Multiscale 2-Mapper – Exploratory Data Analysis Guided by the First Betti Number

Halley Fritze

Topological Data Visualization Workshop

June 9th , 2025

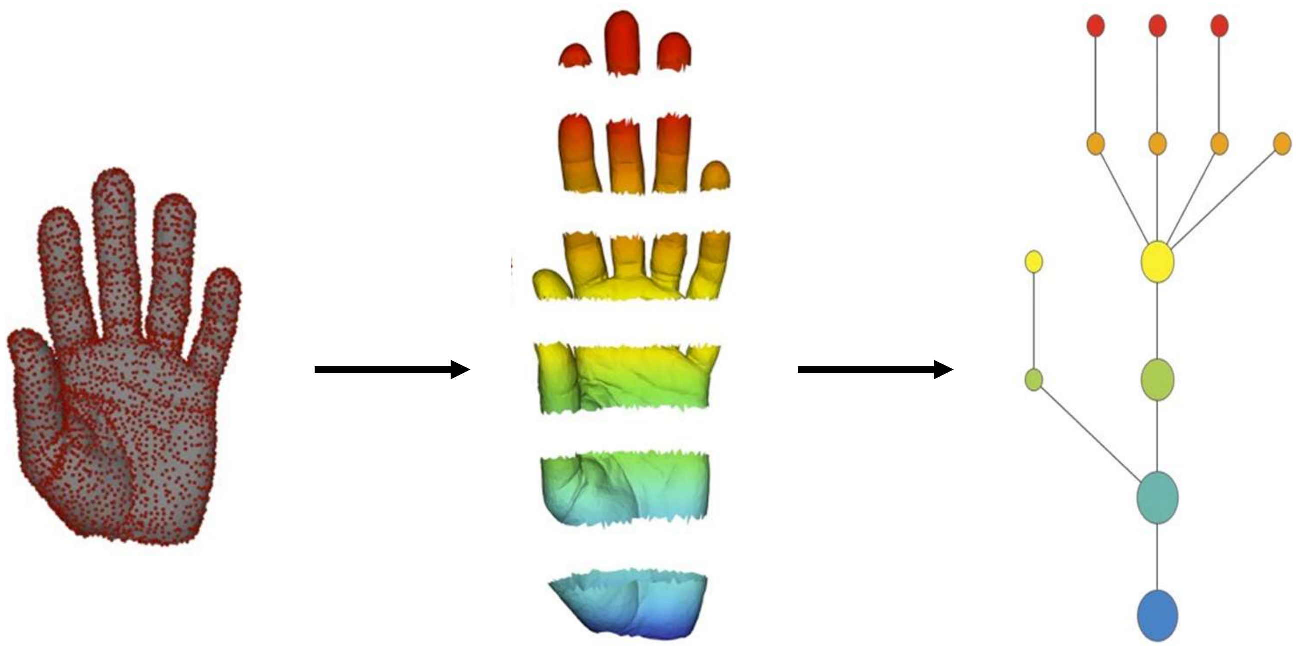
Mapper

The *Mapper Graph* is an unsupervised soft clustering algorithm which clusters data in the form of a graph.

