KTH Computer Science and Communication

Towards a topological machine learning

Mikael Vejdemo-Johansson
AUS-ICMS

Overview of existing work

- Stability of persistence
- Persistence of random point clouds
- Distance to measure
- Persistence as ML feature vectors
- Barcode means

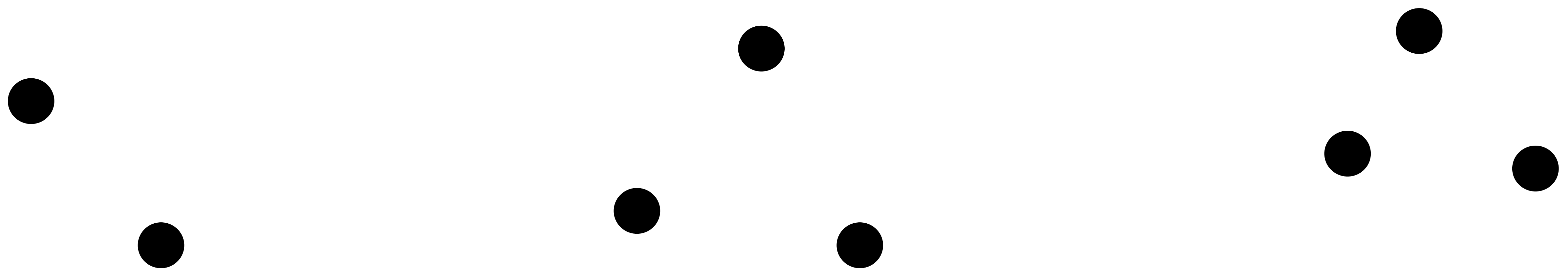
KTH Computer Science and Communication

Point cloud cohomology: persistence

- Just computing cohomology: not useful
  Discrete points have no interesting structure.
- Instead:
    1. Cover each point with a ball
    2. Compute cohomology of the union of balls

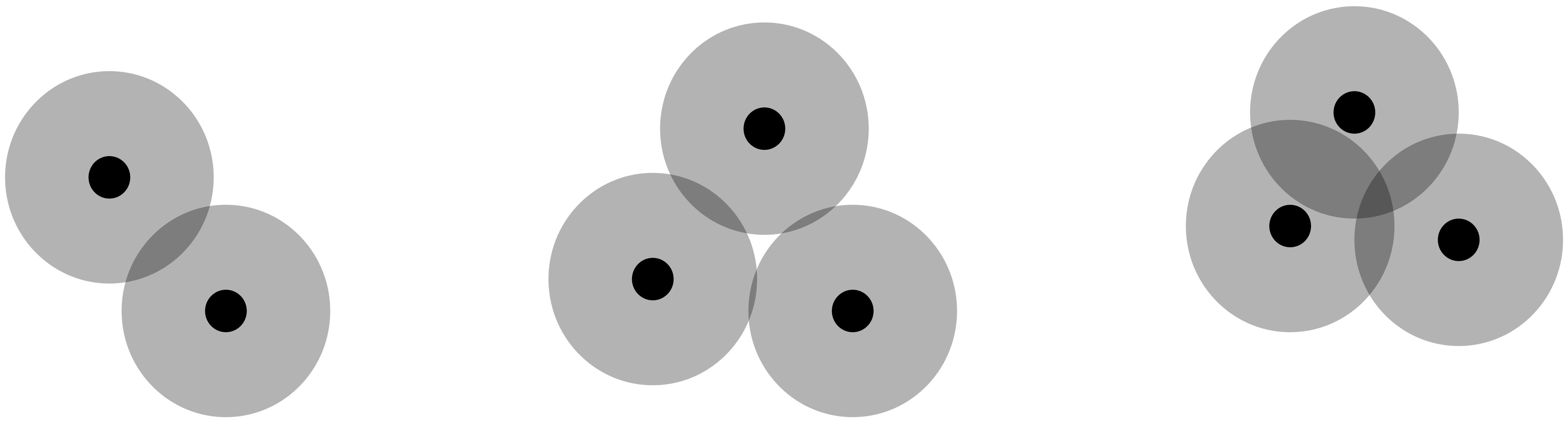
KTH Computer Science and Communication

Encoding geometry

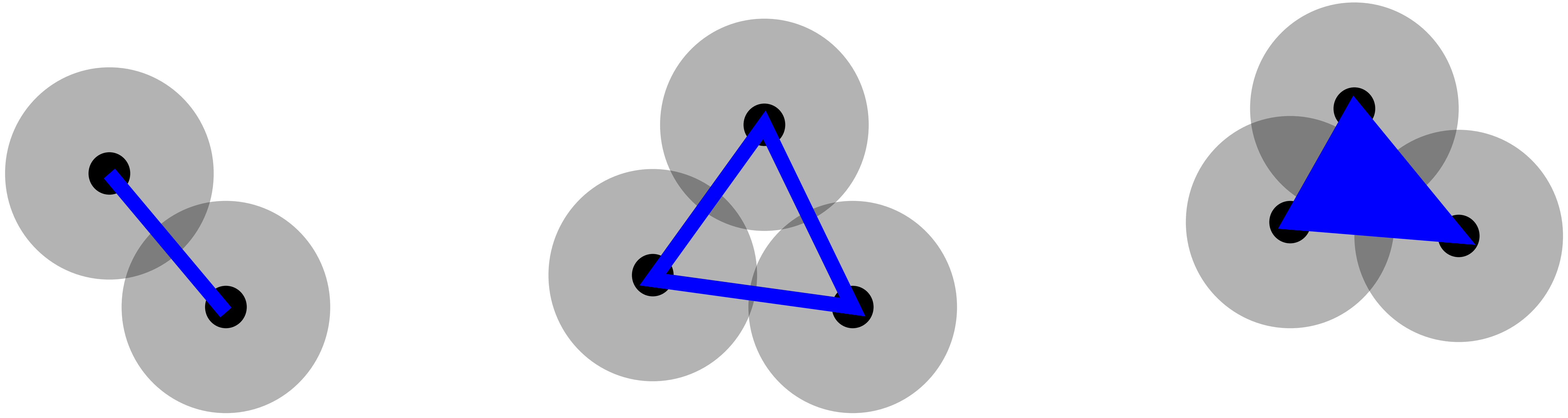


KTH Computer Science and Communication

Encoding geometry

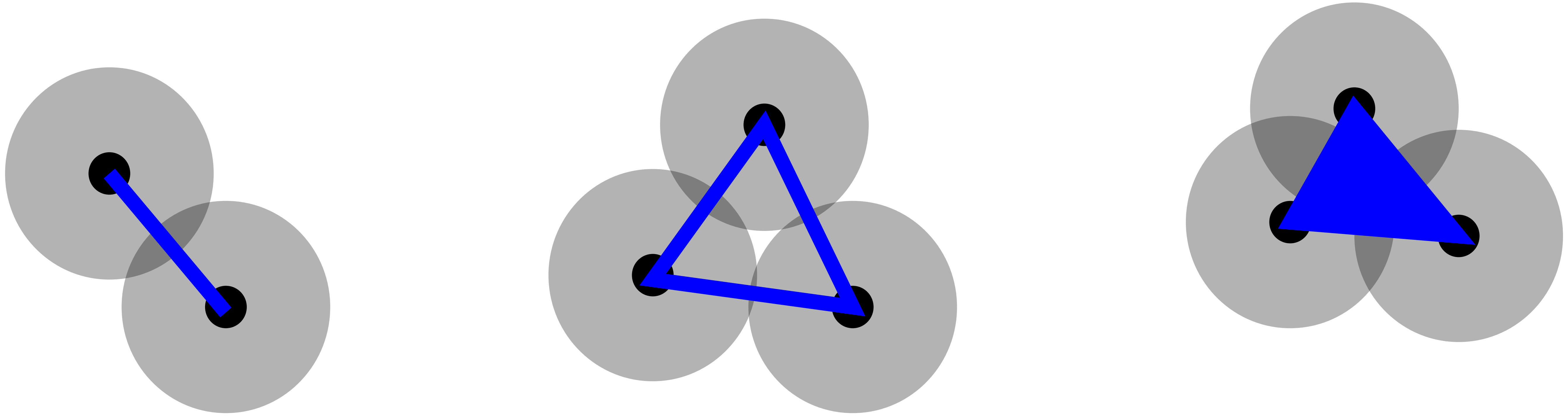


Encoding geometry



KTH Computer Science and Communication

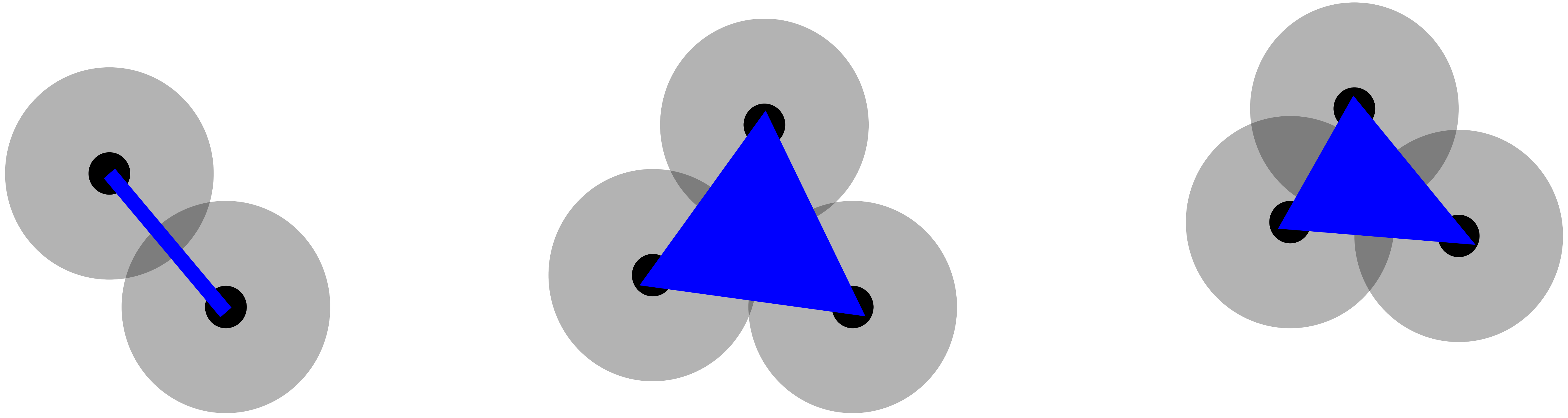
Encoding geometry



The Čech *complex* of radius r has a simplex for data points $x_0,\ldots,x_n$ whenever all the balls of radius r centered around the data points share an intersection. d_i removes x_i. This generalizes the *single linkage graph*.

KTH VETENSKAP OCH KONST
KTH Computer Science and Communication

Encoding geometry



The Vietoris-Rips complex is completely defined by the single linkage graph: includes a simplex whenever there is a clique in the single linkage graph.
Face maps just like in the Čech case.

KTH Computer Science
and Communication

Functoriality: algebraic continuity

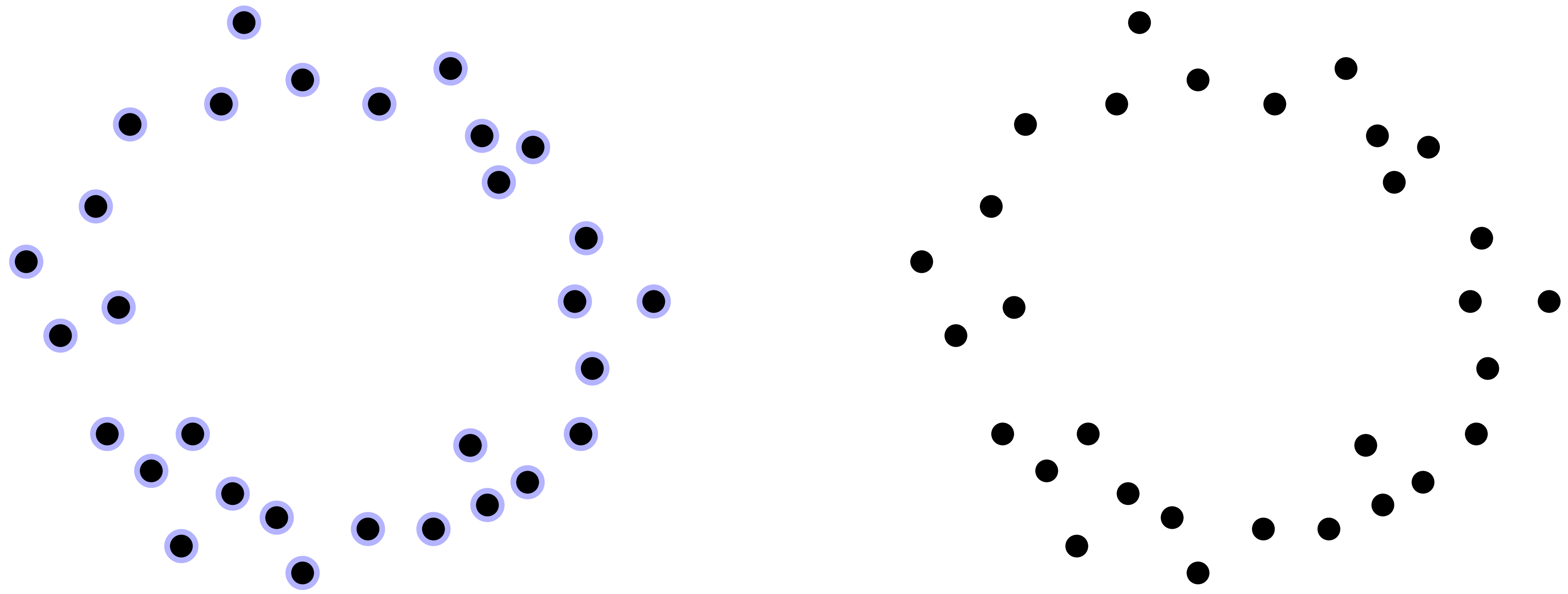
- If we increase the radius, no cells vanish
  the complex can only ever increase.

  For any $r < r'$ there are inclusion maps
  $\check{C}_r(X) \to \check{C}_{r'}(X)$

- Computing cohomology is a functor — continuous in an algebraic sense:
  if there is a map $X \to Y$, then there is an induced map
  $H^1Y \to H^1X$

