Topological Data Analysis

Lecture 13

Topological Data Analysis of Time Series

Oleg Kachan

TDA for time series analysis

Dimensionality

- univariate
- multivariate

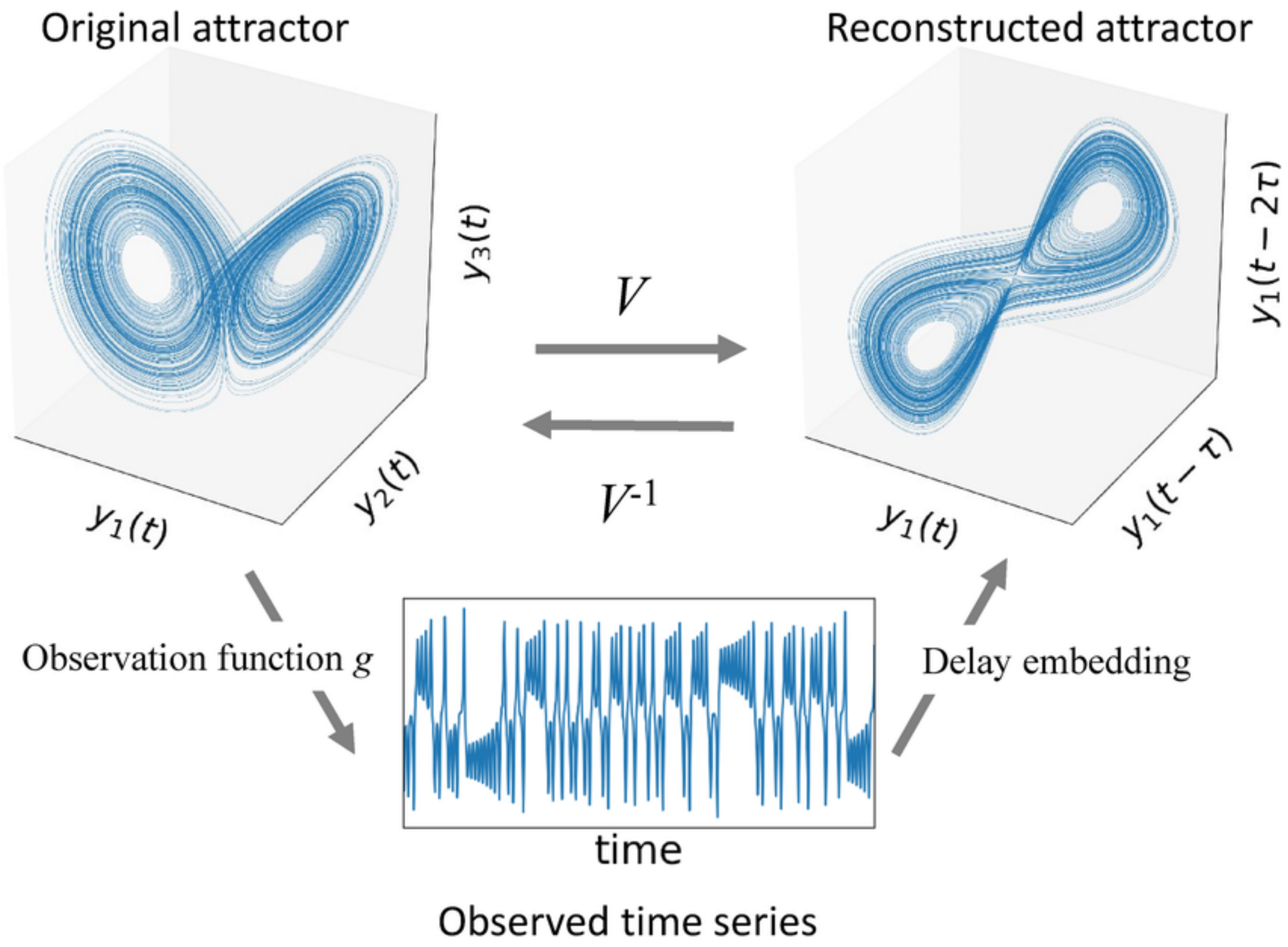
Periodicity

- non-periodic
- periodic

Applications

- medical
- industrial
- financial

Dynamical systems reconstruction



Takens, 1981

Given a time series measurements $g : \mathscr{A} \to \mathbb{R}$ from a D-dimensional attractor $\mathscr{A}$ reconstruction map

$$f : \mathbb{R} \to \mathbb{R}^d$$

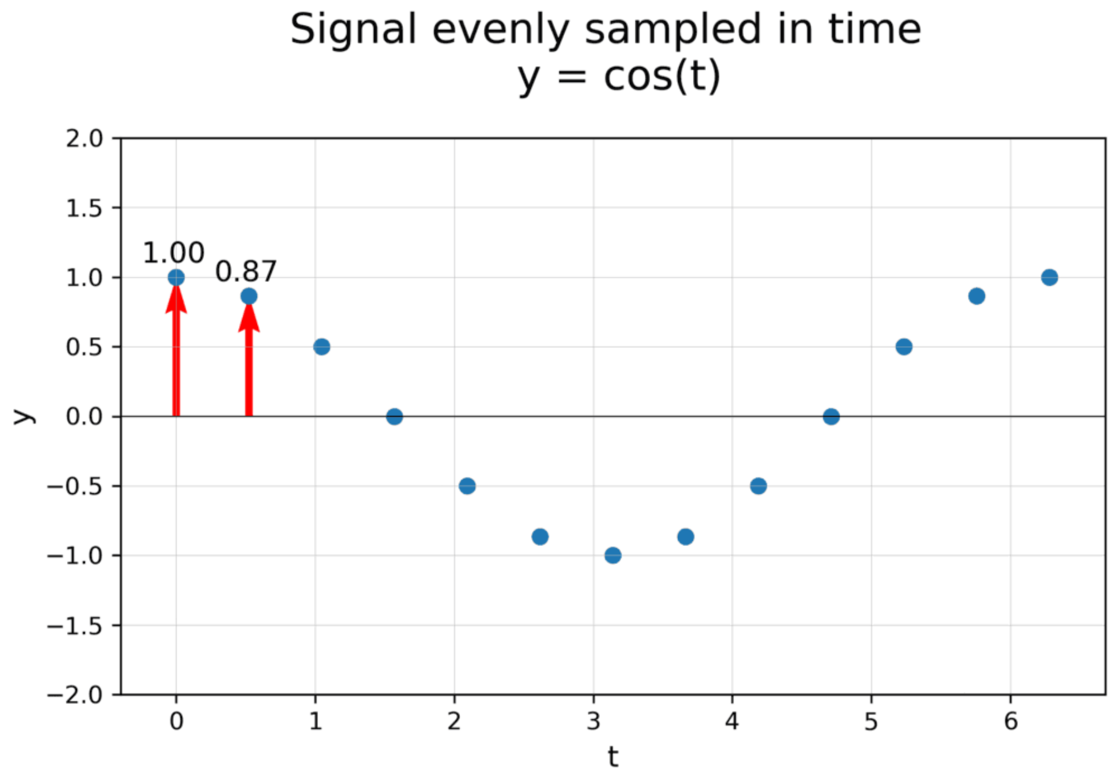
would be topology-preserving for d > 2D+1.

Dynamical Systems and Turbulence
Takens F. *(1981)*

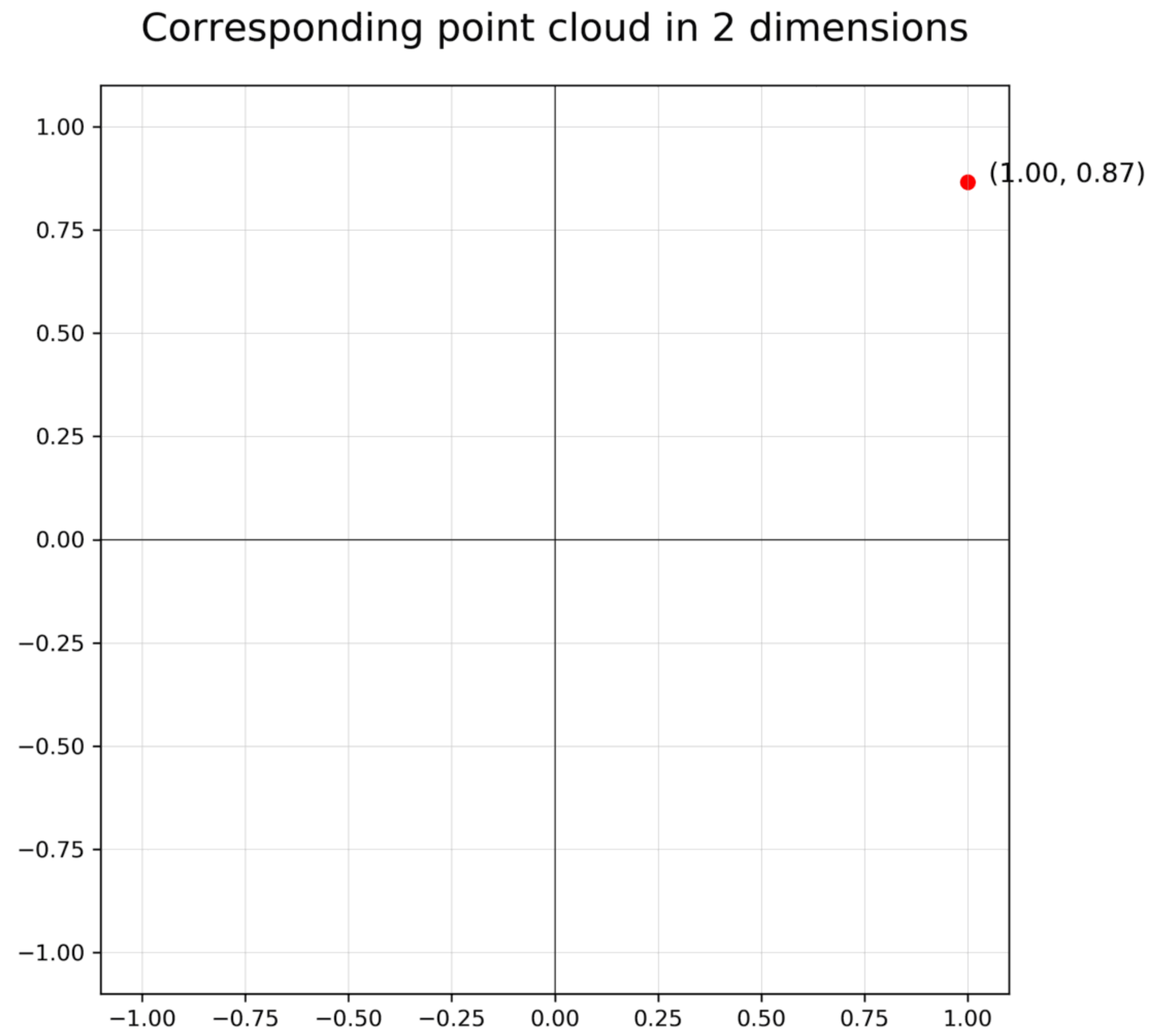
Delay embedding

Given time series $\{x_t\}_{t=1}^{N}$ a delay embedding, given embedding dimension d and lag τ is defined as

$$DX_t : \mathbb{R} \to \mathbb{R}^d \qquad x_t \mapsto (x_{t-d\tau}, \ldots, x_{t-2\tau}, x_{t-\tau})$$



Time series



Delay embedding

giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration
Tauzin et al. arXiv:2004.02551 *(2020)*

Delay embedding

Given time series $\{x_t\}_{t=1}^{N}$ a delay embedding, given embedding dimension d and lag τ is defined as

$$DX_t : \mathbb{R} \to \mathbb{R}^d \qquad\qquad x_t \mapsto (x_{t-d\tau}, \ldots, x_{t-2\tau}, x_{t-\tau})$$

