Reeb Graph and Mapper
Motivations

get a higher-level understanding of the structure of data

avoid paying the algorithmic price of persistence

visualize topology on the data directly

exhibit relations between clusters, variables, etc.
Motivations

get a higher-level understanding of the structure of data

avoid paying the algorithmic price of persistence

visualize topology on the data directly

exhibit relations between clusters, variables, etc.

principle: summarize the topological structure of a map $f : X \to \mathbb{R}$ through a graph
Mapper in the continuous setting

Input:
- continuous function (‘filter’, ‘lens’...) $f : X \rightarrow \mathbb{R}$
- cover $\mathcal{I}$ of $\text{im}(f)$ by open intervals: $\text{im} f \subseteq \bigcup_{I \in \mathcal{I}} I$

Method:
- Compute pullback cover $\mathcal{U}$ of $X$: $\mathcal{U} = \{f^{-1}(I)\}_{I \in \mathcal{I}}$
- Refine $\mathcal{U}$ by separating each of its elements into its various connected components $\rightarrow$ connected cover $\mathcal{V}$
- The Mapper is the nerve of $\mathcal{V}$:
  - 1 vertex per element $V \in \mathcal{V}$
  - 1 edge per intersection $V \cap V' \neq \emptyset$, $V, V' \in \mathcal{V}$
  - 1 $k$-simplex per $(k + 1)$-fold intersection $\bigcap_{i=0}^{k} V_i \neq \emptyset$, $V_0, \cdots, V_k \in \mathcal{V}$
Mapper in the continuous setting

\[ X \xrightarrow{f} \mathbb{R} \]

\[ \mathbb{I} \]
Mapper in the continuous setting

\[ f: X \rightarrow \mathbb{R} \]
Mapper in the continuous setting
Mapper in the continuous setting
Mapper in practice

Input:  
- point cloud $P$ with metric $d_P$
- continuous function (‘filter’, ‘lens’...) $f : P \to \mathbb{R}$
- cover $\mathcal{I}$ of $\text{im}(f)$ by open intervals: $\text{im} f \subseteq \bigcup_{I \in \mathcal{I}} I$

Method:  
- Compute neighborhood graph $G = (P, E)$
  - Compute pullback cover $\mathcal{U}$ of $P$: $\mathcal{U} = \{f^{-1}(I)\}_{I \in \mathcal{I}}$
  - Refine $\mathcal{U}$ by separating each of its elements into its various connected components in $G \to$ connected cover $\mathcal{V}$
- The Mapper is the nerve of $\mathcal{V}$: (intersections materialized by data points)
  - 1 vertex per element $V \in \mathcal{V}$
  - 1 edge per intersection $V \cap V' \neq \emptyset$, $V, V' \in \mathcal{V}$
  - 1 $k$-simplex per $(k + 1)$-fold intersection $\bigcap_{i=0}^{k} V_i \neq \emptyset$, $V_0, \ldots, V_k \in \mathcal{V}$
Mapper in practice

$X$ \quad $\delta$

$f$

$G = \delta$-neighborhood graph
Mapper in practice

\( G = \delta \)-neighborhood graph
Mapper in practice

$X$  $\delta$

$f$

$\mathbb{R}$  $\mathcal{I}$

$V$

$G = \delta$-neighborhood graph
in practice, the result may be different from the continuous setting due to the neighborhood graph

Mapper in practice

\[ G = \delta \text{-neighborhood graph} \]
Mapper in practice

Parameters:

- filter \( f : P \rightarrow \mathbb{R} \)

- cover \( \mathcal{I} \) of \( \text{im}(f) \) by open intervals

- neighborhood size \( \delta \)
Mapper in practice

Parameters:

- filter $f : P \rightarrow \mathbb{R}$

- cover $\mathcal{I}$ of $\text{im}(f)$ by open intervals

- neighborhood size $\delta$

geometric scale

range scale
Mapper in practice

Parameters:

- filter \( f : P \to \mathbb{R} \)