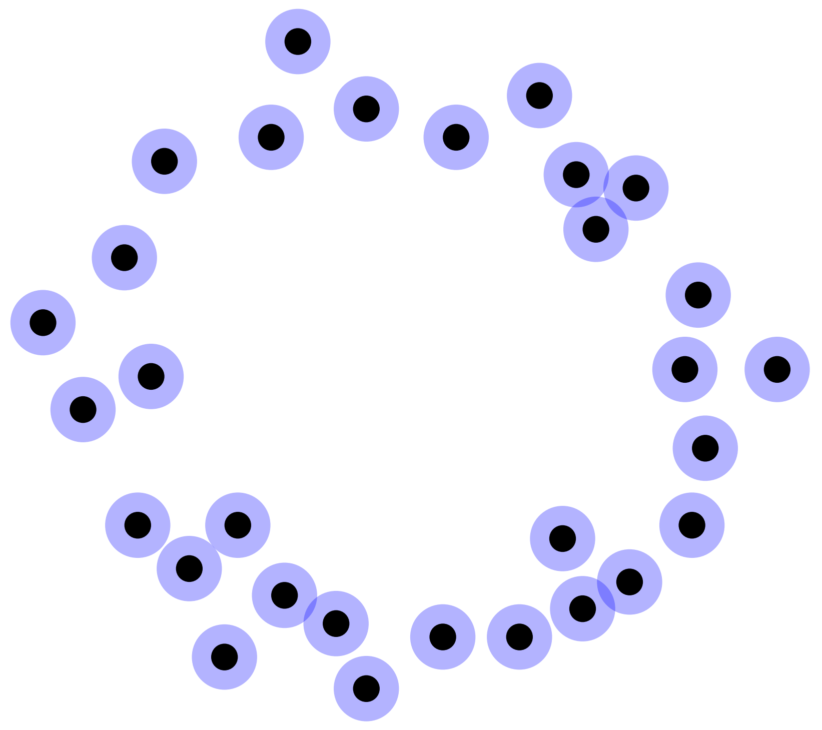
KTH Computer Science
and Communication

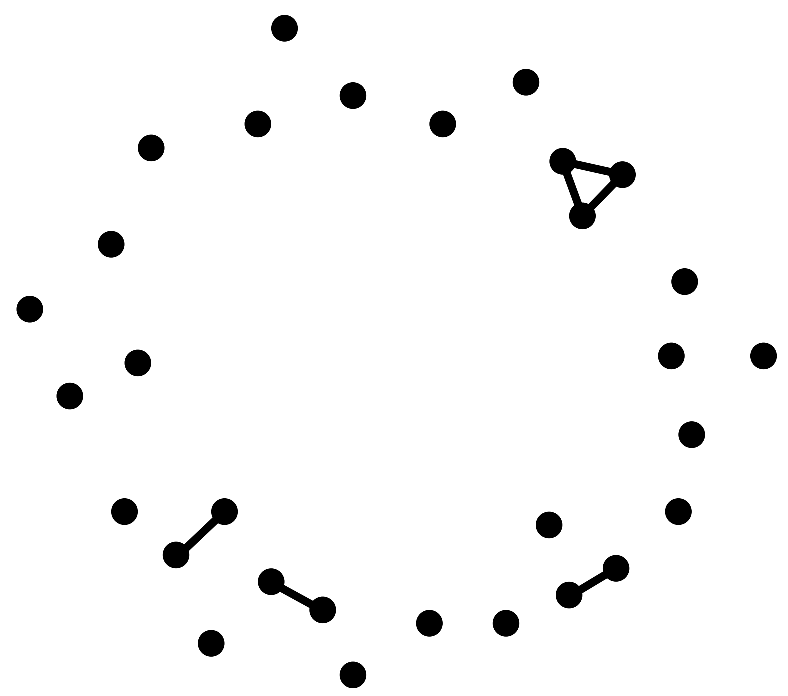
Persistent cohomology: tracking maps



KTH Computer Science and Communication

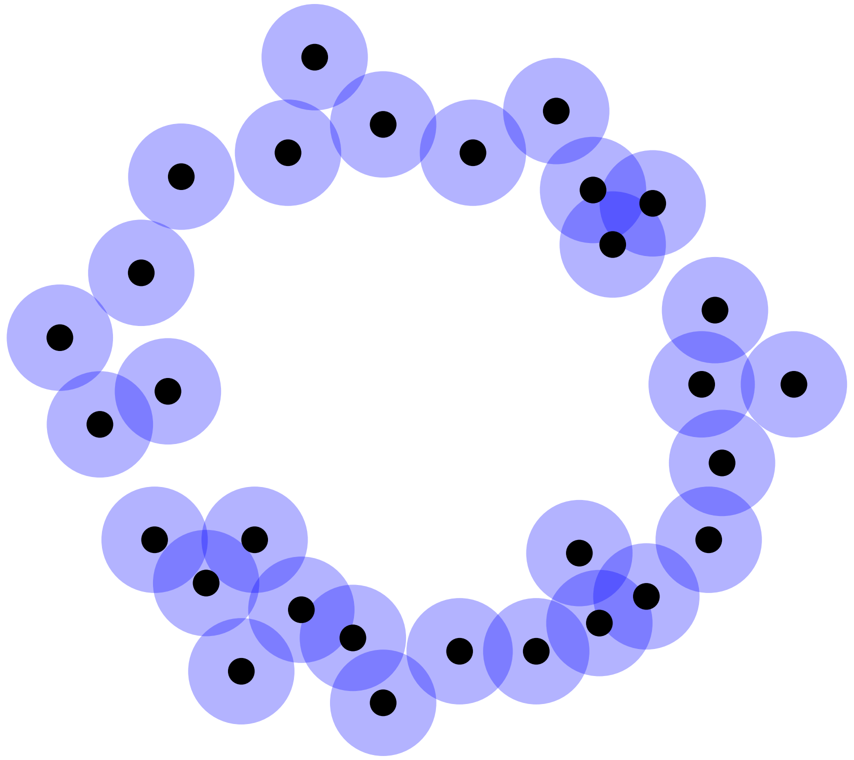
Persistent cohomology: tracking maps

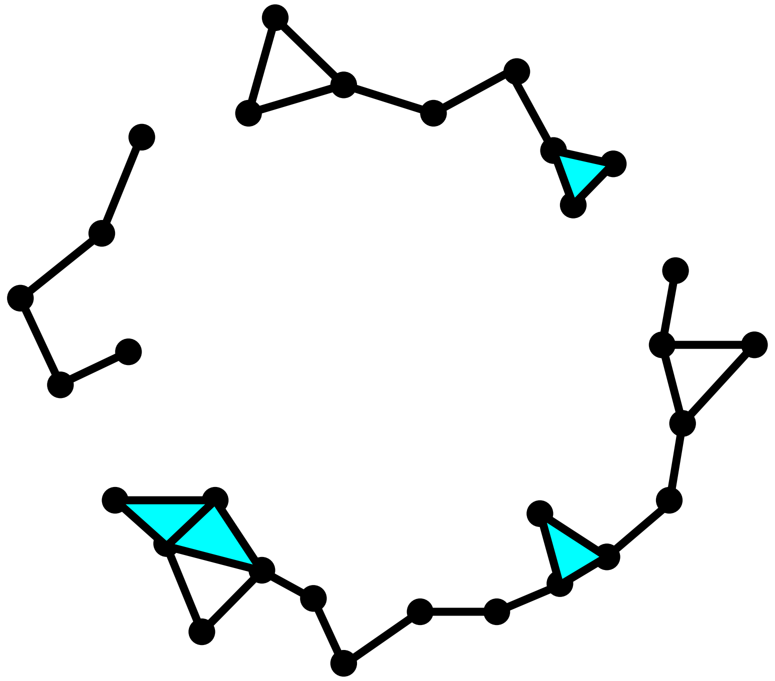




1 loop

Persistent cohomology: tracking maps

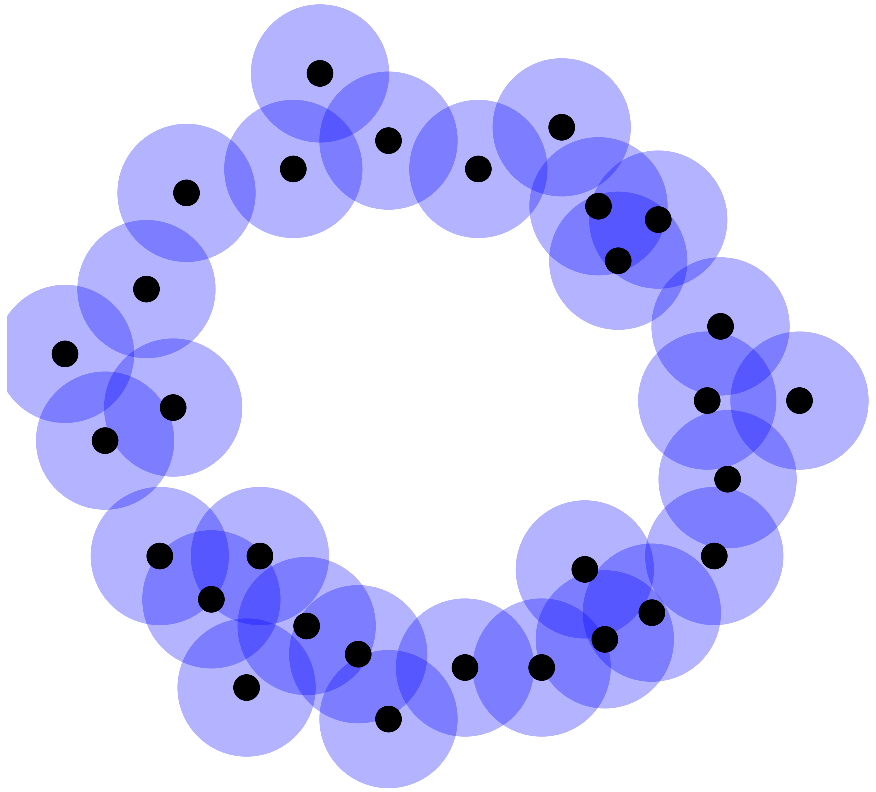


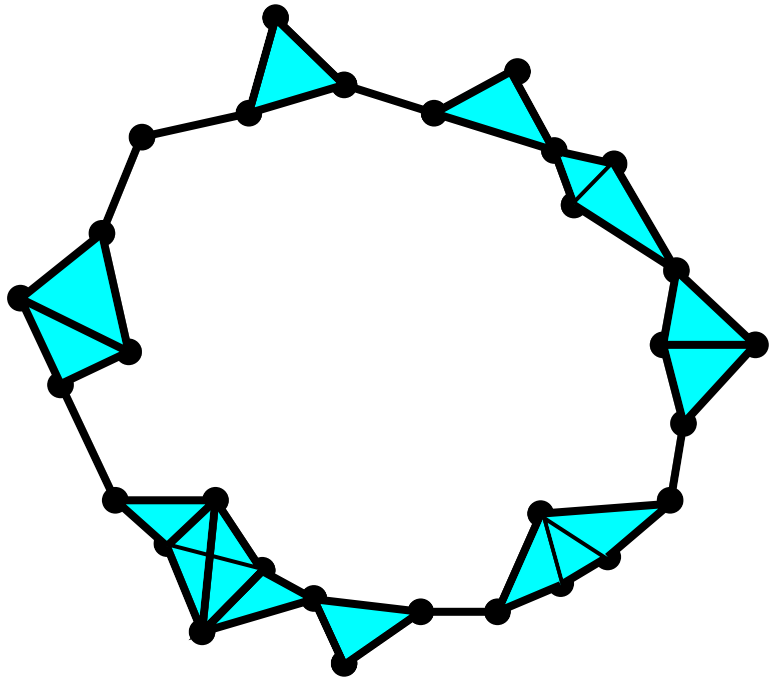


3 loops

KTH Computer Science and Communication

Persistent cohomology: tracking maps





1 loop

KTH Computer Science
and Communication

Key point:
are loops the same?

- The induced map come to our rescue:
- From a sequence of spaces

  $X_1 \hookrightarrow X_2 \hookrightarrow X_3 \hookrightarrow \ldots \hookrightarrow X_n$

  cohomology produces a sequence of vector spaces

  $H^1X_n \to \ldots \to H^1X_3 \to H^1X_2 \to H^1X_1$

KTH Computer Science and Communication

Algebra glues features

- Theorem (Gabriel, 1972): If M is a collection of vector spaces with linear maps along a path, M decomposes into interval modules.
- Decomposition produces birth/death pairs just like our barcodes.

Interval decompositions produce salient features

- Decomposing a diagram such as

  $H^1X_n \to \ldots \to H^1X_3 \to H^1X_2 \to H^1X_1$

  produces a direct sum diagram where each summand has the shape

  $0 \to \ldots \to 0 \to \mathbb{k} \to \ldots \to \mathbb{k} \to 0 \to 0 \to 0$

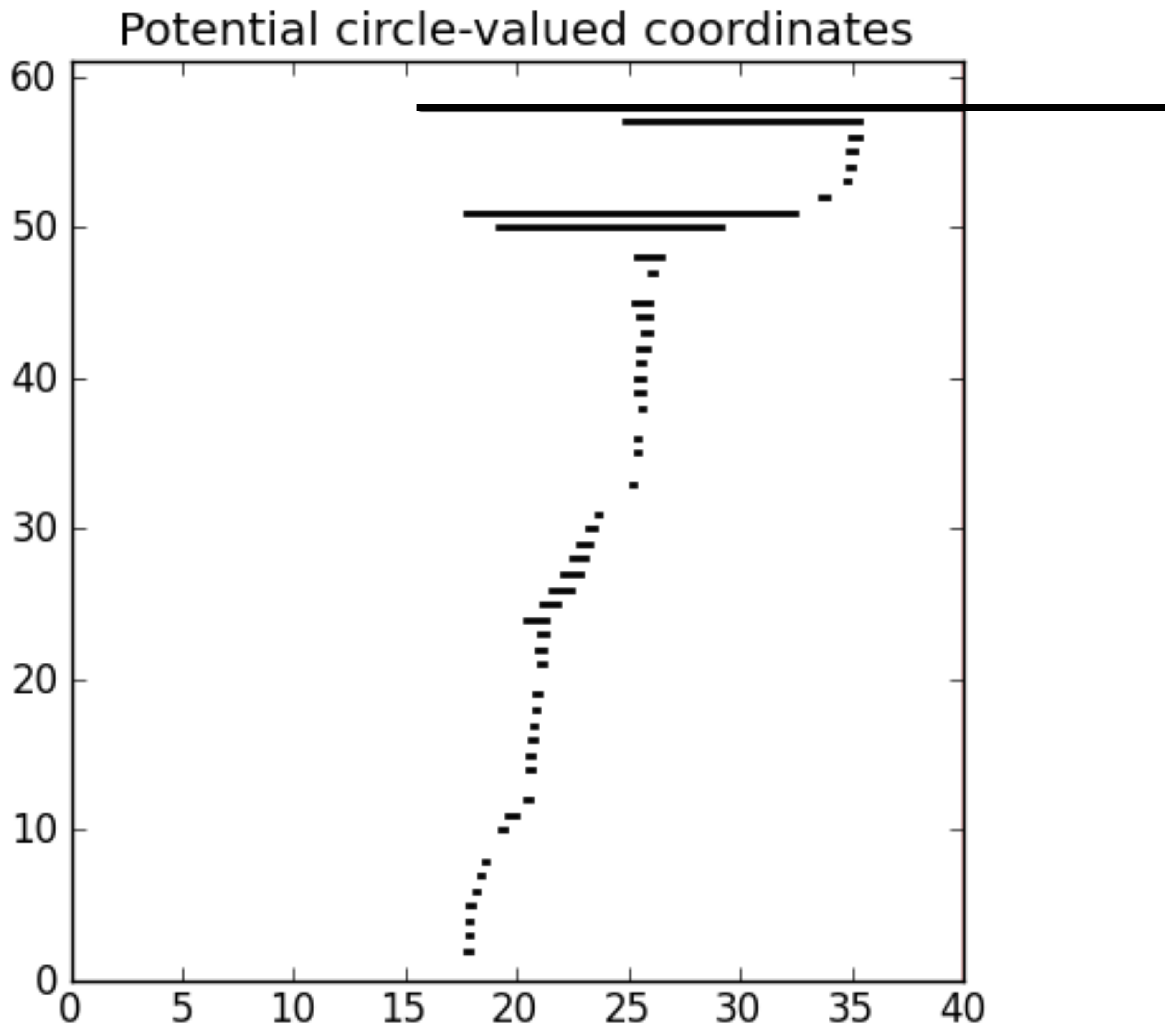
  where all maps $0 \to 0$ and $\mathbb{k} \to \mathbb{k}$ are isomorphisms.

- Each such summand is a choice of a (higher dimensional) essential circle coordinate across parameter values that persists across different values.

KTH Computer Science
and Communication

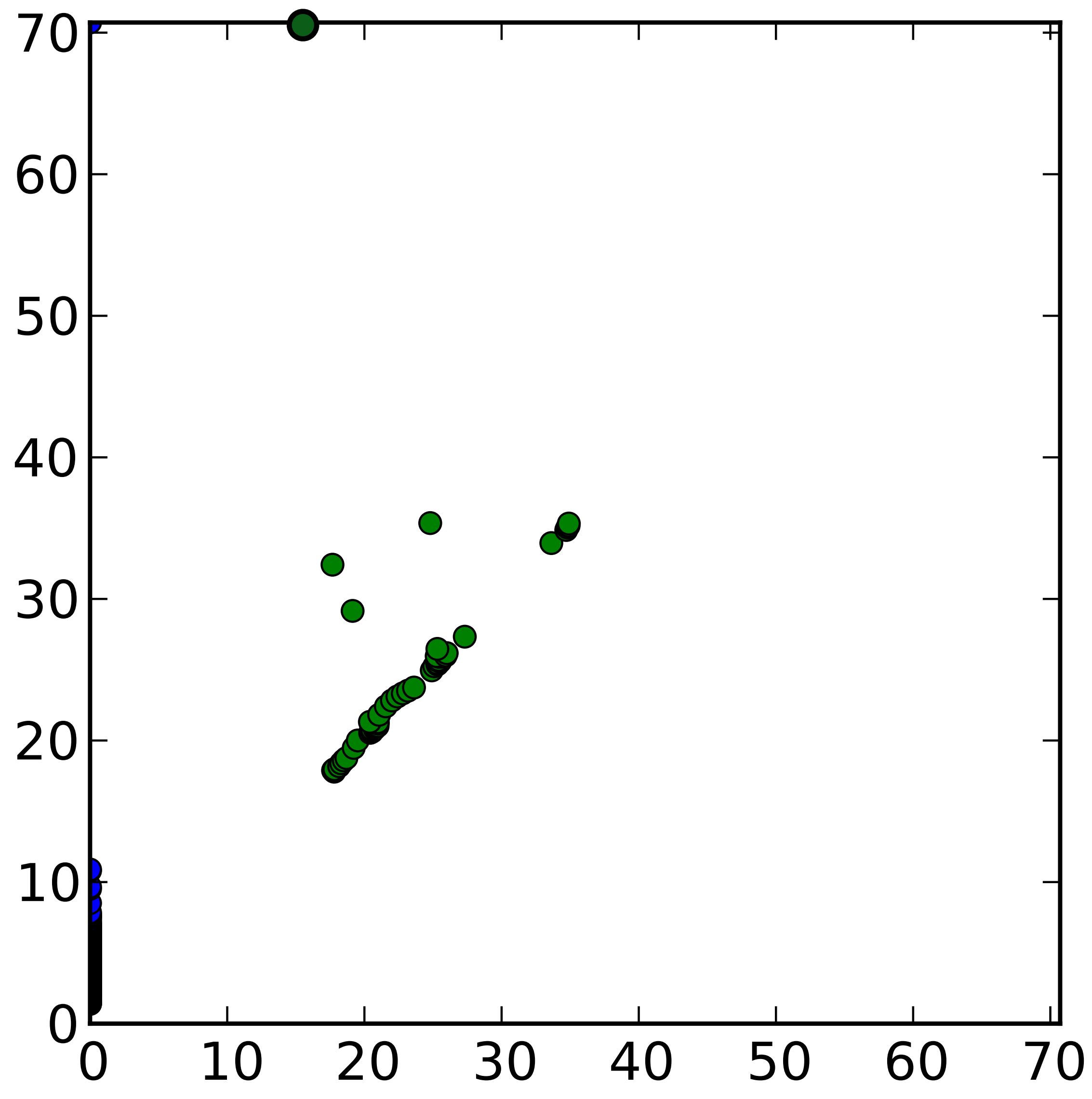
Barcodes and persistence diagrams

- Visualization tools for topological information
- Displays each interval in the decomposition



