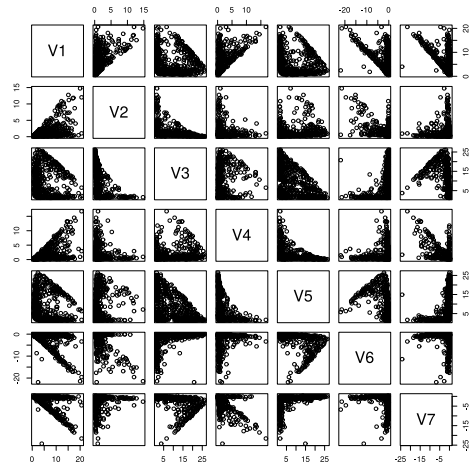
Structural Analysis of Multivariate Point Clouds using Simplicial Chains

Bastian Rieck Heike Leitte

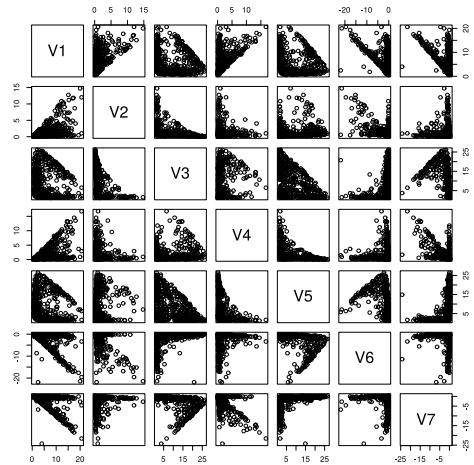
Interdisciplinary Center for Scientific Computing
Heidelberg University

Motivation



SPLOM

Motivation

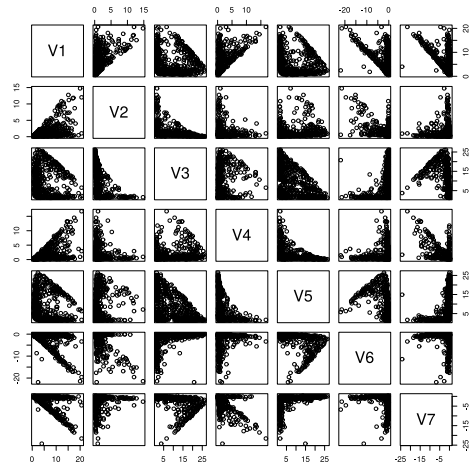


SPLOM

Common workflow:

- Use e.g. SPLOM to make sense of data.

Motivation

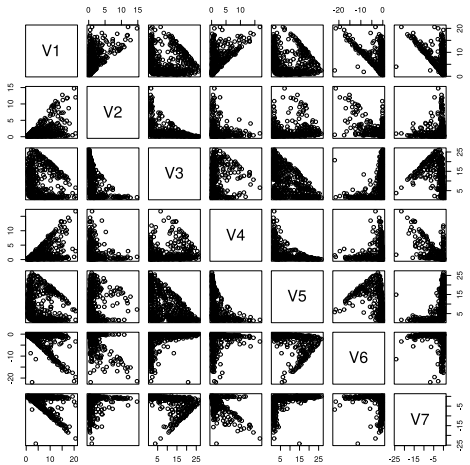


SPLOM

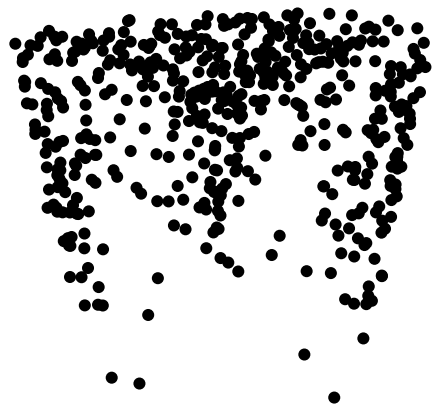
Common workflow:

- Use e.g. SPLOM to make sense of data.
- Use dimensionality reduction methods to obtain visualization for exploratory data analysis.

Motivation



SPLOM

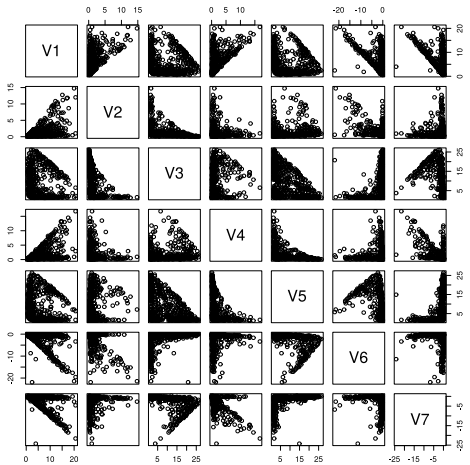


MDS

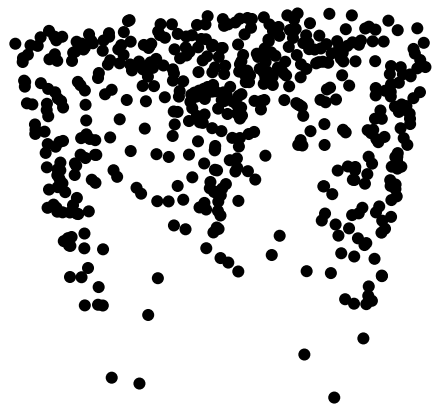
Common workflow:

- Use e.g. SPLOM to make sense of data.
- Use dimensionality reduction methods to obtain visualization for exploratory data analysis.

Motivation



SPLOM



MDS

Isomap

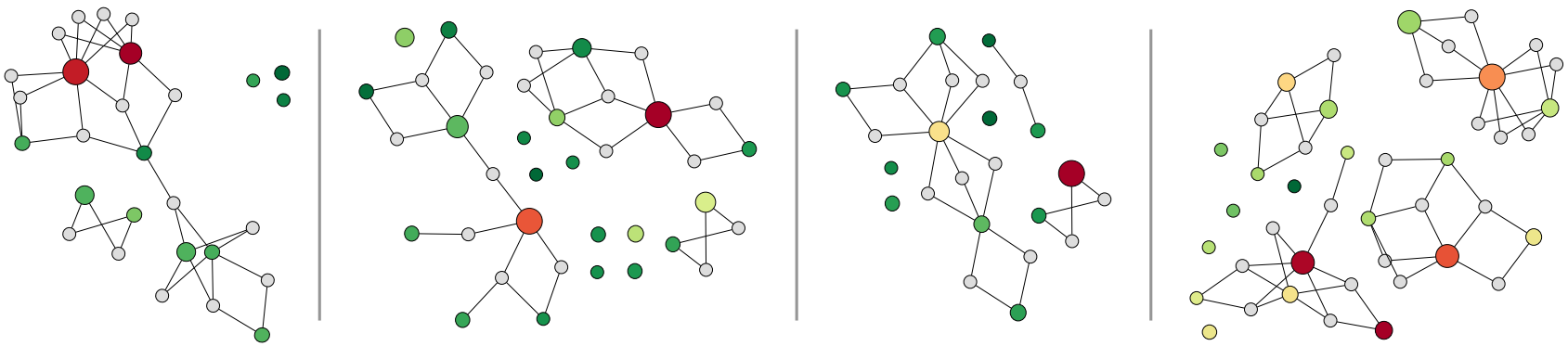
Common workflow:

- Use e.g. SPLOM to make sense of data.
- Use dimensionality reduction methods to obtain visualization for exploratory data analysis.

Beyond comparing projection scatterplots?

Typical goal

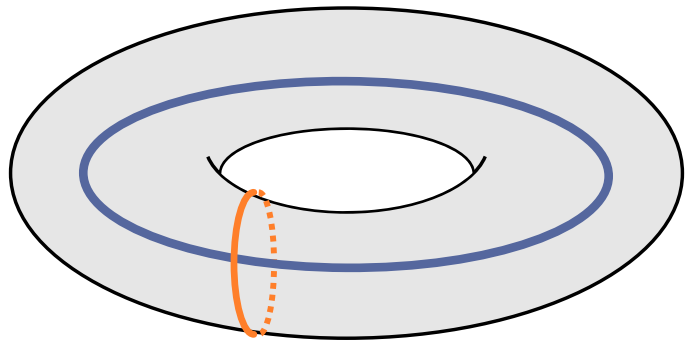
Quantify changes & differences



Our approach

- Combine geometrical (*distance-preserving*) and topological (*structure-preserving*) methods.
- Visualize internal connectivity of a data set as a graph

Topology

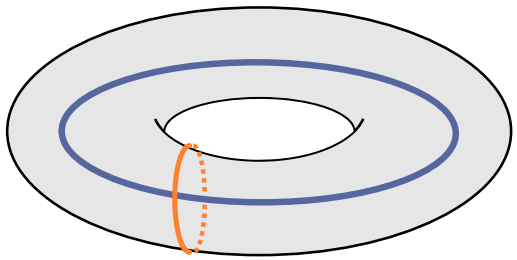


$\mathfrak{b}_1 = 2, \mathfrak{b}_2 = 1$

- Describe data sets by high-dimensional "holes".
- Hole $\approx$ inhomogeneous region in the data.
- Detection using *persistent homology*.
- Description using *simplicial chains*