- cover \( \mathcal{I} \) of \( \text{im}(f) \) by open intervals

- neighborhood size \( \delta \)

\( \rightarrow \) uniform cover \( \mathcal{I} \):

  - resolution / granularity: \( r \) (diameter of intervals)

  - gain: \( g \) (percentage of overlap)
Mapper in applications

breast cancer subtype
Mapper in applications

recovery from spinal cord injuries
Mapper in applications

protein folding pathways
Mapper in applications

- Diagnosis of pulmonary embolism
- Recovery from spinal cord injuries
- Breast cancer subtype
- Protein folding pathways
Mapper in applications

implicit networks in the US house of representatives

protein folding pathways

recovery

Histopathology

Lowest values

recovery

protein folding pathways

diagnosis of pulmonary embolism
Mapper in applications

classification of NBA players

implicit networks in the US house of representatives

recovery

protein folding pathways
Mapper in applications

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

classification of NBA players

implicit networks in the US house of representatives

recovery from spinal cord injuries

breast cancer subtype

Black Magic
Mapper in applications

Extracting insights from the shape of complex data using topology, Lum et al., Nature, 2013


Using Topological Data Analysis for Diagnosis Pulmonary Embolism, Rucco et al., arXiv preprint, 2014

Topological Methods for Exploring Low-density States in Biomolecular Folding Pathways, Yao et al., J. Chemical Physics, 2009

CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes, Sarikonda et al., J. Autoimmunity, 2013

Innate and adaptive T cells in asthmatic patients: Relationship to severity and disease mechanisms, Hinks et al., J. Allergy Clinical Immunology, 2015
Choice of parameters

Parameters:

- filter $f : P \rightarrow \mathbb{R}$

- cover $\mathcal{I}$ of $\text{im}(f)$ by open intervals

- neighborhood size $\delta$

→ uniform cover $\mathcal{I}$:
  - resolution / granularity: $r$ (diameter of intervals)
  - gain: $g$ (percentage of overlap)
Choice of parameters

How to choose $r, g$ and $\delta$?

→ in practice: trial-and-error (or some vague procedure)

Choice of parameters

Illustration: $P \subset \mathbb{R}^2$ sampled from known probability distribution
Choice of parameters

Illustration: $P \subset \mathbb{R}^2$ sampled from known probability distribution

$f = \text{density estimator}, \ r = 1/30, \ g = 20\%$

$\delta = \text{percentage of the diameter of} \ X$

- $\delta = 1\%$
- $\delta = 10\%$
- $\delta = 30\%$
Choice of parameters

Illustration: \( P \subset \mathbb{R}^2 \) sampled from known probability distribution

\[ f = \text{abscissa}, \quad r = 1/30, \quad g = 10\% \]

\( \delta = \text{percentage of the diameter of } X \)

\[ \delta = 1\% \]

+ small cluster removal
Choice of parameters

Structure and Stability of the 1-Dimensional Mapper. Carrière, O. 2016

→ clarifies formally the roles of $r$ and $g$ in the continuous setting

→ gives sufficient conditions on $\delta$ to get approximation results

→ also gives a notion of distance and stability for Mappers...
Choice of parameters

**Structure and Stability of the 1-Dimensional Mapper.**
Carrière, O. 2016

→ clarifies formally the roles of $r$ and $g$ in the continuous setting

→ gives sufficient conditions on $\delta$ to get approximation results

→ also gives a notion of distance and stability for Mappers...

Other publications:
- de Silva, Munch, Patel. *Categorified Reeb Graphs*. 2015
Extended Persistence
Extended Persistence

\[ d_0 = b_2 \]
\[ d_{1h} = b_{1v} \]
\[ d_{1v} = b_{1h} \]
\[ d_2 = b_0 \]
Extended Persistence
- ordinary / relative
- extended

- the upward branch gives a relative $H_0$ point plus a symmetric ordinary $H_1$ point
- the downward branch gives an ordinary $H_0$ point plus a symmetric relative $H_2$ point