Shape analysis & machine learning for time series

V. Guerrini[1,2] and T. Germain[1,2],

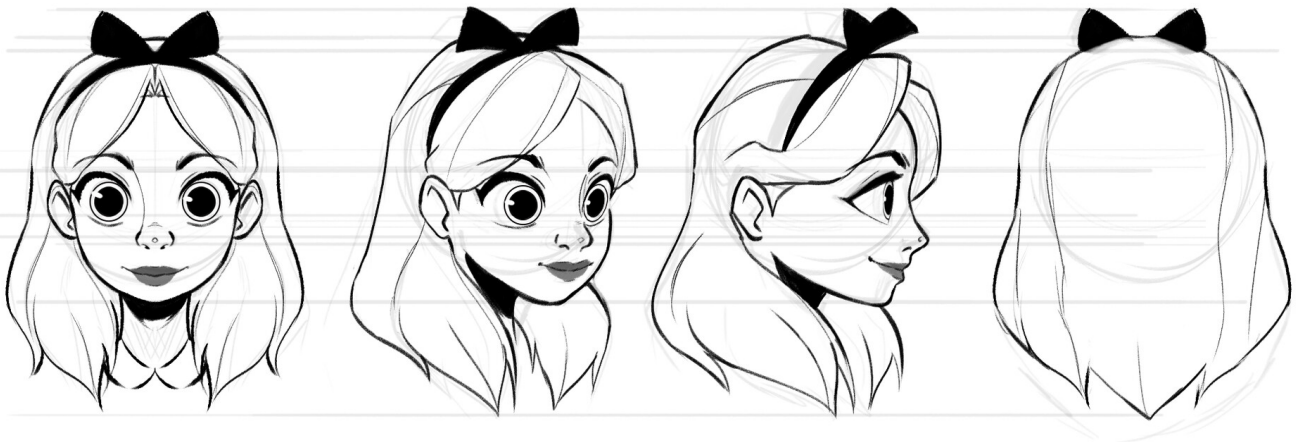
C. Truong[1,2] and L. Oudre[1,2]

[1]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190, Gif-sur-Yvette, France
[2]Université de Paris, CNRS, Centre Borelli, F-75005 Paris, France

Duke, July 2024

Motivation

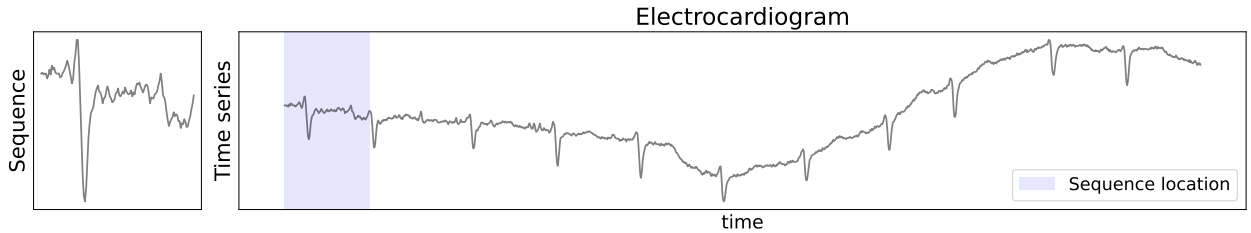
- The word *"shape"* commonly refers to the appearance of an object.



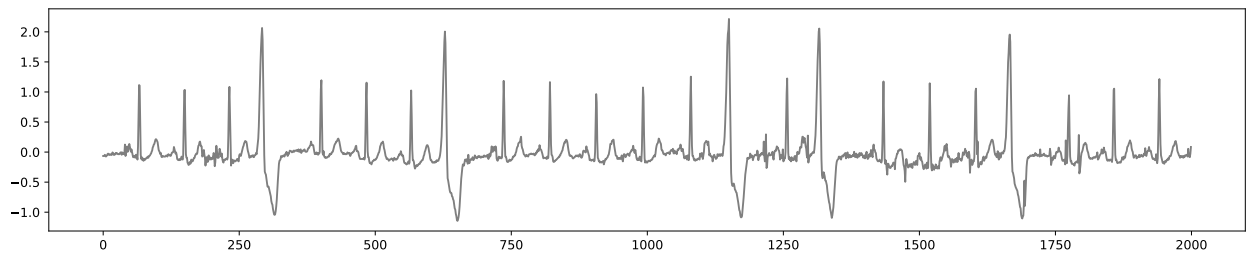
- From a geometric perspective, *shape* refers to geometric properties of an object that are invariant to some source of variability/deformations.
- Shape-based methods benefit from **good generalization** properties and they have links with **contrastive learning**, **domain adptation**.
- Shape analysis has applications in computer vision and computational anatomy but can be extended to time series.

Shape-related problems in time series

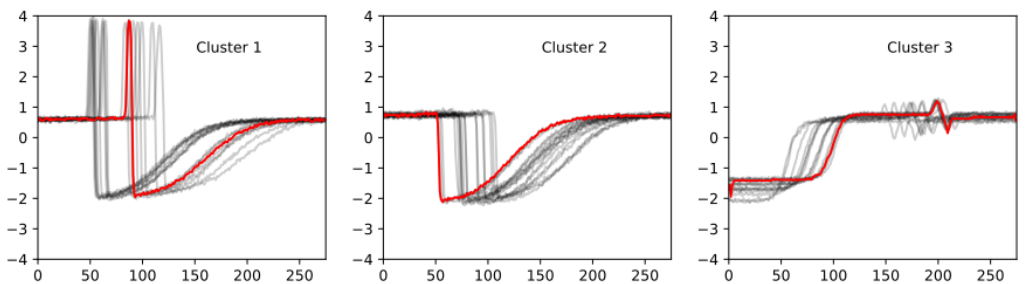
1. **Similarity search**



2. **Motif discovery**



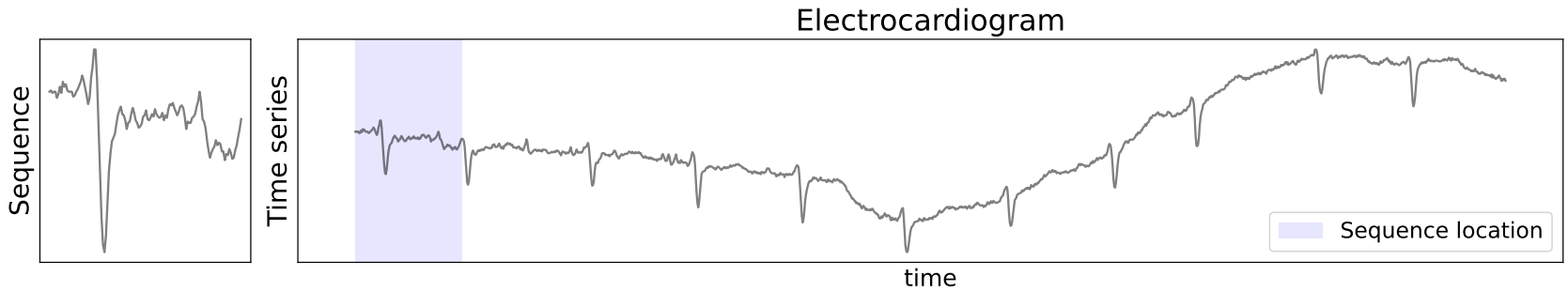
3. **Clustering**



Similarity search in time series

Definition

Similarity search consists in retrieving occurrences of a query pattern in a time series.