Choosing τ

Optimal delay is the first minimum of the mutual information between original and delayed time series [Pereira15]

$$\tau^* = \arg\min_{\tau} I(X_t, X_{t-\tau})$$

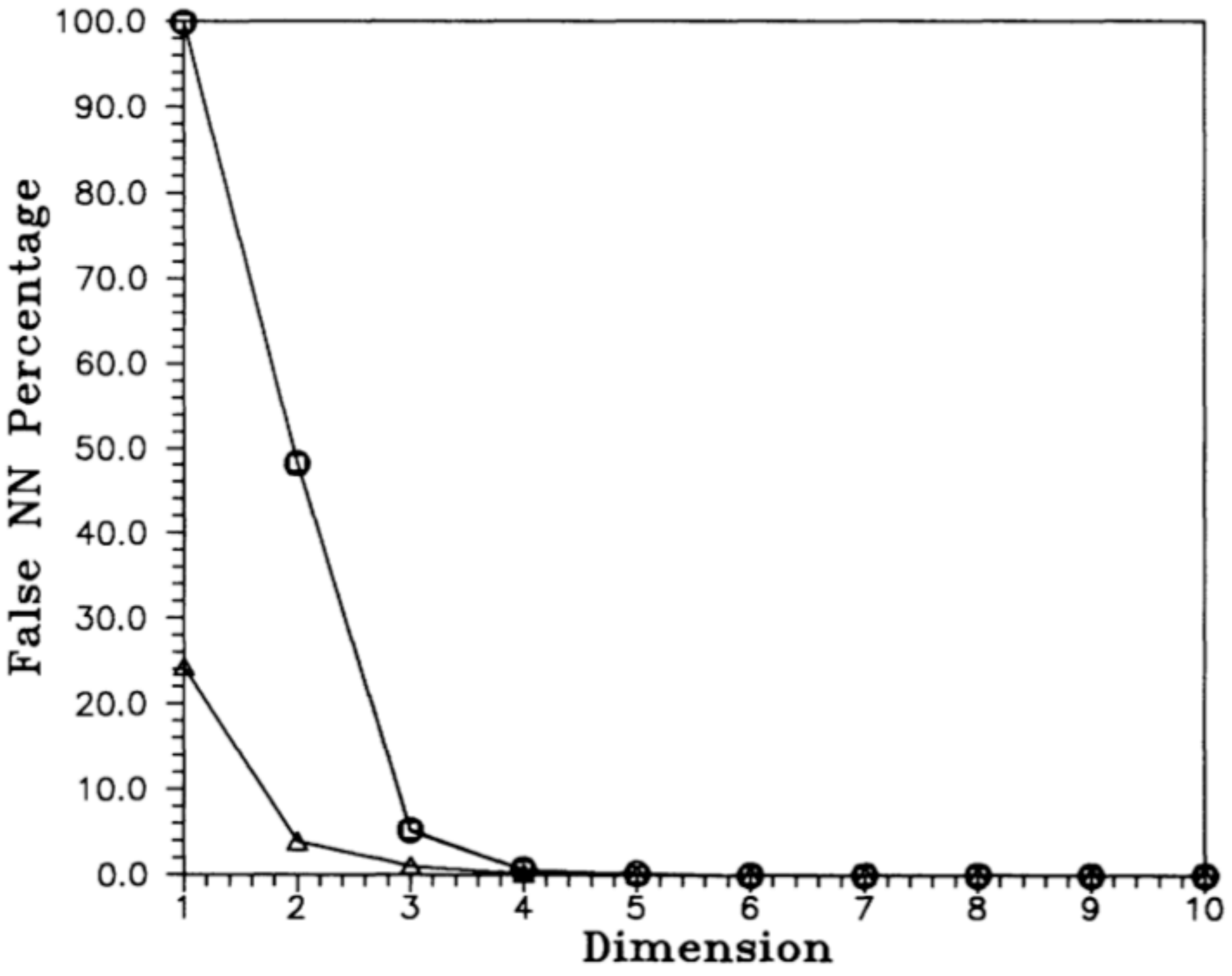
Persistent Homology for Time Series and Spatial Data Clustering
Pereira and de Mello, Expert Systems with Applications *(2015)*

Delay embedding

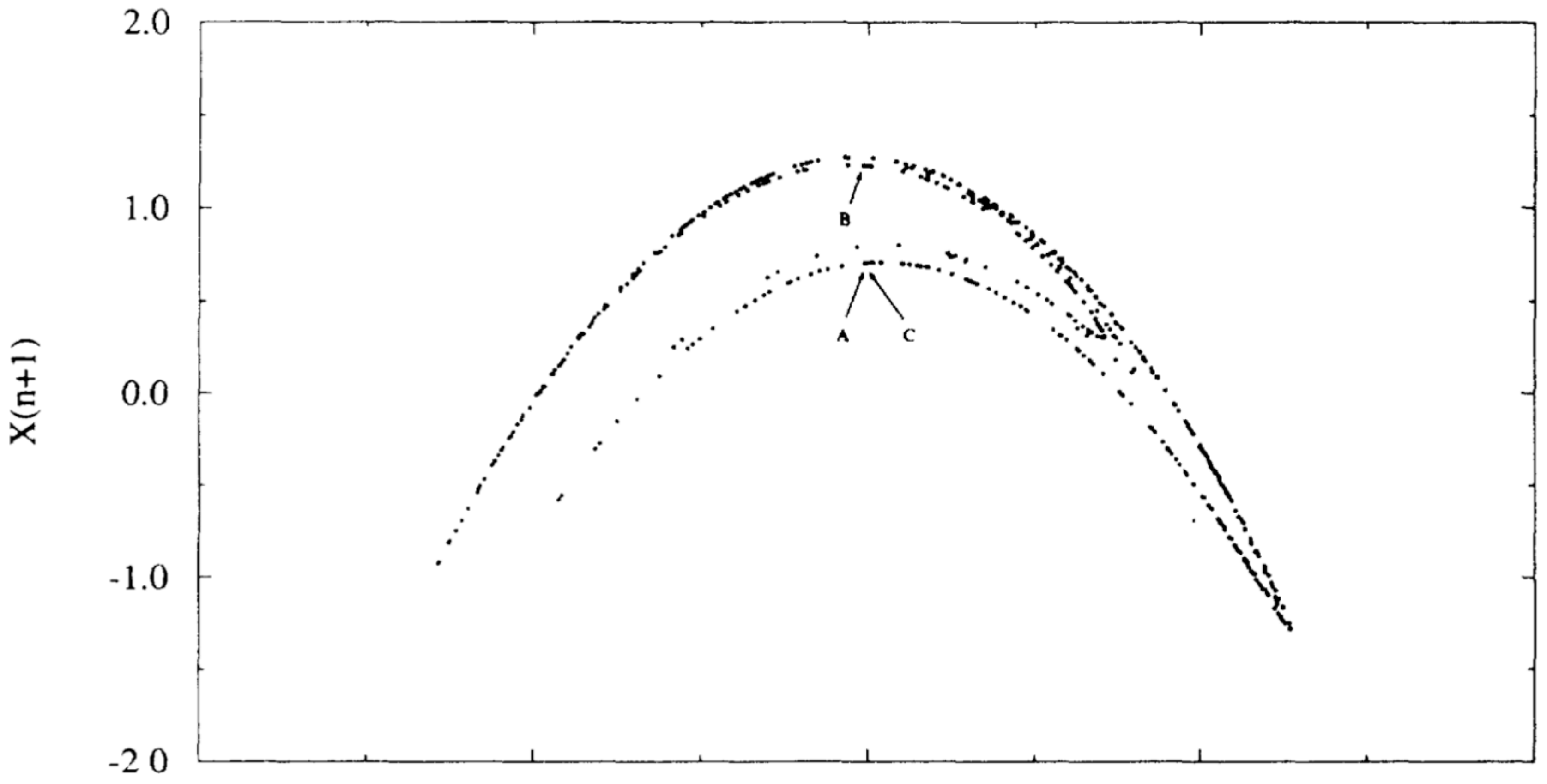
Given time series $\{x_t\}_{t=1}^{N}$ a delay embedding, given embedding dimension d and lag τ is defined as

$$DX_t : \mathbb{R} \to \mathbb{R}^d \qquad\qquad x_t \mapsto (x_{t-d\tau}, \ldots, x_{t-2\tau}, x_{t-\tau})$$

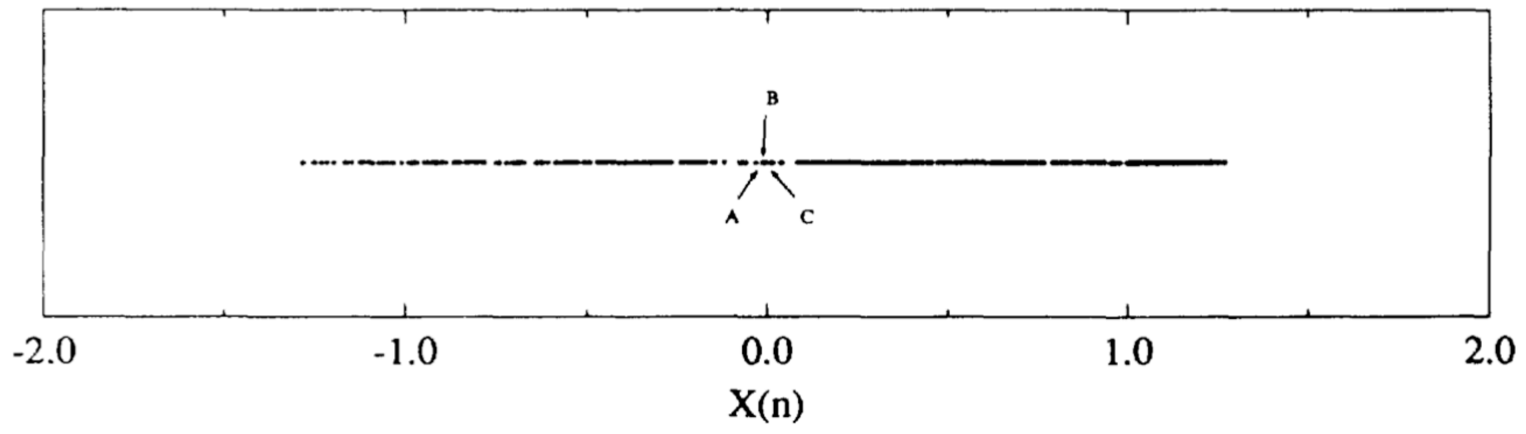
Choosing d



True neighbors in $\mathbb{R}^2$



False neighbors in $\mathbb{R}$

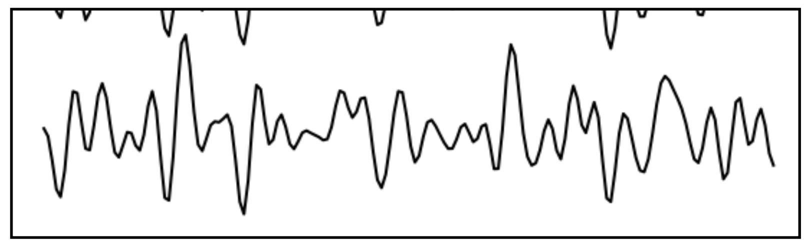


Determining Embedding Dimension for Phase-space Reconstruction using a Geometrical Construction
Kennel et al. *Physical Review A (1992)*

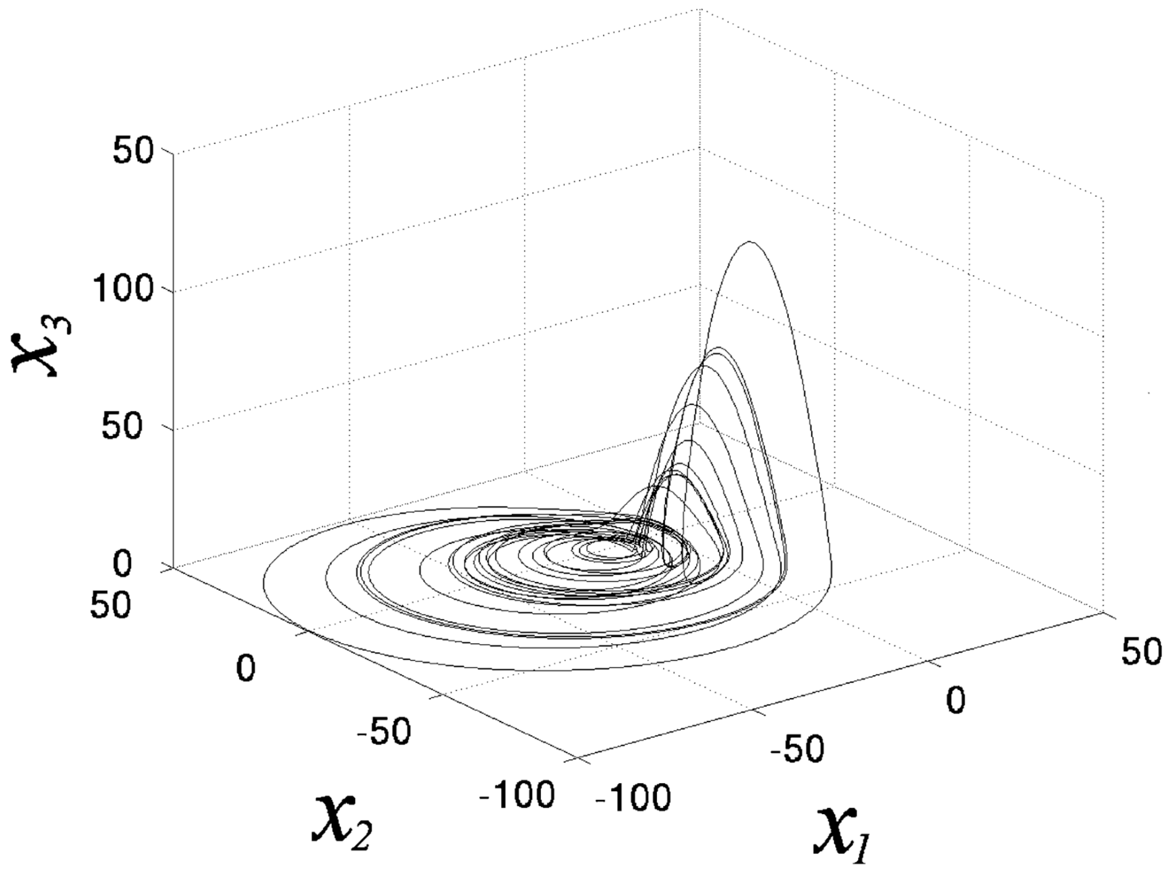
Delay embedding

Given time series $\{x_t\}_{t=1}^{N}$ a delay embedding, given embedding dimension d and lag τ is defined as