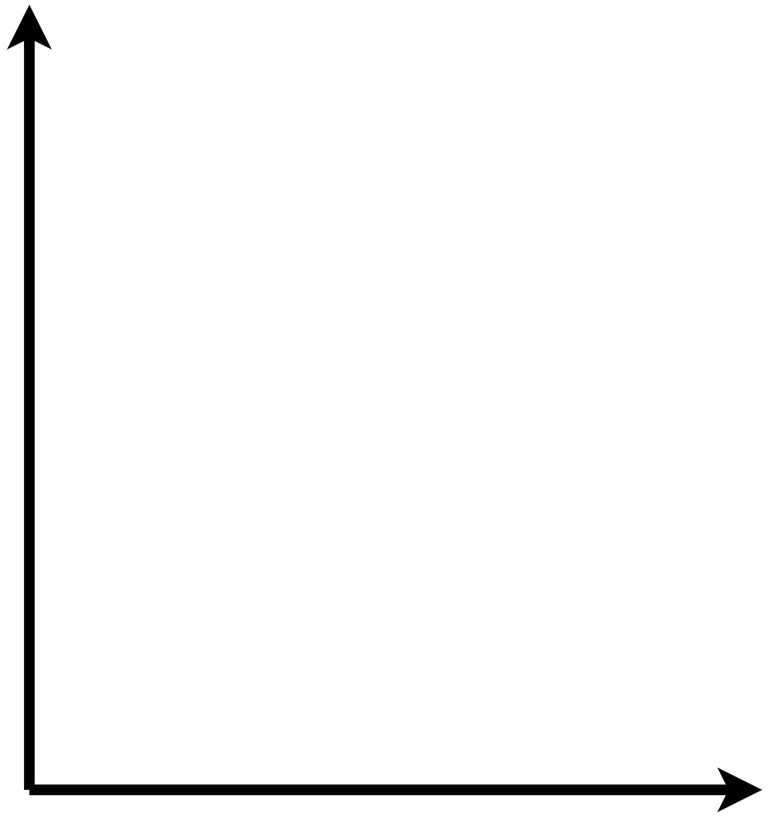
Barcodes and persistence diagrams

- Visualization tools for topological information
- Displays each interval in the decomposition