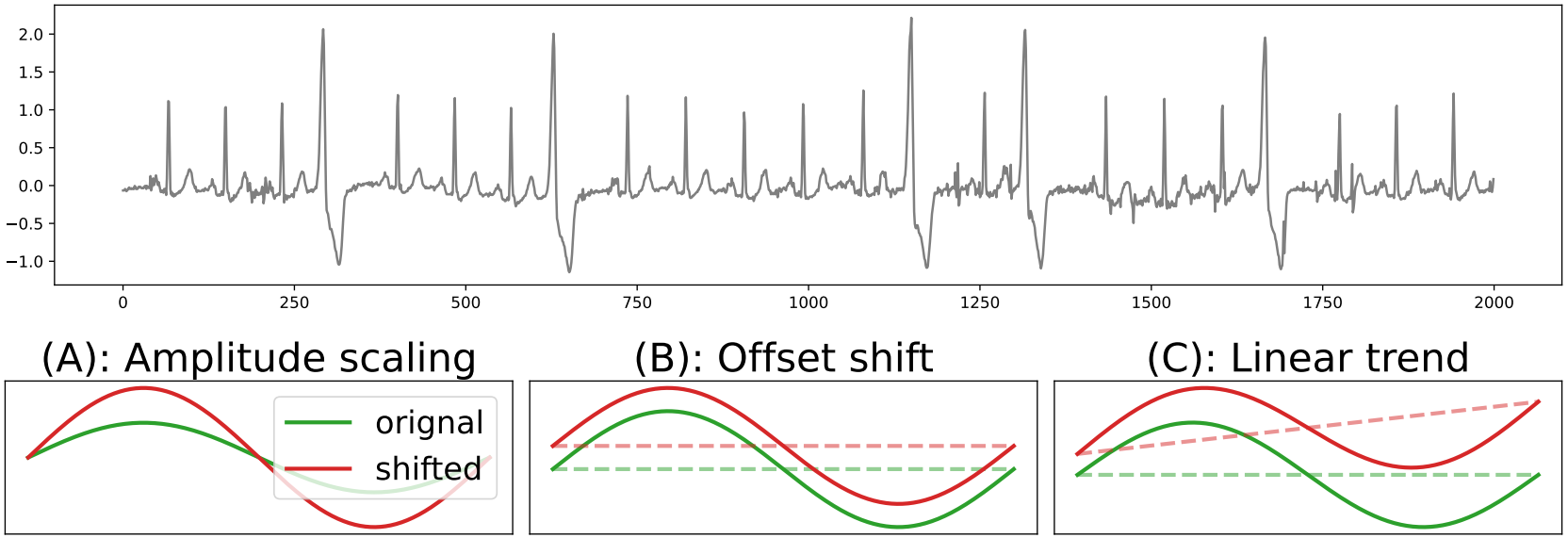
Main goal

Defining distance functions between subsequences that take into account sources of variability and are computationally efficient.

Hypothesis:

- A single long time series
- No time warping variations

A distance example: LT-normalized Euclidean distance



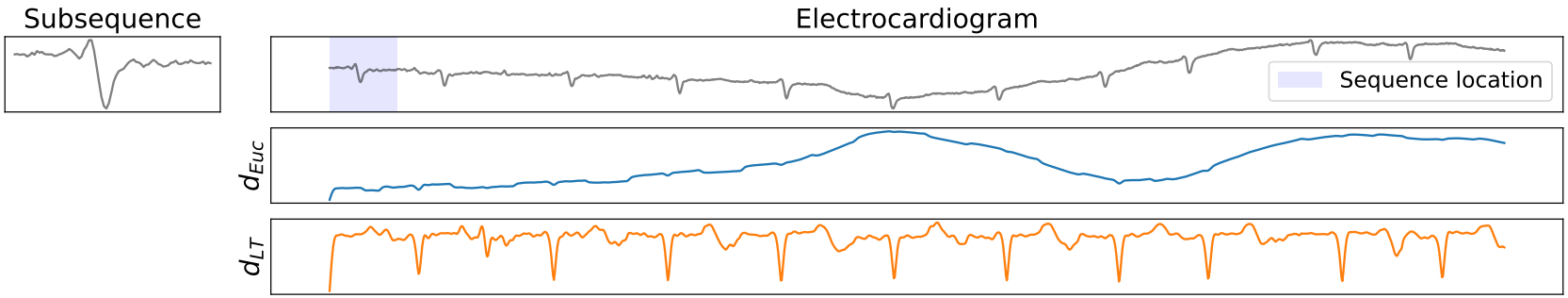
Definition (LT-normalized Euclidean distance)

The LT-normalized (Euclidean) distance between $x \in \mathbb{R}^w$ and $y \in \mathbb{R}^w$ is:

$$d_{LT}(x,y) = \left\| \frac{x - (\alpha_x \mathbf{t} + \beta_x \mathbf{1})}{\|x - (\alpha_x \mathbf{t} + \beta_x \mathbf{1})\|} - \frac{y - (\alpha_y \mathbf{t} + \beta_y \mathbf{1})}{\|y - (\alpha_y \mathbf{t} + \beta_y \mathbf{1})\|} \right\|$$

where $\mathbf{t} = (0, \ldots, w-1)$, $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^w$ and (α_x, β_x) are solutions of the linear regression problem $\underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \|x - (a\mathbf{t} + b\mathbf{1})\|^2$.

A distance example: LT-normalized Euclidean distance



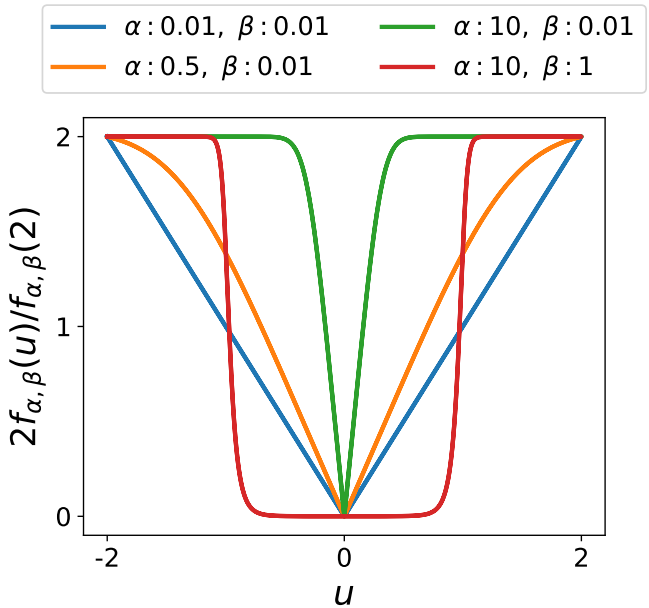
(α, β)-rectified distance

Let $\alpha \in \mathbb{R}_+^*$ and $\beta \in]0, 2[$, the (α, β)-rectified distance between $x \in \mathbb{R}^w$ and $y \in \mathbb{R}^w$ is:

$$d_{\alpha,\beta}(x, y) = 2f_{\alpha,\beta}(d_{LT}(x, y))/f_{\alpha,\beta}(2)$$

with

$$f_{\alpha,\beta}(x) = \sqrt{\tanh(\alpha\beta^2) + \tanh(\alpha(x^2 - \beta^2))}$$



Graph computation complexity

For any sequence $Q \in \mathbb{R}^w$ and time series $S \in \mathbb{R}^n$, the distance profile between Q and S is computed in $\mathcal{O}(n \log(n))$.

Motif discovery in time series

Definition

Motif Discovery consists of finding repeated patterns and locating their occurrences in a time series without prior knowledge about their shape or location.

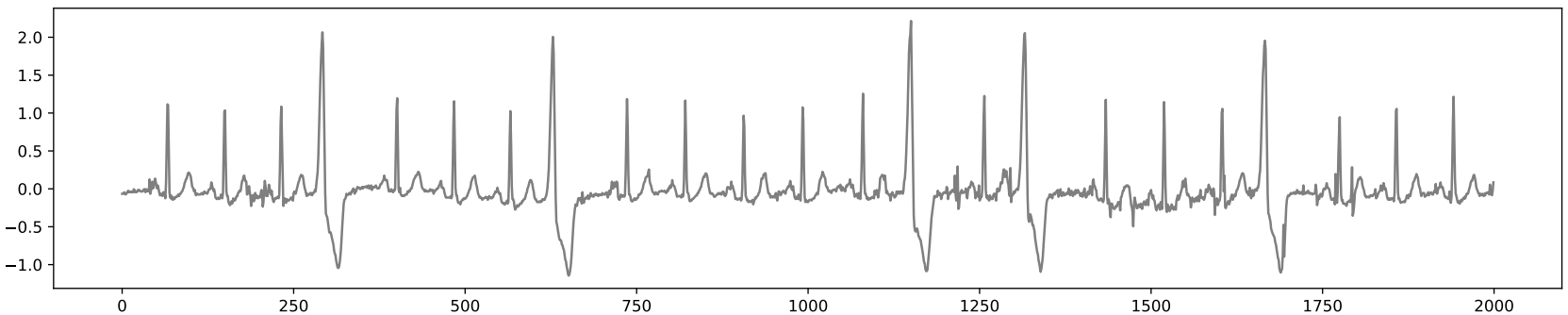
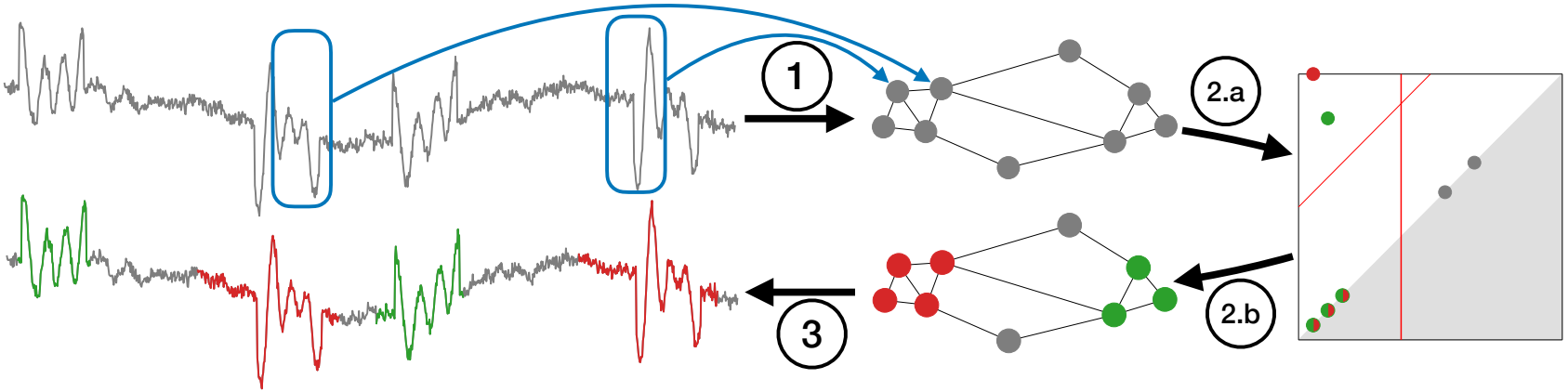


