated

Synthetic Experiment

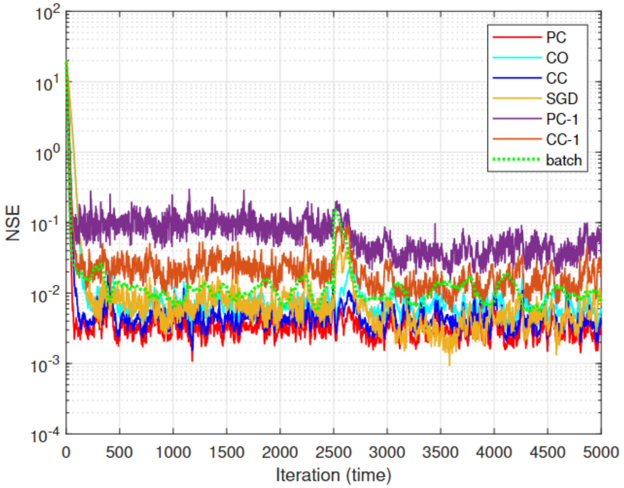
$$\mathrm{NSE}(\hat{\mathbf{s}}_t, \mathbf{s}_t^\star) = \frac{\|\hat{\mathbf{s}}_t - \mathbf{s}_t^\star\|_2^2}{\|\mathbf{s}_t^\star\|_2^2}.$$



(a) **TV-GGM.** $N=18$, $\alpha=\beta=10^{-2}$, $\gamma=99.9\times10^{-2}$



(b) **TV-SEM.** $N=28$, $\alpha=\beta=0.1\times10^{-2}$, $\lambda=0.5$, $\gamma=99\times10^{-2}$



(c) **TV-SBM.** $N=28$, $\alpha=\beta=0.1\times10^{-2}$, $\lambda_1=10$, $\lambda_2=10$, $\gamma=99\times10^{-2}$

TABLE I: Average time (expressed in seconds) required to compute the PC and the CVX solution at each time instant.

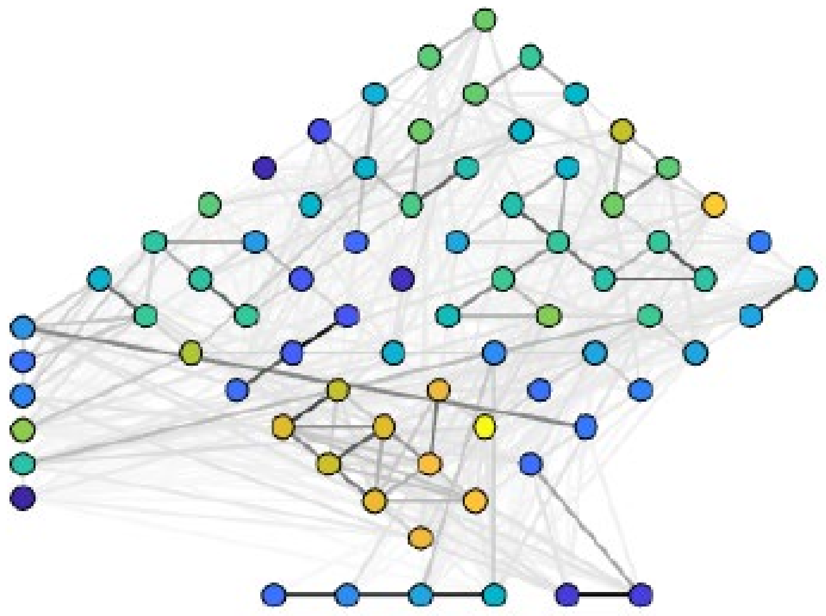
	PC	CVX
TV-GGM	0.110×10^{-2}	3.6
TV-SEM	0.824×10^{-2}	2.0
TV-SBM	0.023×10^{-2}	3.6