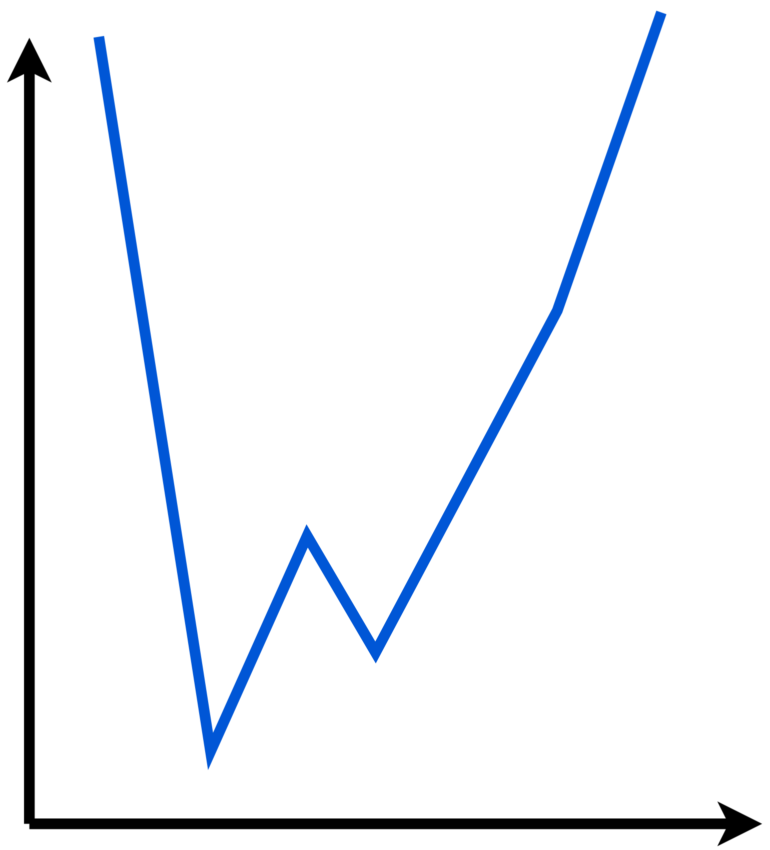
KTH Computer Science
and Communication

Barcodes are stable

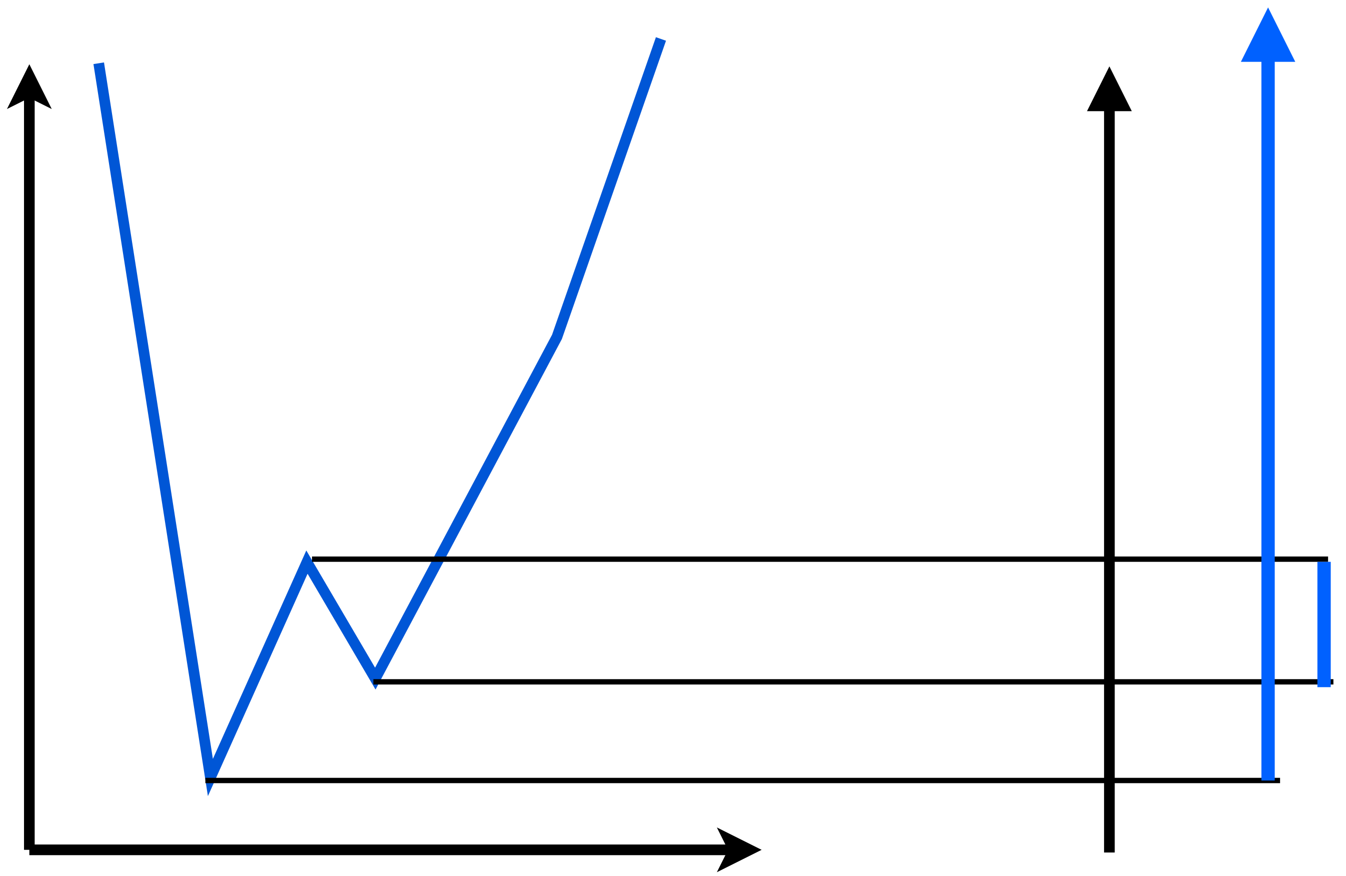


KTH Computer Science and Communication

Barcodes are stable

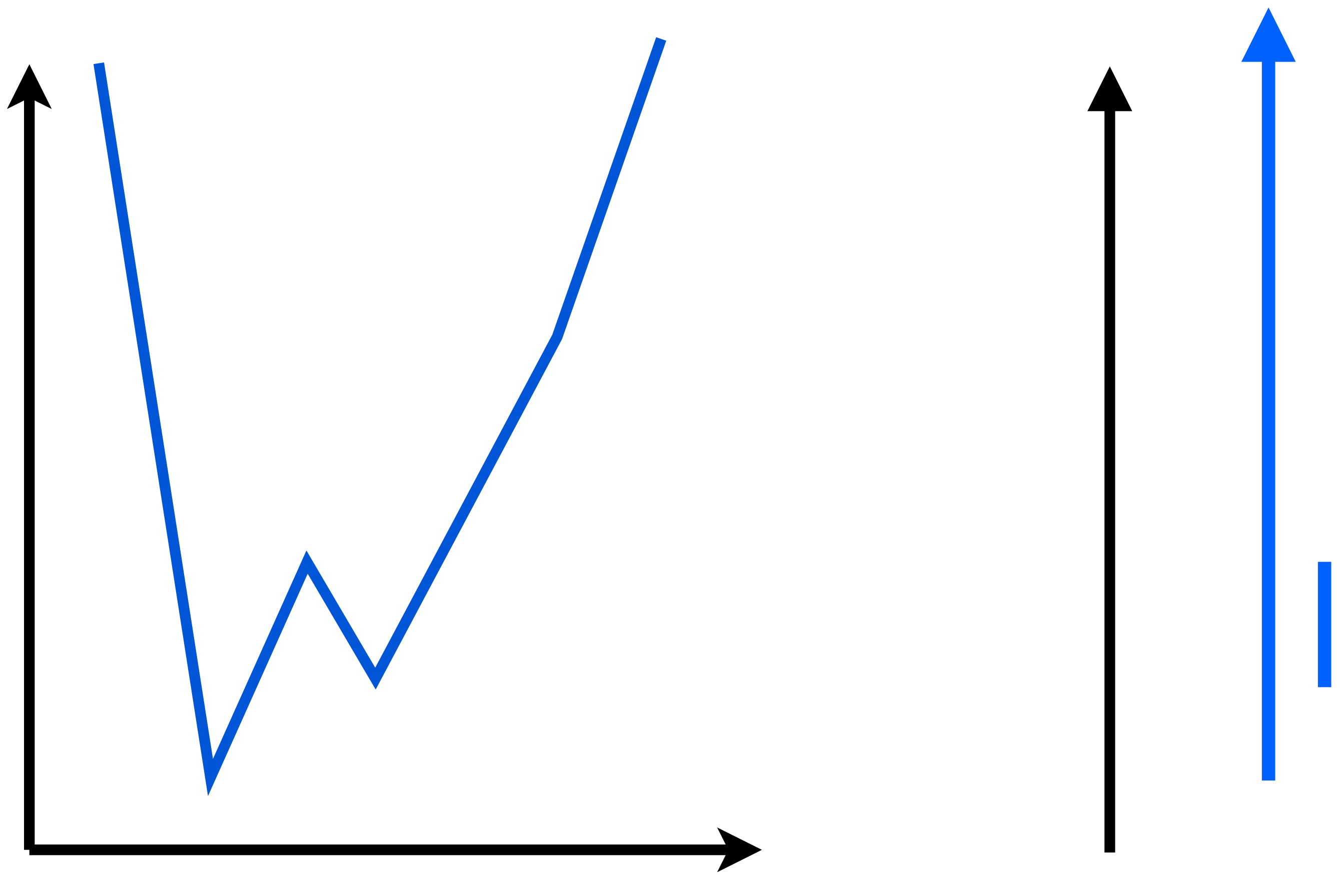


Barcodes are stable

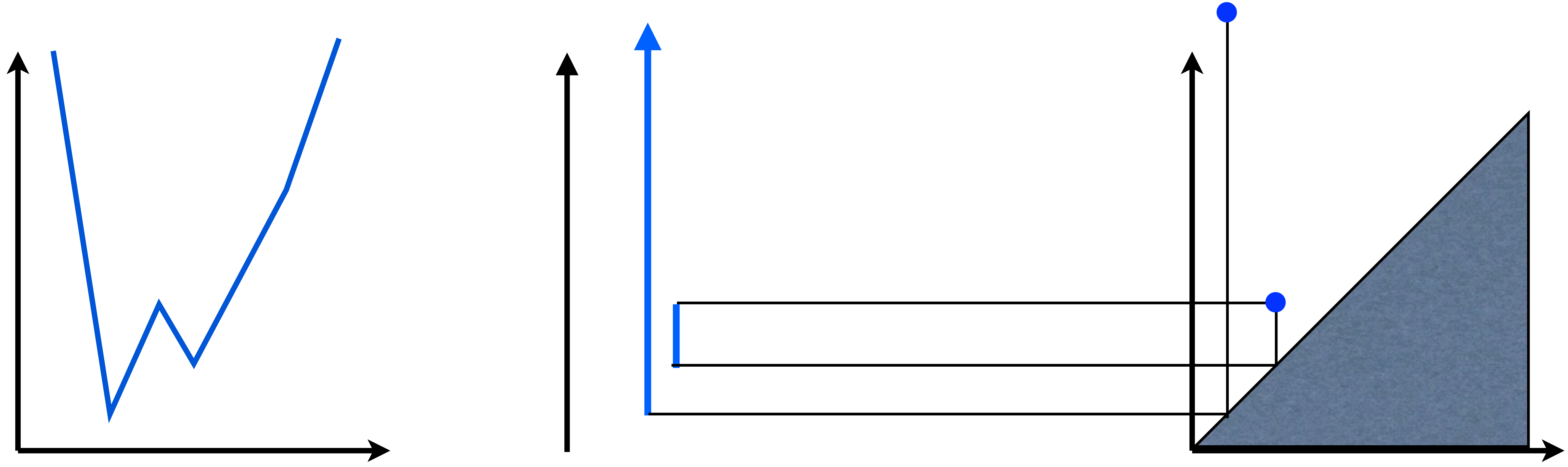


KTH Computer Science and Communication

Barcodes are stable

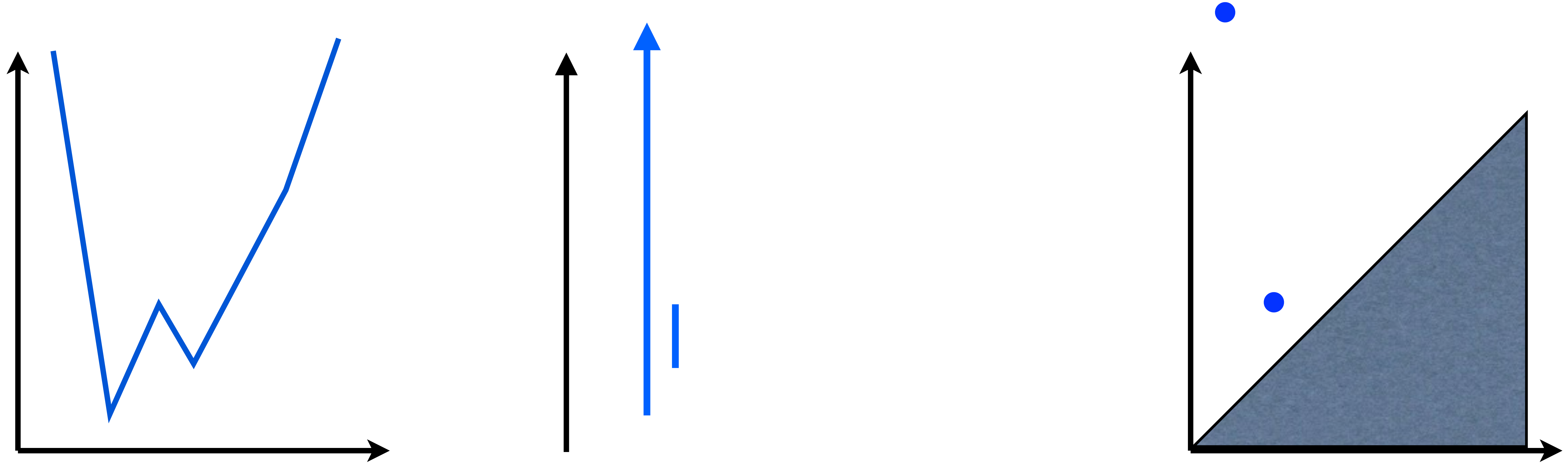


Barcodes are stable



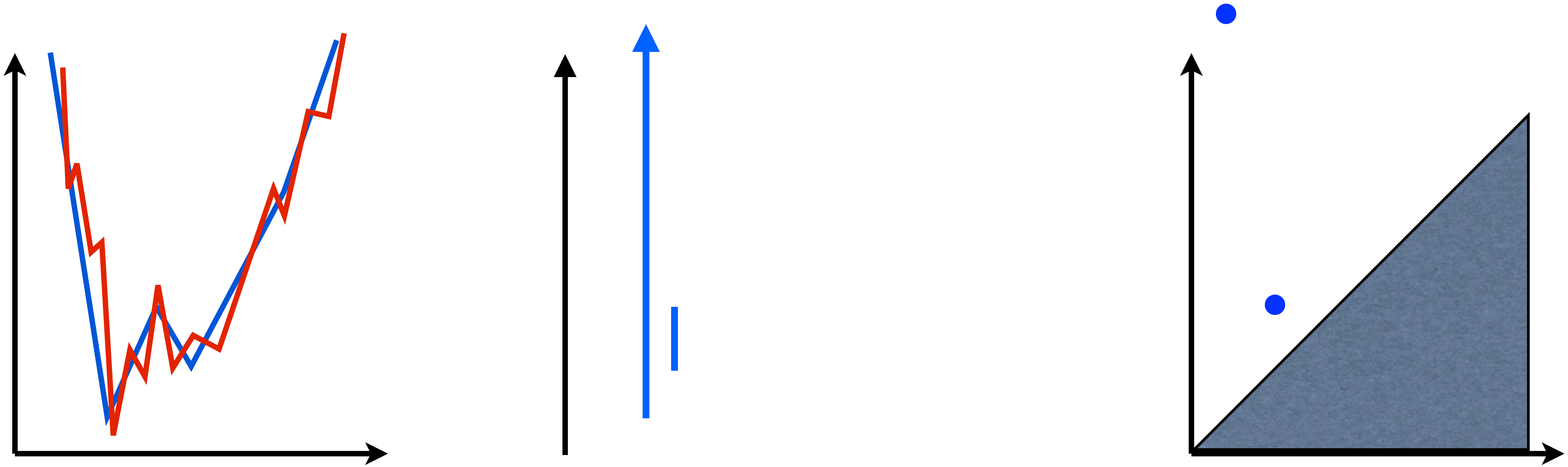
KTH Computer Science and Communication

Barcodes are stable



KTH Computer Science
and Communication

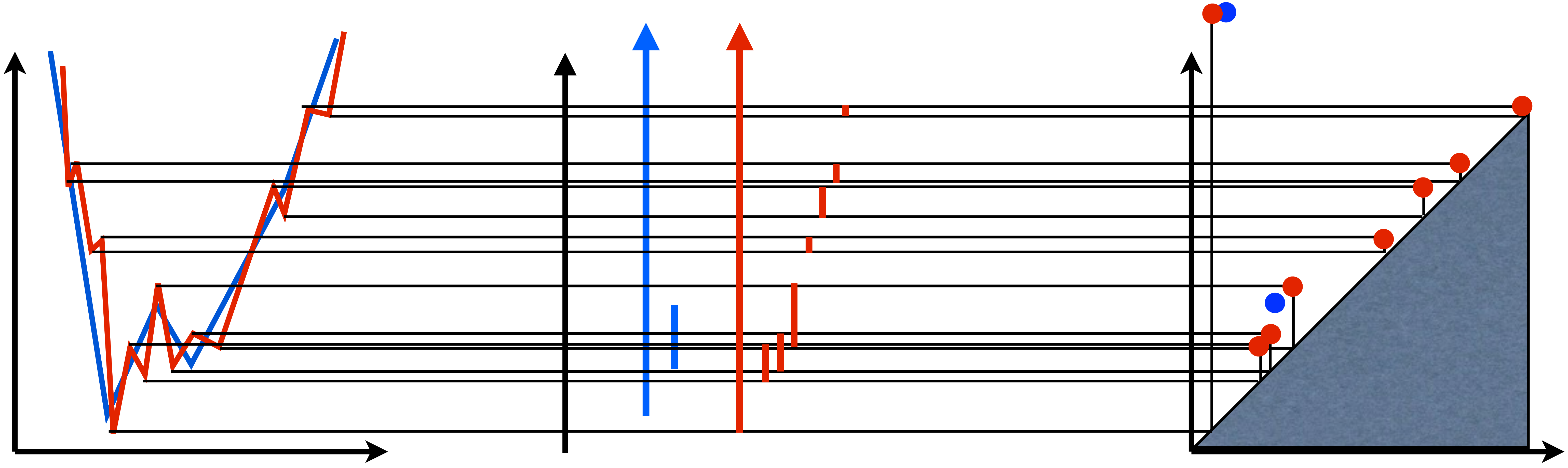
Barcodes are stable



KTH VETENSKAP OCH KONST

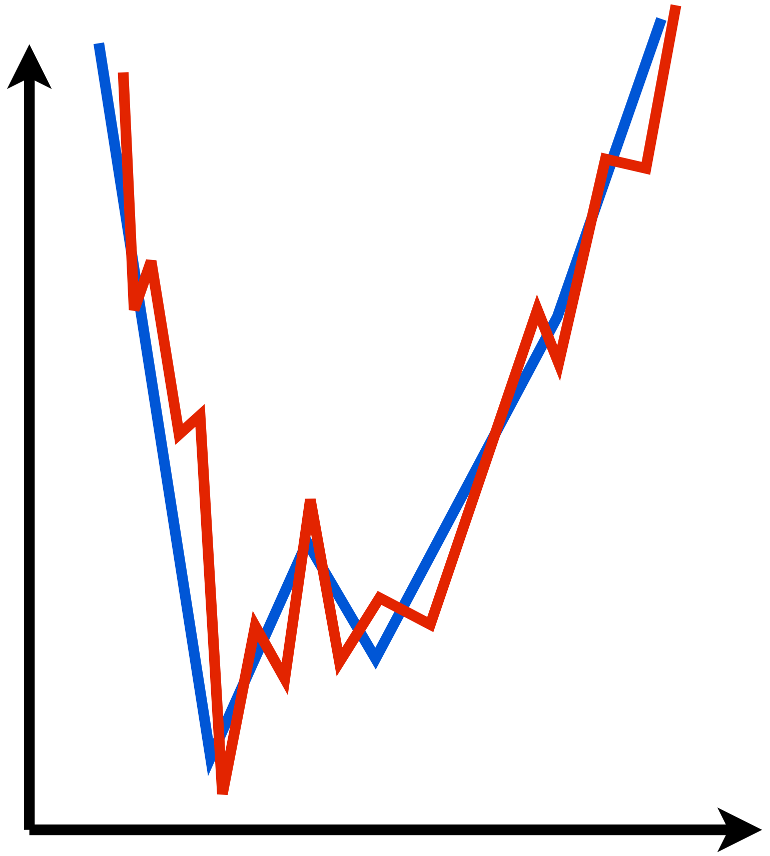
KTH Computer Science and Communication

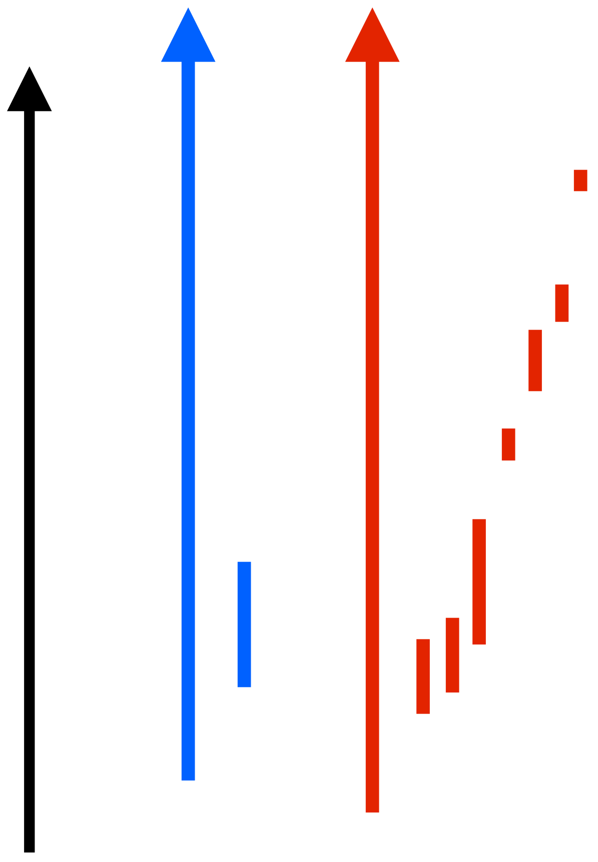
Barcodes are stable

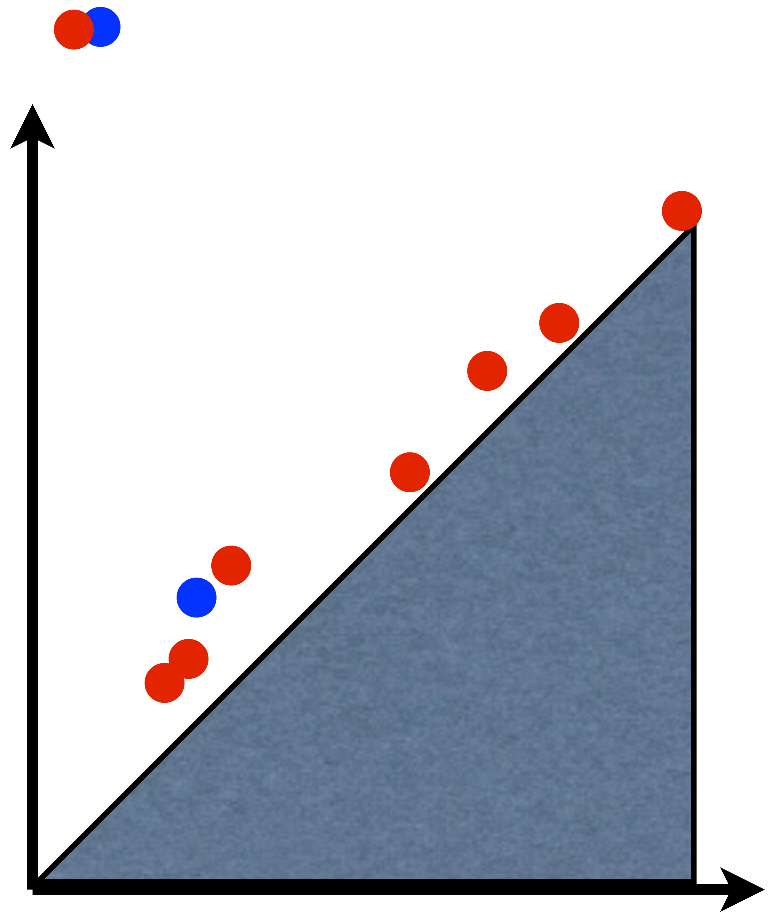


KTH Computer Science and Communication

Barcodes are stable

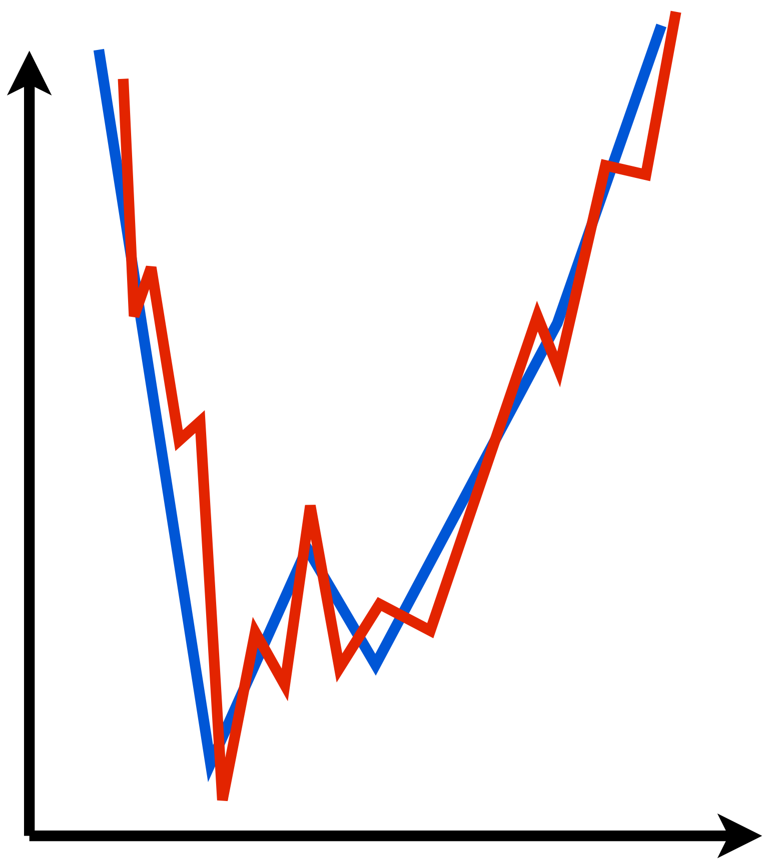




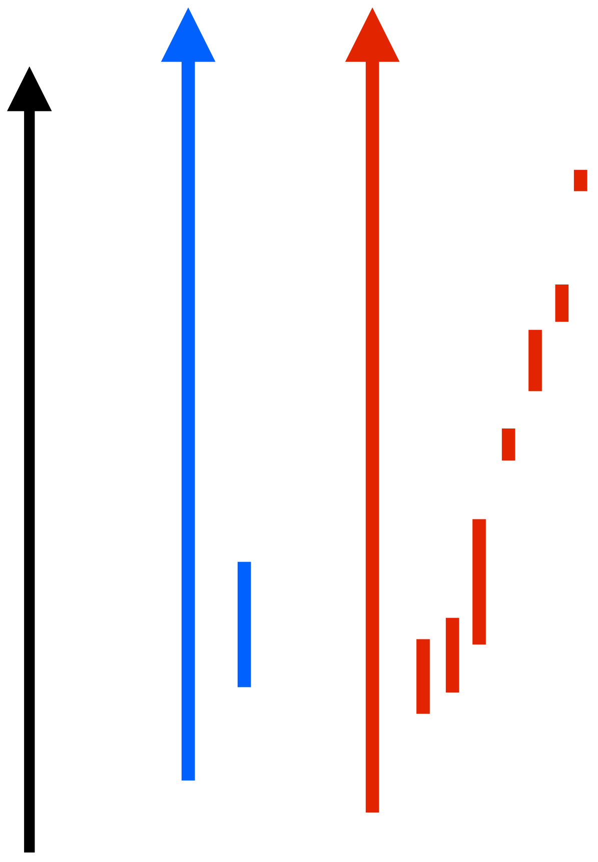


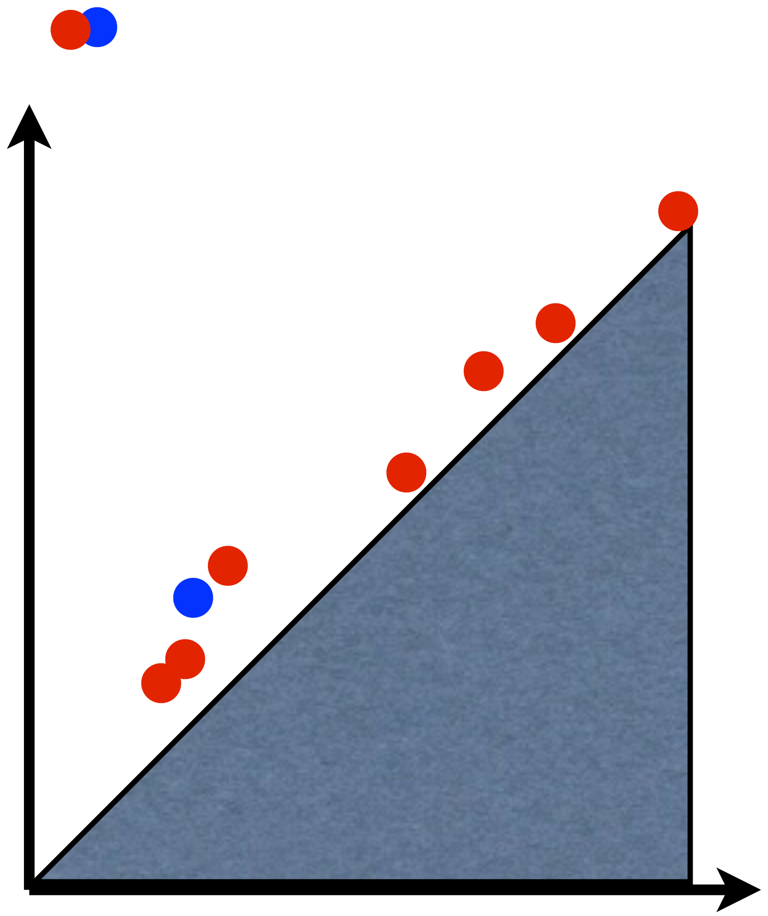
KTH Computer Science and Communication

Barcodes are stable



L_∞ distance of functions





upper bound on maximal distance between points

upper bound on Wasserstein distance

isometric to *interleaving distance*

Stability of persistence

- Persistence is:
  the evolution of the topology of sublevel sets of a function on a (sample from a) manifold.
- Stability of persistence:
  the mapping from the function-and-manifold to the persistence diagram is
  - Continuous
  - 1-Lipschitz
- Therefore: what persistence computes is related to the original topology. Small features go away with small perturbations.

KTH Computer Science and Communication

Persistence of random points

- Zomorodian:
  Random flag complex. β_n and β_{n+2} (mostly) don't overlap
- Adler et.al.:
  Excursion sets of random fields: Snap, crackle, pop.
- Kahle:
  Phase transitions for persistence of random points in $\mathbb{R}^d$ for a wide family of probability distributions.

Distance to measure

- Consistent, stable and easy to compute estimator of topological features of an underlying probability distribution.
- Connect points to their nearest neighbors. Work with the resulting filtered simplicial complex.
- Chazal, Guibas, Oudot, Skraba (2013)

KTH Computer Science and Communication

Machine learning features

- Adcock, Carlsson, Carlsson:
  Classify functions on barcodes to make it easier to pick features
- Berwald, Gidea, V-J:
  Classify different modes of dynamics using ML on persistence barcodes
- Topology as dimensionality reduction.
- Bubenik: Persistence landscapes

KTH Computer Science and Communication

Mean barcode Confidence intervals

- Fasy, Lecci, Rinaldo, Wasserman, Balakrishnan, Singh: Confidence sets for persistence diagrams
- Turner, Mileyko, Mukherjee, Harer: Fréchet means for distributions of persistence diagrams
- Munch, Bendich, Turner, Mukherjee, Mattingly, Harer: Probabilistic Fréchet means and statistics on vineyards
- Mileyko, Mukherjee, Harer: Probability measures on the space of persistence diagrams

New directions

- Machine learning provides tools that can make persistent homology work better.
- I will describe my latest research ideas next: using density estimation to build streaming topological learning.