Figure: ECG with premature ventricular contraction (PVC)

Hypothesis:

- A single long univariate time series
- May contain motifs of different length

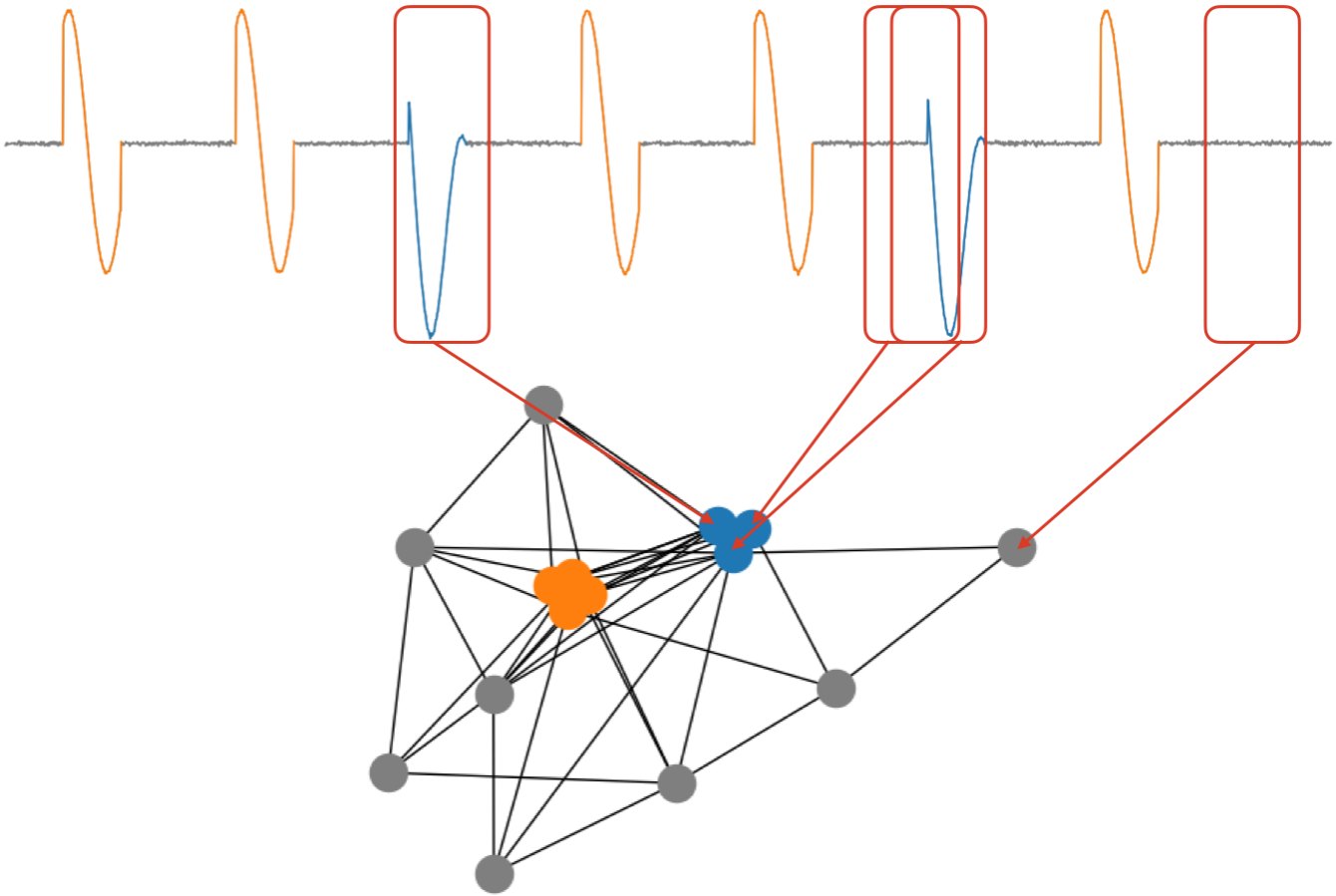
PEPA algorithm overview



Algorithm steps:

1. From time series to graph
2. Graph clustering with persistent homology
   a. From graph to persistent diagram
   b. From persistent diagram to clusters
3. From cluster to motifs sets

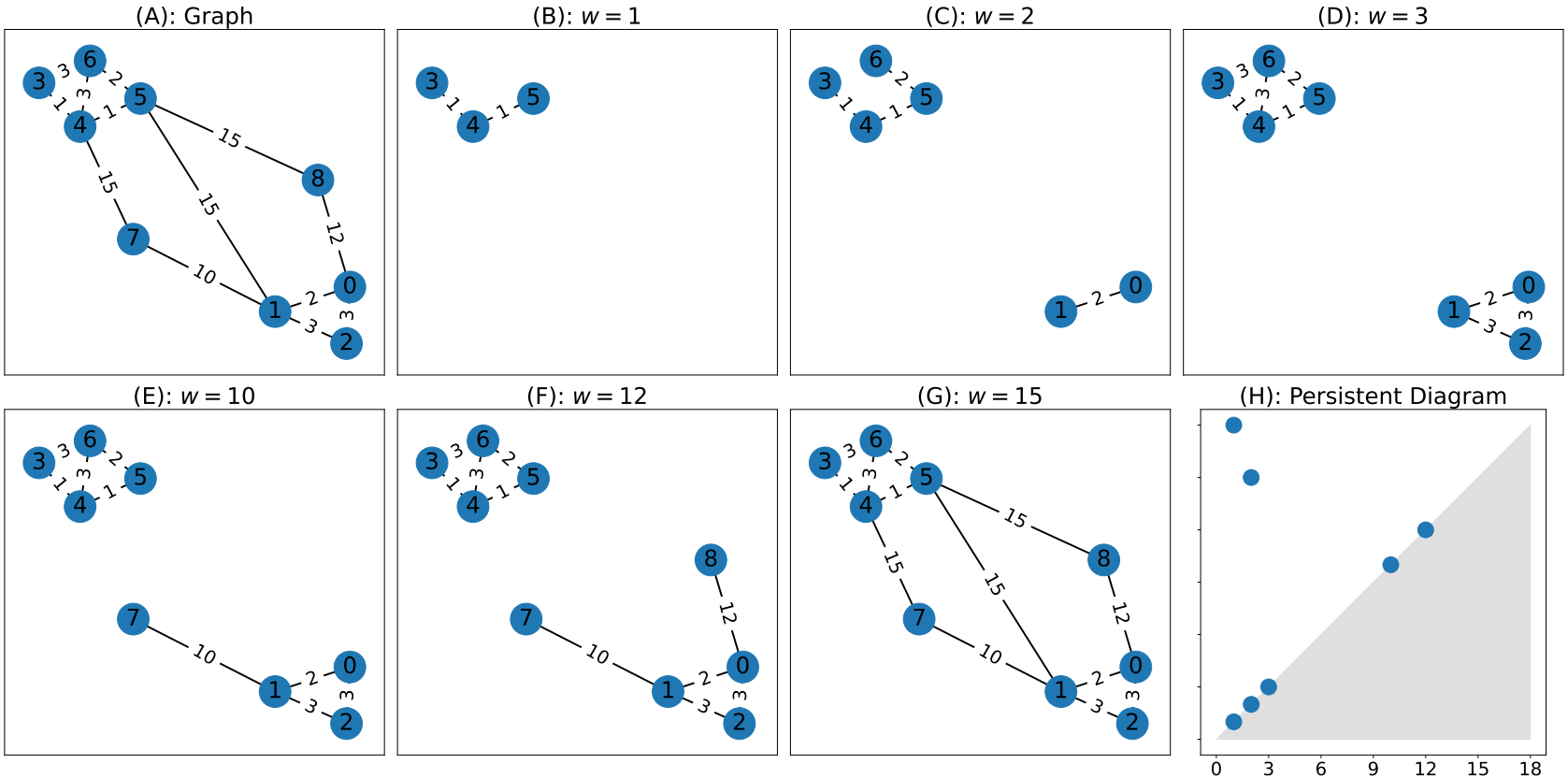
From time series to graph



Main idea

Subsequences that overlap the same motif are close to each other and far from any other subsequences.