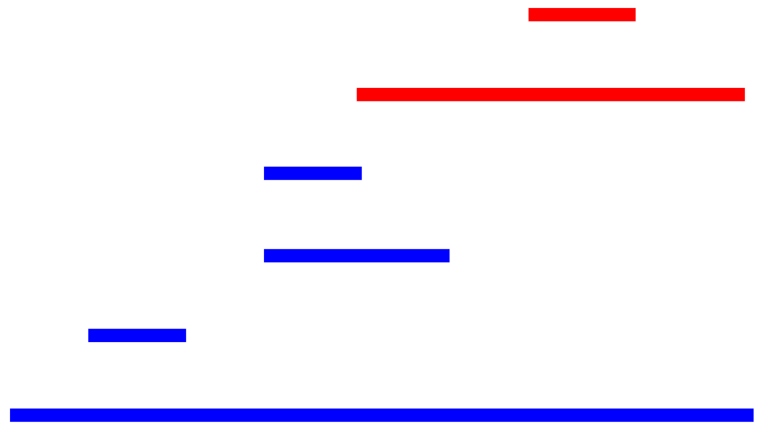
$$DX_t : \mathbb{R} \to \mathbb{R}^d \qquad\qquad x_t \mapsto (x_{t-d\tau}, \ldots, x_{t-2\tau}, x_{t-\tau})$$



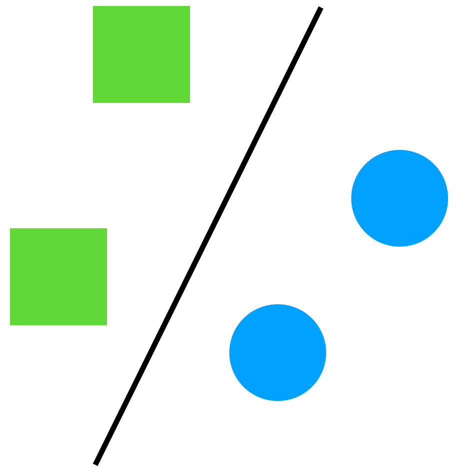
Time series



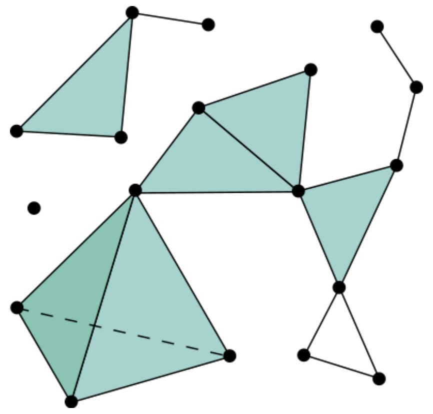
Delay embedding



Persistence diagram

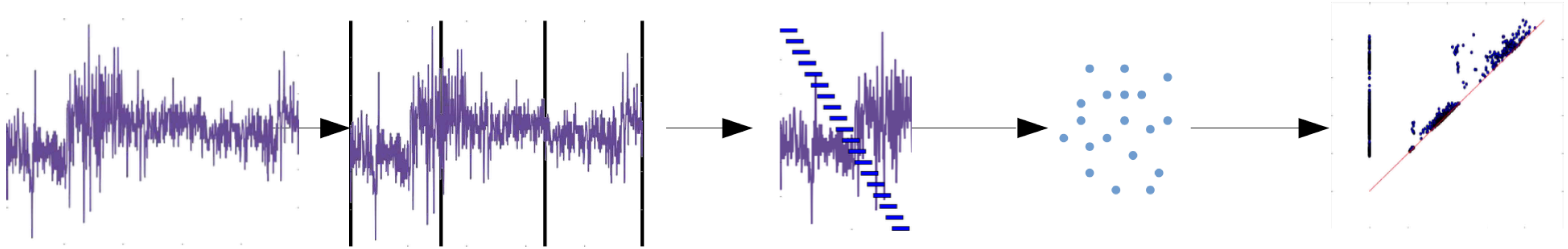


Vector representation
ML task



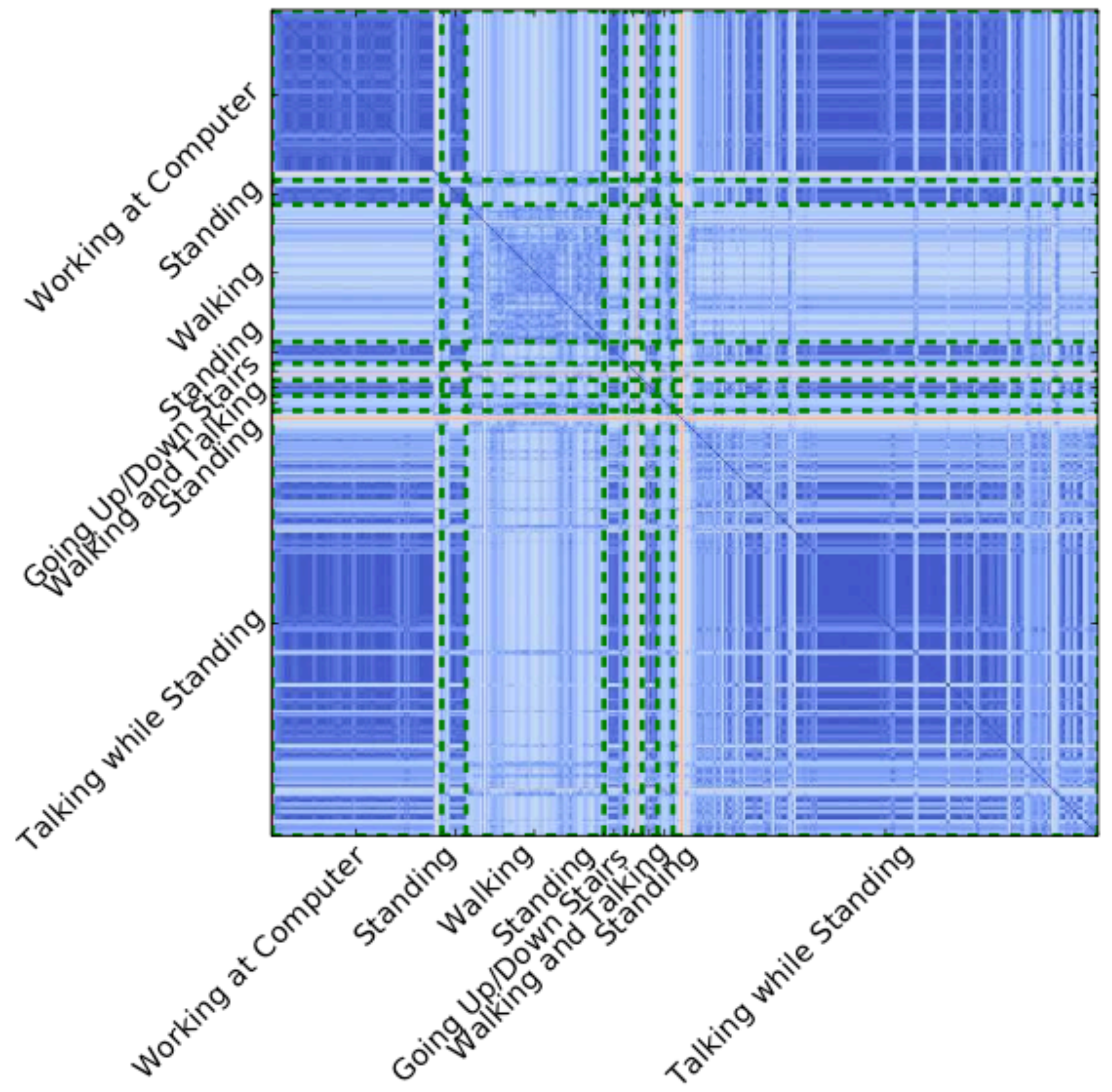
Vietoris-Rips filtration

Delay embedding



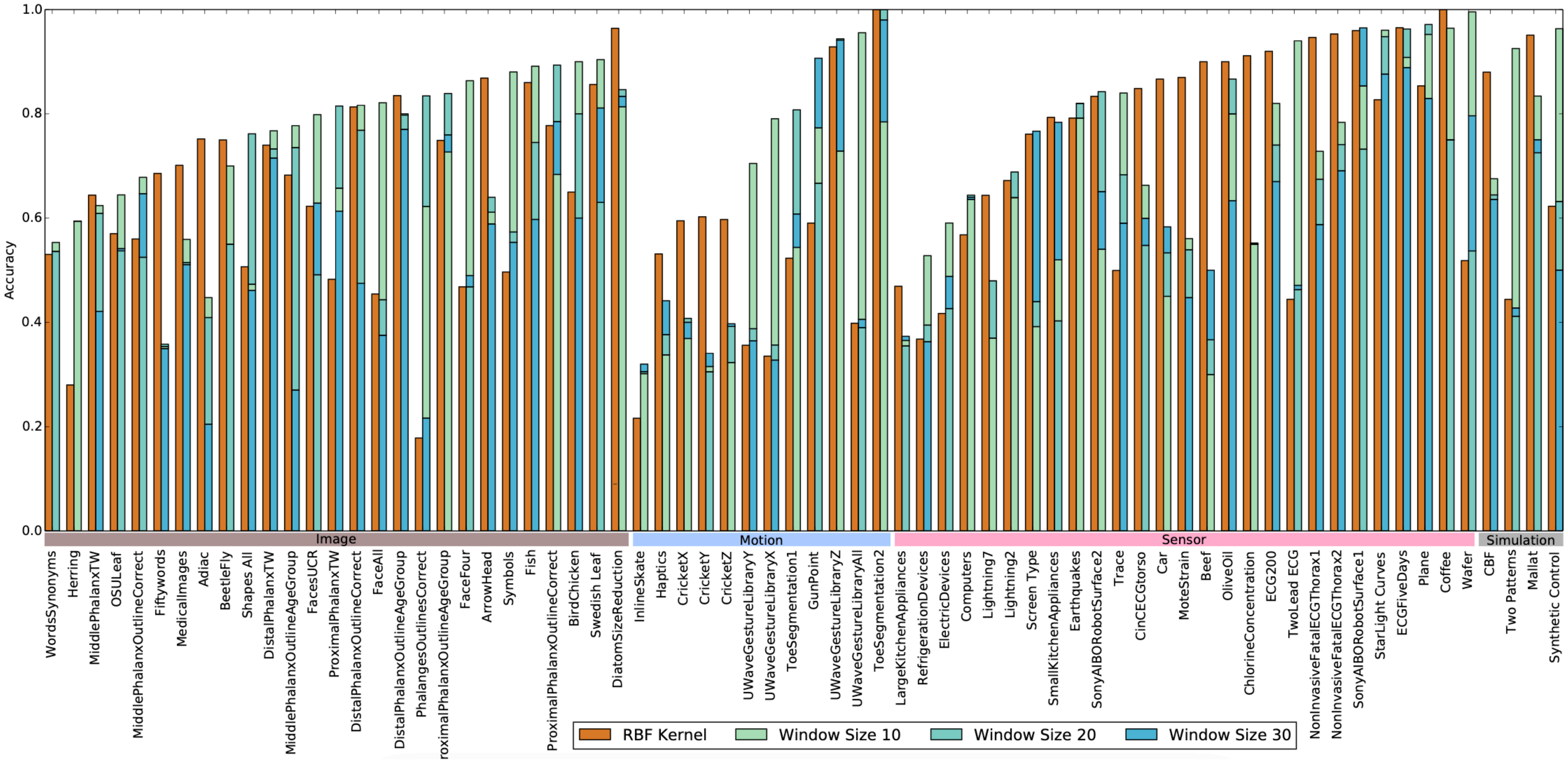
Time series | Segmentation | Delay embedding | Persistence diagram

Dataset	Records	Dim	Classes	Seg. Size	Win. Size	Features
Activity [5]	1,926,896	3	7	1000, 2000	50,100,150	$PD^{0,1}$
Walking [5]	149,322	3	22	200	15	$PD^{0,1}$
Mocap [8]	225,551*	62	2*	150	15	$PD^{0,1}$, CI
Kitchen [8]	17,726,895	9	5	1000	120	$PD^{0,1}$
EEG Eye [28]	14,980	16	2	300	20	$PD^{0,1}$
PAMAP [27]	2,872,533	52	19	1000	20,30,40,50	$PD^{0,1}$
Bird Sounds [13]	24,050,377	1	16	800	15,30,45	$PD^{0,1}$



On Time-series Topological Data Analysis: New Data and Opportunities
Seversky et al. *International Conference on Computer Vision and Pattern Recognition (2017)*

Delay embedding



Single-source Continuous Signals:

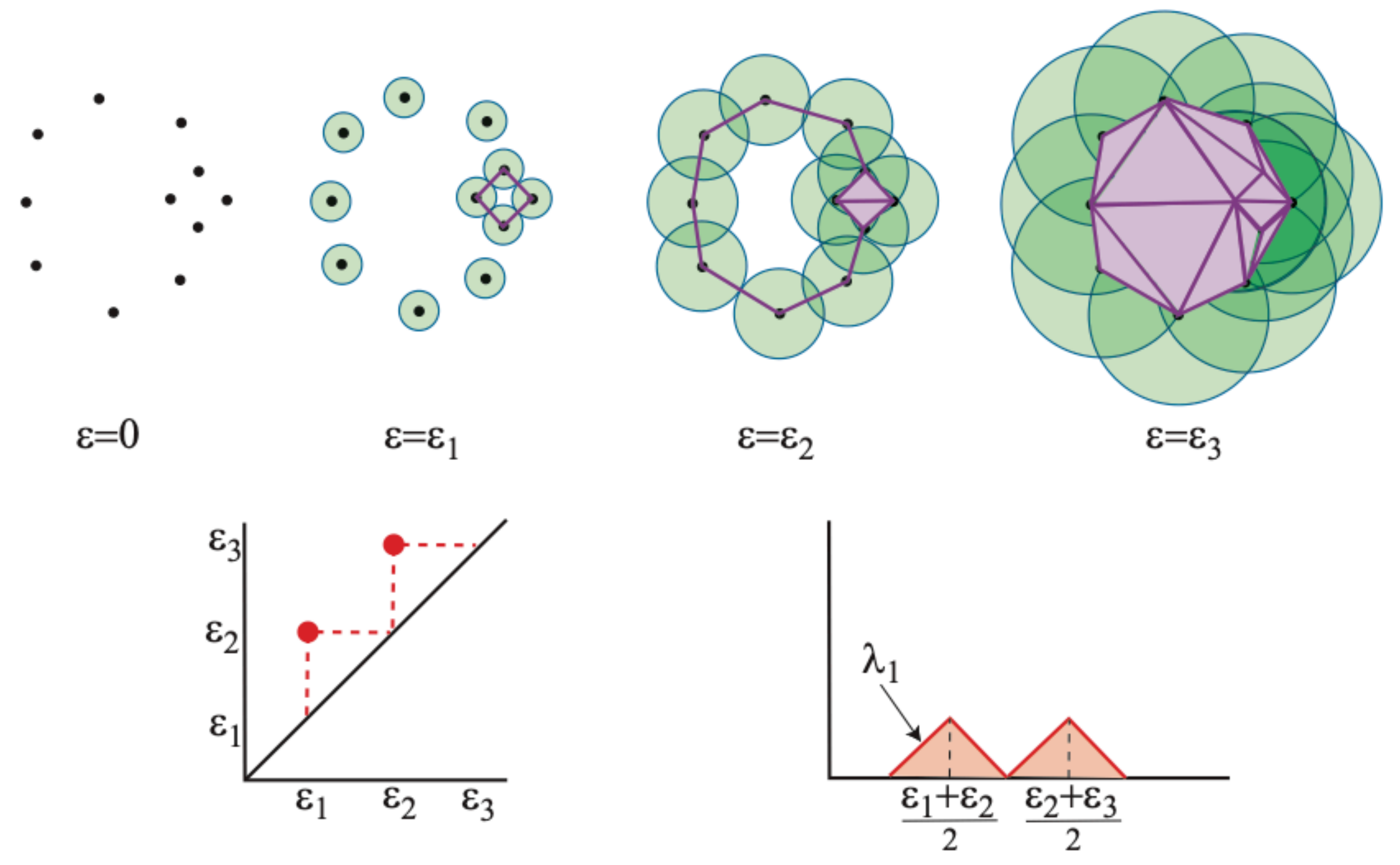
- **Activity** [5]: Three-dimensional accelerometer measurements capturing movements for 15 human participants and 7 daily activities.
- **Walking** [5]: Three-dimensional accelerometer measurements of 22 people walking along predefined and free-form paths.
- **MOCAP** [8]: Motion capture data capturing multiple types of human motion.
- **Kitchen** [8]: Motion capture data capturing multiple prescribed kitchen tasks.
- **EEG** [28]: EEG sensor data capturing eye state.
- **PAMAP** [27]: Physical Activity Monitoring for Aging People: human motion measurements for 19 different activity types.
- **Bird Sounds** [13]: Audio recordings of different bird species with multiple recordings per species type.

Multi-source Signals:

- **UCR** [6]: 47 time-series datasets broadly categorized by type: Image, Motion, Sensor, & Simulation. The segment size is set to the signal length and the window size is varied. The same features, distances, and kernels are computed, as outlined in Table 1.

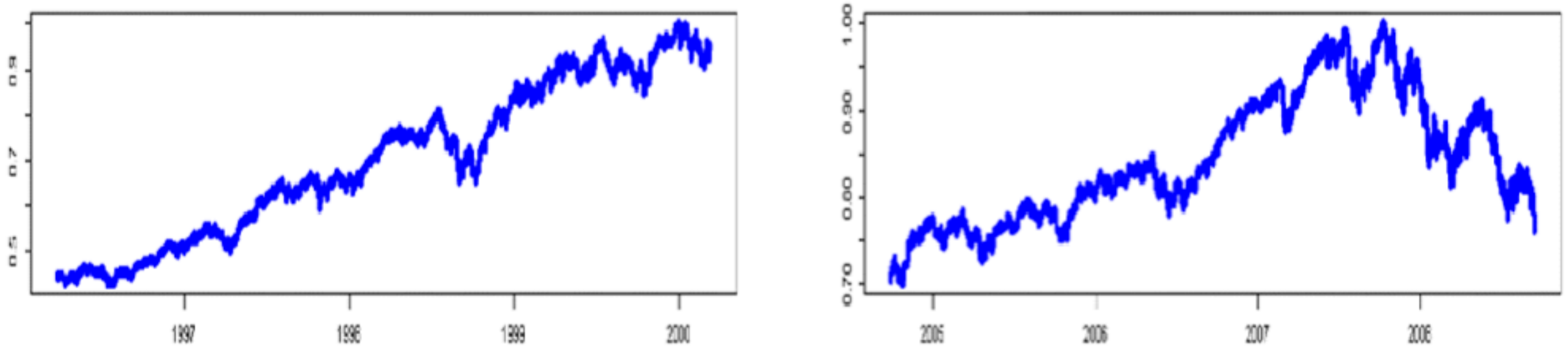
On Time-series Topological Data Analysis: New Data and Opportunities
Seversky et al. *International Conference on Computer Vision and Pattern Recognition (2017)*

Delay embedding

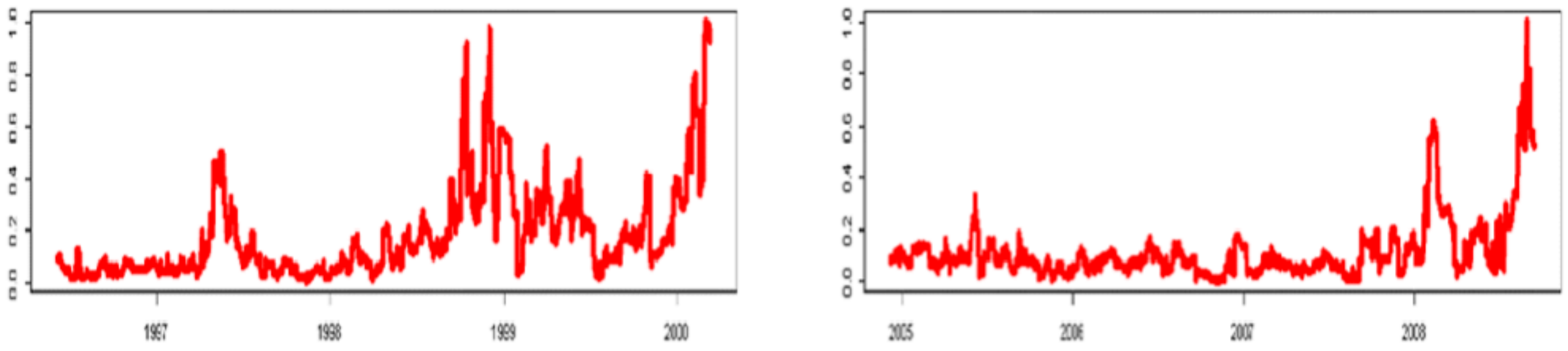


Persistence landscapes

Dotcom crash (2000) **Lehman Brothers crash (2008)**



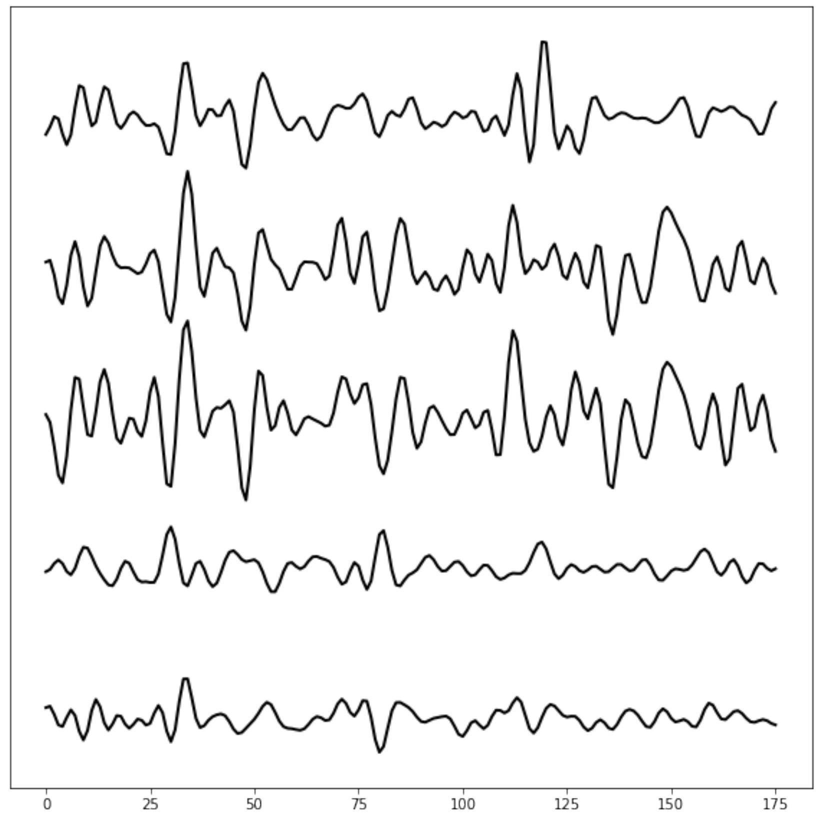
S&P 500



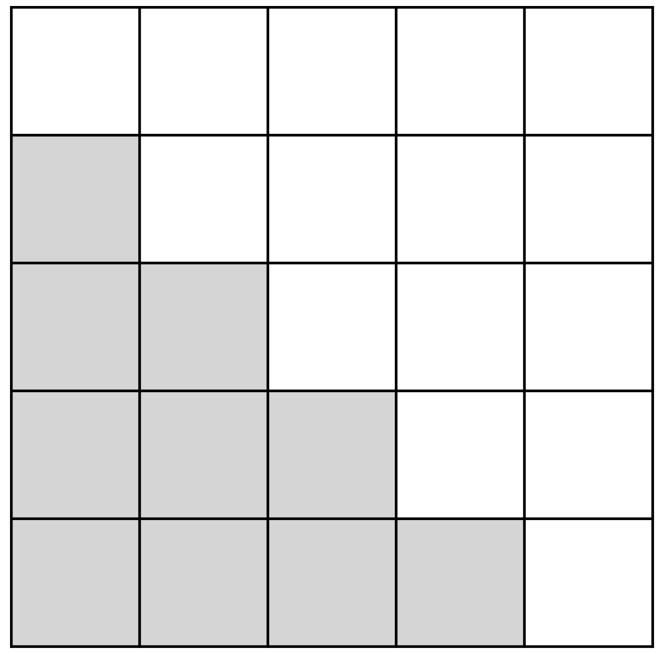
L1 norm of persistence landscapes, d = 50

Topological Data Analysis of Financial Time Series: Landscapes of Crashes
Gidea et al. *Psysica A (2017)*