TU Delft

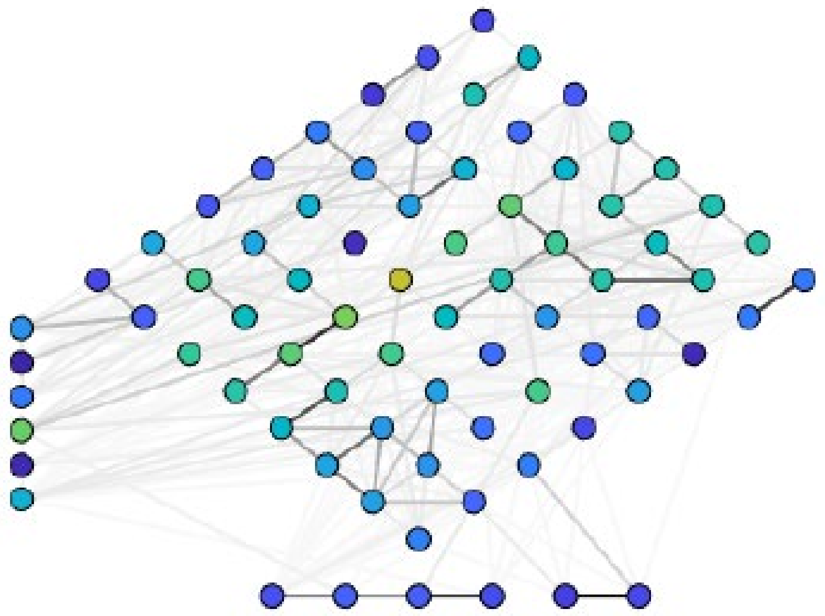
Epilepsy Study

$N = 76$ electrodes

Pre-ictal



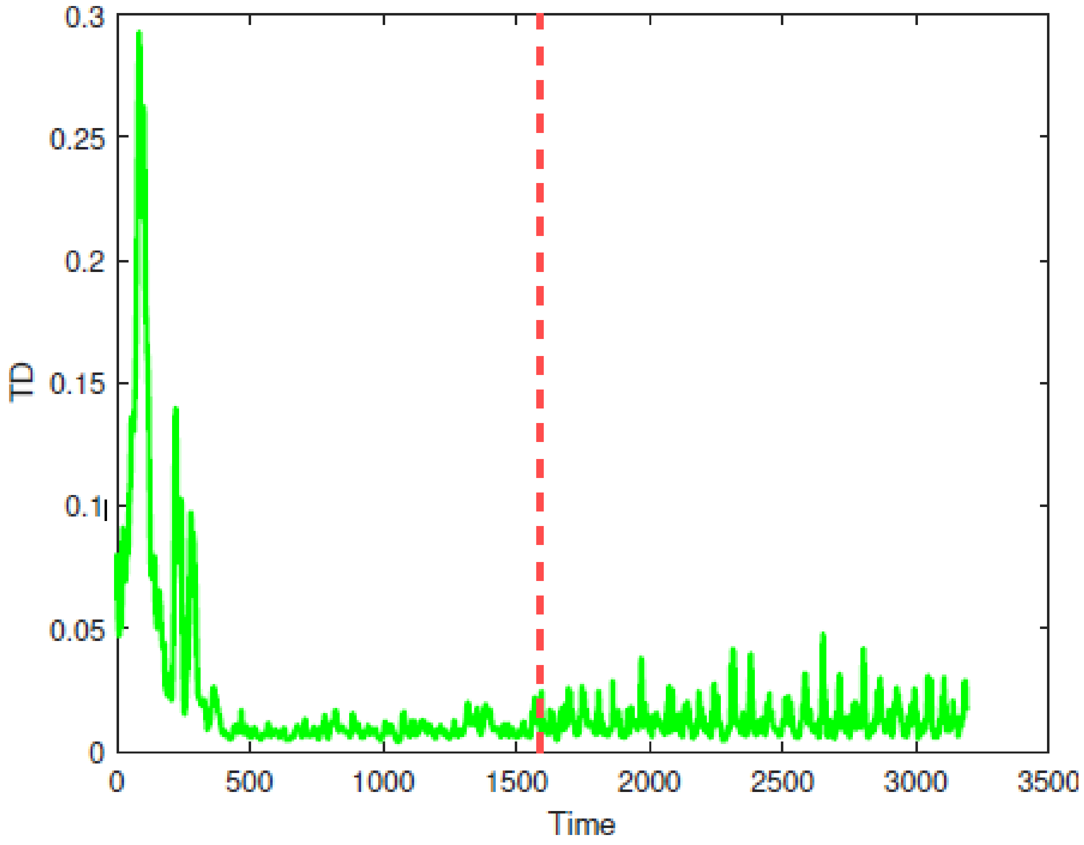
Ictal

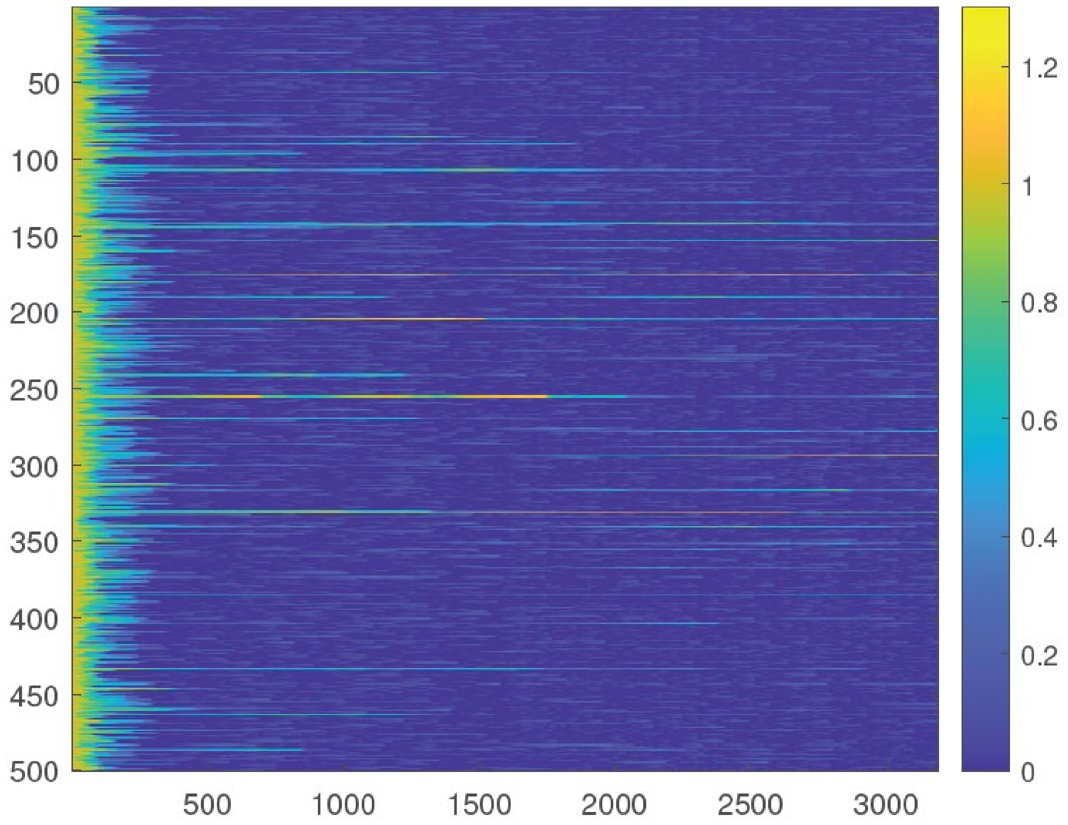


TU Delft

Epilepsy Study

$$\mathrm{TD}(t) := \|\hat{\mathbf{s}}_t - \hat{\mathbf{s}}_{t-1}\|_2$$





TU Delft

Q&A

Alberto Natali

a.natali@tudelft.nl

TUDelft