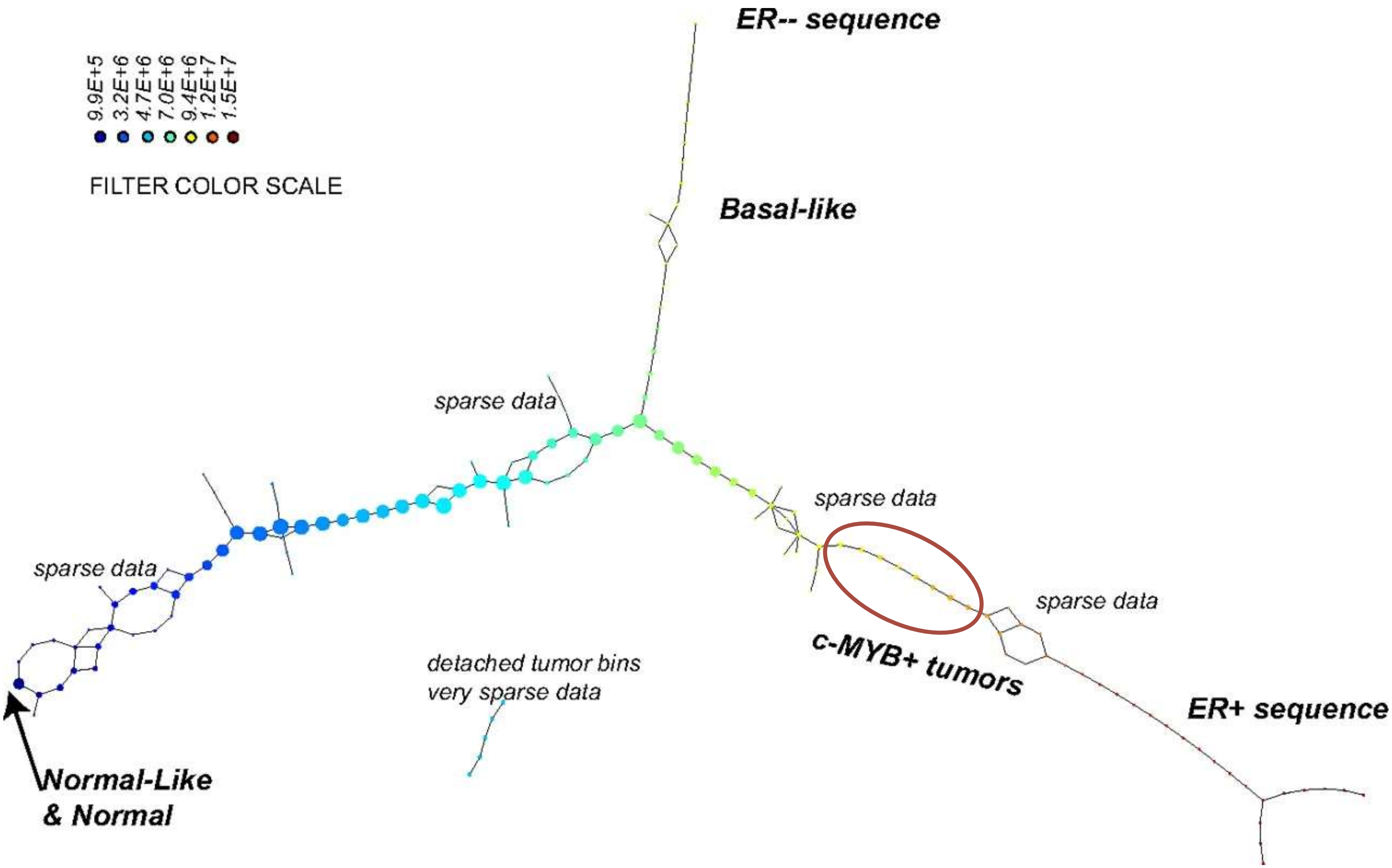
G. Singh, F. Mémoli, and G. E. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, 4th Symposium on Point Based Graphics, PBG@Eurographics 2007, Prague, Czech Republic, September 2-3, 2007, Eurographics Associa- tion, 2007, 91–100, doi: 10.2312/SPBG/SPBG07/091-100.