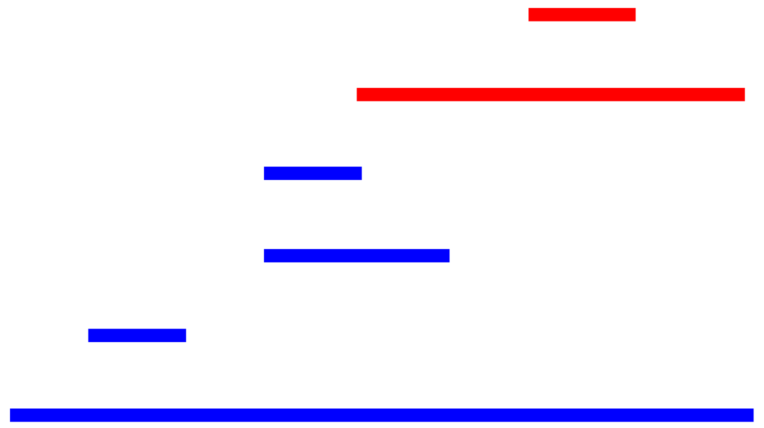
Similarity matrix filtration



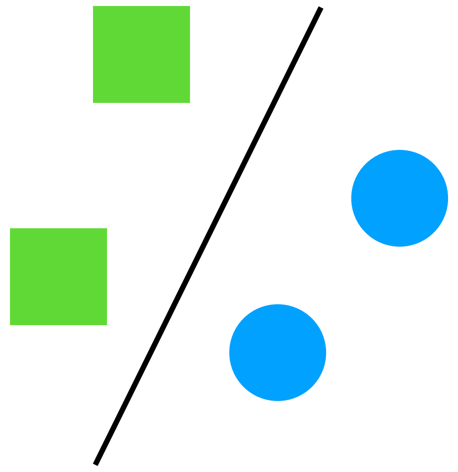
Time series



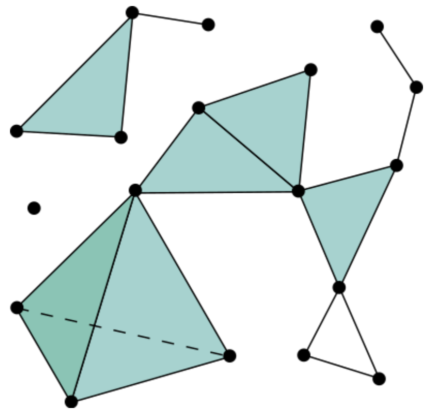
Similarity matrix



Persistence diagram



Vector representation
ML task

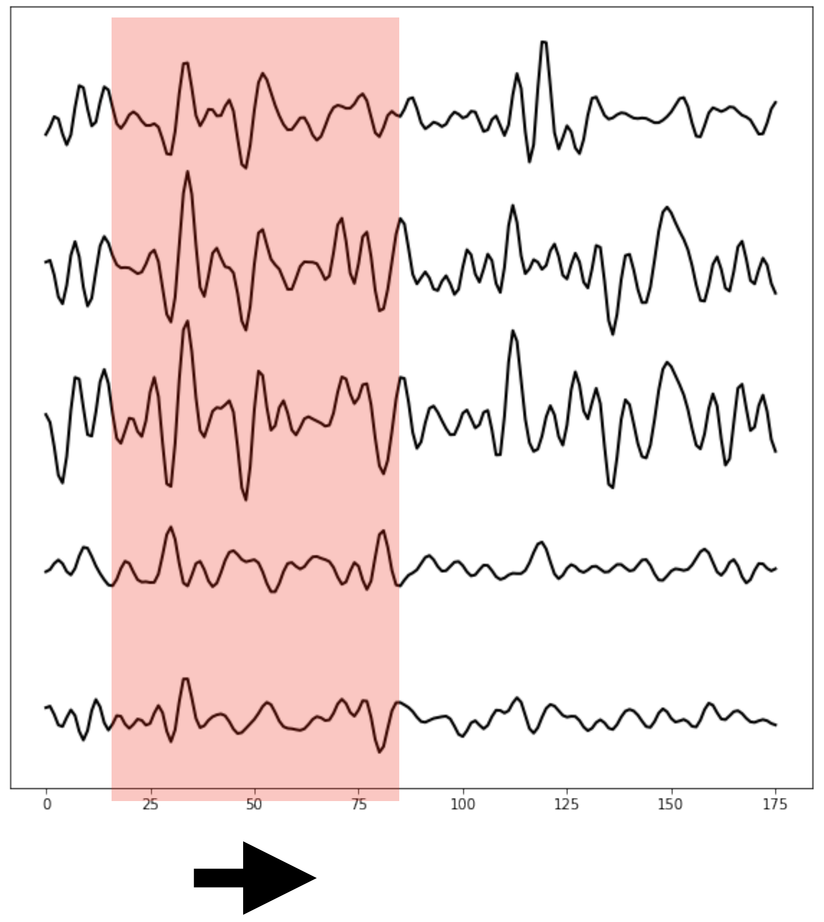


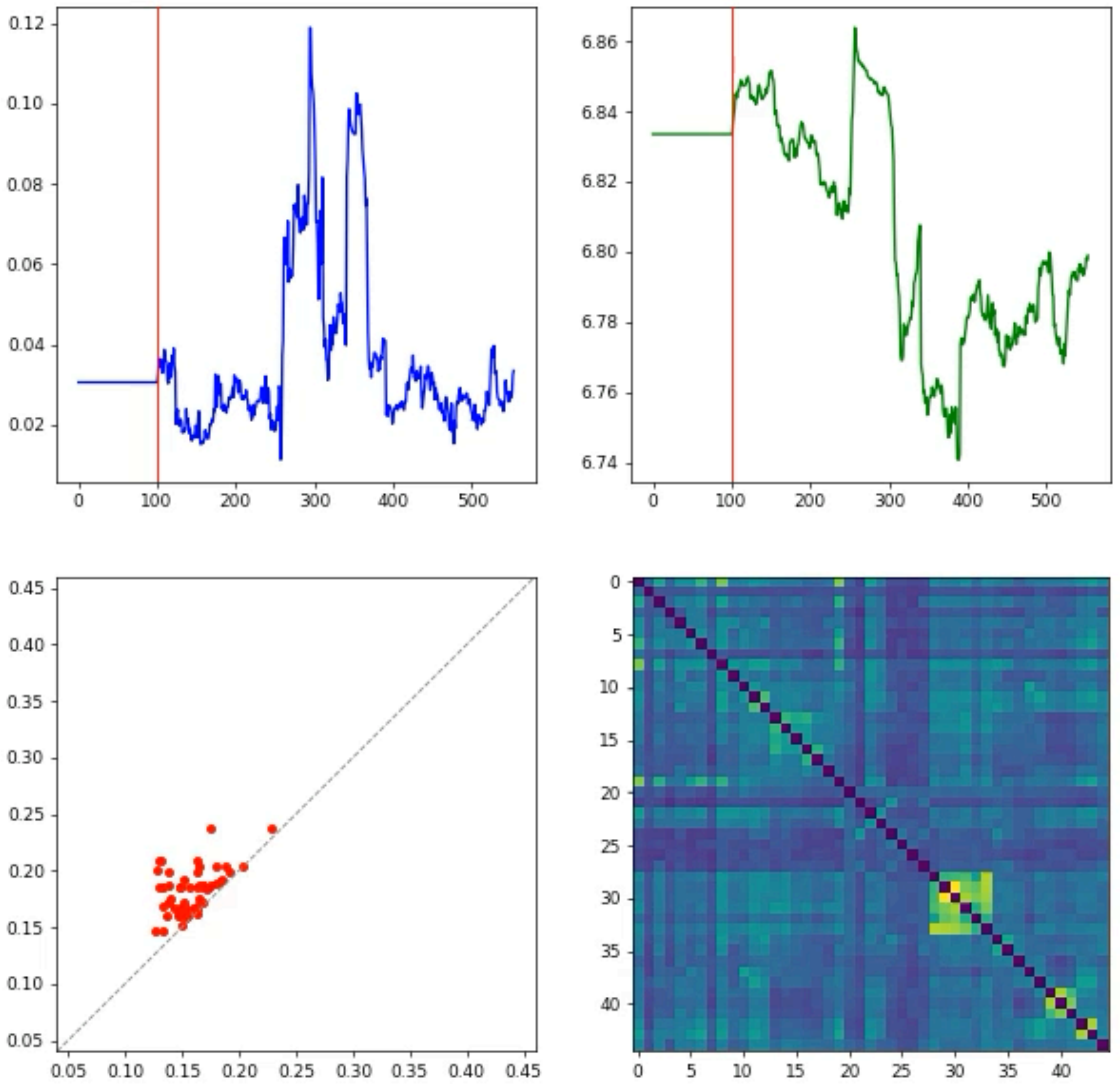
Vietoris-Rips filtration

$$f(\sigma) = \begin{cases} 0, & \dim(\sigma) = 0, \\ a_{ij}, & \dim(\sigma) = 1, \\ \max(a_{ij}, a_{ik}, a_{jk}), & \dim(\sigma) \geq 2. \end{cases}$$