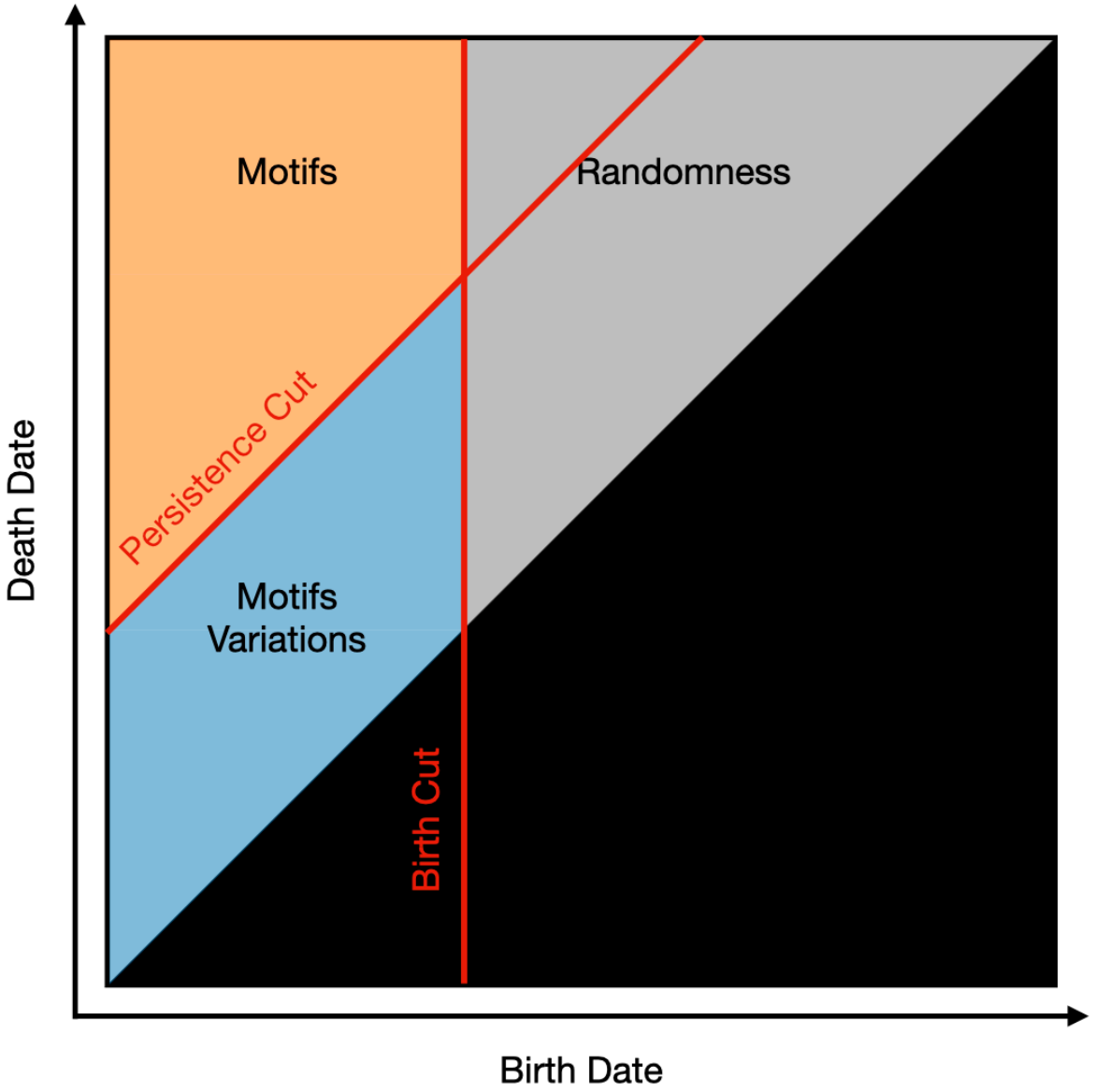
Graph clustering through persistent homology



Connected subgraph tracking rule:

- Birth Date: distance at which its first edge appears.
- Death Date: distance at which the subgraph gets connected an older subgraph.

Graph clustering through persistent homology



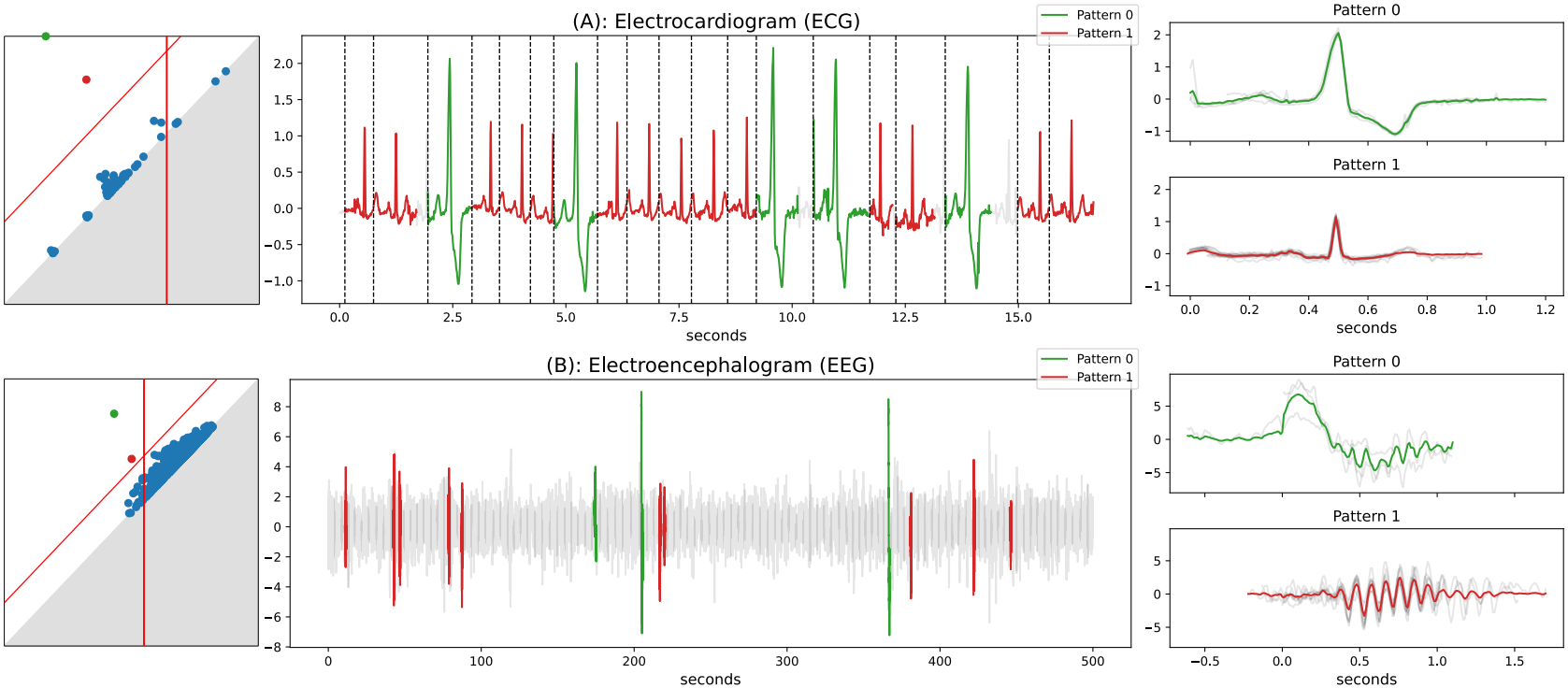
From persistence diagram to motif sets

Input: the graph of a time series, $\mathcal{G}_S^K$.
Parameters: persistent cut, birth cut.