Mapper

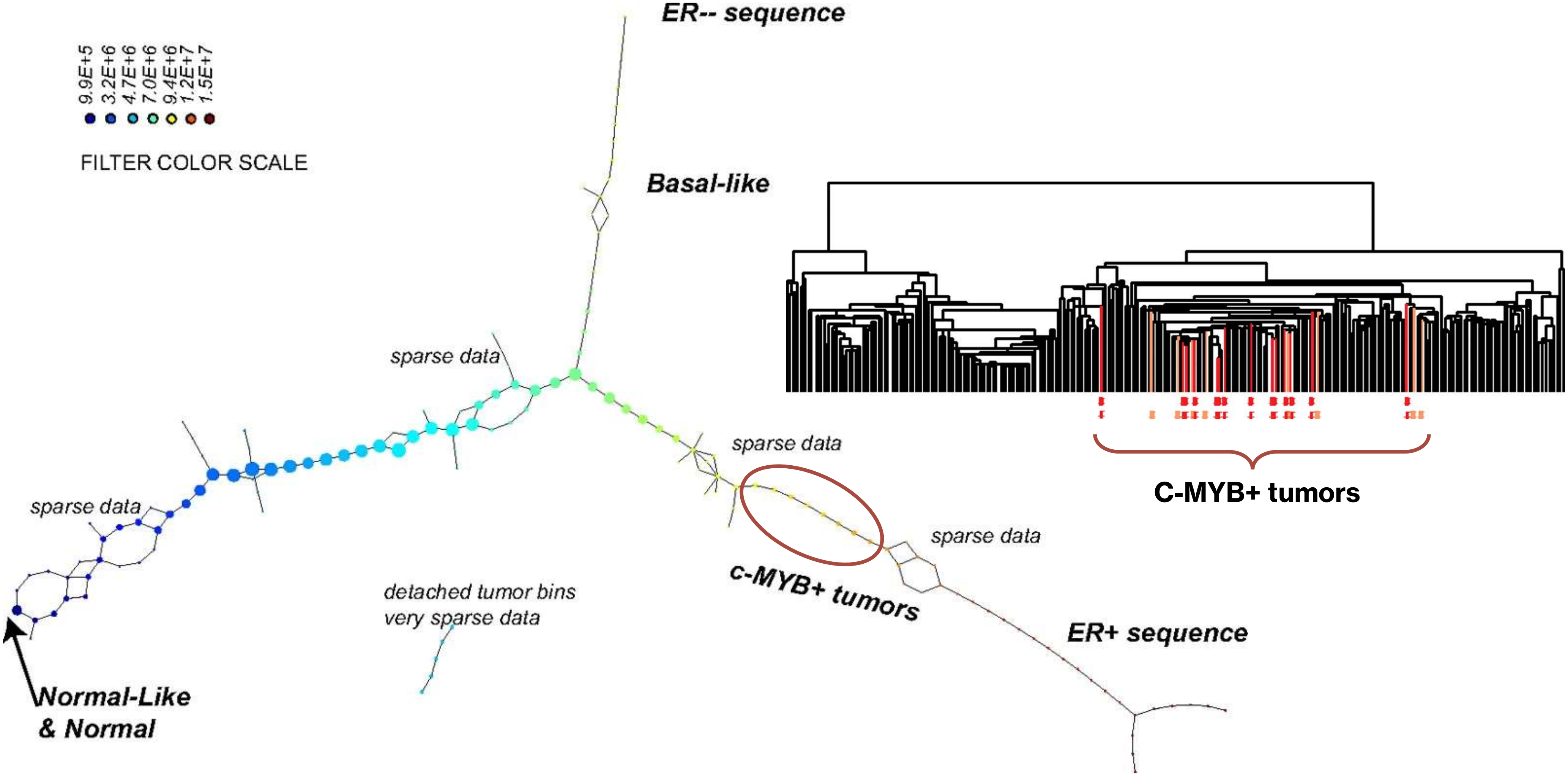
The *Mapper Graph* is an unsupervised soft clustering algorithm which clusters data in the form of a graph.



G. Singh, F. Mémoli, and G. E. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, 4th Symposium on Point Based Graphics, PBG@Eurographics 2007, Prague, Czech Republic, September 2-3, 2007, Eurographics Associa- tion, 2007, 91–100, doi: 10.2312/SPBG/SPBG07/091-100.



