Homological Analysis of Sensors from Power Plants

by

Luciano Melodia

13th September 2021

Affiliations

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Professorship for Evolutionary Data Management
Friedrich-Alexander University Erlangen-Nürnberg
Martensstrasse 3, 91058 Erlangen
Germany

This project is a cooperation between:

SIEMENS
energy

Content

- **Classification of Power Plant Sensor Data:**
  - Labeling System.
  - Structure of the Argument.

Content

- **Classification of Power Plant Sensor Data:**
  - Labeling System.
  - Structure of the Argument.
- **Theoretical Background:**
  - Geometry of $\mathbb{SW}_{M,\tau}f(t)$.
  - Persistent Homology of $\mathbb{SW}_{M,\tau}f(t)$.
  - Remark: Homology of $\mathbb{T}^n$.

Content

- **Classification of Power Plant Sensor Data:**
  - Labeling System.
  - Structure of the Argument.
- **Theoretical Background:**
  - Geometry of $\mathbb{SW}_{M,\tau} f(t)$.
  - Persistent Homology of $\mathbb{SW}_{M,\tau} f(t)$.
  - Remark: Homology of $\mathbb{T}^n$.
- **Experimental Setup:**
  - Homological Analysis.
  - Neural Network.

Content

- **Classification of Power Plant Sensor Data:**
  - Labeling System.
  - Structure of the Argument.
- **Theoretical Background:**
  - Geometry of $\mathbb{SW}_{M,\tau}f(t)$.
  - Persistent Homology of $\mathbb{SW}_{M,\tau}f(t)$.
  - Remark: Homology of $\mathbb{T}^n$.
- **Experimental Setup:**
  - Homological Analysis.
  - Neural Network.
- **Experimental Results:**
  - Results.
  - Summary.
  - Closing Thoughts.

Classification of Power Plant Sensor Data

Labeling System

Components of the power plant reference designation system
(germ. *Kraftwerkskennzeichensystem*, abbreviated as **KKS**):

- **Overall system:**
  Counting the overall systems.

Labeling System

Components of the power plant reference designation system
(germ. *Kraftwerkskennzeichensystem*, abbreviated as **KKS**):

- **Overall system:**
  Counting the overall systems.
- **Function:**
  Main group and subgroups of functional units.

Labeling System

Components of the power plant reference designation system (germ. *Kraftwerkskennzeichensystem*, abbreviated as **KKS**):

- **Overall system:**
  Counting the overall systems.
- **Function:**
  Main group and subgroups of functional units.
- **Aggregate:**
  Aggregate is part of a subgroup and itself a group of units.

Labeling System

Components of the power plant reference designation system
(germ. *Kraftwerkskennzeichensystem*, abbreviated as **KKS**):

- **Overall system:**
  Counting the overall systems.
- **Function:**
  Main group and subgroups of functional units.
- **Aggregate:**
  Aggregate is part of a subgroup and itself a group of units.
- **Operating resources:**
  Operating equipment or signal indicator in the aggregate.

Labeling System

Schema of the KKS:

$$\overbrace{\text{L or D}}^{\text{Overall system }(\texttt{OS})} \quad \overbrace{\text{(D)LLLDD}}^{\text{Function }(\texttt{F})} \quad \overbrace{\text{LLDDD(L)}}^{\text{Aggregate }(\texttt{A})} \quad \overbrace{\text{LLDD}}^{\text{Operating resources }(\texttt{OR})}$$

Labeling System

Schema of the KKS:

$$\overbrace{\text{L or D}}^{\text{Overall system }(\texttt{OS})} \quad \overbrace{\text{(D)LLLDD}}^{\text{Function }(\texttt{F})} \quad \overbrace{\text{LLDDD(L)}}^{\text{Aggregate }(\texttt{A})} \quad \overbrace{\text{LLDD}}^{\text{Operating resources }(\texttt{OR})}$$

Example:

Block.

Main group 2L:

2nd steam, water, gas circuit.

Subgroup (2L)A:

Feedwater system.

Subgroup (2LA)C:

Feedwater pumping system.

Counter (2LAC)03:

3rd feedwater pumping system.

Main group C:

Direct measurement.

Subgroup (C)T:

Temperature measurement.

Counter (CT)002:

2nd temperature measurement.

Main group Q:

Control equipment.

Subgroup (Q)T:

Immersion sleeves.