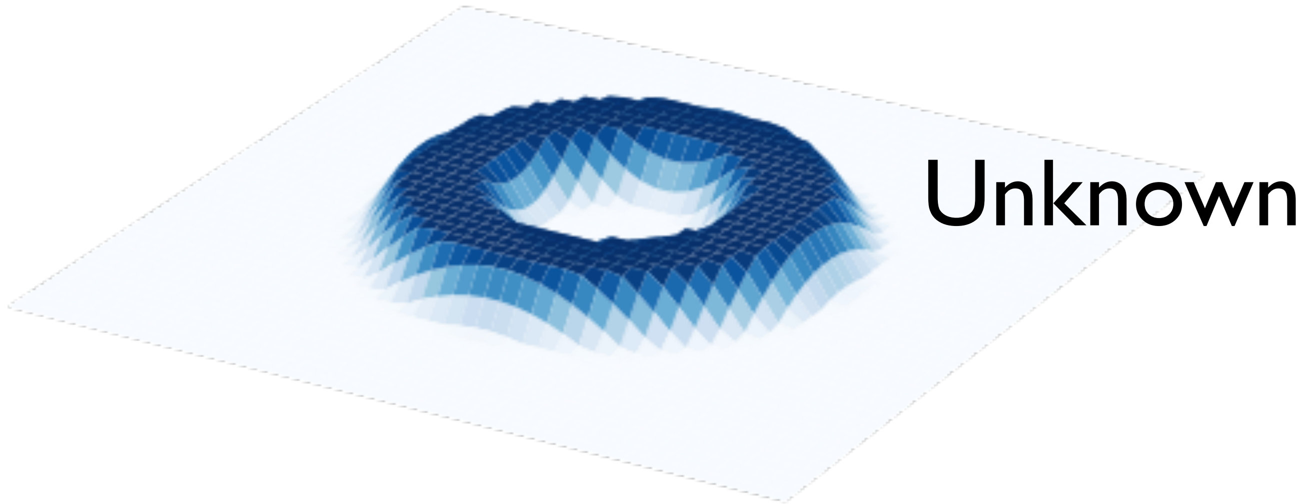
Streaming persistence

KTH Computer Science and Communication

- Inspiring example:
  - Tape accelerometer to a shoe.
  - Measure signal:
    different closed curves for different steps/gaits.
  - Build a topological model of this recurrence.
- Problem:
  - Want to run data capture for a long time.
  - Persistence behaves badly [$O(n^6)$] with size.
  - 100 runs with 100 points faster than one run with 10 000 points.
- We need to control input size to persistence.