M. Nicolau, A. J. Levine, and G. Carlsson, Topology based data anal- ysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival (en), Proc. Natl. Acad. Sci. U. S. A. 108 (2011), no. 17, 7265–7270.



M. Nicolau, A. J. Levine, and G. Carlsson, Topology based data anal- ysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival (en), Proc. Natl. Acad. Sci. U. S. A. 108 (2011), no. 17, 7265–7270.

Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f: X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

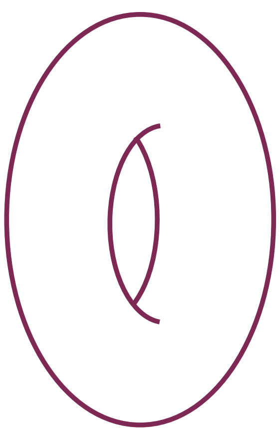
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$

Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f: X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$

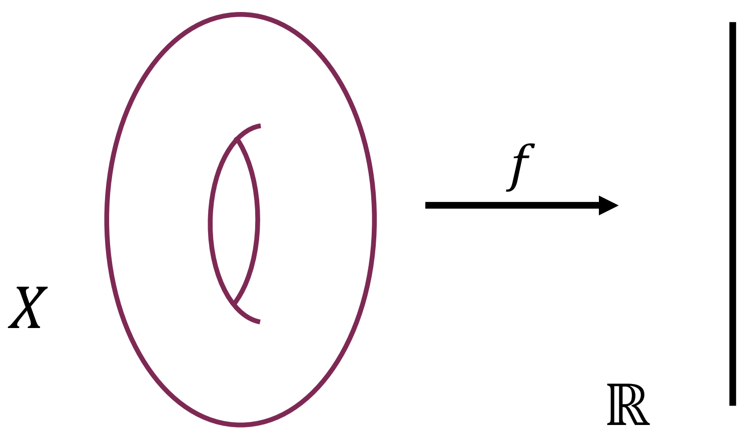
X



Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f: X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

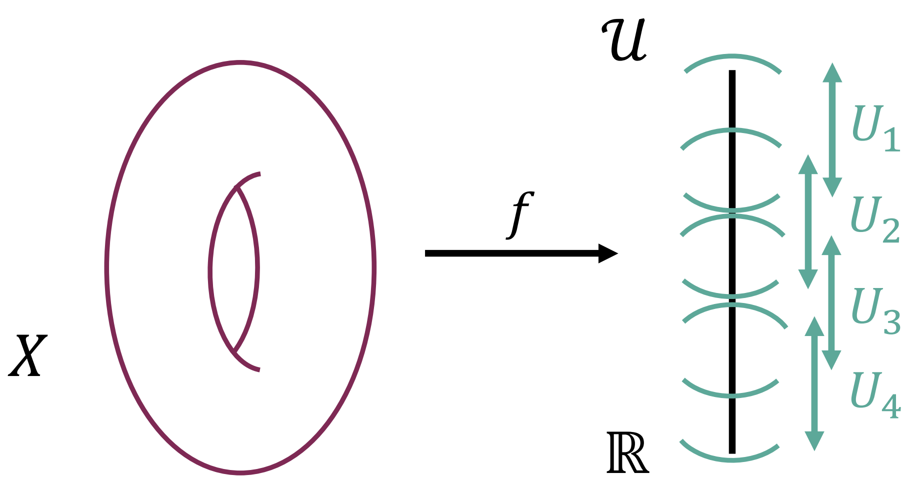
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$



Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f: X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