Counter (QT)12:

12th immersion sleeve.

$$\overbrace{1}^{\text{Block.}} \quad \overbrace{\text{(2)LAC03}} \quad \overbrace{\text{CT002(-)}} \quad \overbrace{\text{QT12}}$$

Structure of the Argument

Smooth manifold assumption:

1. Let $(t_i) := \{t_i\}_{i=0}^{n}$, with $t_i \in T$ be a time series, and T a strictly totally ordered set.

Structure of the Argument

Smooth manifold assumption:

1. Let $(t_i) := \{t_i\}_{i=0}^{n}$, with $t_i \in T$ be a time series, and T a strictly totally ordered set.
2. Let $f : T \rightarrow \mathbb{R}$ be a function, then $f \circ (t_i) = (f(t_i))$ is certainly a time series.

Structure of the Argument

Smooth manifold assumption:

1. Let $(t_i) := \{t_i\}_{i=0}^{n}$, with $t_i \in T$ be a time series, and T a strictly totally ordered set.
2. Let $f : T \longrightarrow \mathbb{R}$ be a function, then $f \circ (t_i) = (f(t_i))$ is certainly a time series.
3. On an open interval, (a, b), there exists a polynomial function, $p : (a, b) \longrightarrow \mathbb{R}$, approximating $(f(t_i))$ with ϵ-error, or in other words arbitrarily well.

Structure of the Argument

Smooth manifold assumption:

1. Let $(t_i) := \{t_i\}_{i=0}^{n}$, with $t_i \in T$ be a time series, and T a strictly totally ordered set.
2. Let $f : T \longrightarrow \mathbb{R}$ be a function, then $f \circ (t_i) = (f(t_i))$ is certainly a time series.
3. On an open interval, (a, b), there exists a polynomial function, $p : (a, b) \longrightarrow \mathbb{R}$, approximating $(f(t_i))$ with ϵ-error, or in other words arbitrarily well.
4. For a smooth function $p : (a, b) \longrightarrow \mathbb{R}$ its graph $\mathcal{G}p := \{(t_i, p(t_i)) \mid t_i \in (a, b)\}$ is a smooth manifold with atlas $\varphi : \mathcal{G}p \longrightarrow \mathbb{R}$, $\varphi(t_i, p(t_i)) \longmapsto t_i$. $\mathcal{G}p \cong \mathbb{R}$ as smooth manifolds, thus higher homology groups of $\mathcal{G}p$ are trivial.

Theoretical Background

Geometry of $\mathbb{SW}_{M,\tau}f(t)$

The *sliding-window embedding* is given by

$$\mathrm{SW}_{M,\tau}f(t) = [f(t)\ f(t+\tau)\ \cdots\ f(t+M\tau)]^\top\,, \tag{1}$$

where τ is called *step size* or *time delay*, $M\tau$ is called *window size* and $M+1$ is the dimension of the embeddings' space. The *sliding-window point cloud associated with T* is

$$\mathbb{SW}_{M,\tau}f := \{\mathrm{SW}_{M,\tau}f(t_i) \mid t_i \in T\}. \tag{2}$$

Geometry of $\mathbb{SW}_{M,\tau}f(t)$

The *sliding-window embedding* is given by

$$\mathrm{SW}_{M,\tau}f(t) = [f(t)\ f(t+\tau)\ \cdots\ f(t+M\tau)]^{\top}, \tag{1}$$

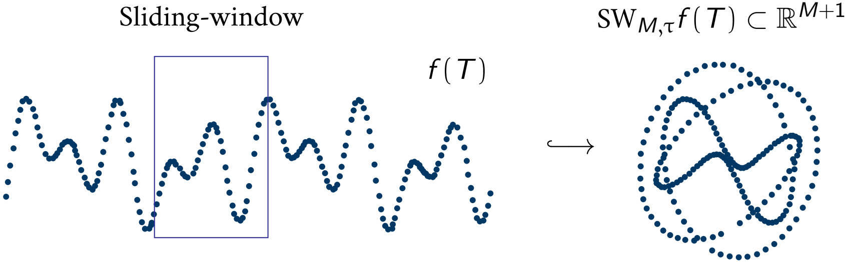
where τ is called *step size* or *time delay*, $M\tau$ is called *window size* and $M+1$ is the dimension of the embeddings' space. The *sliding-window point cloud associated with T* is

$$\mathbb{SW}_{M,\tau}f := \{\mathrm{SW}_{M,\tau}f(t_i) \mid t_i \in T\}. \tag{2}$$