The need for geometrical information

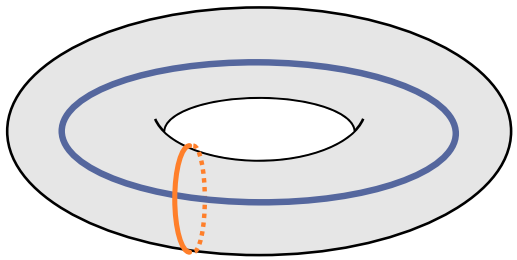
Without geometrical information:



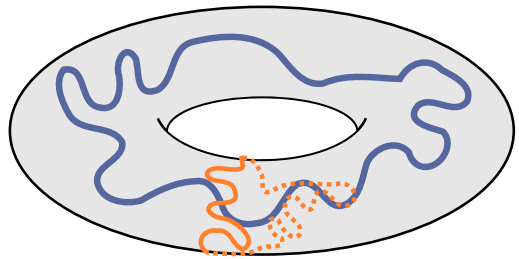
What we want

The need for geometrical information

Without geometrical information:



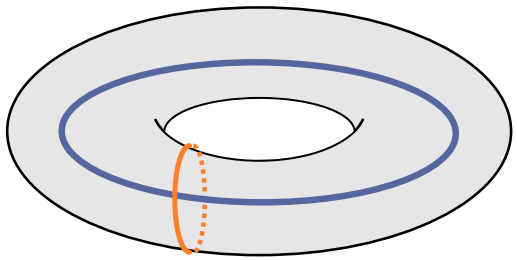
What we want



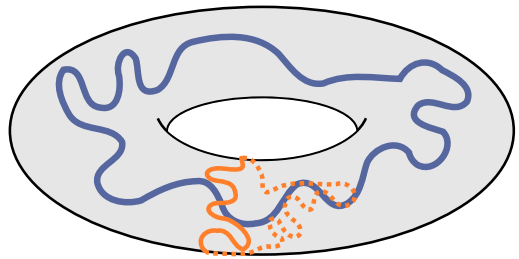
What we get

The need for geometrical information

Without geometrical information:

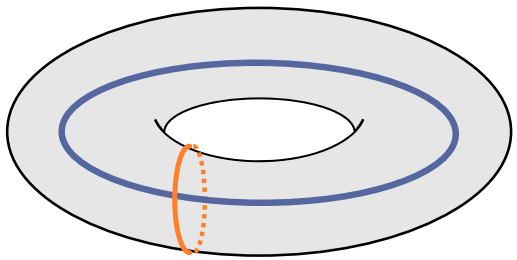


What we want



What we get

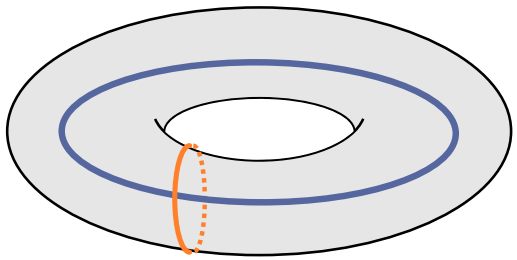
With geometrical information:



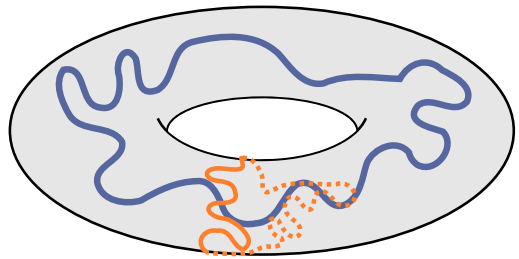
What we want

The need for geometrical information

Without geometrical information:

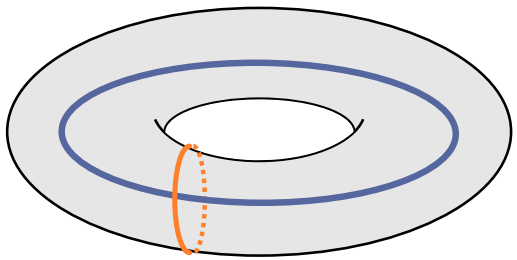


What we want

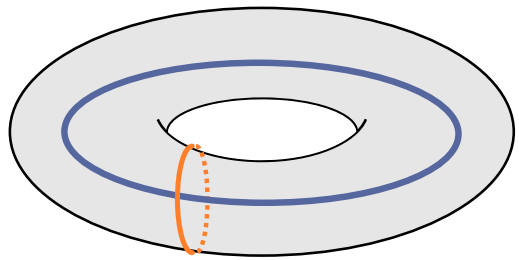


What we get

With geometrical information:

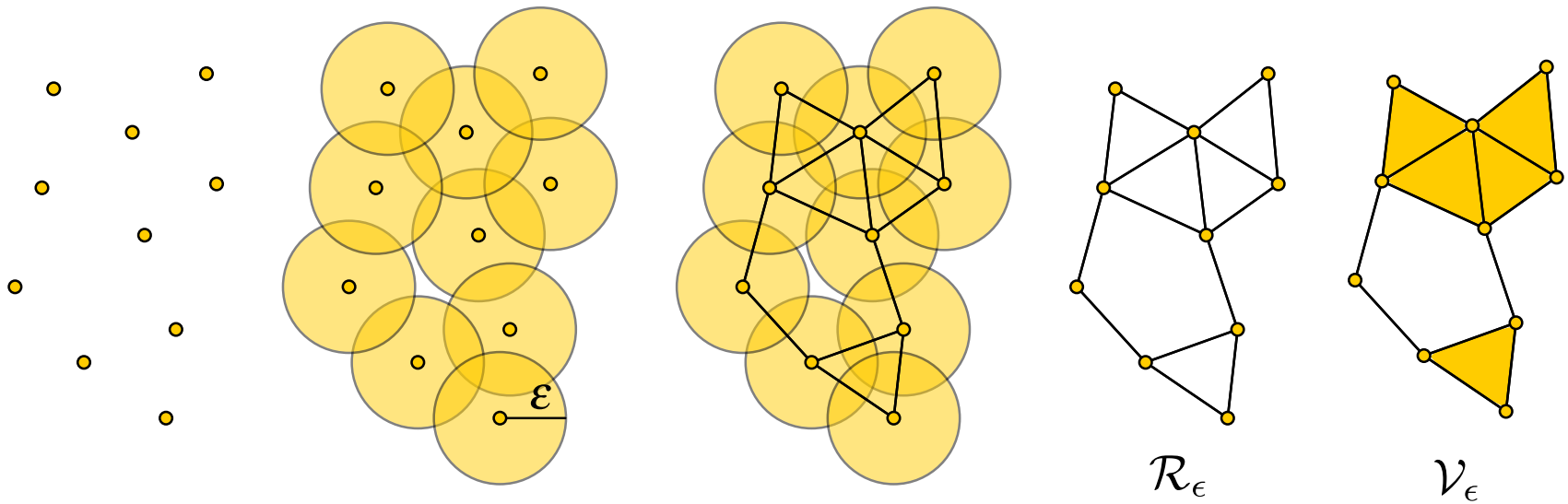


What we want

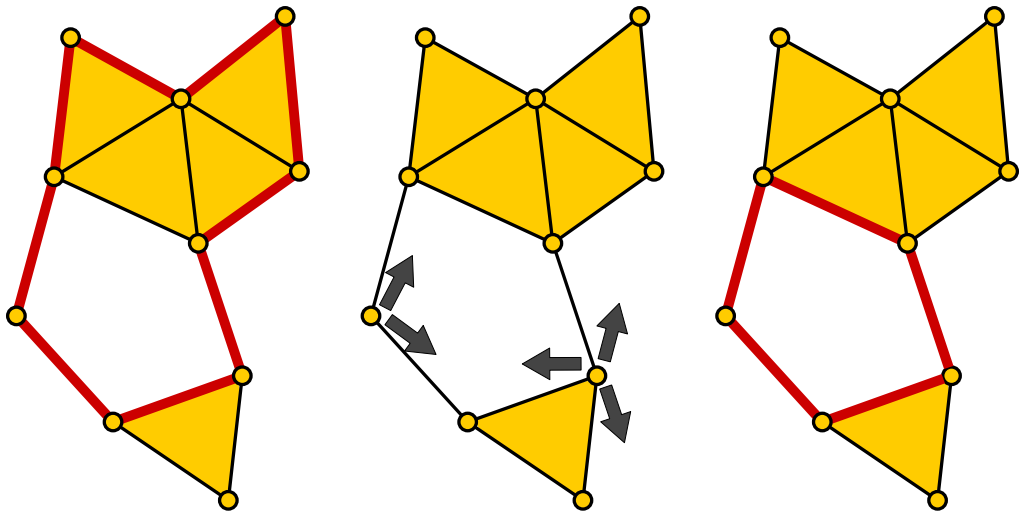


What we get

Finding loops



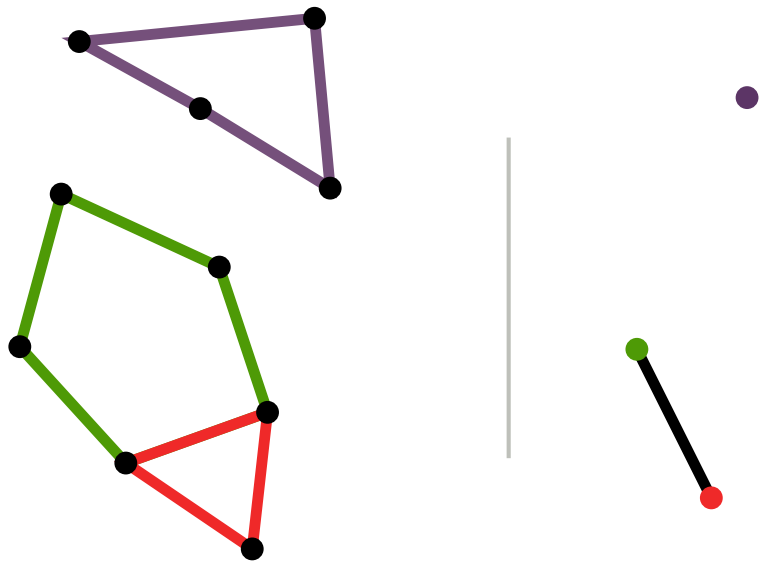
Localization of loops



- Obtain unlocalized loops by traversal.
- Solve all-pairs shortest path problem.
- Obtain localized loops by traversal.