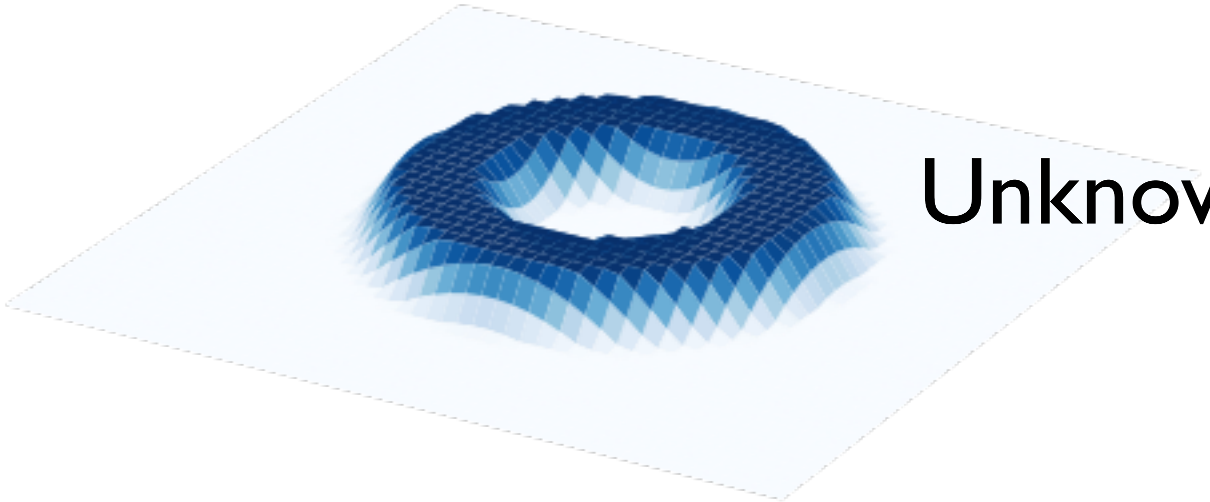
KTH Computer Science and Communication

Streaming persistence



Streaming persistence

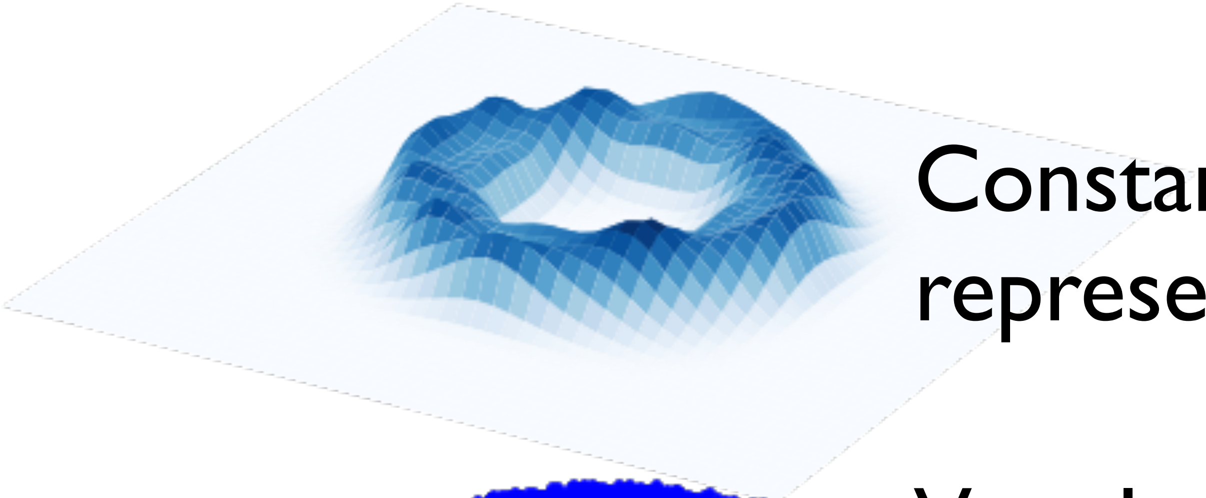
Very large
Appears gradually



Unknown

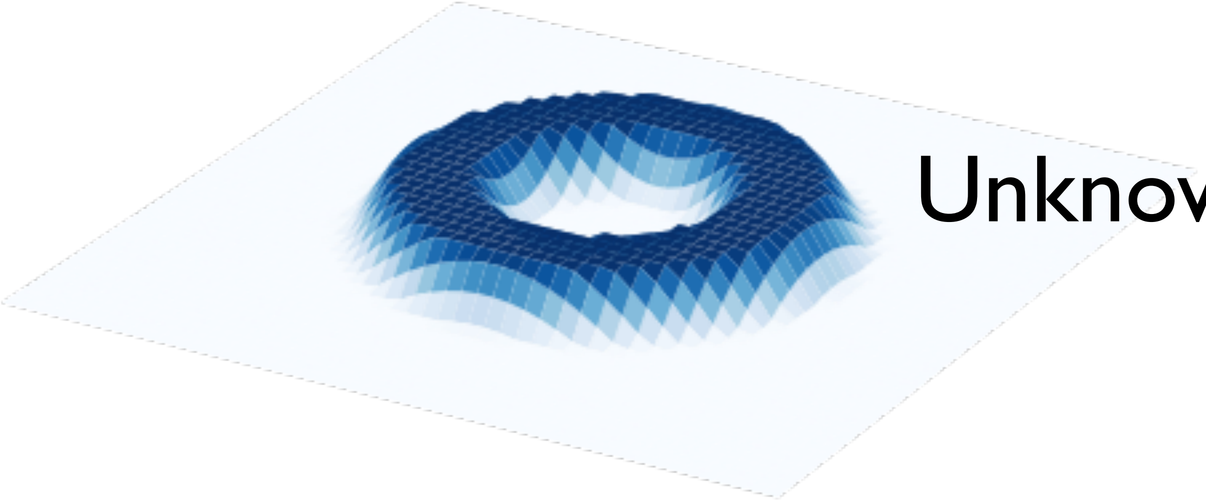
Streaming persistence

KTH Computer Science and Communication



Constant size
representation

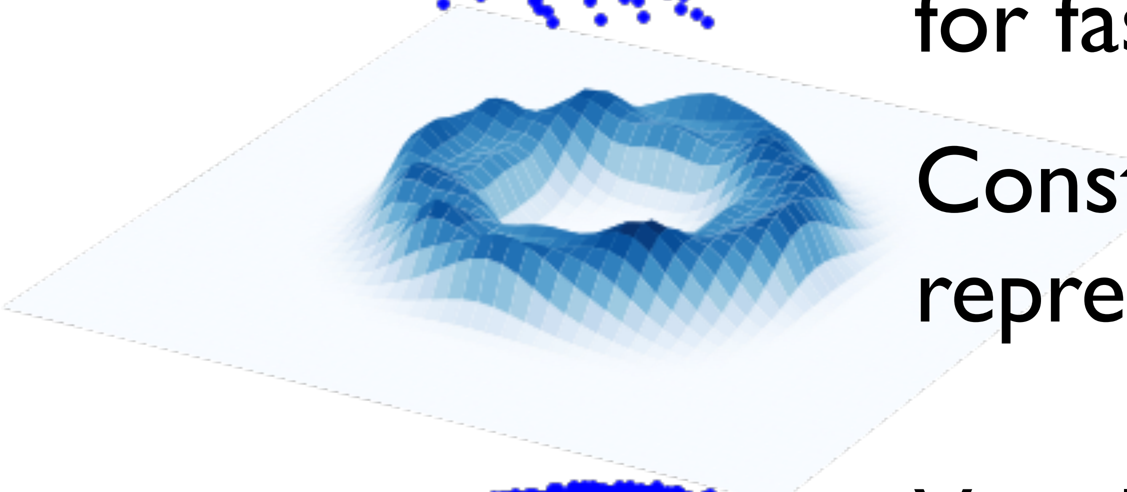
Very large
Appears gradually



Unknown

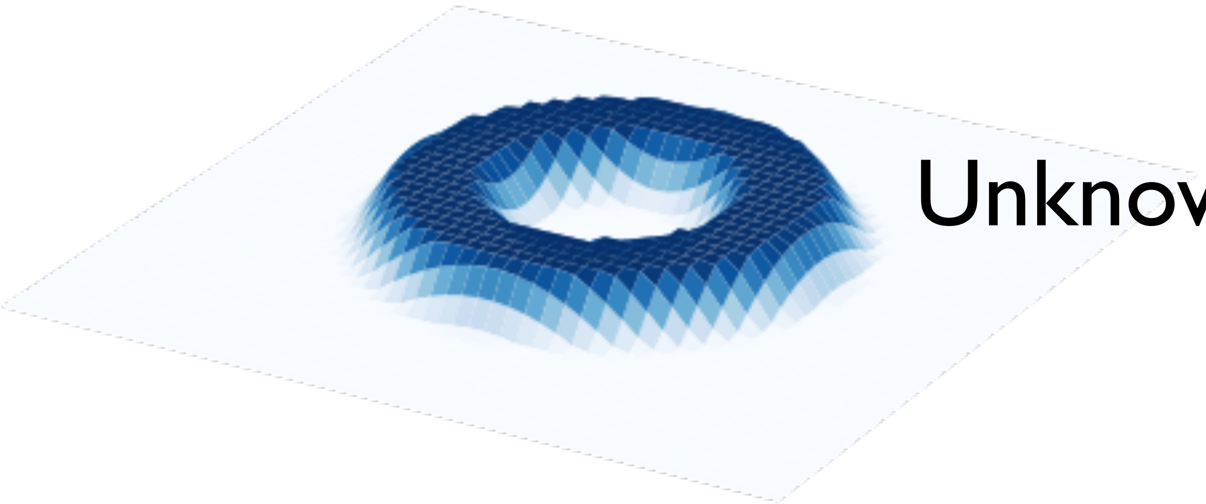
Streaming persistence

Small enough
for fast persistence



Constant size
representation

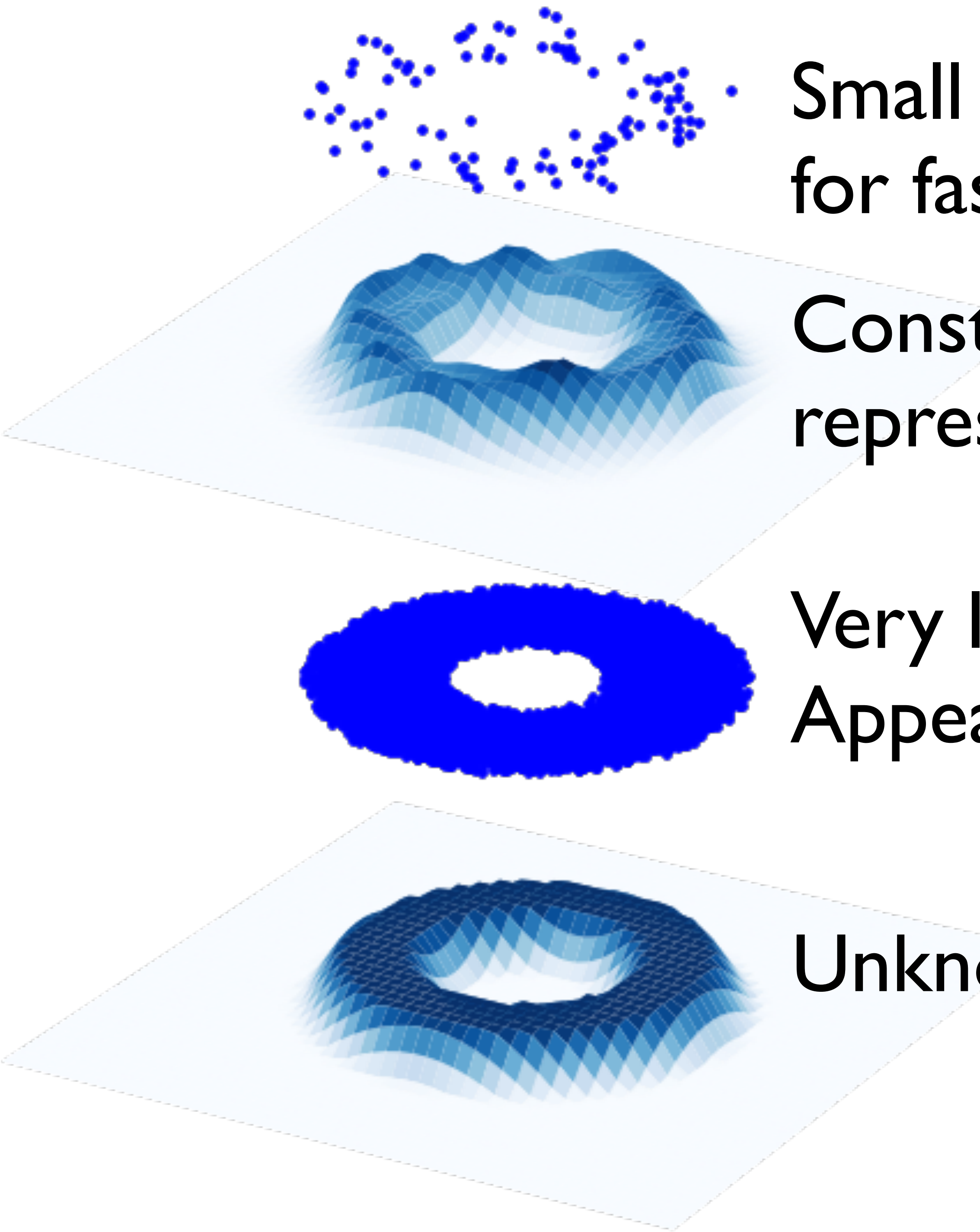
Very large
Appears gradually



Unknown

KTH Computer Science and Communication

Streaming persistence

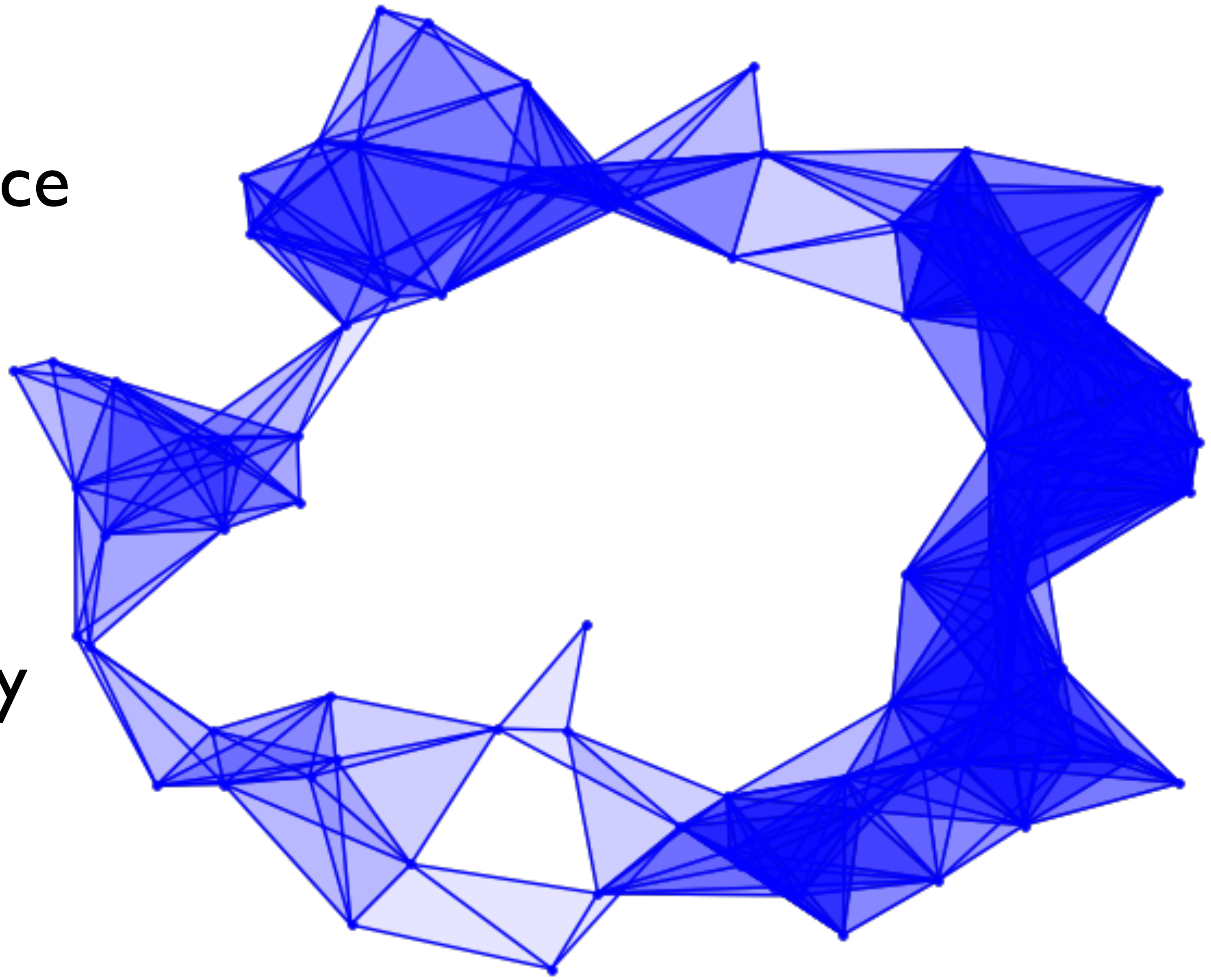


Small enough
for fast persistence

Constant size
representation

Very large
Appears gradually

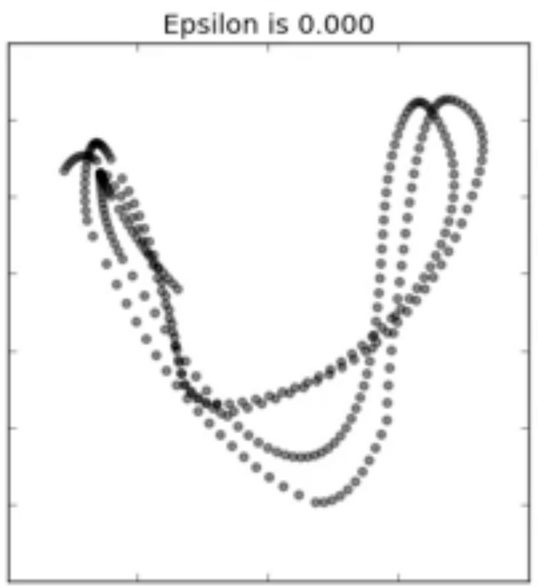
Unknown

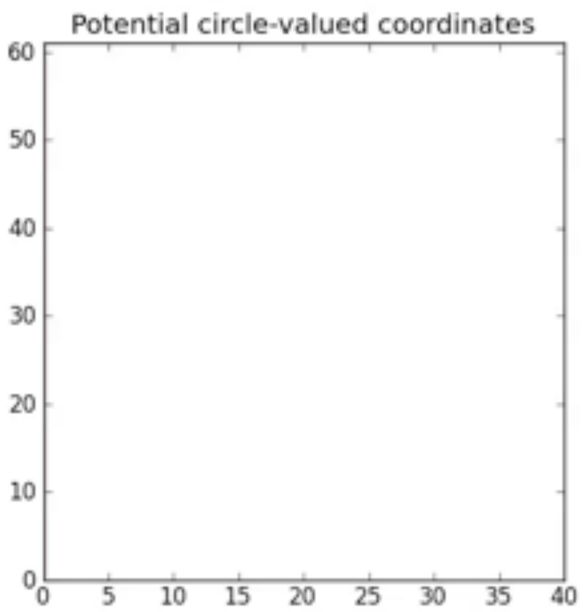


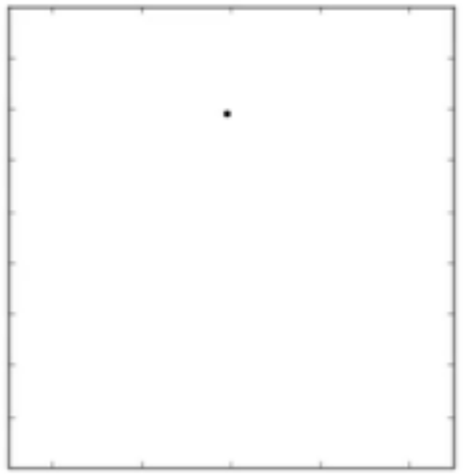
KTH Computer Science
and Communication

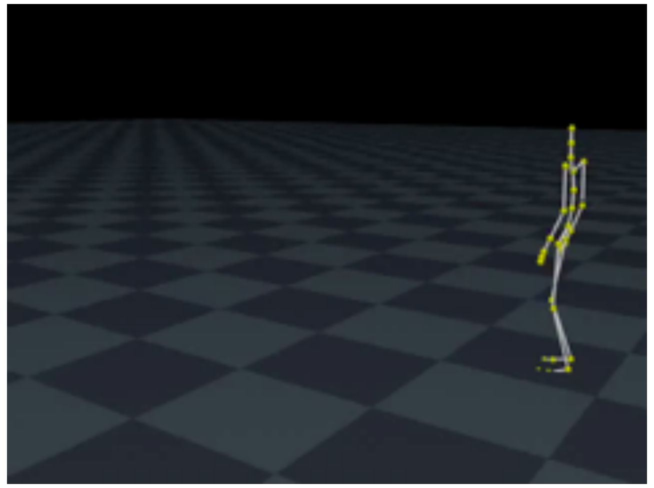
Cohomology

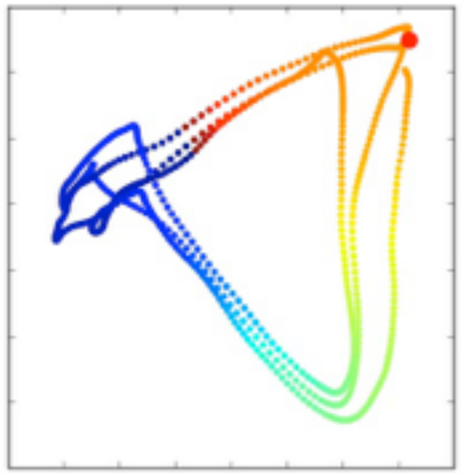
- Produces a topological system for generating circle-valued coordinate functions.
- This improves on state-of-the-art for analyzing recurrence.
- Applicable in motion capture and gait analysis.

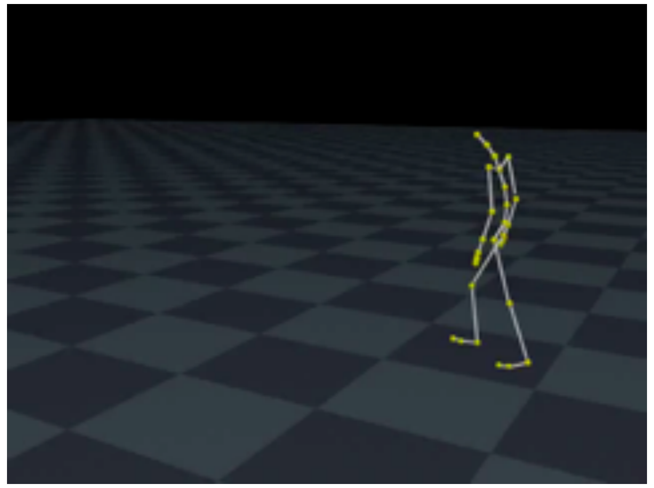






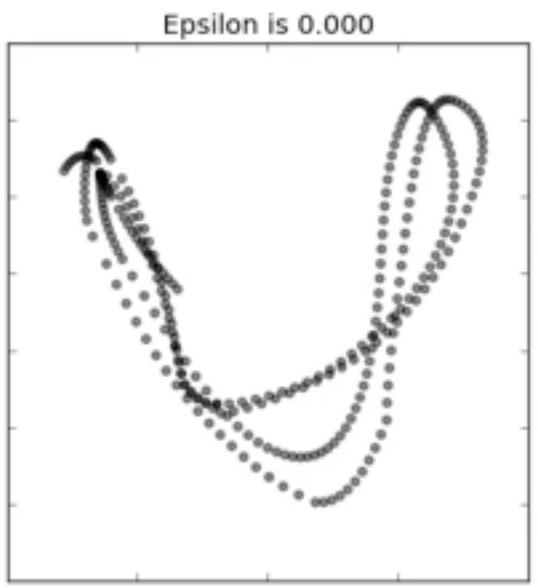


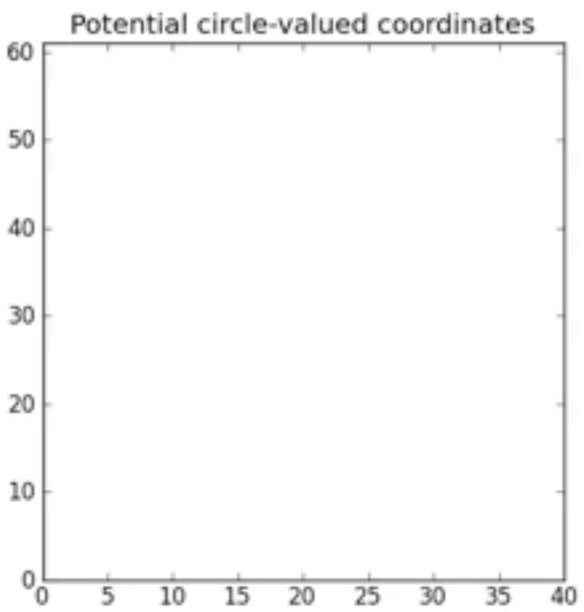




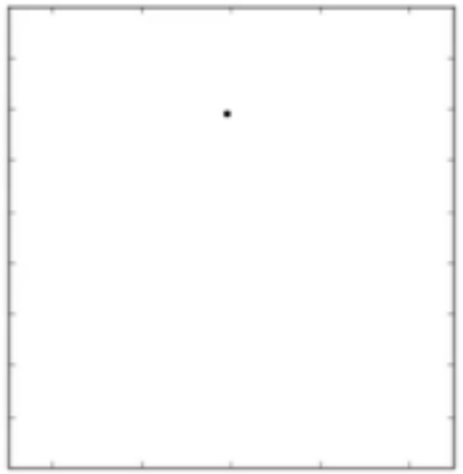
KTH Computer Science and Communication

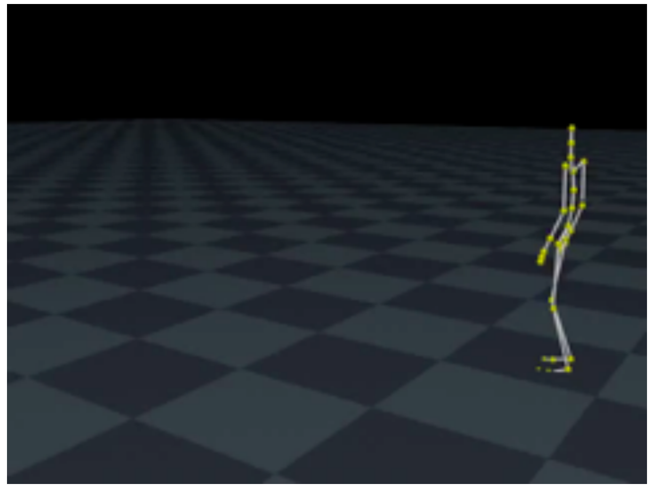
Cohomology

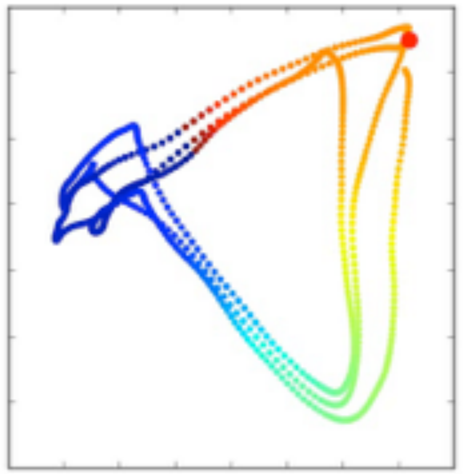


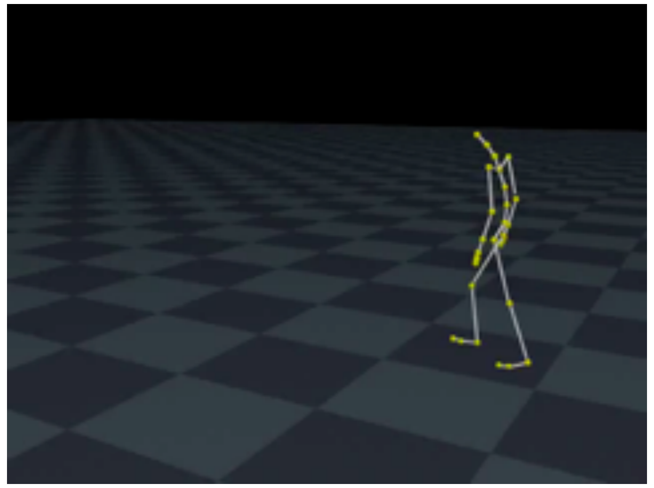


- Produces a topological system for generating circle-valued coordinate functions.
- This improves on state-of-the-art for analyzing recurrence.
- Applicable in motion capture and gait analysis.