Periodicity of $f(T)$
Period $f(t_i + 2\pi/L) = f(t_i)$
Number of harmonics N
Number of (non-)commensurate frequencies N

Circularity of $\mathbb{SW}_{M,\tau}f(T) \subset \mathbb{R}^{M+1}$
Roundness $M\tau = \frac{M}{M+1}\frac{2\pi}{L}$
Ambient dimension $M \geq 2N$
Intrinsic dimension $\subset \mathbb{S}^1_1 \times \cdots \times \mathbb{S}^1_N$

Illustration following Perea



Remark: Homology of $\mathbb{T}^n$

Let $\mathbb{T}^2 \simeq \mathbb{S}^1 \times \mathbb{S}^1$ and $\mathbb{Z}_p := \mathbb{Z}/(p\mathbb{Z})$ with p prime.
We use that

$$H_0(\mathbb{S}^1; \mathbb{Z}_p) = H_1(\mathbb{S}^1; \mathbb{Z}_p) = \mathbb{Z}_p, \tag{3}$$

$$H_i(\mathbb{S}^1; \mathbb{Z}_p) = 0, \text{ for } i > 1. \tag{4}$$

Remark: Homology of $\mathbb{T}^n$

Let $\mathbb{T}^2 \cong \mathbb{S}^1 \times \mathbb{S}^1$ and $\mathbb{Z}_p := \mathbb{Z}/(p\mathbb{Z})$ with p prime.
We use that

$$H_0(\mathbb{S}^1; \mathbb{Z}_p) = H_1(\mathbb{S}^1; \mathbb{Z}_p) = \mathbb{Z}_p, \tag{3}$$

$$H_i(\mathbb{S}^1; \mathbb{Z}_p) = 0, \text{ for } i > 1. \tag{4}$$

Thus, we get

$$H_1(\mathbb{T}^2; \mathbb{Z}_p) = \mathbb{Z}_p \oplus \mathbb{Z}_p, \tag{5}$$

$$H_2(\mathbb{T}^2; \mathbb{Z}_p) = \mathbb{Z}_p, \tag{6}$$

$$H_i(\mathbb{T}^2; \mathbb{Z}_p) = 0, \text{ for } i > 2. \tag{7}$$

Remark: Homology of $\mathbb{T}^n$

The homology groups of a sphere are torsion free. As we work in a field of coefficients, we can apply Künneth's formula, because all modules over a field are free.

Remark: Homology of $\mathbb{T}^n$

The homology groups of a sphere are torsion free. As we work in a field of coefficients, we can apply Künneth's formula, because all modules over a field are free.
Thus, we can generalize for $\mathbb{T}^n \cong \mathbb{S}^1_1 \times \cdots \times \mathbb{S}^1_n$:

$$H_k(\mathbb{T}^n; \mathbb{Z}_p) = \bigoplus_{i_1+\cdots+i_r=k} H_{i_1}(\mathbb{S}^1; \mathbb{Z}_p) \otimes \cdots \otimes H_{i_r}(\mathbb{S}^1; \mathbb{Z}_p), \tag{8}$$

$$H_k(\mathbb{T}^n; \mathbb{Z}_p) = \mathbb{Z}_p^{\binom{n}{k}}. \tag{9}$$

In fact, we have now a relation between the dimension of the embedding (if it is a hyper-torus) and its homology groups.

Remark: Homology of $\mathbb{T}^n$

Recall, that $\beta_k := \operatorname{rank} H_k(X; \mathbb{F})$.

n	$\mathbb{T}^n$	β_0	β_1	β_2	β_3	β_4	β_5
0	one-point-space	1	0	0	0	0	0
1	circle	1	1	0	0	0	0
2	2-torus	1	2	1	0	0	0
3	3-torus	1	3	3	1	0	0
4	4-torus	1	4	6	4	1	0
5	5-torus	1	5	10	10	5	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Experimental Setup

Heuristic Choice of Parameters

- We examined about $18 \cdot 10^3$ different signals from *four different combined cycle gas turbine power plants* with a total of *two gas turbines, two boilers* for steam generation and *one steam turbine*.