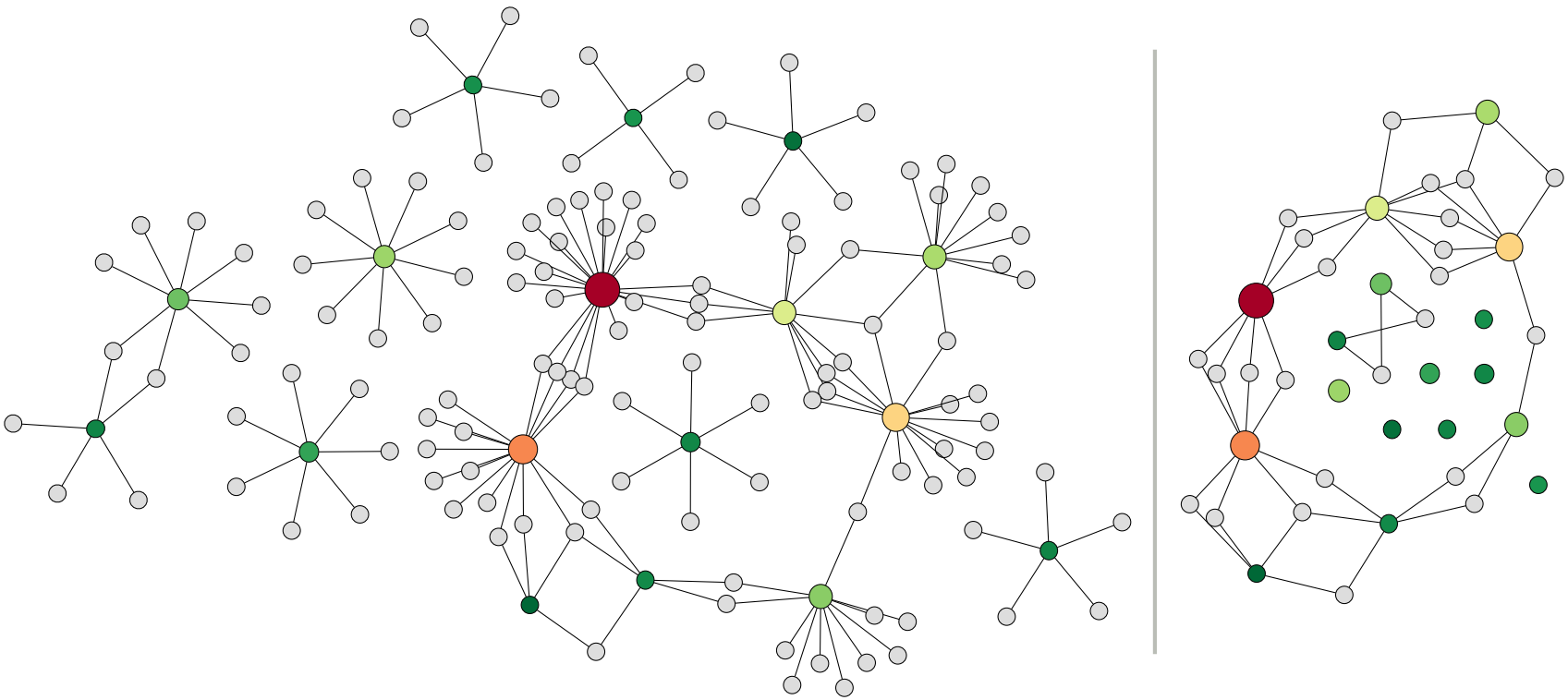
The simplicial chain graph

Idealized example



The simplicial chain graph

In practice

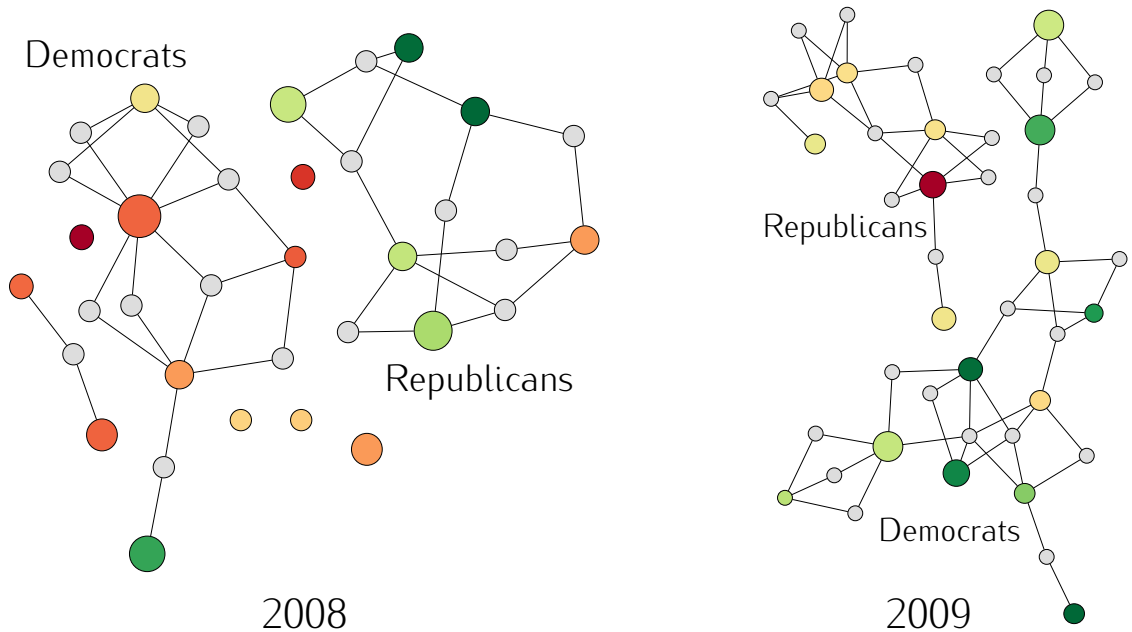


Data set: Voting data

Alice	+1	+1	...	+1
Bob	−1	−1	...	−1
Carol	+1	−1	...	0
Dave	0	0	...	0