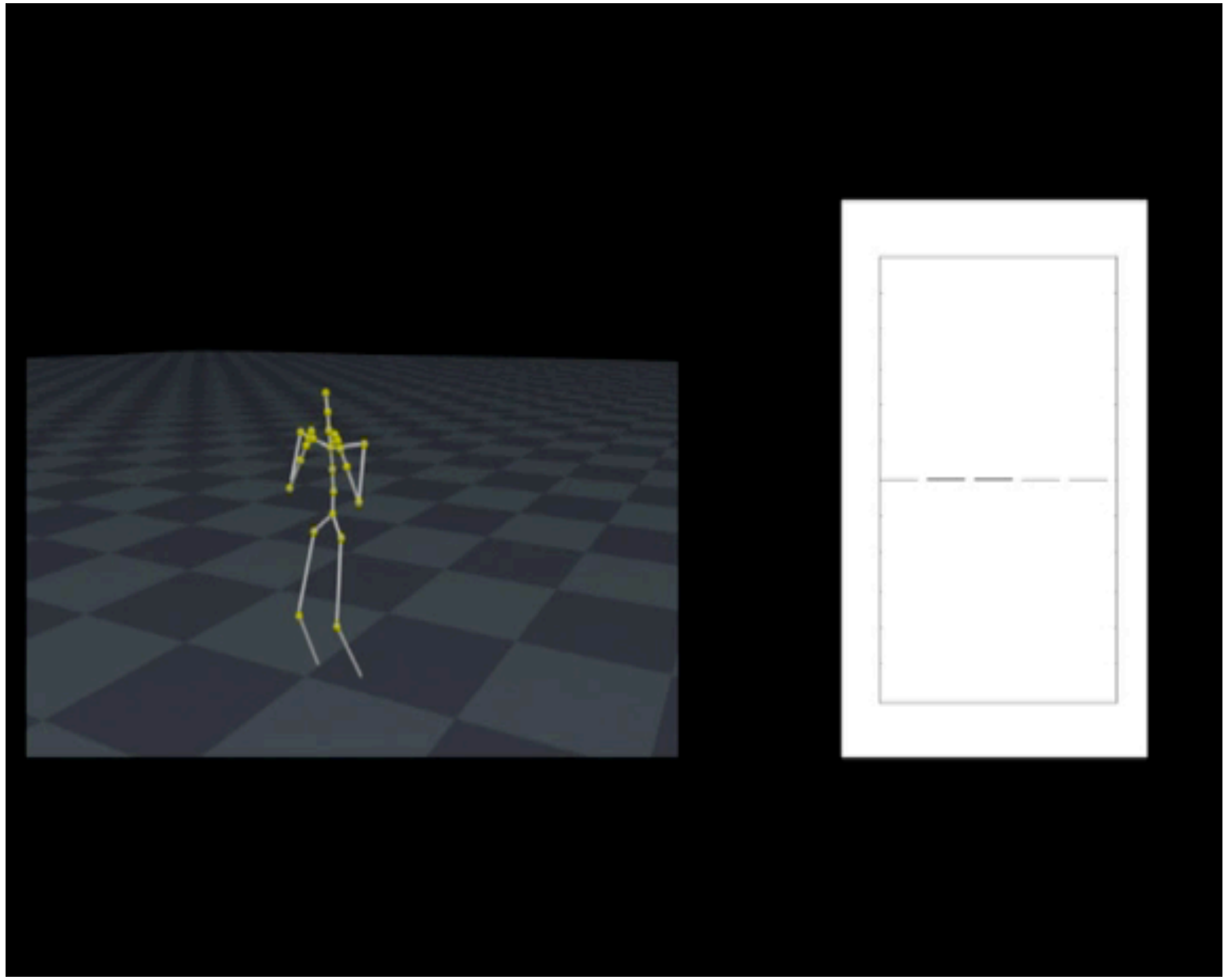




KTH Computer Science
and Communication

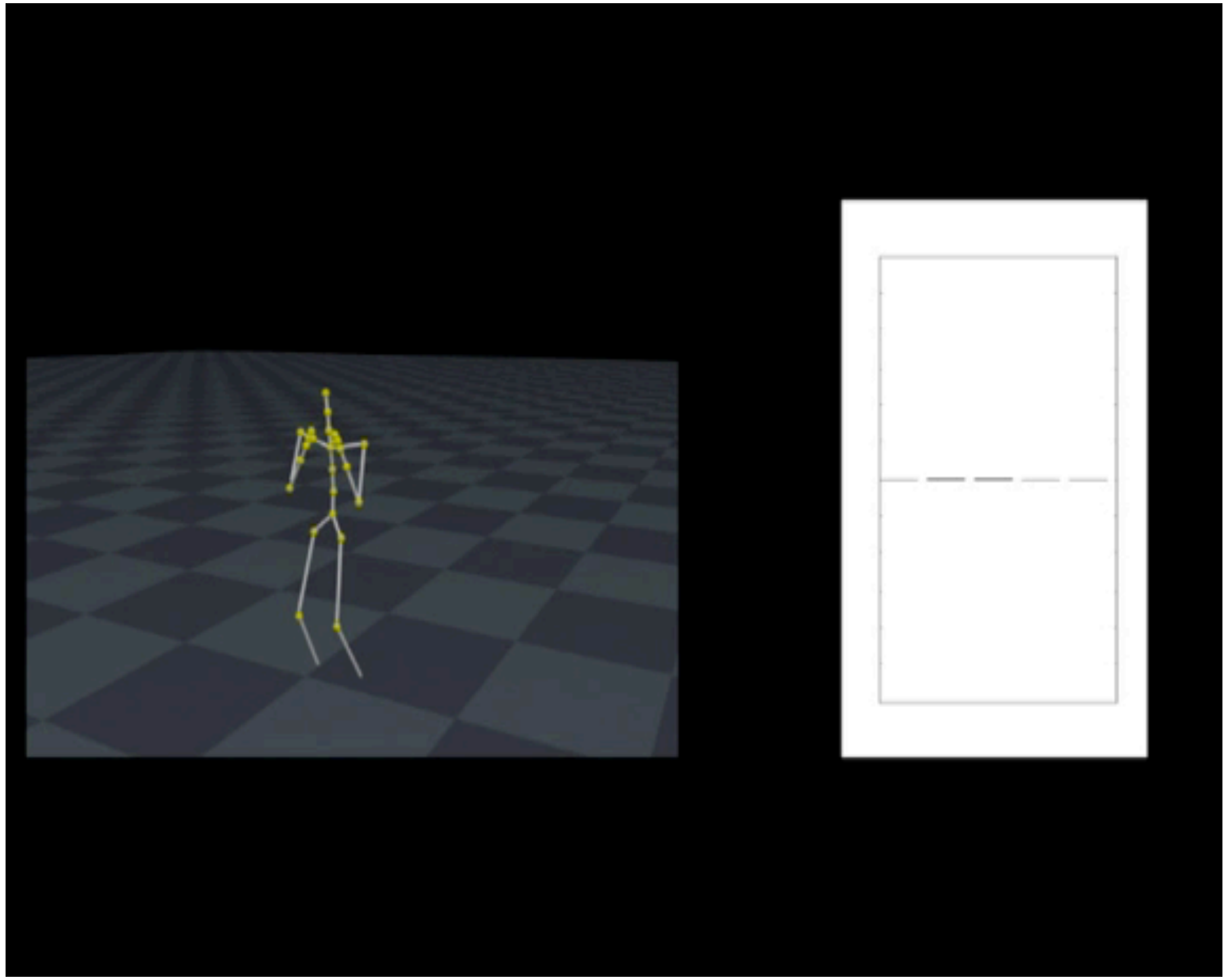
Indicator / feature functions on recurrence

- The functions generated by cohomology work as indicators of different recurrent features.

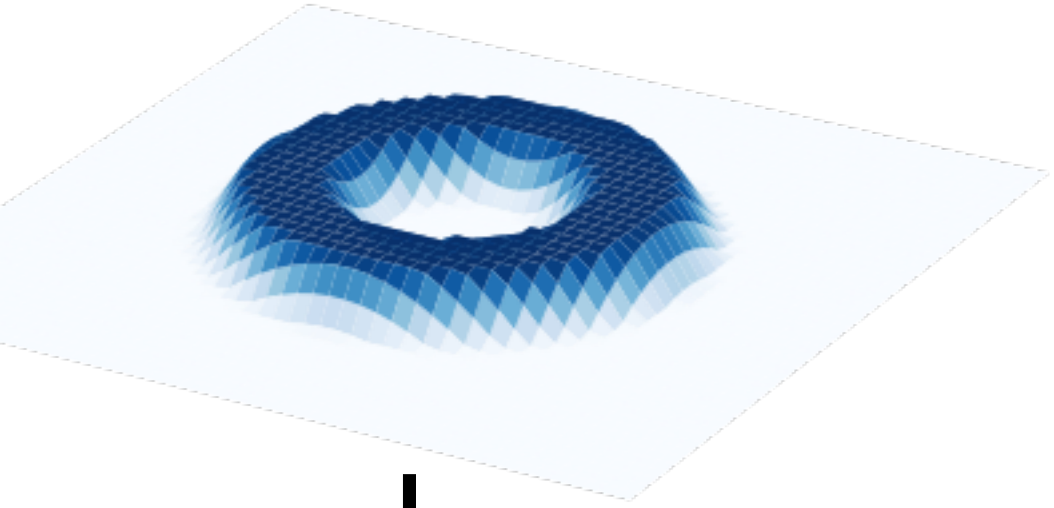


Indicator / feature functions on recurrence

- The functions generated by cohomology work as indicators of different recurrent features.