Procedure

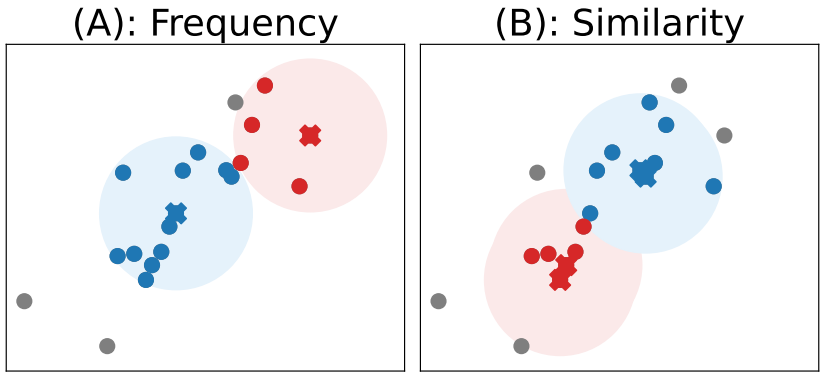
1. Perform filtration and prevent subgraphs from merging when their persistence exceeds the persistence cut.
2. Remove all subsequences whose date of birth is greater than the birth cut.
3. For each subgraph (motif), merge subsequences that are time-adjacent (occurrences) while preventing overlapping between occurrences.

Examples



Web App

State-of-the-art



Two main families of algorithms:

- **Frequency-based:** algorithms identify sets of subsequences that represent the most frequently repeated patterns.
- **Similarity-based:** algorithms identify sets of subsequences that represent repeated patterns with minimal variability between occurrences, regardless of frequency.

Observation: Most algorithms rely on three core parameters: the number of motifs, the motif length, and a similarity threshold.
➔ In practice, the setting often results from a trial-and-error strategy.

Scientific challenges

The **choice of methods** faces several challenges:

- There is **no unanimity** on how to formally define a motif.
- How do you identify the **number of patterns** and **the length of their occurrences**?
- Pattern occurrences may have **different lengths**.
- Lack of **metrics** to assess method performance.

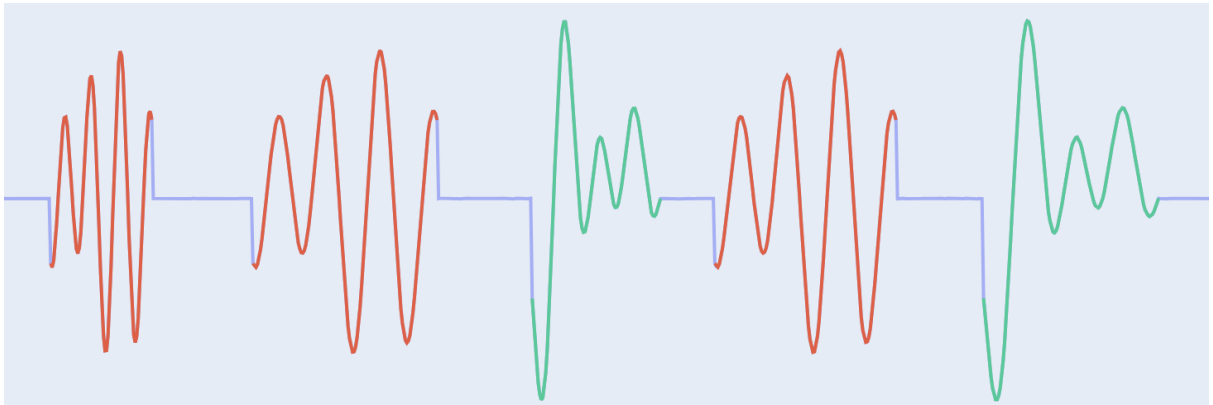


Figure: Time series with 2 patterns of variable lengths.

Setting up of a benchmark

Setting up a **benchmark** for pattern detection in time series.

- **State of the art** of the task, **taxonomy** of methods according to their nature and strengths.
- Formalization of **metrics** for the task.
- Setting up a **large dataset**, pre-processed and adapted to the task.
- Performance testing of **12 algorithms**, representatives of the state of the art, on real data.
- Setting up experiments to identify the **strengths** of each algorithm.

Question of major interest which we do not answer in this work:

How can these methods be adapted to the search of patterns in **multivariate** time series ?

Experiment: comparison with state-of-the-art

Protocol:

- task: motif set discovery
- datasets: 6 real-worlds, 3 synthetics
- metrics: f1-score
- algorithms: PEPA, A-PEPA (adaptative version), and 6 competitors.

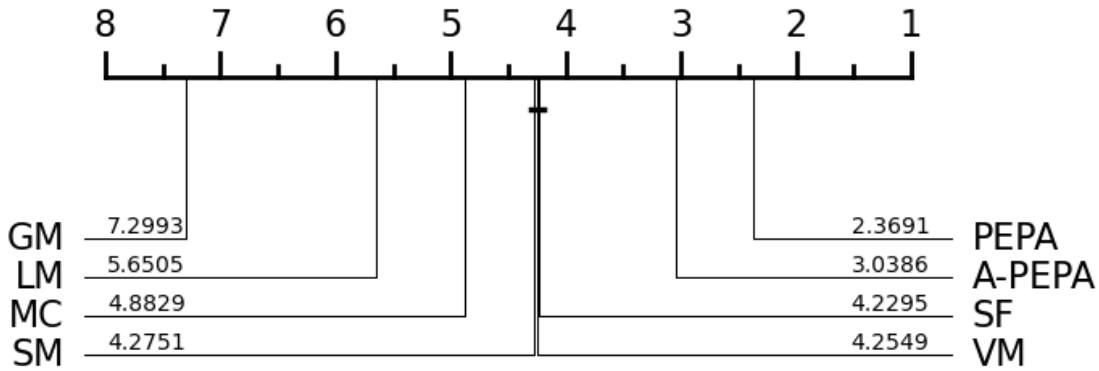
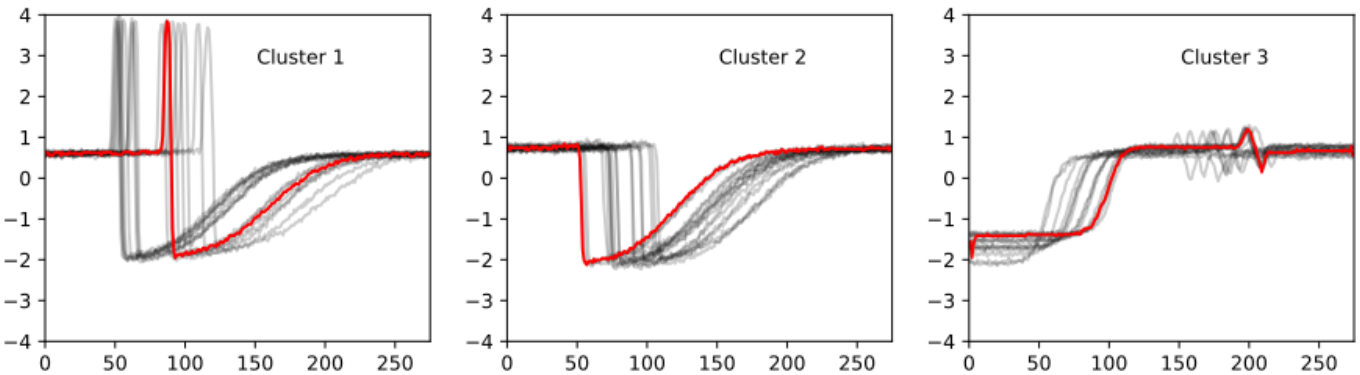


Figure: Critical difference diagram (f1-score based rank, Friedman's test & Nemenyi post-hoc test). PEPA and A-PEPA performs significantly better than other algorithms.

Shape-based clustering

Definition

Grouping time series with similar shape together.