- Voting data from the United States House of Representatives.
- About 430 points of dimensions 600–900.
- Data sets from 1990–2011.

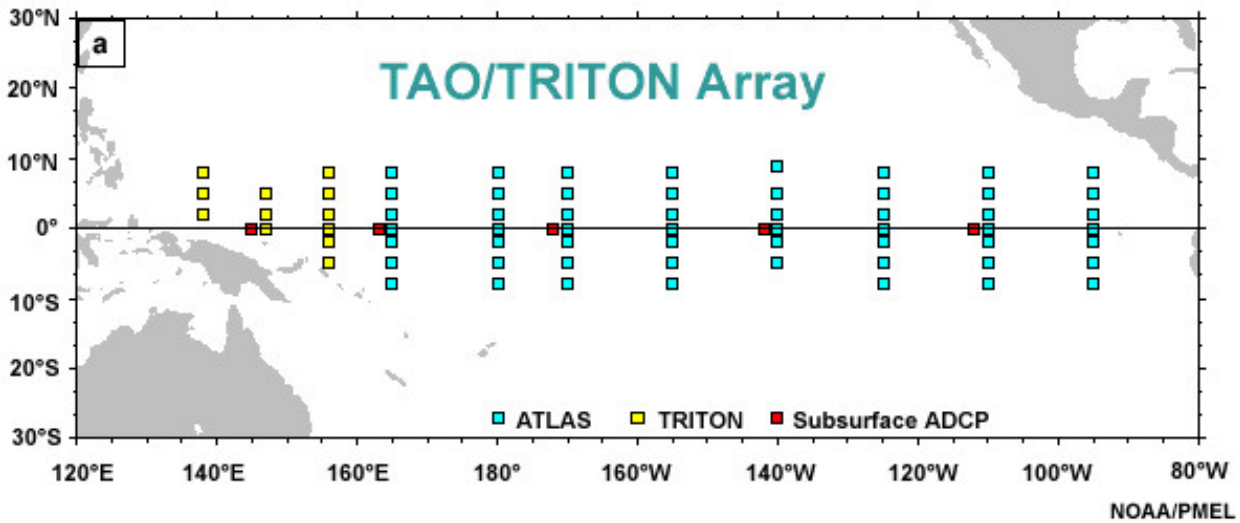
Data set: Voting data



Boundaries between parties are defined by "dissenters" from the party line.