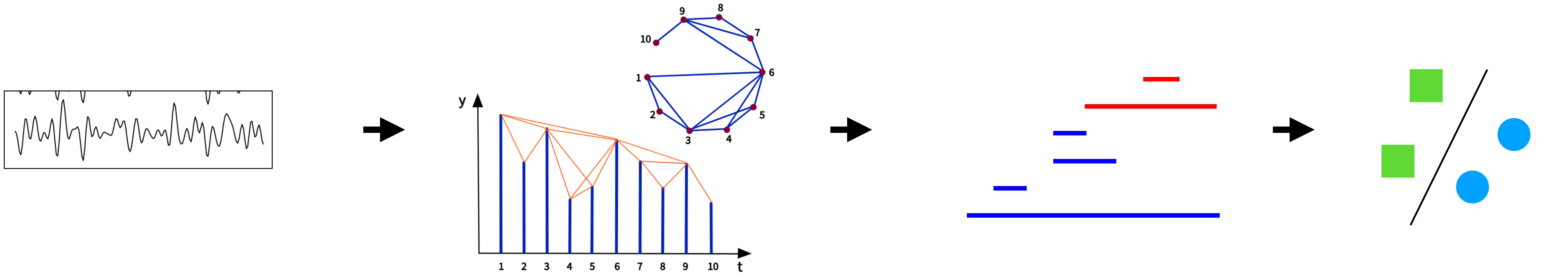
Similarity matrix filtration

Dynamics





Visibility graph



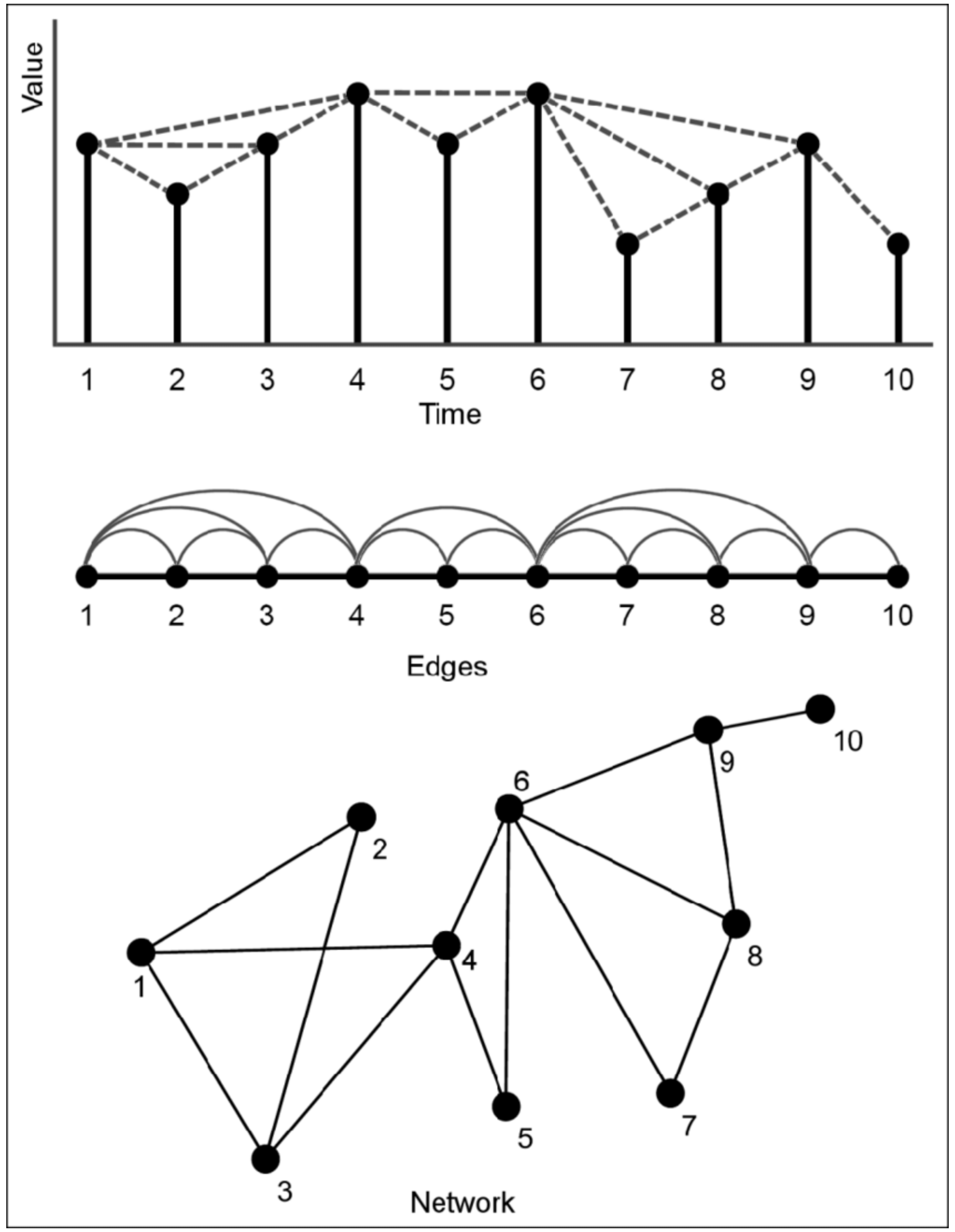
Time series

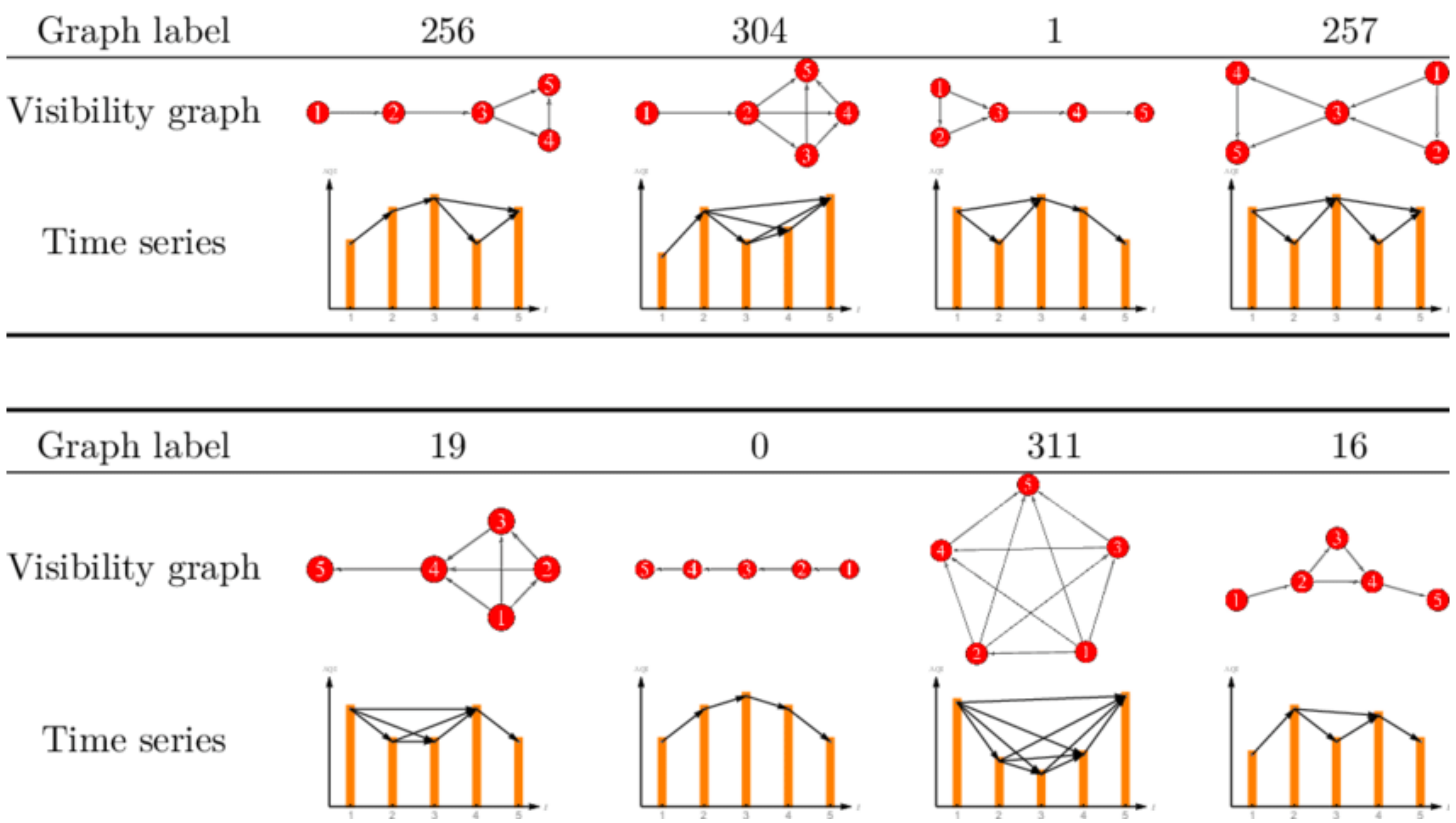
Visibility graph

Persistence diagram

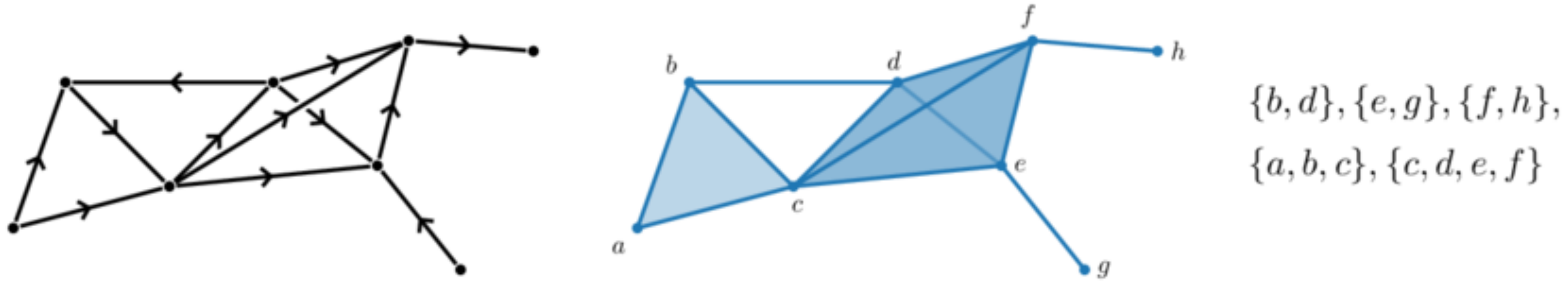
Vector representation
ML task

Visibility graph



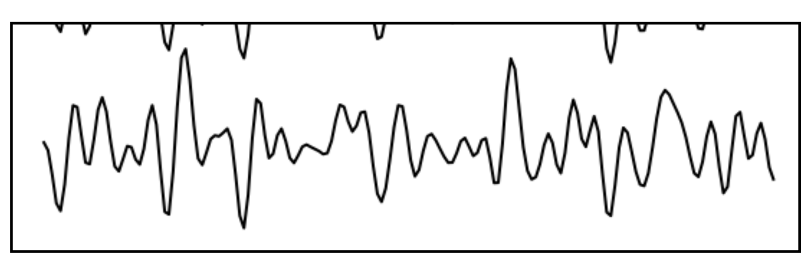


Directed clique complex [Luetgehetmann2019]

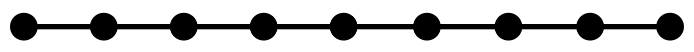


Computing Persistent Homology of Directed Flag Complexes
Luetgehetmann et al. *arXiv:1906.10458 (2019)*

Lower star filtration



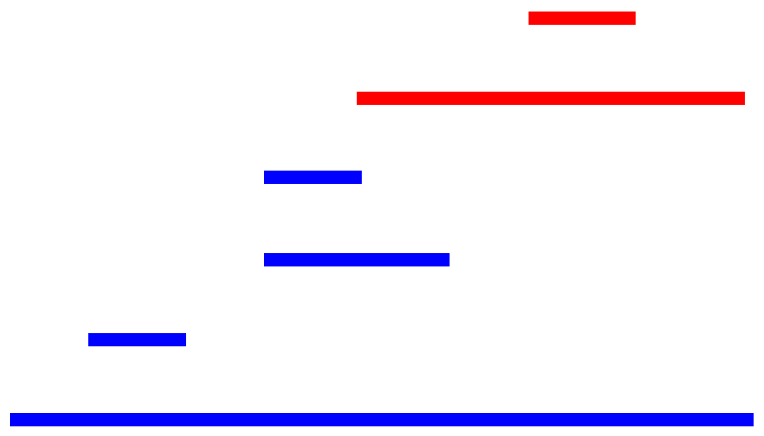
Time series



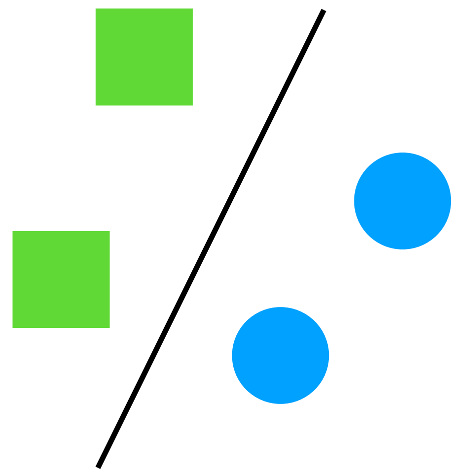
Linear graph

$$f(\sigma) = \begin{cases} x_t, & \dim(\sigma) = 0, \\ \max(x_t, x_{t+1}), & \dim(\sigma) = 1. \end{cases}$$