Idea for a pipeline for online topological learning



Online density estimation

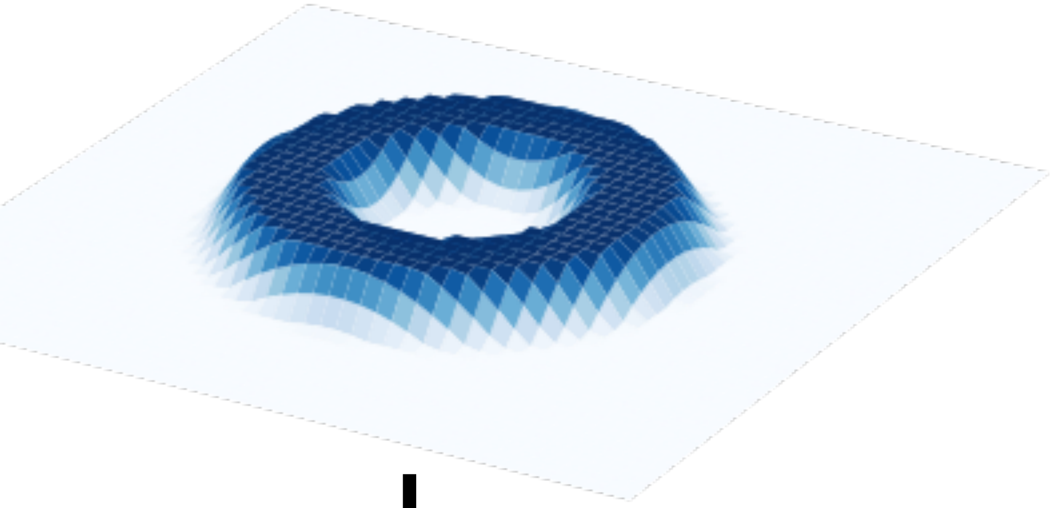
$\varphi(a_t)$

$\varphi(a_t)$

$\varphi(a_t)$

KTH Computer Science
and Communication

Idea for a pipeline for online topological learning



Online density estimation

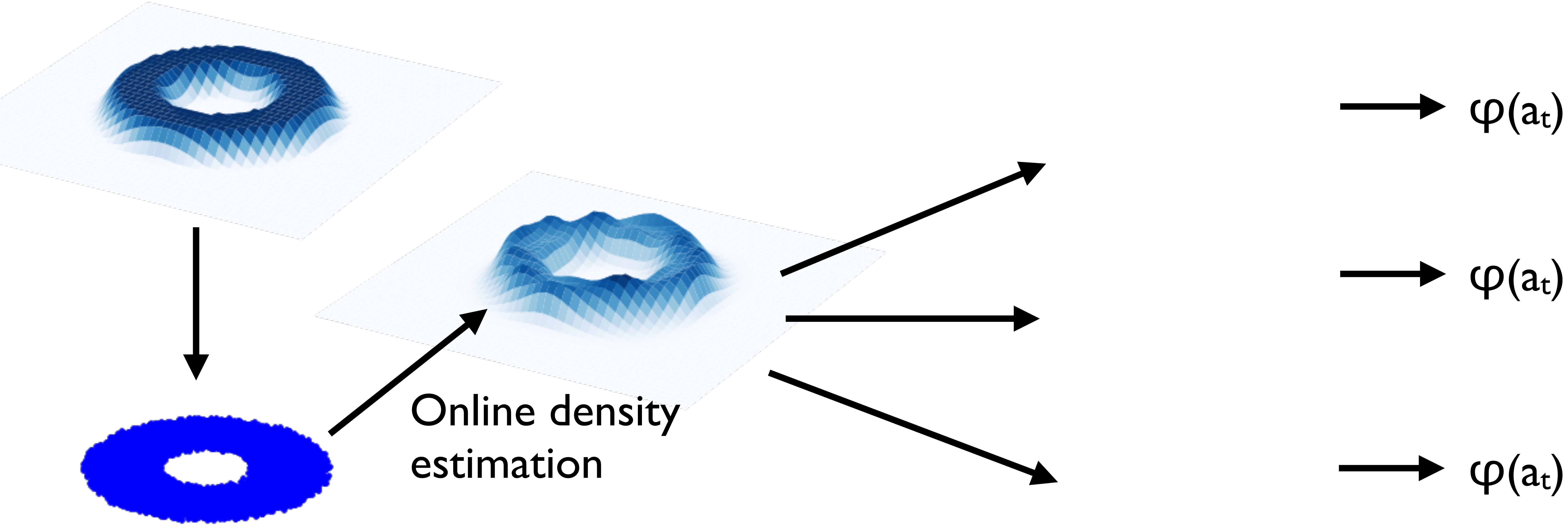
$\varphi(a_t)$

$\varphi(a_t)$

$\varphi(a_t)$

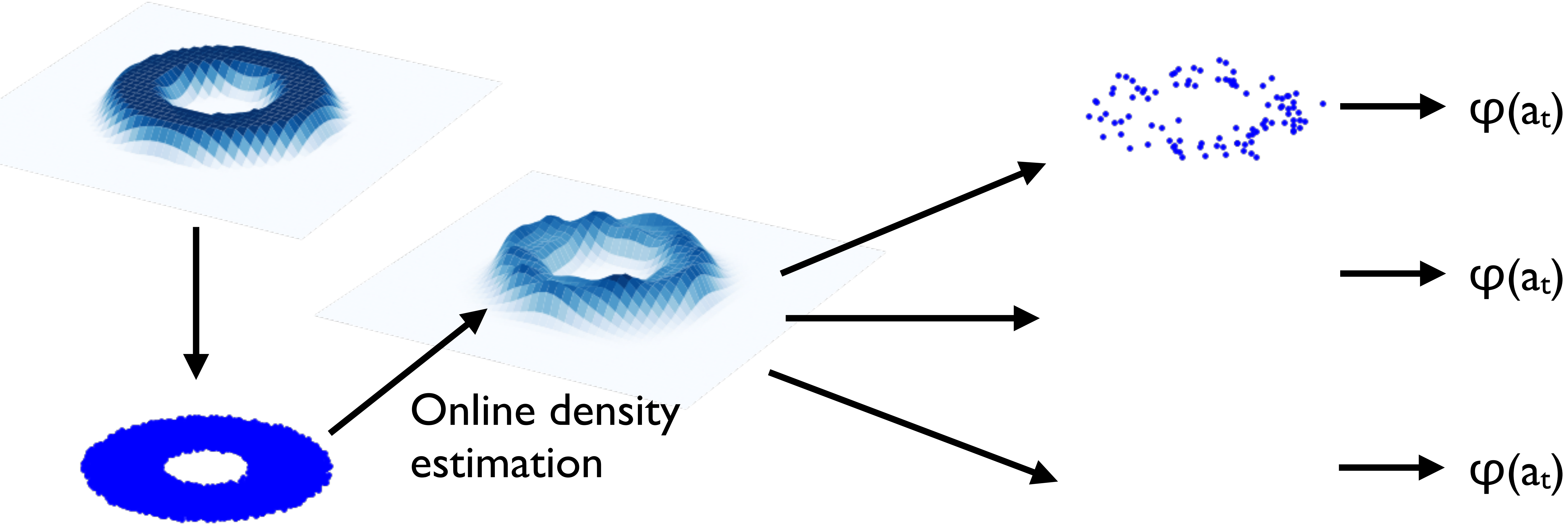
KTH Computer Science and Communication

Idea for a pipeline for online topological learning

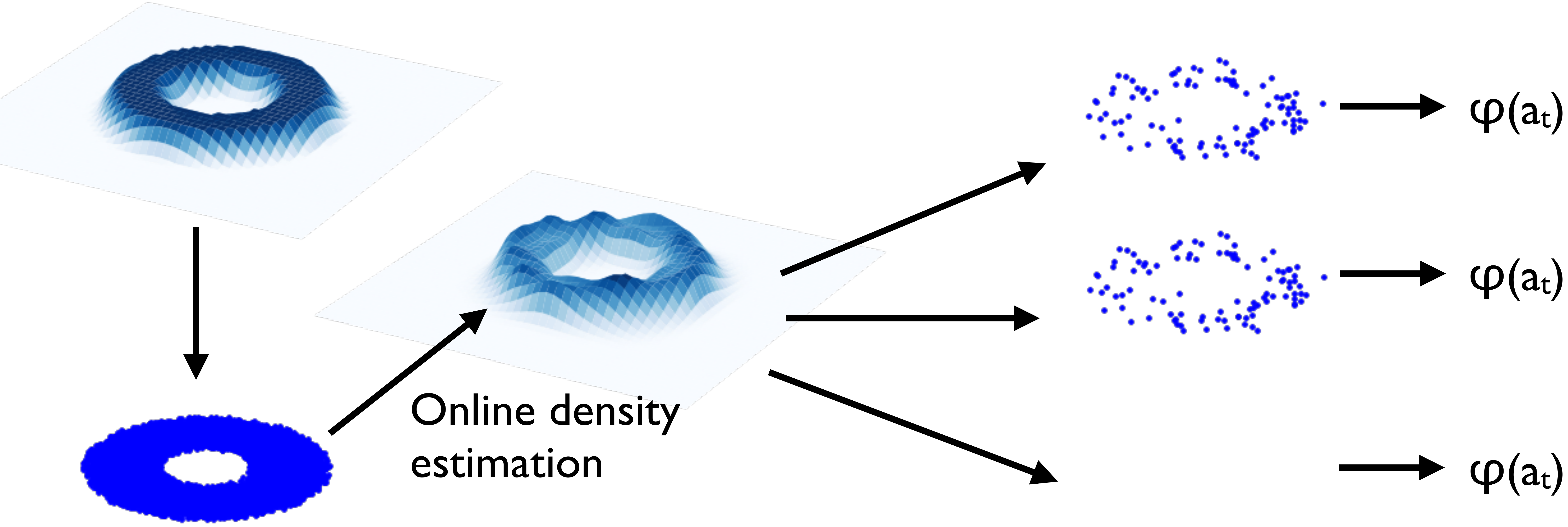


KTH Computer Science and Communication

Idea for a pipeline for online topological learning

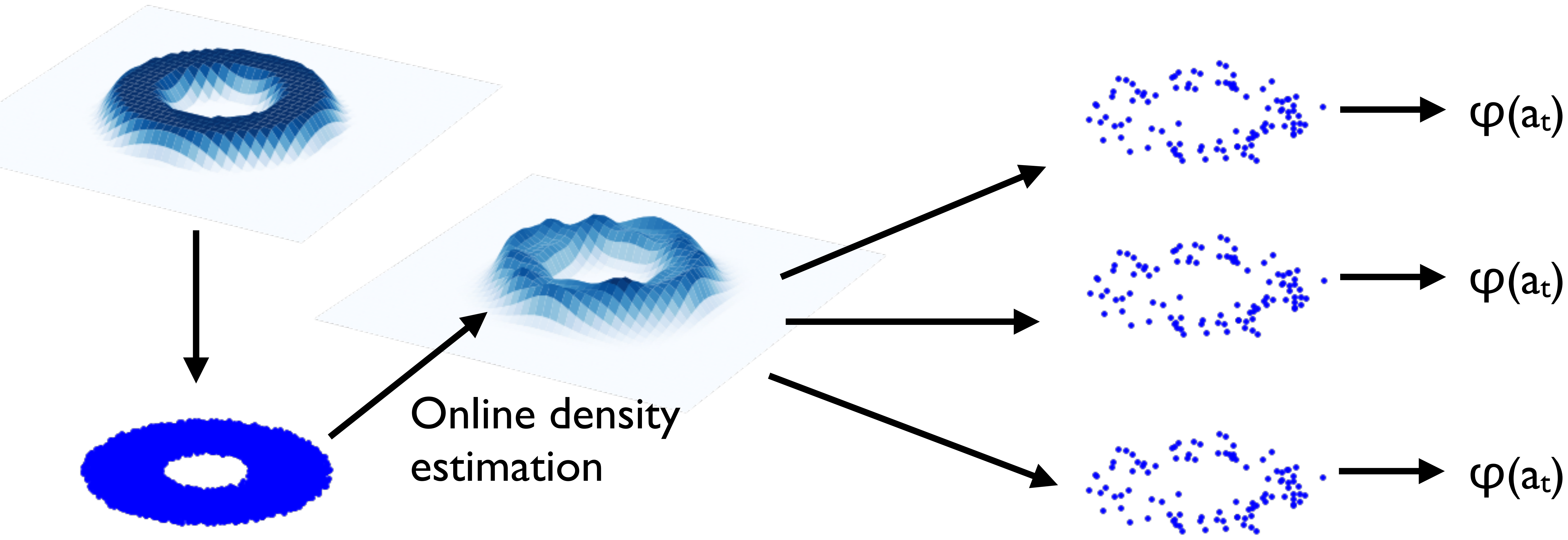


Idea for a pipeline for online topological learning

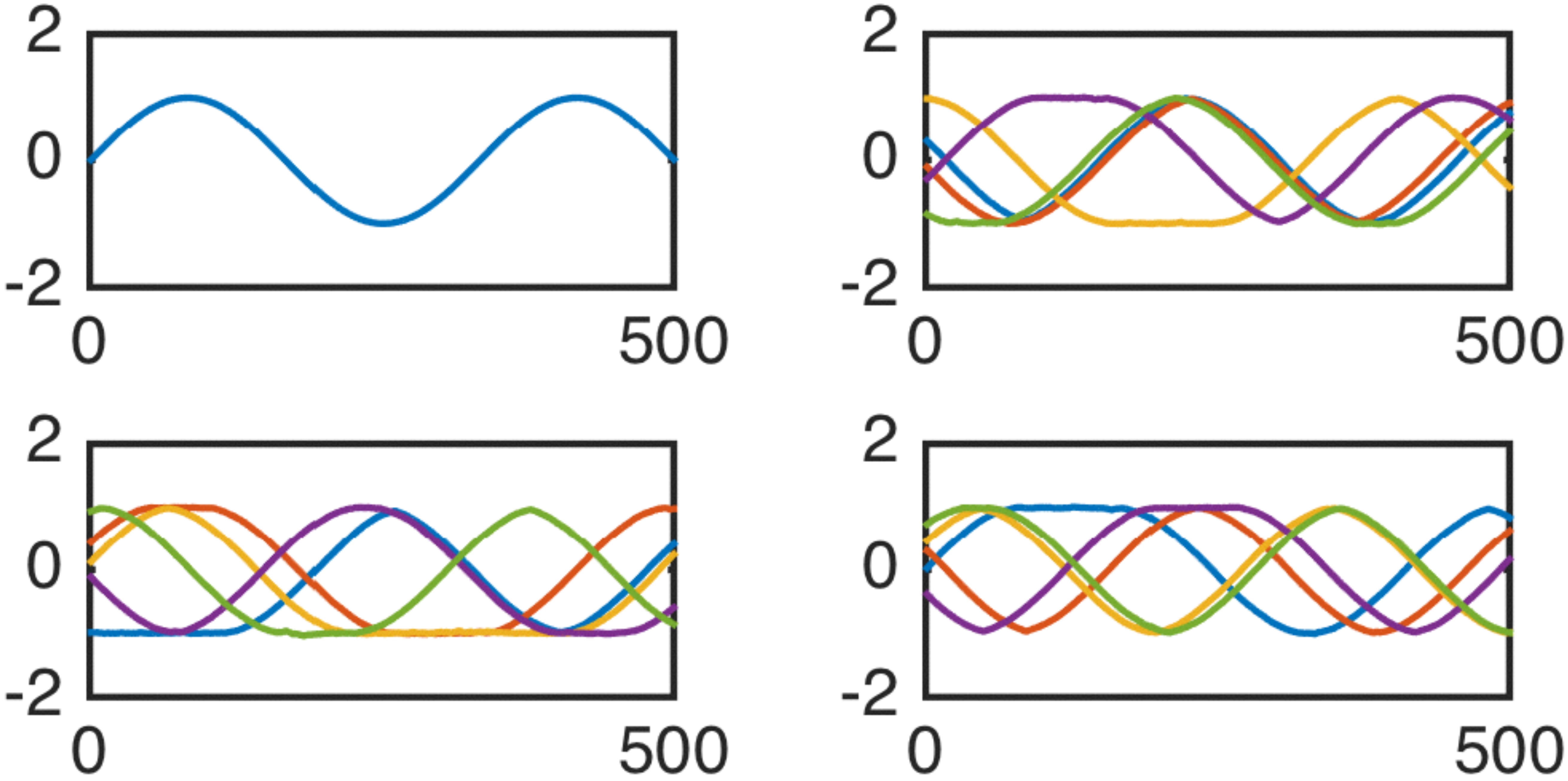


KTH Computer Science and Communication

Idea for a pipeline for online topological learning



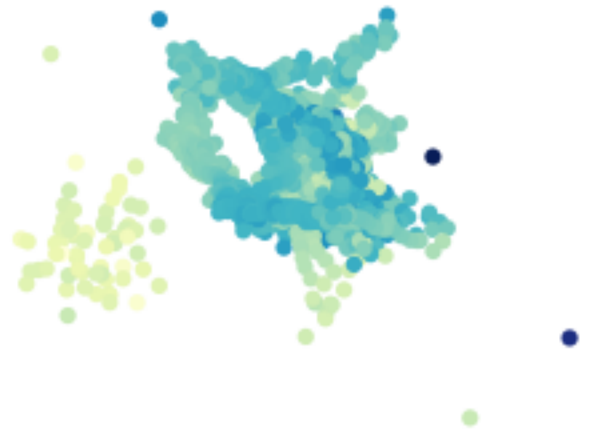
Current state



1-d seems doable: learned curves qualitatively similar

KTH Computer Science and Communication

Current state

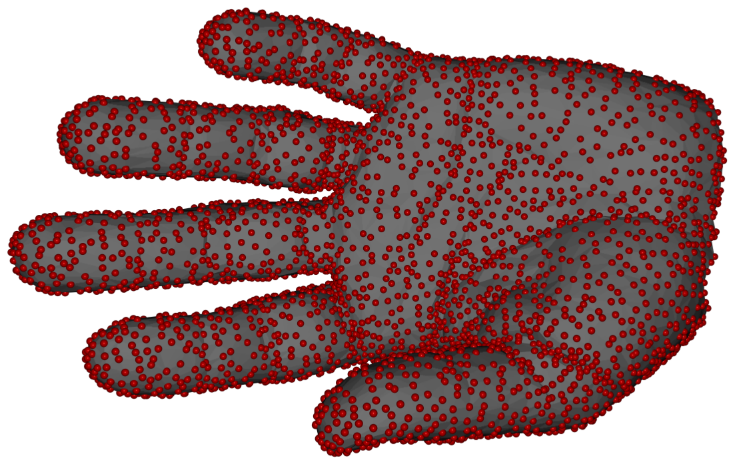


3d with Gaussian Mixture Models? Not quite…

Next up: try GPLVM
Gaussian Process Latent Variable Models

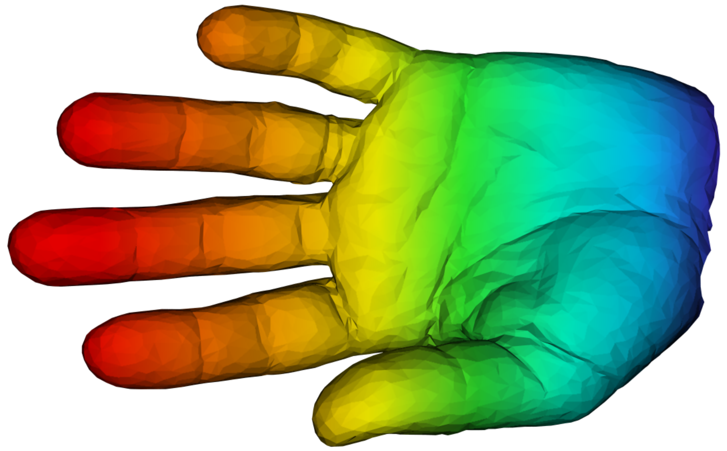
If anyone has any good ideas — **tell me**!

Mapper



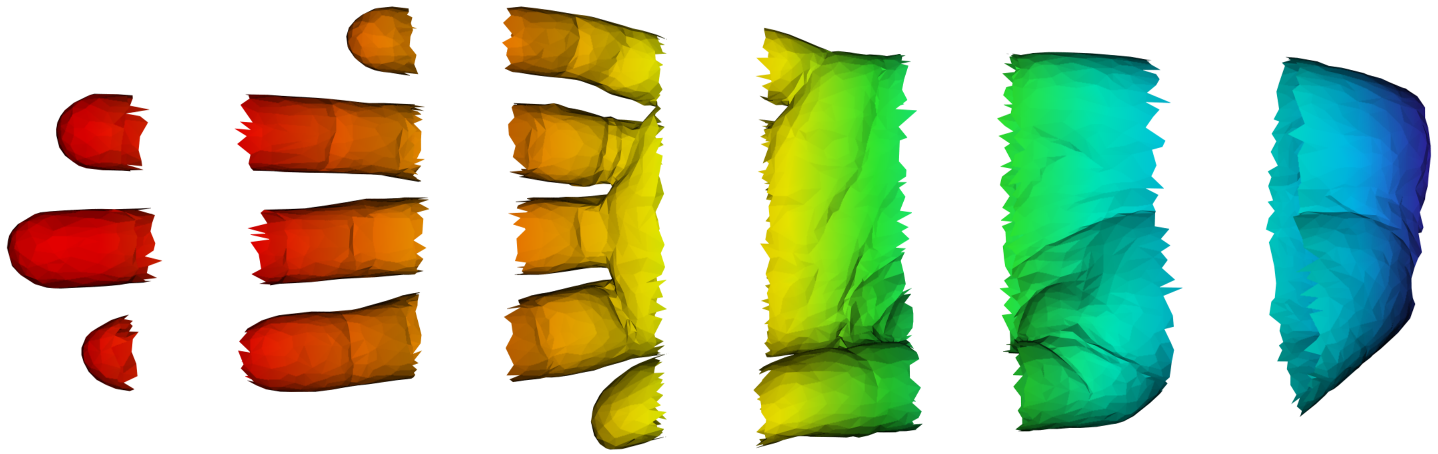
KTH Computer Science and Communication

Mapper

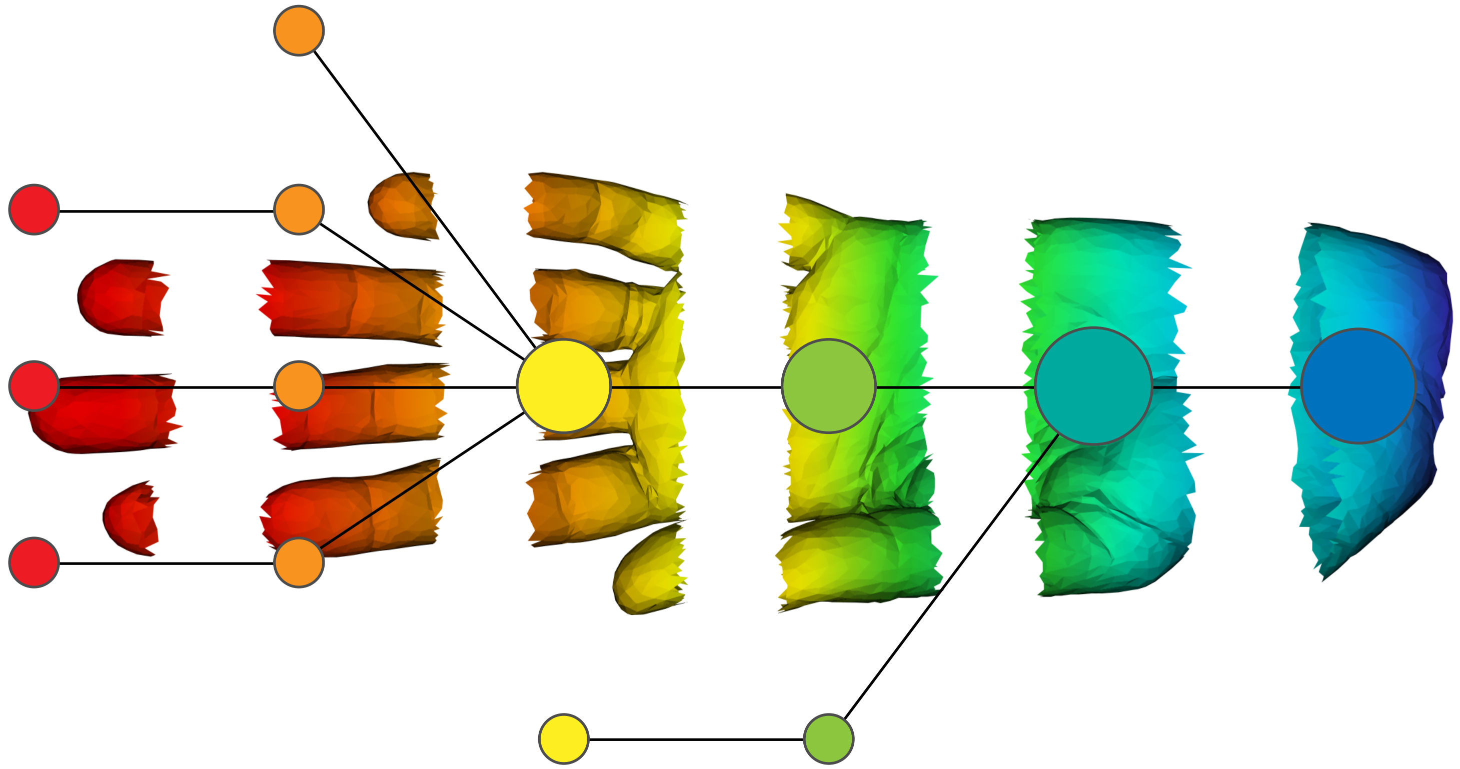


KTH Computer Science and Communication

Mapper

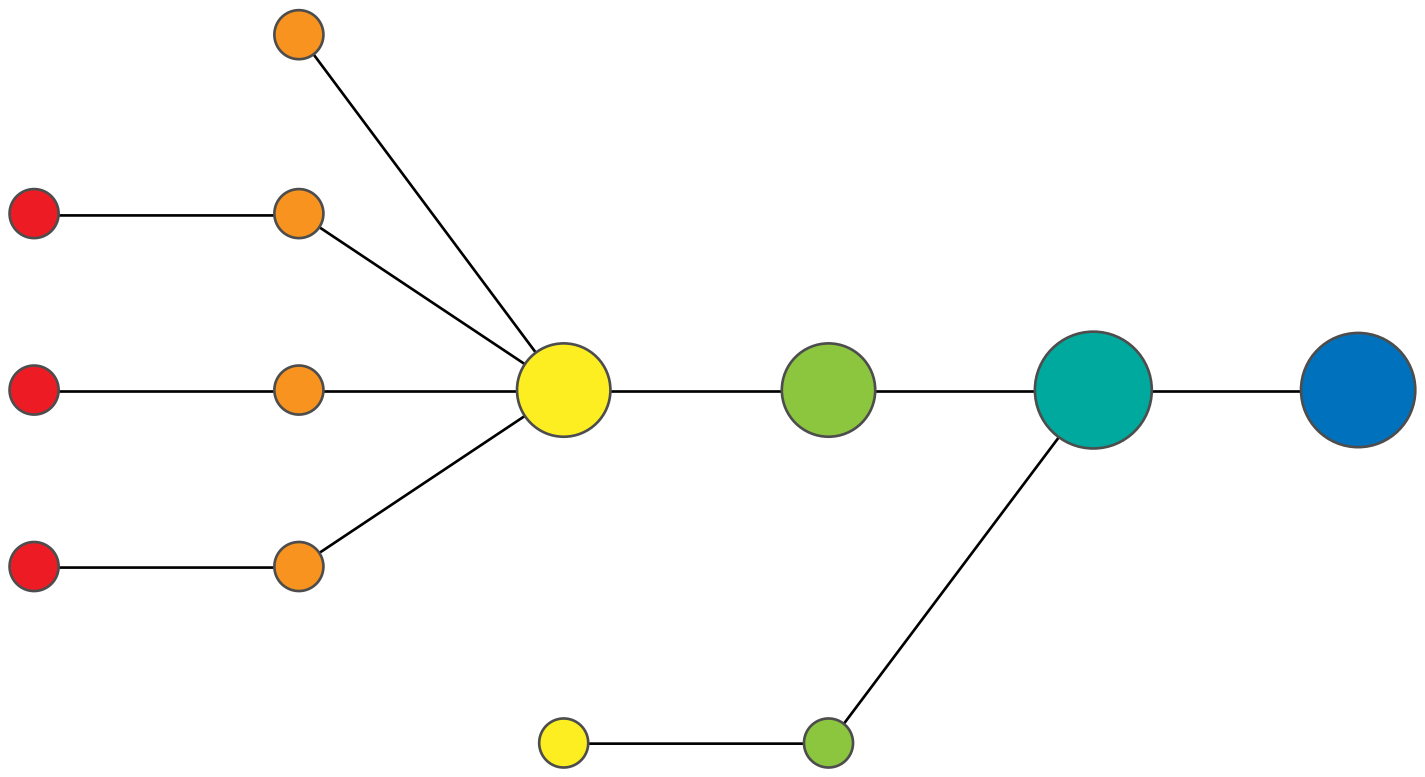


Mapper



KTH Computer Science and Communication

Mapper



Mapper

- Builds automatically a discrete model.
- Stable under permutation of data.
  Under permutation of filter functions.
- Persistent topology along varying feature functions:
  active research field.
- Strong potential for novel model creation for ML application.
  Going beyond Data Analysis.

KTH Computer Science
and Communication

Thank you for listening

شكرا لإصغائكم