Heuristic Choice of Parameters

- We examined about $18 \cdot 10^3$ different signals from *four different combined cycle gas turbine power plants* with a total of *two gas turbines, two boilers* for steam generation and *one steam turbine*.
- *Time delay* is set to $\tau = 1$.

Heuristic Choice of Parameters

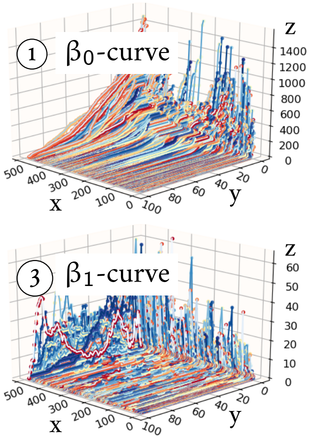
- We examined about $18 \cdot 10^3$ different signals from *four different combined cycle gas turbine power plants* with a total of *two gas turbines, two boilers* for steam generation and *one steam turbine*.
- *Time delay* is set to $\tau = 1$.
- *Embedding dimension* is set to $M = 5$ using the *false nearest neighbor algorithm*.
  Distribution of optimal dimension per signal:
  $M = 2 : 4.345,\ M = 3 : 2.594,\ M = 4 : 3.877,\ M = 5 : 7.347.$

Heuristic Choice of Parameters

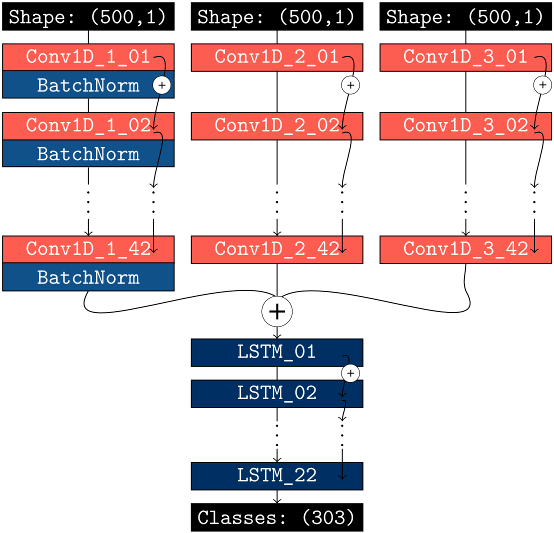
- We examined about $18 \cdot 10^3$ different signals from *four different combined cycle gas turbine power plants* with a total of *two gas turbines, two boilers* for steam generation and *one steam turbine*.
- *Time delay* is set to $\tau = 1$.
- *Embedding dimension* is set to $M = 5$ using the *false nearest neighbor algorithm*.
  Distribution of optimal dimension per signal:
  $M = 2 : 4.345,\ M = 3 : 2.594,\ M = 4 : 3.877,\ M = 5 : 7.347.$
- Time series with *persistence entropy* ≥ 0.98 on the *persistence diagrams of* $\mathbb{SW}_{M,\tau} f$ *associated with* T_j have been removed.

Homological Analysis

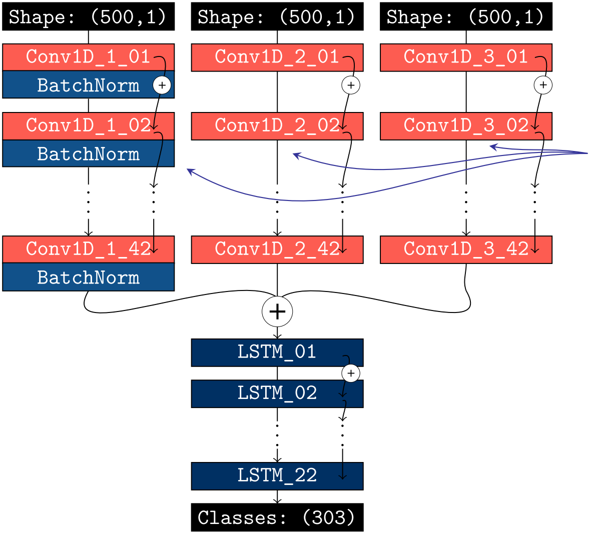
Persistence representations of the
heating medium system of a gas turbine power plant:



Neural Network

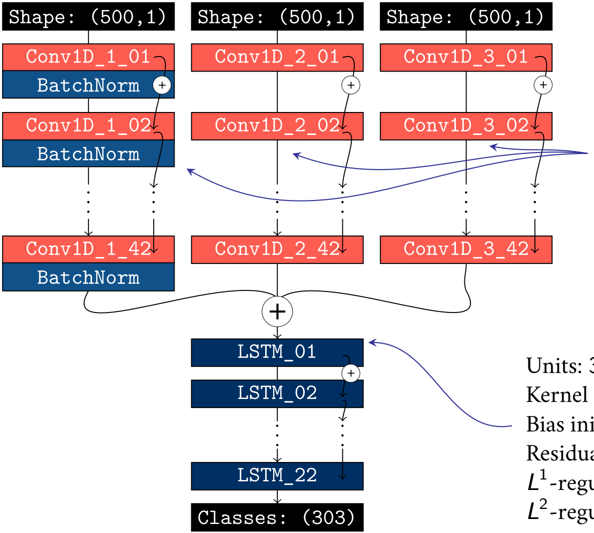


Neural Network



Filters: 64,
Kernel-size: 3,
Kernel init.: *Glorot normal*,
Bias init.: *Zeros*,
Padding: *Causal*,
Residual: $\mathcal{C}^1$,
L^1-regularization: 0.001,
L^2-regularization: 0.01.

Neural Network



Filters: 64,
Kernel-size: 3,
Kernel init.: *Glorot normal*,
Bias init.: *Zeros*,
Padding: *Causal*,
Residual: $\mathcal{C}^1$,
L^1-regularization: 0.001,
L^2-regularization: 0.01.

Units: 32,
Kernel init.: *Glorot normal*,
Bias init.: *Zeros*,
Residual: $\mathcal{C}^1$,
L^1-regularization: 0.001,
L^2-regularization: 0.01.

Experimental Results

Results

OS	F	A	OR	Accuracy	F1	Precision	Recall
$\mathcal{C}^0$-ConvNet without topological features:							
✓	✓	✓	✓	0.4821 ±0.0031	0.5677 ±0.0033	0.6912 ±0.0029	0.4816 ±0.0037
✓	✗	✗	✗	0.7129 ±0.0102	0.7904 ±0.0092	0.9010 ±0.0097	0.7041 ±0.0088
✓	✓	✗	✗	0.5691 ±0.0037	0.6830 ±0.0058	0.8699 ±0.0065	0.5622 ±0.0052
✓	✓	✓	✗	0.5426 ±0.0055	0.6681 ±0.0036	0.8682 ±0.0048	0.5429 ±0.0029
$\mathcal{C}^0$-ConvNet:							
✓	✓	✓	✓	0.6142 ±0.0047	0.6212 ±0.0077	0.7681 ±0.0082	0.5216 ±0.0073
✓	✗	✗	✗	**0.8316** ±0.0121	**0.8511** ±0.0063	**0.9327** ±0.0163	**0.7827** ±0.0039
✓	✓	✗	✗	0.7024 ±0.0091	0.7567 ±0.0101	0.8756 ±0.0109	0.6663 ±0.0094
✓	✓	✓	✗	0.6291 ±0.0078	0.7376 ±0.0065	0.8726 ±0.0056	0.6389 ±0.0077
$\mathcal{C}^1$-ConvNet:							
✓	✓	✓	✓	**0.6383** ±0.0085	**0.6566** ±0.0055	**0.7849** ±0.0074	**0.5597** ±0.0076
✓	✗	✗	✗	0.8221 ±0.0028	0.8497 ±0.0023	0.9267 ±0.0033	0.7846 ±0.0018
✓	✓	✗	✗	**0.7284** ±0.0019	**0.7670** ±0.0027	**0.8826** ±0.0017	**0.6782** ±0.0066
✓	✓	✓	✗	**0.6524** ±0.0009	**0.7276** ±0.0028	**0.8821** ±0.0032	**0.6192** ±0.0025

Summary

- The best classification results are about 64% for the *entire KKS* (`OS F A OR`), about 65% for the *aggregate* (`OS F A`), 73% for the *functional level* (`OS F`), and 83% for the *entire system* (`OS`).

Summary

- The best classification results are about 64% for the *entire KKS* (`OS F A OR`), about 65% for the *aggregate* (`OS F A`), 73% for the *functional level* (`OS F`), and 83% for the *entire system* (`OS`).
- For all experiments it holds that precision $\gg$ recall. Thus, the exactness of our classifier is relatively huge in comparison to its average completeness per class.