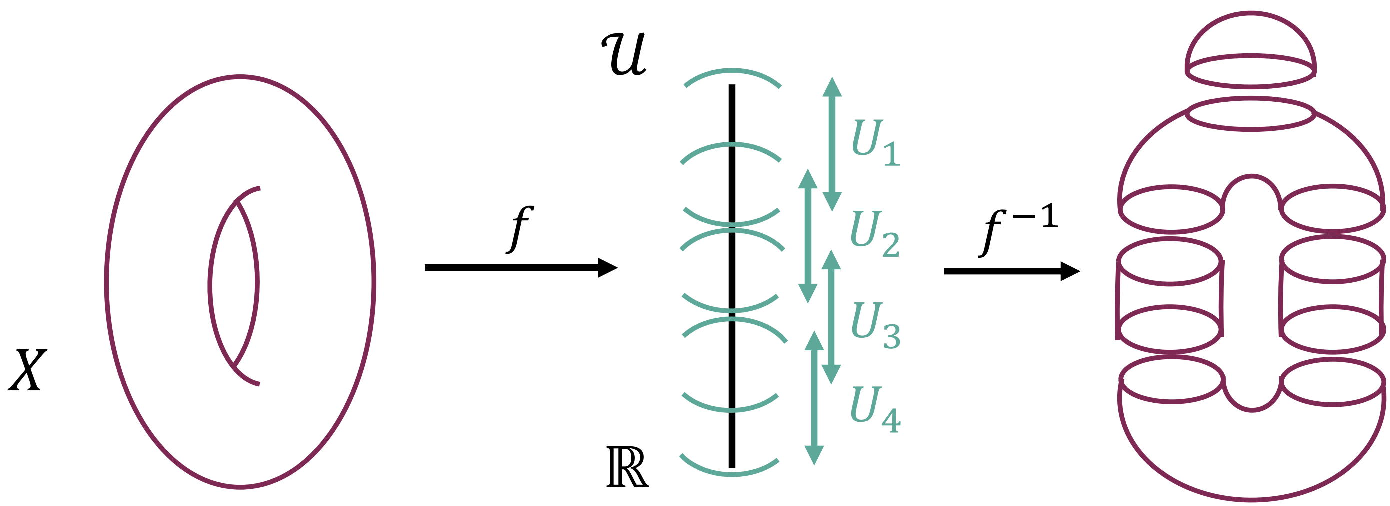
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$



Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f: X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

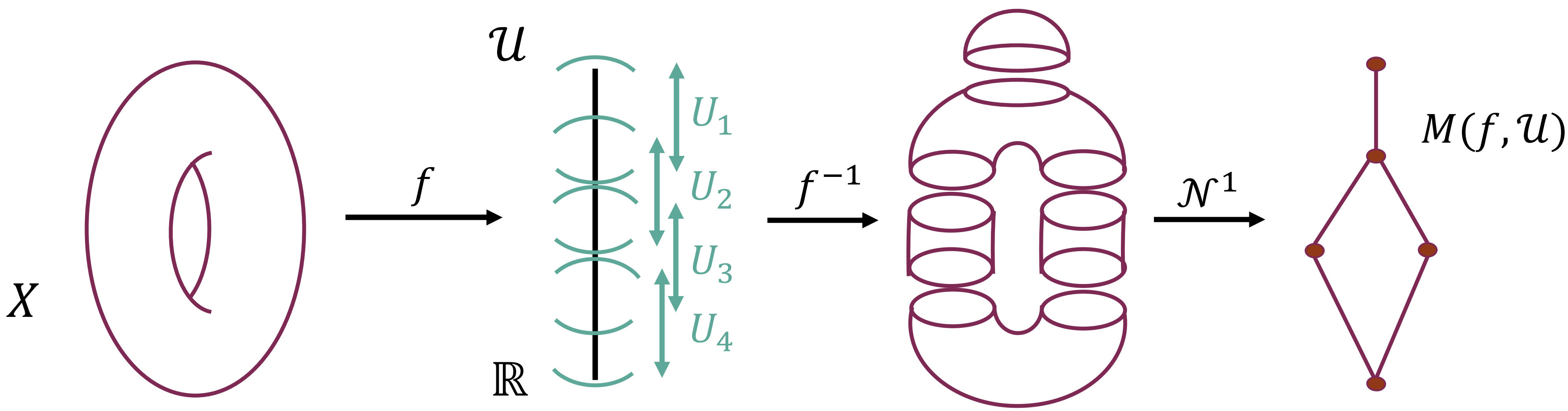
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$