Lower star filtration



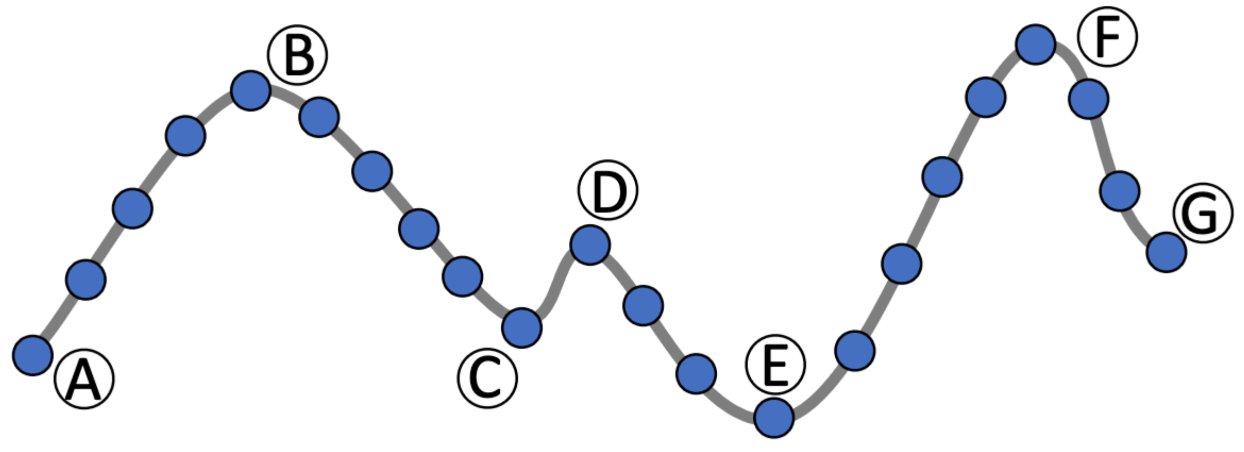
Persistence diagram



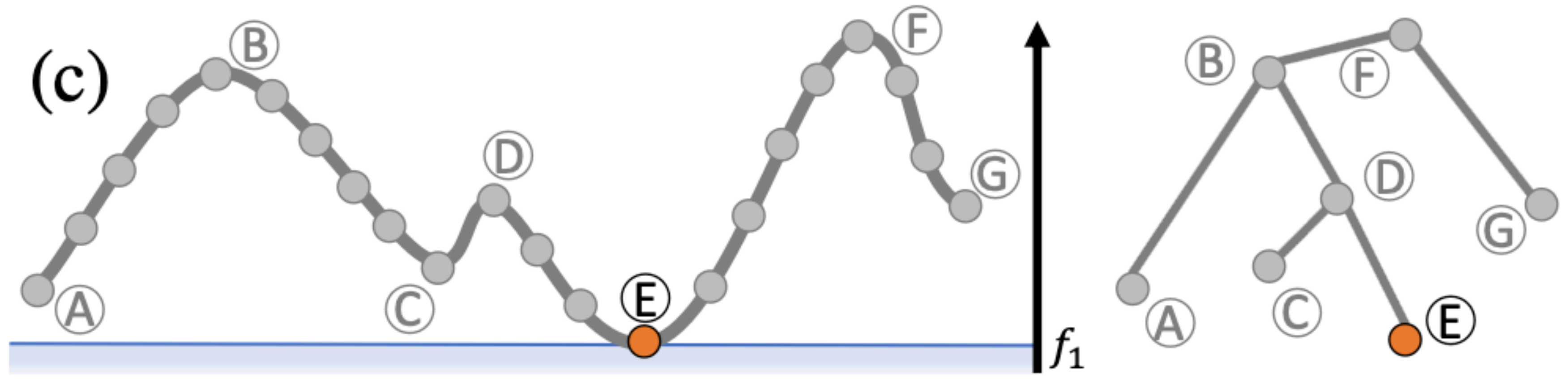
Vector representation

ML task

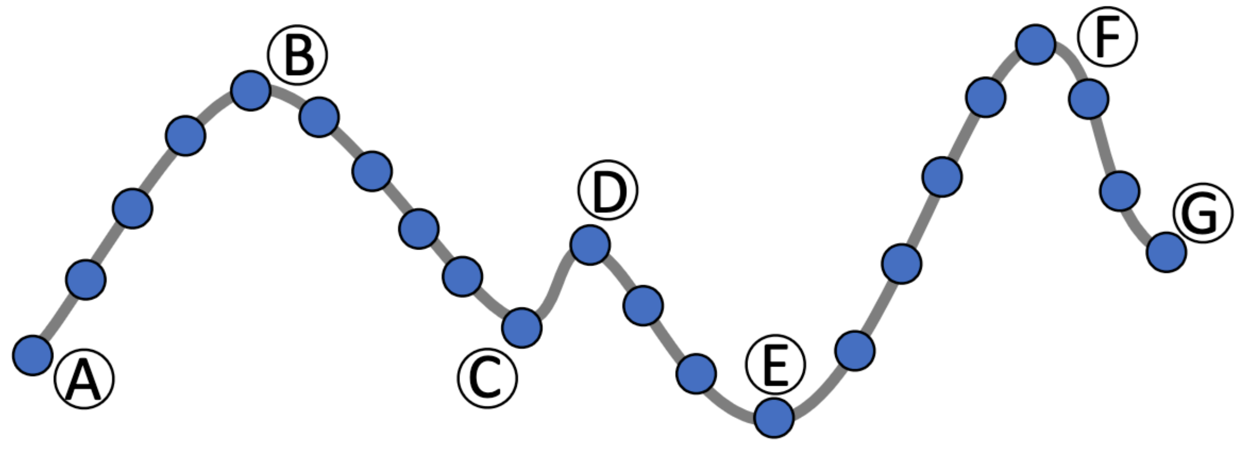
Lower star filtration



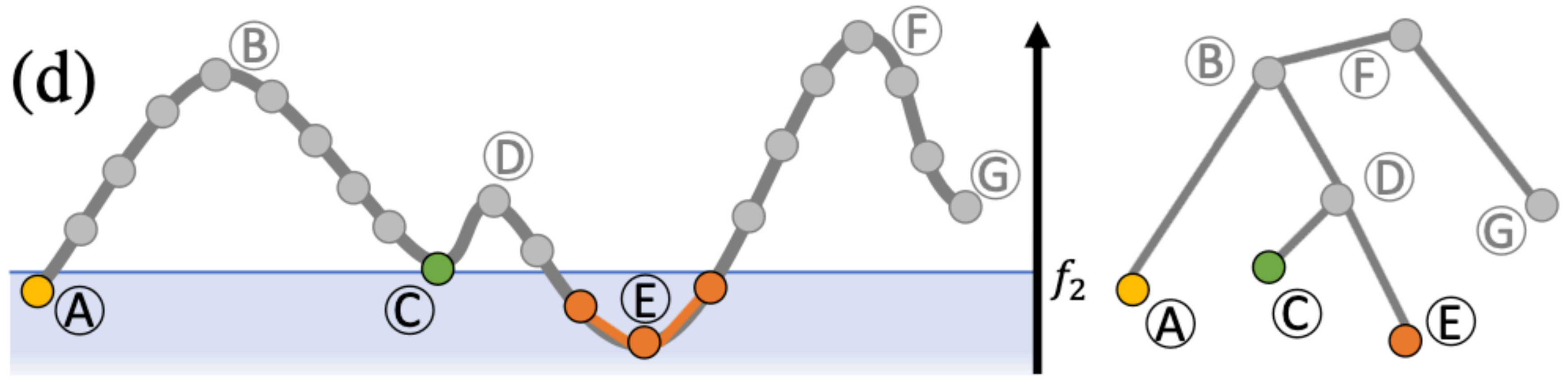
Time series



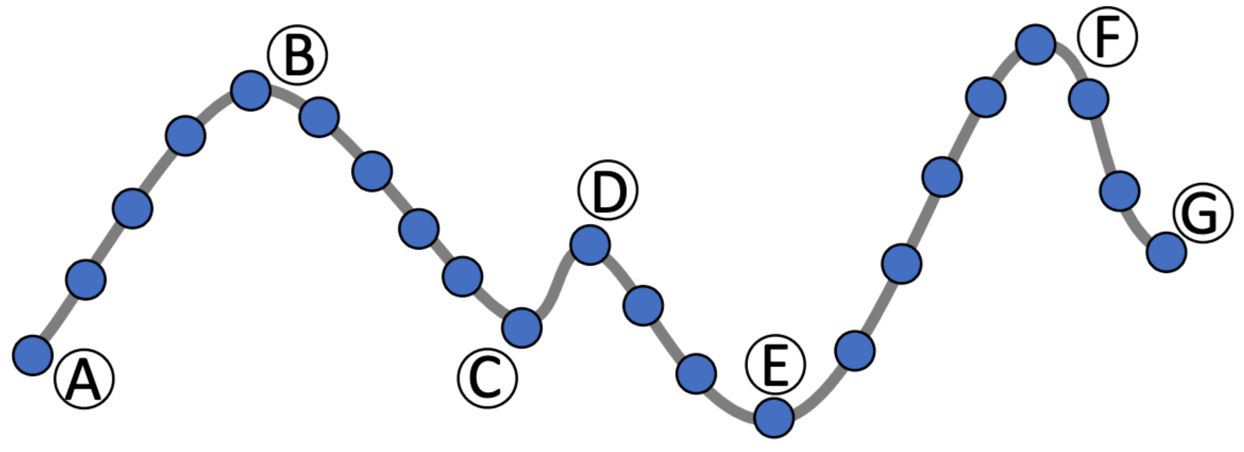
Lower star filtration



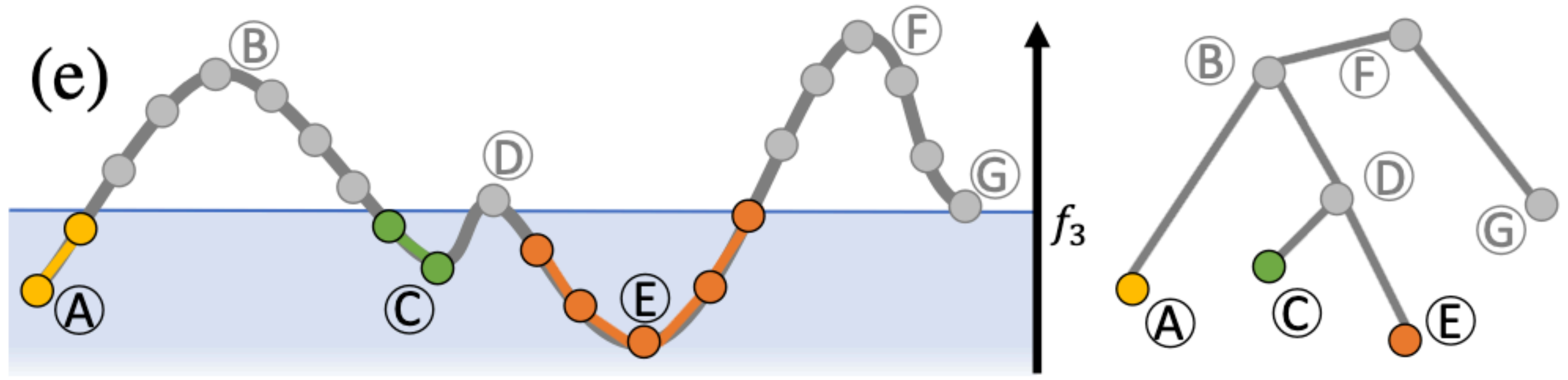
Time series



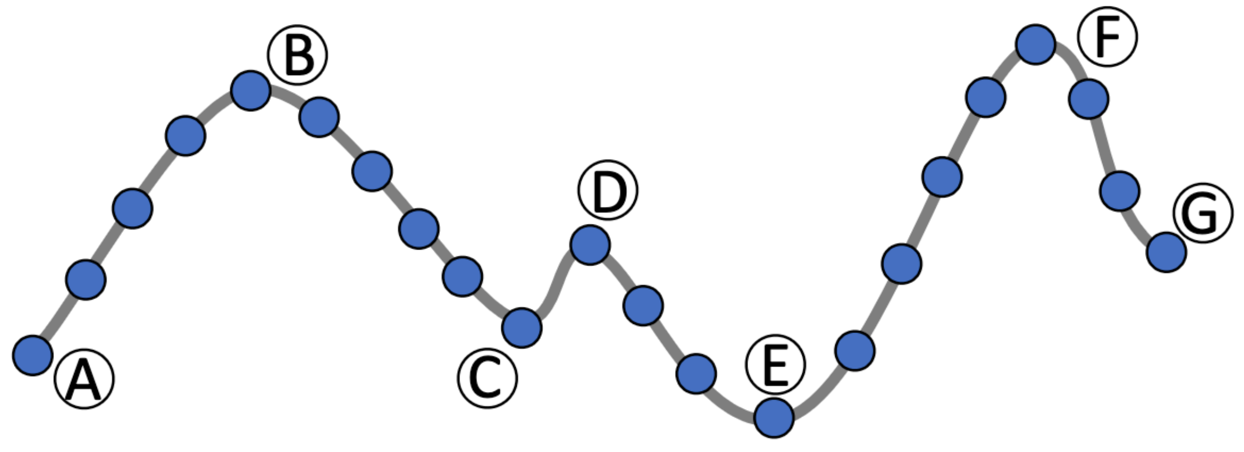
Lower star filtration



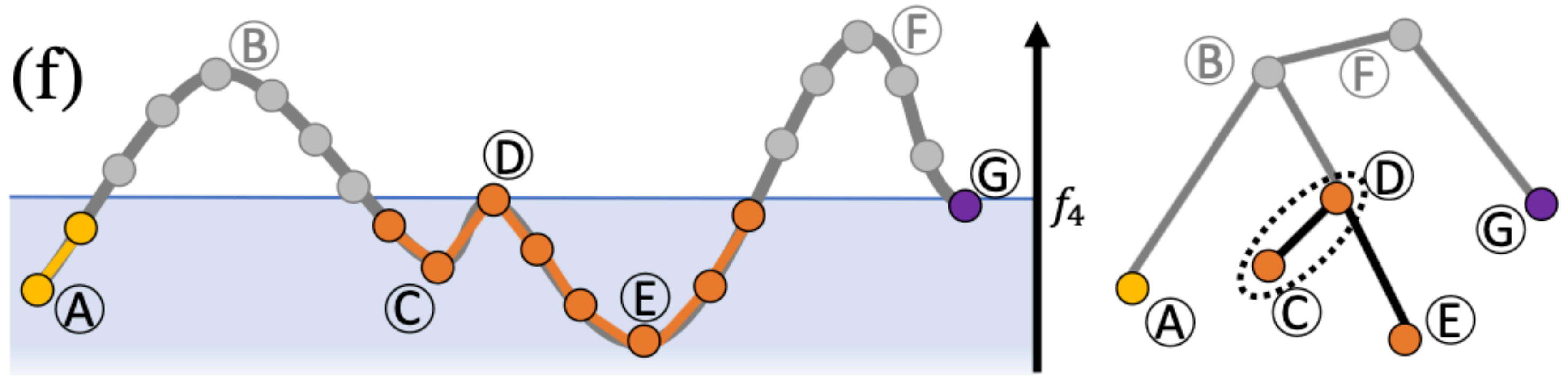
Time series



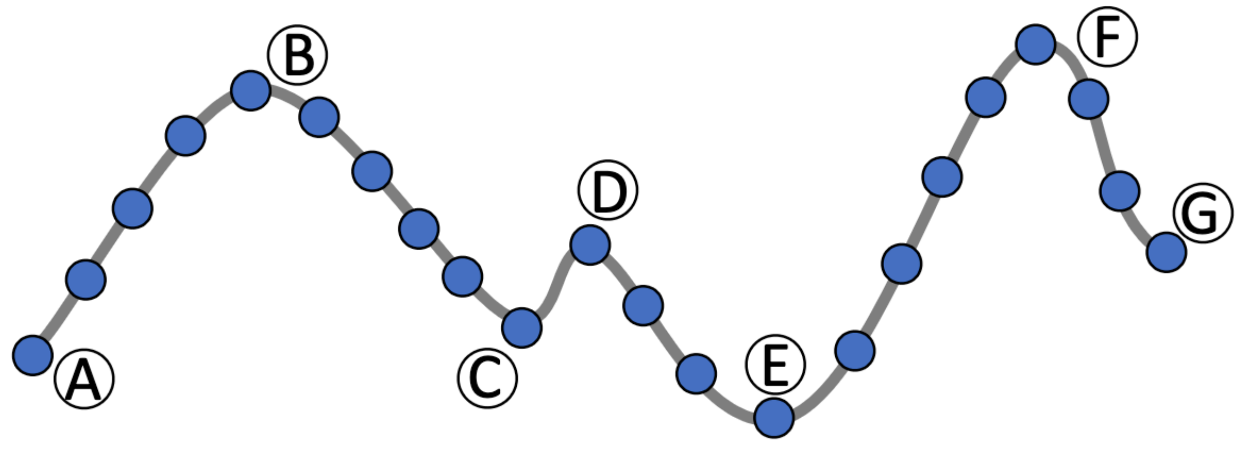
Lower star filtration



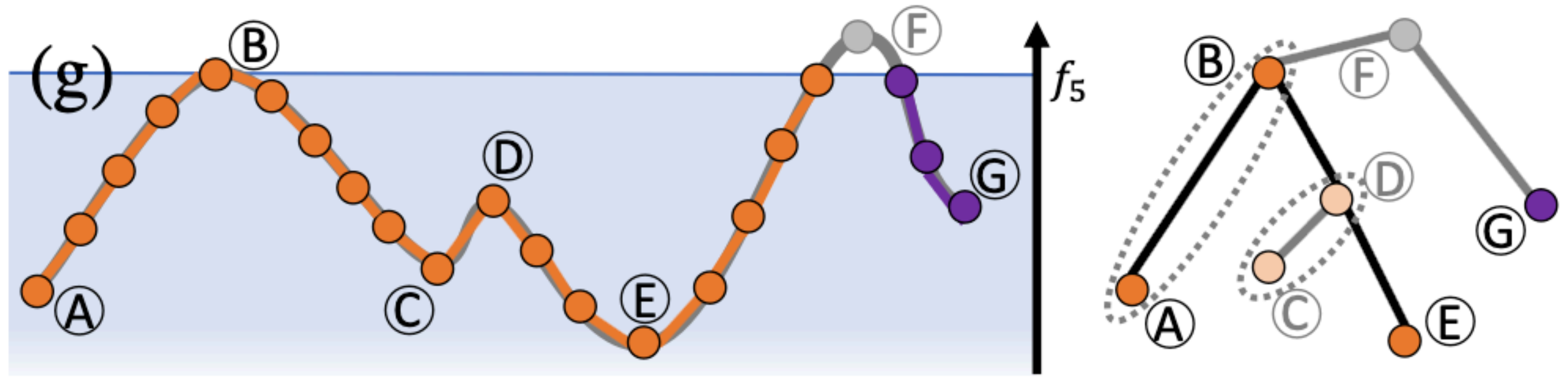
Time series



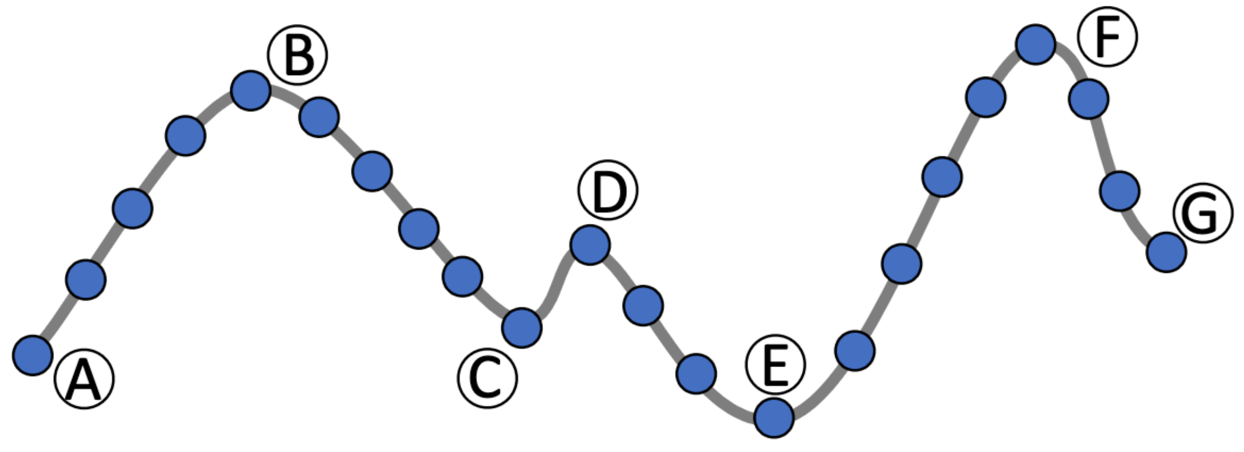
Lower star filtration



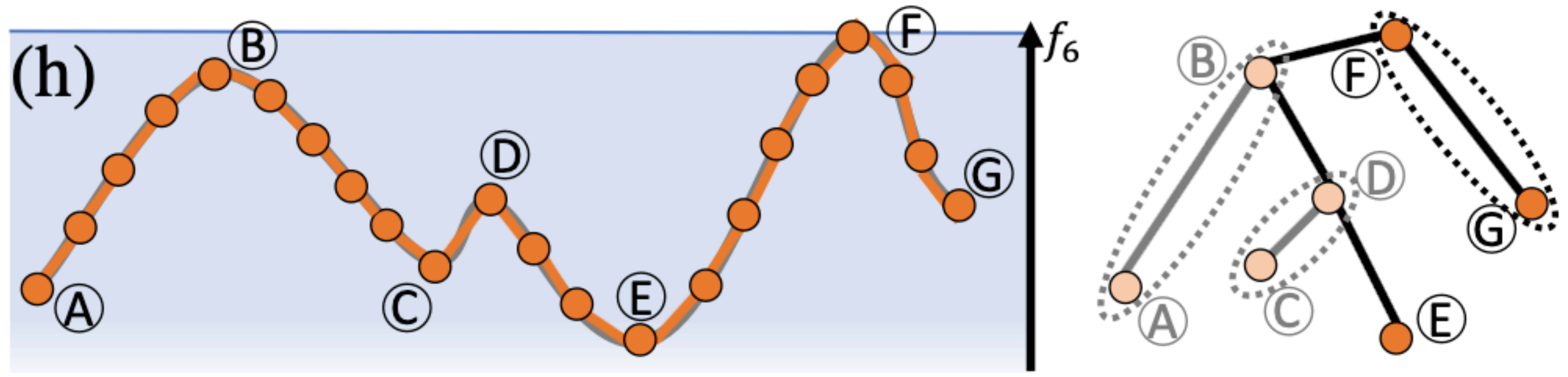
Time series



Lower star filtration



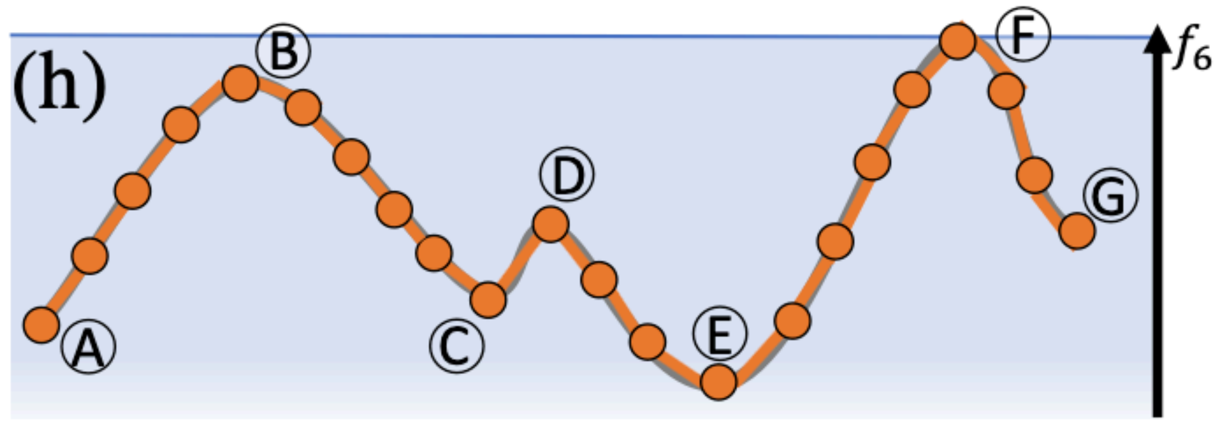