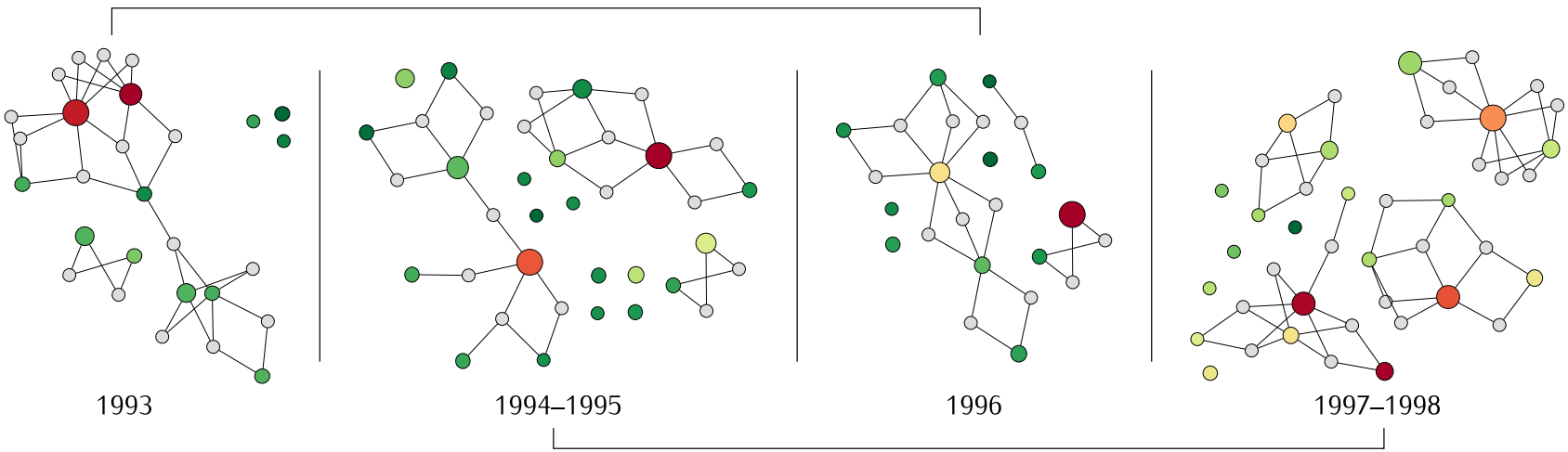
Data set: TAO

Tropical atmosphere ocean project



- El Niño phenomenon.
- Continuous data stream from all buoy moorings.
- 5-dimensional feature space: Wind velocities, humidity, air temperature, and sea surface temperature.

Data set: TAO

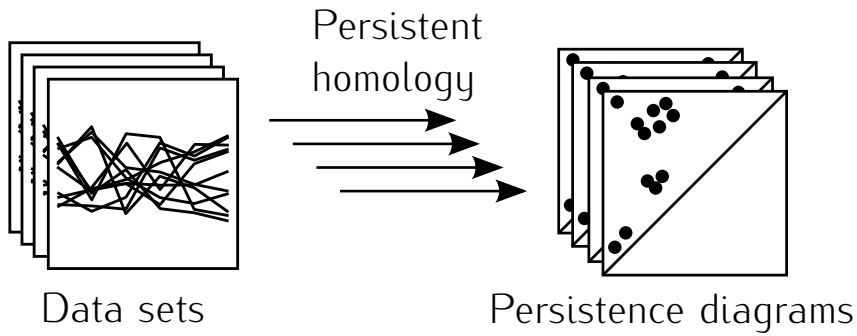


- 1993, 1996: No El Niño phenomenon in data set.
- 1994–1995, 1997–1998: El Niño phenomenon.