Summary

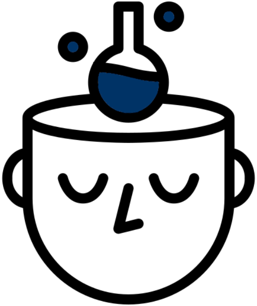
- The best classification results are about 64% for the *entire KKS* (OS F A OR), about 65% for the *aggregate* (OS F A), 73% for the *functional level* (OS F), and 83% for the *entire system* (OS).
- For all experiments it holds that precision $\gg$ recall. Thus, the exactness of our classifier is relatively huge in comparison to its average completeness per class.
- We have shown that residual connections improve classification results for all labels except for the *overall system* (OS) assignment.

Summary

- The best classification results are about 64% for the *entire KKS* (`OS F A OR`), about 65% for the *aggregate* (`OS F A`), 73% for the *functional level* (`OS F`), and 83% for the *entire system* (`OS`).
- For all experiments it holds that precision $\gg$ recall. Thus, the exactness of our classifier is relatively huge in comparison to its average completeness per class.
- We have shown that residual connections improve classification results for all labels except for the *overall system* (`OS`) assignment.
- The use of β_0 and β_1-curves improved the expected value of the classification results for all label variants studied.

Conclusion

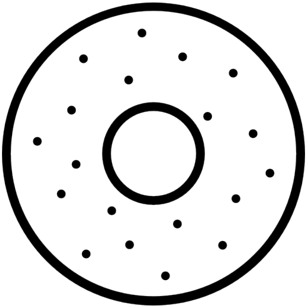
Closing Thoughts



Other experiments *performed by some of our students* show that the **`OR`-entity** achieves the **highest accuracy** in predicting the constituent identifiers in all models tested, followed by `A`, `F`, and `OS`.

This is promising since we have already demonstrated an **accuracy of** 83% **for `OS`.**

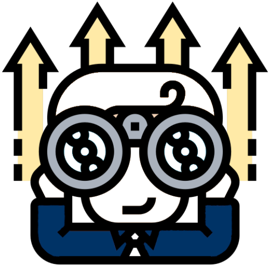
Closing Thoughts



Since the **signal is embedded in a torus**, one could construct neural network layers operating on a given **Lie group** $(\mathbb{S}^1_1 \times \cdots \times \mathbb{S}^1_p \simeq \mathbb{T}^p) \times \mathbb{R}^q$ and perform **parallel transport**.

The required smooth manifold can be derived from the persistence diagram.

Closing Thoughts



Further experiments shall be performed <u>without</u> using the corresponding **numbers of the aggregates** and **functional units**. This would result in much higher accuracy and would be sufficient for practical use.

References I

- **Melodia L.**, Lenz, R.: Estimate of the Neural Network Dimension Using Algebraic Topology and Lie Theory. Image Mining. Theory and Applications VII, pp.15-29 (2020).
- **Melodia L.**, Lenz R.: Persistent Homology as Stopping-Criterion for Voronoi Interpolation. Proceedings of the International Workshop on Combinatorial Image Analysis, pp. 29–44 (2019).

References II

- Perea, J.: Topological Time Series Analysis. Notices of the American Mathematical Society 66, (2019).
- Perea, J.: Persistent Homology of Toroidal Sliding Window Embeddings. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6435 – 6439 (2016).
- Perea J.: Topological Time Series Analysis. Lecture 2: Persistent Homology of Sliding Window Point Clouds. https://www-m15.ma.tum.de/foswiki/pub/M15/Allgemeines/SummerSchool2016/perea_lect2.pdf.
- Perea J.: Topological Time Series Analysis. Day 1: Geometry of Sliding Window Embeddings. https://www-m15.ma.tum.de/foswiki/pub/M15/Allgemeines/SummerSchool2016/perea_lect1.pdf.
- Perea, J., Harer, J.: Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. Foundations of Computational Mathematics 15, pp. 799 – 838 (2015).

Thank You!

Have I piqued your interest?
Drop me a line:

`luciano.melodia@fau.de`!

And please ★ our repository:
`https://github.com/karhunenloeve/TwirlFlake`.

The icons used on these slides were kindly provided by
`https://flaticons.com` and `https://fontawesome.com`.
We express our gratitude and appreciation for this!