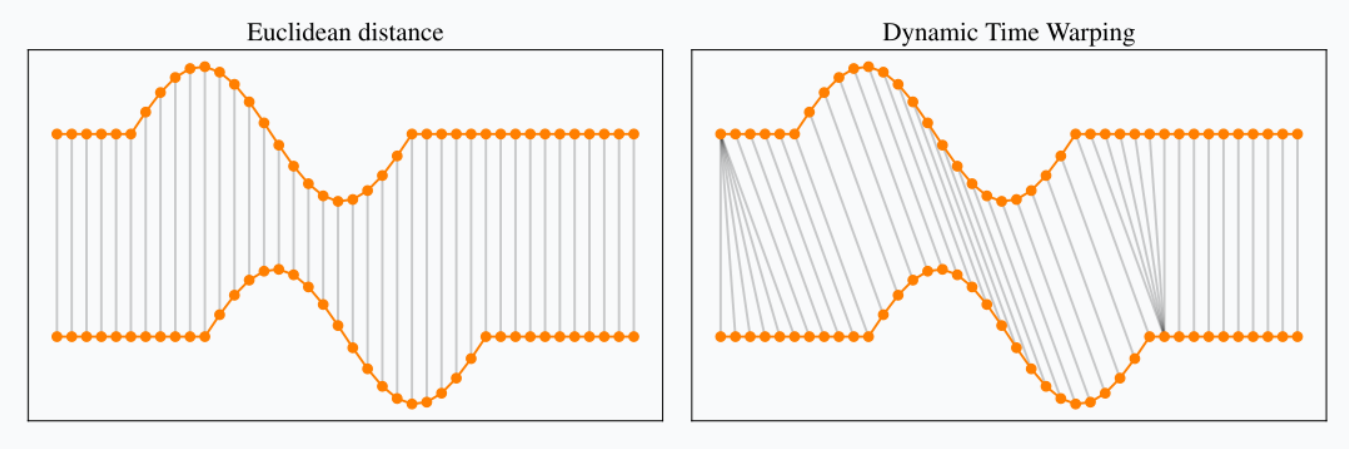
Hypothesis:

- Clustering at the sequence level
- Time warping made possible

Dynamic Time Warping distance (DTW)

Definition

Dynamic Time Warping (DTW) is a distance function that is invariant to time parametrization.



Main idea

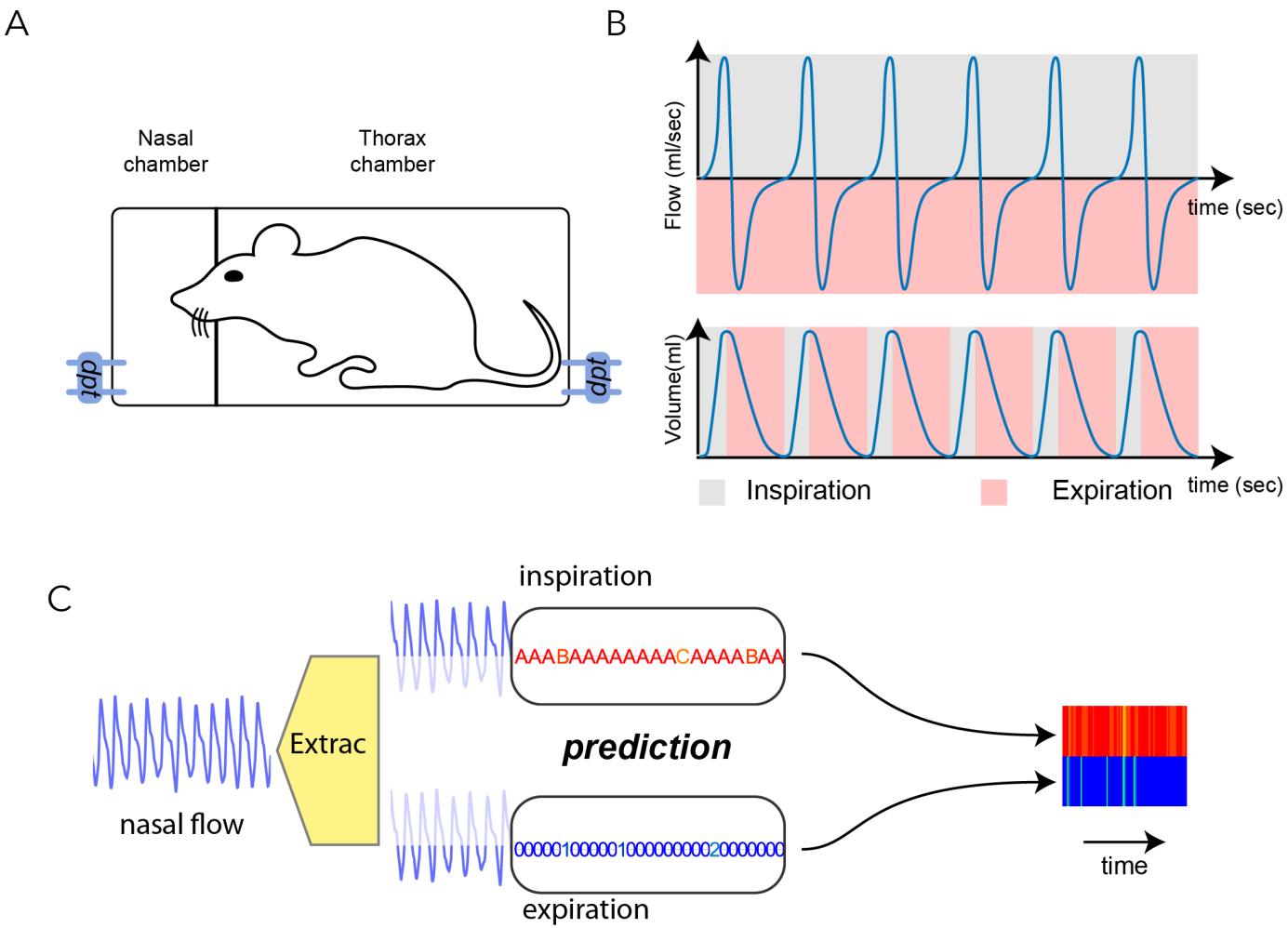
DTW learns the alignment between two curves before computing the euclidean distance between realigned curves.

➔ Computing DTW-Fréchet means is possible and leads to a Kmean-DTW algorithm.

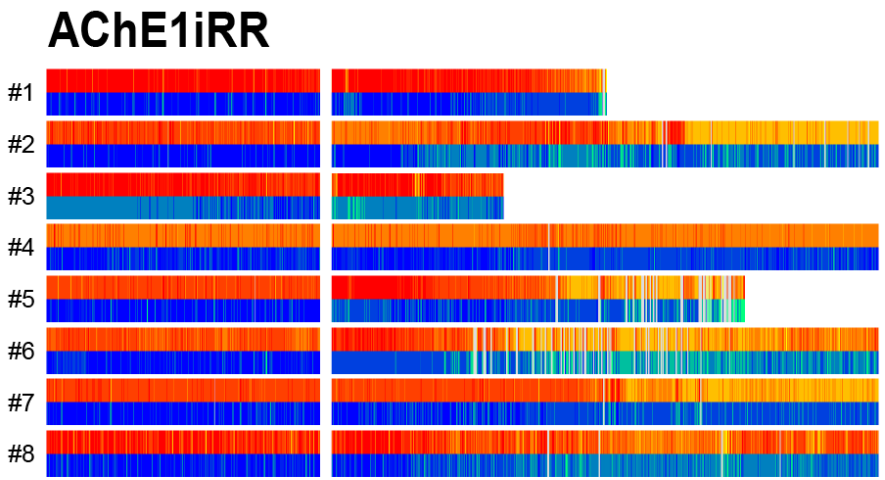
An application example: Mice breathing behavior analysis

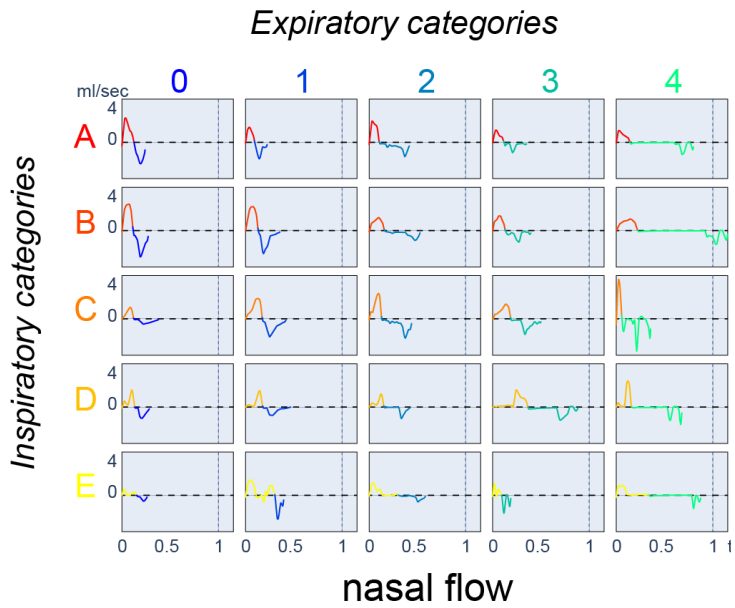
Objective

From plethysmogram signals exploring mice breathing behavior changes when exposed to irritant molecules.