Lessons learned

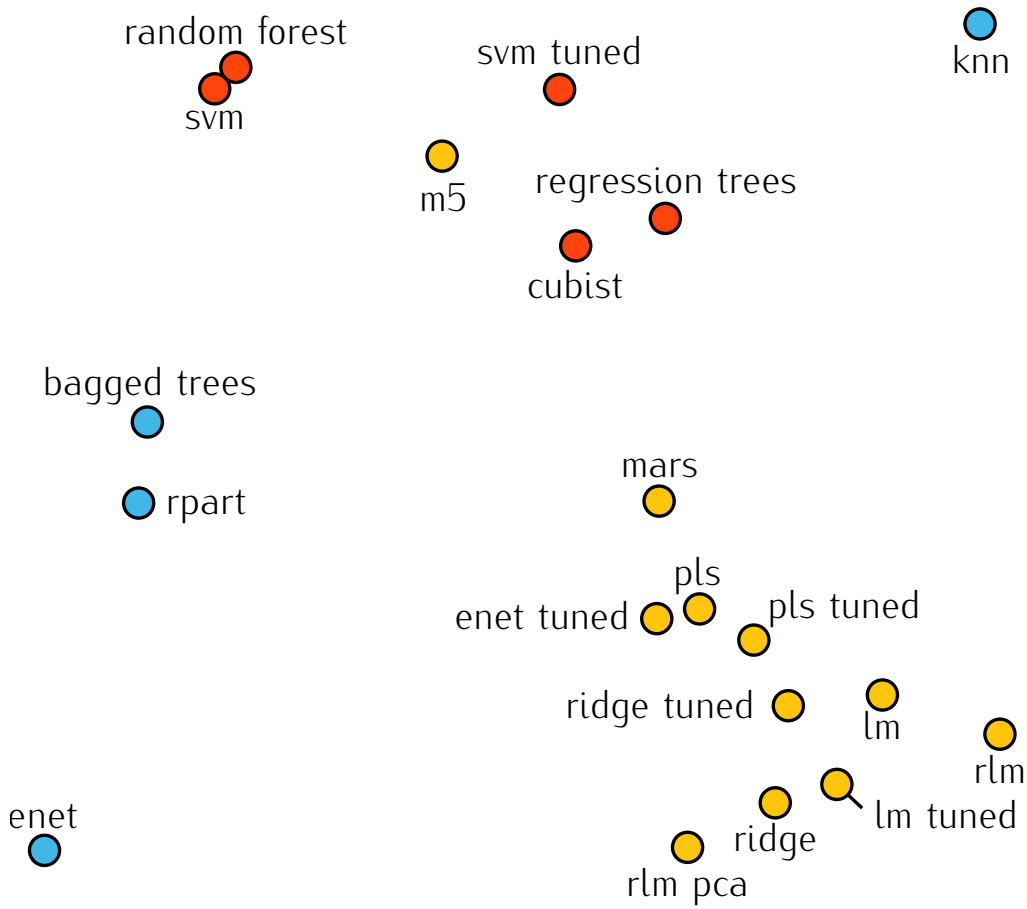
- Features obtained via persistent homology are suitable for comparative analysis.
- Visualization of internal point cloud structure quickly becomes too abstract.
- Thus: Aim to quantify differences in persistent homology using well-defined metrics.

Summarize data even further



Application

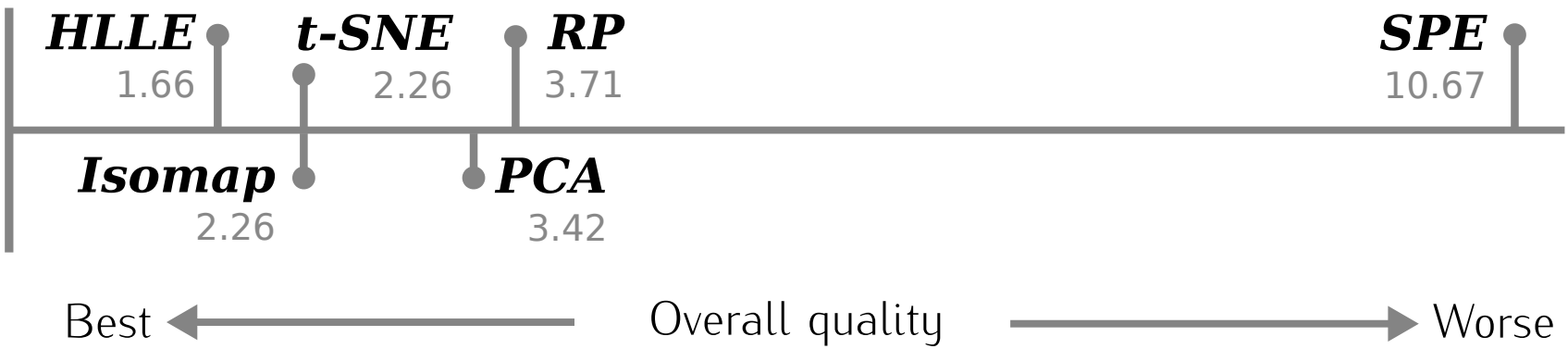
Quantifying differences in models for multivariate data



Application

A framework for comparing dimensionality reduction methods

- Define data descriptors on data set.
- Calculate their persistent homology.
- Compare their “topological fingerprints”.



Conclusion

- Simplicial chain graphs as a visual metaphor for structures in multivariate point clouds.
- Aspects: Comparison and quantifying differences graphically.
- Recent work concentrates on metric quantification using persistent homology.
- Open question: What “interesting” structures in high-dimensional space can we capture?