Time series



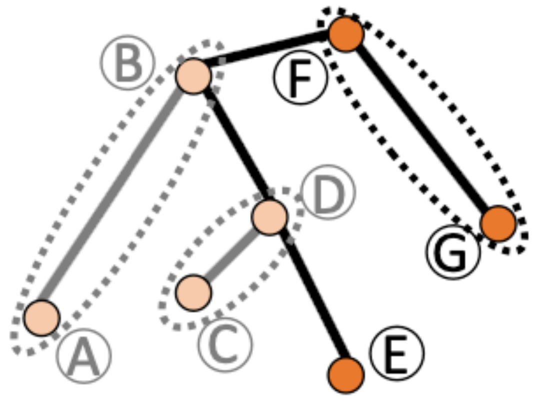
Persistence pairing

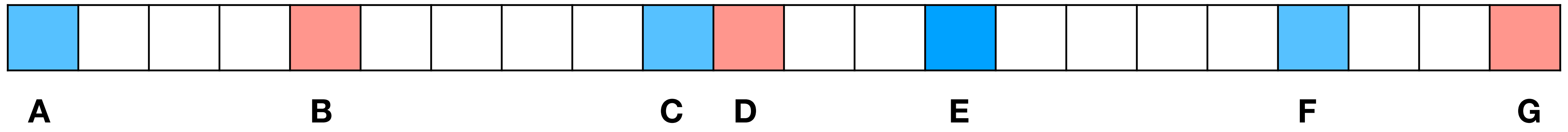
Lower star filtration

Time series

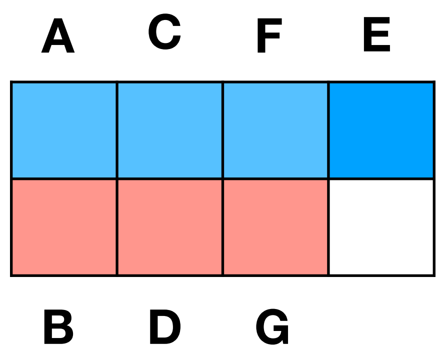


Persistence diagram





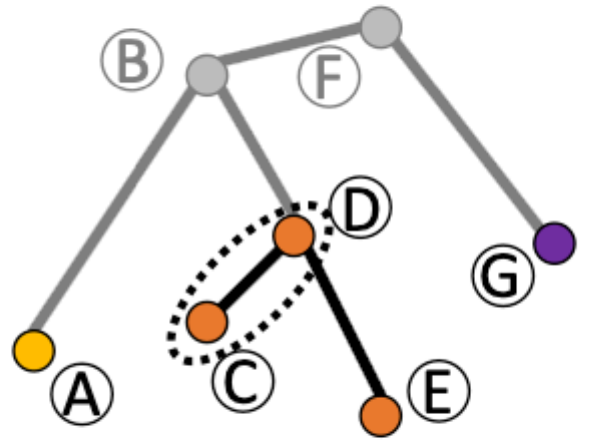
Birth indices



Death indices

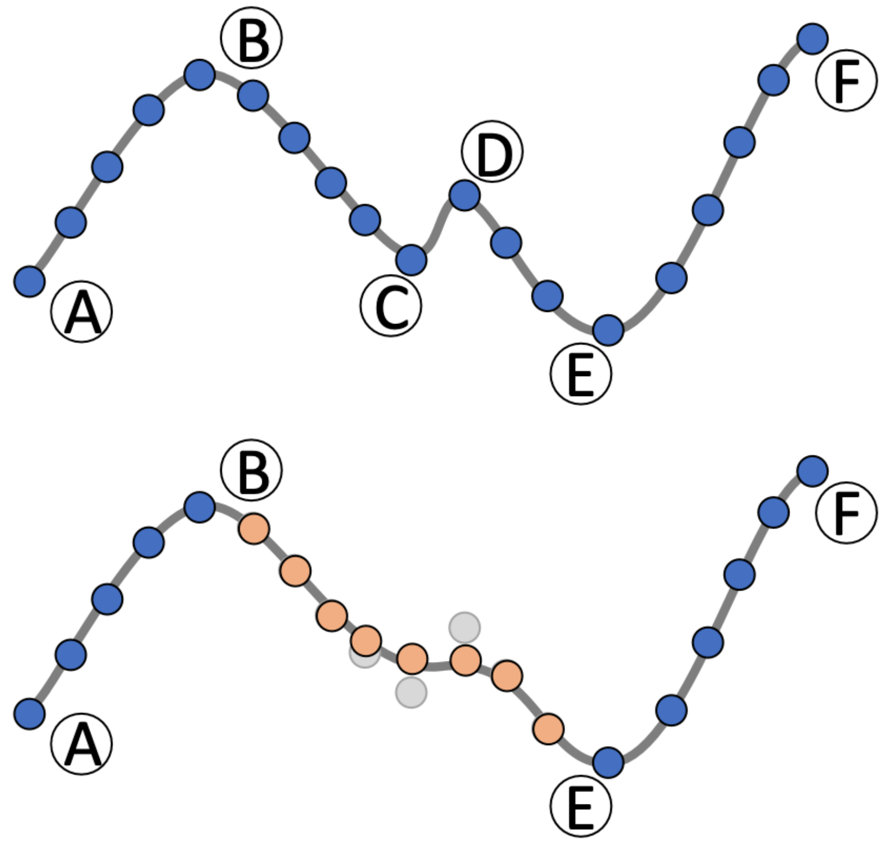
Lower star filtration

Persistence merge tree

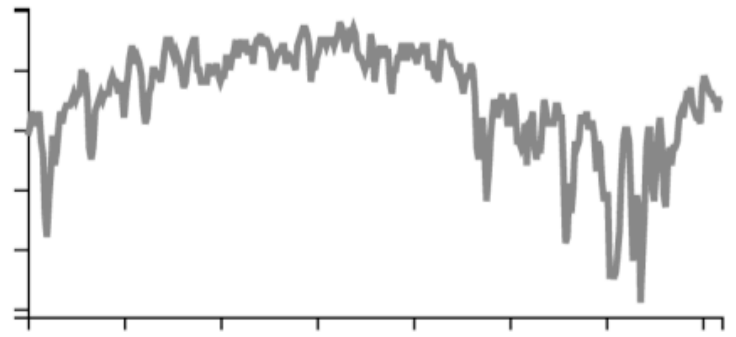


Topological smoothing

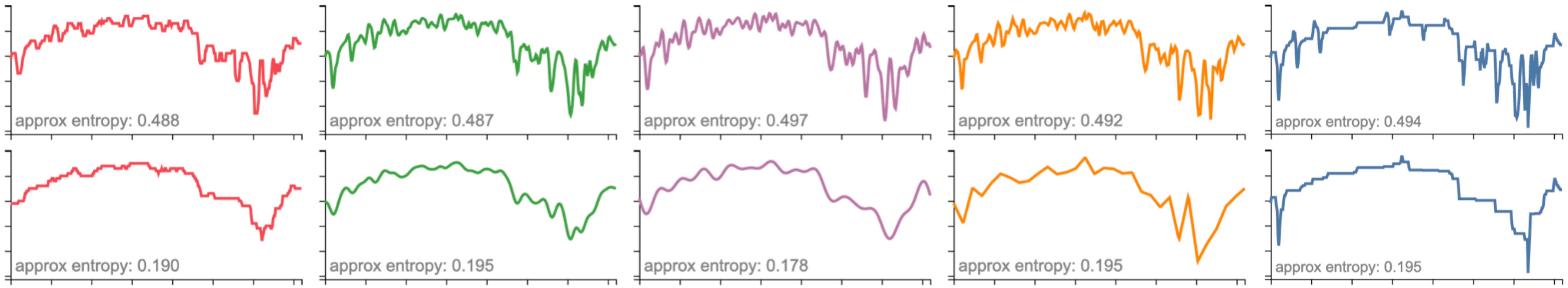
- remove critical pairs with persistence below a threshold t
- enforce monotonicity between neighboring extrema (isotonic regression is used)



Input time series



Smoothed time series



Median **Gaussian** **Cutoff** **Subsampling** **TopoLines**

TopoLines: Topological Smoothing for Line Charts
Rosen et al. *EuroVis (2020)*