An application example: Mice breathing behavior analysis





Benefits:

- Fast to compute
- Ease experiment interpretation.
- Symbolic representation suited for changed point or anomaly detection.

Limitations

- Invariance to time warping is to strong, need to quantify deformations between respiratory cycles.

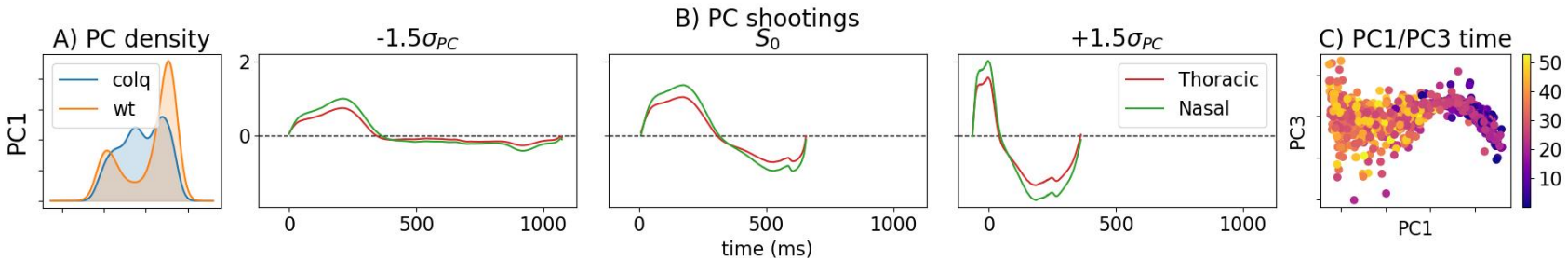
An adaptation of LDDMM to time series.

Main idea

- Time series are represented by deformations of a reference time series. The deformations are parameterized diffeomorphisms.
- Once the deformations and the reference time series are learned, the vectorized representation of individual time series is given by the parametrization of their corresponding deformation.

Benefits

- Vectorized representation of time series (irregularly sampled, variable length, and multivariate)
- Interpretable methods
- Possibility to perform machine learning methods and Kernel-SVM, Kernel-PCA.



Thank you!

Bibliography

- **PEPA algorithm**: T. Germain, C. Truong, and L. Oudre. "Persistence-Based Motif Discovery in Time Series". In: *IEEE Transactions on Knowledge and Data Engineering* (2024)
- **Graph clustering**: Alexandre Bois, Brian Tervil, and Laurent Oudre. "Persistence-based clustering with outlier-removing filtration". In: *Frontiers in Applied Mathematics and Statistics* 10 (2024), p. 1260828
- **Distance**: Thibaut Germain, Charles Truong, and Laurent Oudre. *Linear-trend normalization for multivariate subsequence similarity search. In Proceedings of the International Conference on Data Engineering Workshops (ICDEW), Utrecht, Netherlands*. 2024
- **Web App:** Thibaut Germain, Charles Truong, and Laurent Oudre. *Interactive motif discovery in time series with persistent homology. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Vilnius, Lithuania*. 2024

From time series to graph

Aim: Map a time series $S = (s_1, \ldots, s_n) \in \mathbb{R}^n$ into an undirected weighted graph $\mathcal{G}_S^K = (V, E)$.

Graph specification

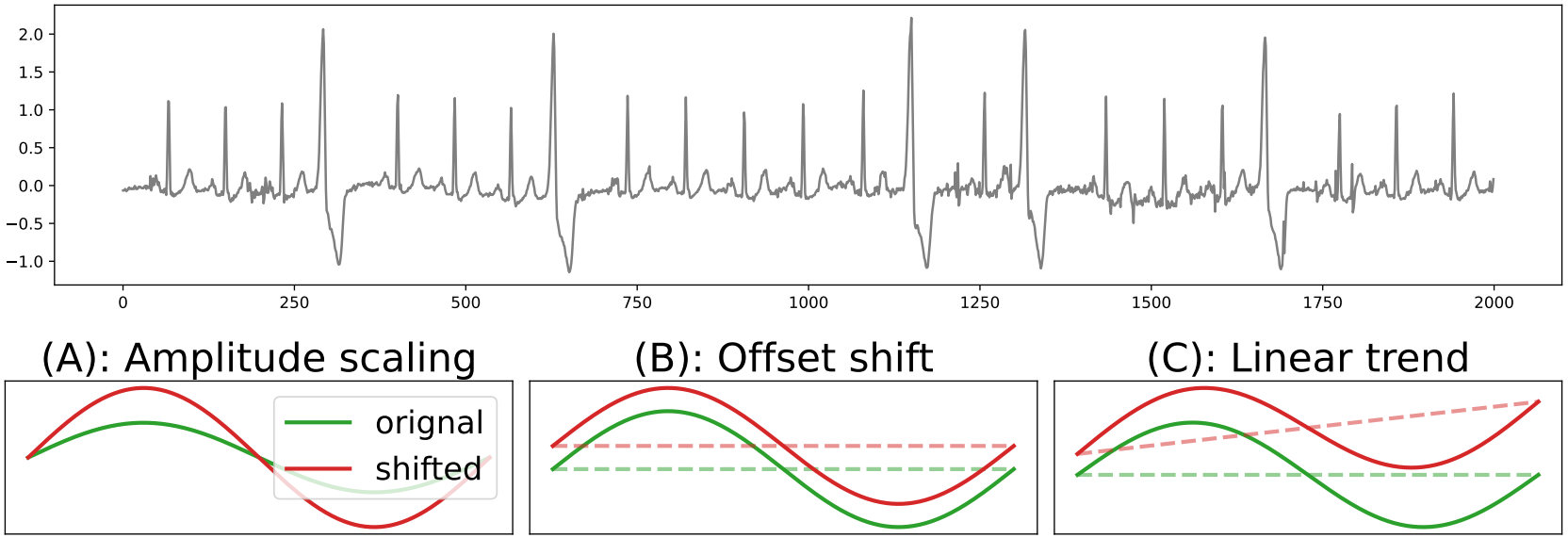
- **Vertices**: $V = (S_i^w)_{i=1\ldots n-w+1}$, subsequences of length l of S
- **Edges**: $E = E_1 \cup E_2$, union of two edges sets:
  - **Similarity set:** each subsequence S_i^w is connected to its K most similar non-overlapping subsequences.

$$E_1 = \bigcup_{i=1}^{n-w+1} \left\{ \left(S_i^w, N_i^k, d_i^k \right) \mid k = 1, \ldots, K \right\}, \tag{1}$$

  - **Time set:** edges connecting time adjacent subsequences.

$$E_2 = \bigcup_{i=1}^{n-w} \{ (S_i^w, S_{i+1}^w, \max \left(d_i^1, d_{i+1}^1 \right) \}. \tag{2}$$

Distance between subsequences



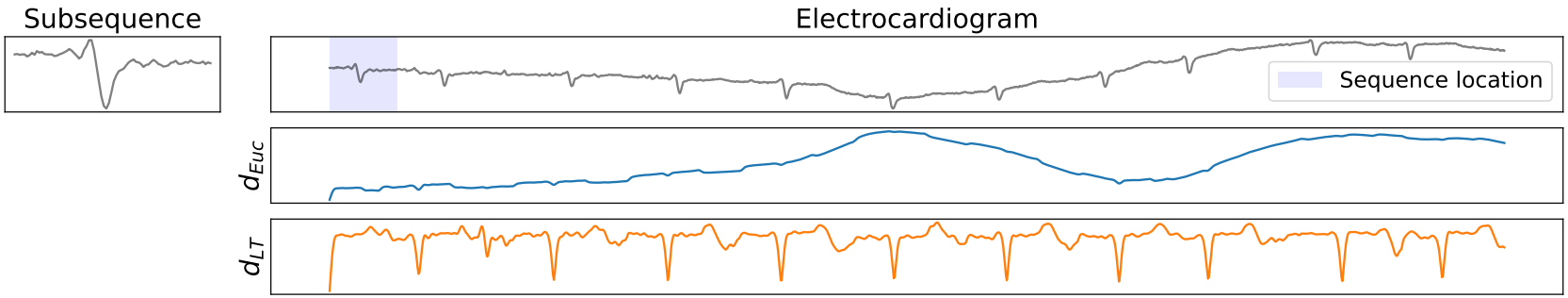
Definition (LT-normalized Euclidean distance)

The LT-normalized (Euclidean) distance between $x \in \mathbb{R}^w$ and $y \in \mathbb{R}^w$ is:

$$d_{LT}(x, y) = \left\| \frac{x - (\alpha_x \mathbf{t} + \beta_x \mathbf{1})}{\|x - (\alpha_x \mathbf{t} + \beta_x \mathbf{1})\|} - \frac{y - (\alpha_y \mathbf{t} + \beta_y \mathbf{1})}{\|y - (\alpha_y \mathbf{t} + \beta_y \mathbf{1})\|} \right\|$$

where $\mathbf{t} = (0, \ldots, w-1)$, $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^w$ and (α_x, β_x) are solutions of the linear regression problem $\underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \|x - (a\mathbf{t} + b\mathbf{1})\|^2$.

Distance between subsequences



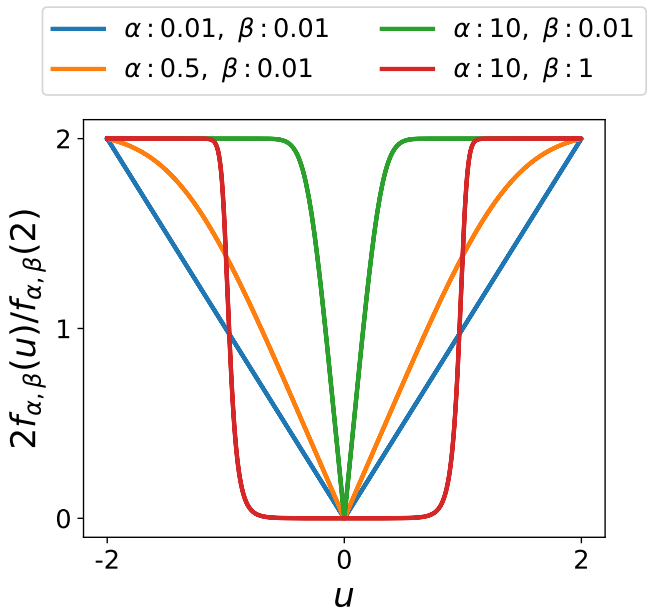
(α, β)-rectified distance

Let $\alpha \in \mathbb{R}_+^*$ and $\beta \in]0, 2[$, the (α, β)-rectified distance between $x \in \mathbb{R}^w$ and $y \in \mathbb{R}^w$ is:

$$d_{\alpha,\beta}(x, y) = 2f_{\alpha,\beta}(d_{LT}(x, y))/f_{\alpha,\beta}(2)$$

with

$$f_{\alpha,\beta}(x) = \sqrt{\tanh(\alpha\beta^2) + \tanh(\alpha(x^2 - \beta^2))}$$



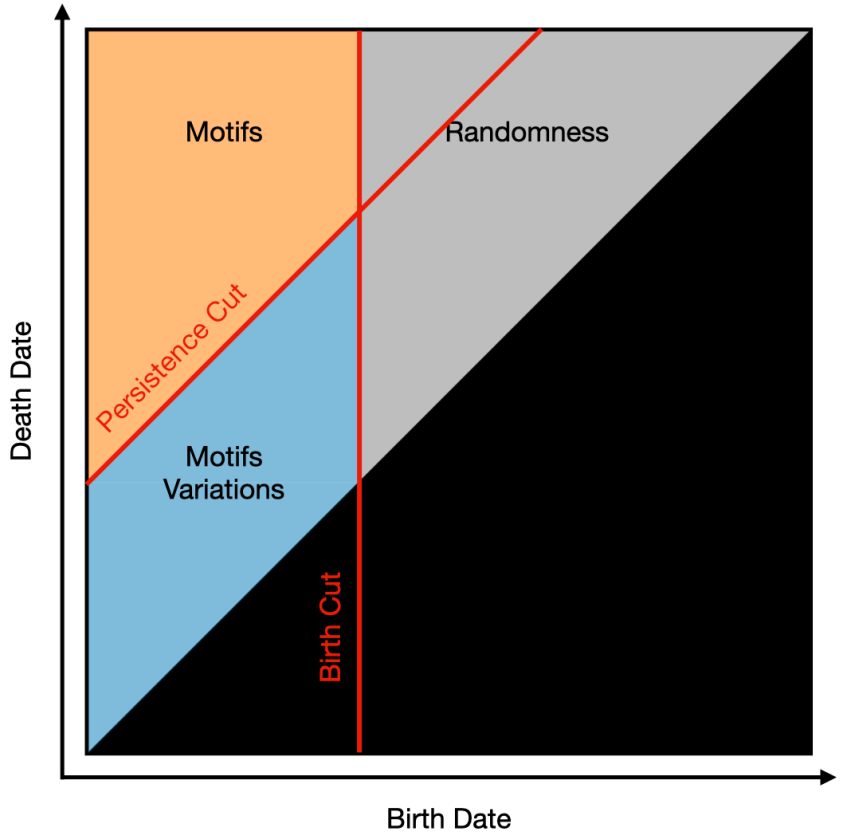
Graph computation complexity

For any time series $S \in \mathbb{R}^n$, the graph $\mathcal{G}_S^K$ is computed in $\mathcal{O}(Kn^2)$.

Graph clustering through persistent homology

Some properties:

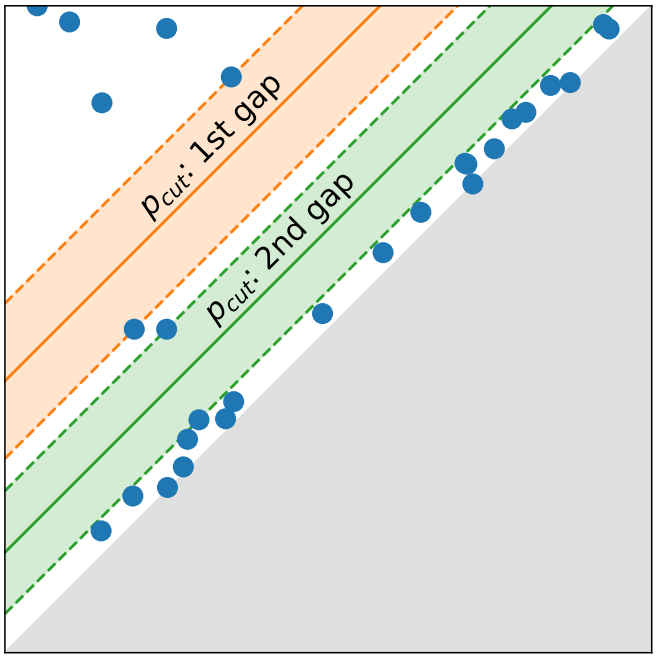
- All nodes (subsequences) have birth and death dates.
- Births and deaths are tracked with an algorithm similar to the Kruskall's algorithm for computing the minimum spanning tree (MST).
- The filtration of $\mathcal{G}_S^K$ can be reduced to the filtration of its MST.



Parameter heuristics

Birth cut:
Otsu method: the cut maximizes the inter-class variance.

Persistence cut:
Fixed cut: the number of motifs is given.
Adaptive cut: based on the persistence gap.



Distance parameter (α, β)**:**
For any $\alpha \in \mathbb{R}_+$ and $\beta \in]0, 2[$, $f_{\alpha,\beta}$ is bijective on $[0, 2]$ and strictly increasing.
Thus, edges' weight order is preserved while their value changes when $f_{\alpha,\beta}$ is modified: the filtration and the MST are preserved.
The optimization criteria is given by:

$$\underset{\alpha,\beta}{\operatorname{argmin}}\, D_{C.S}(B_{\alpha,\beta}, U_{[0,2]})$$

where $D_{C.S}$ is the Cauchy-Schwartz divergence, and $\mathcal{B}_{\alpha,\beta}$ is the kernel density estimation of vertices' birth, and $U_{[0,2]}$ is the density of the uniform distribution on $[0, 2]$.

Experiment: Influence of the subsequence length

Question: Does the subsequence length parameter affect the retrieval of variable-length motif sets?