Mapper

Definition. Let $X \subseteq \mathbb{R}^d$ be a data set. Given a continuous lens $f\colon X \to \mathbb{R}^n$, cover $\mathcal{U} = \{U_i\}_{i \in I}$ of $f(X)$, and clustering algorithm we define the *Mapper graph* as

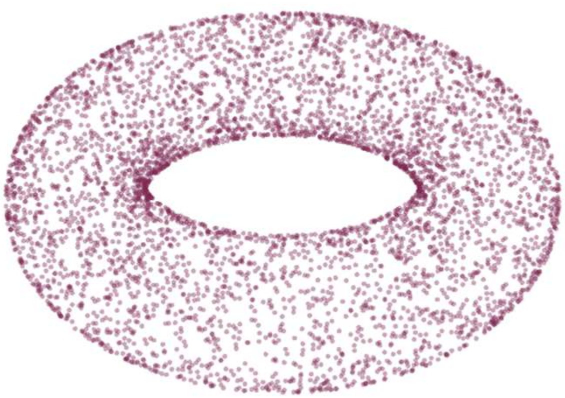
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^1(f^{-1}(\mathcal{U}))$$



Example.

Let $X \subset \mathbb{R}^3$ be a point cloud of a torus.

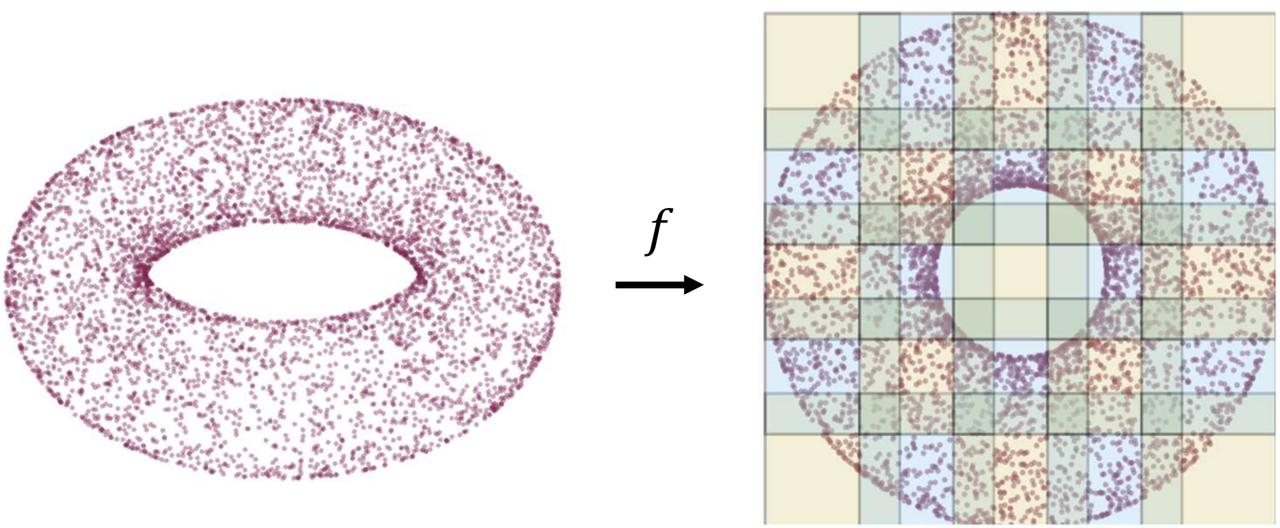
- Choose $f: X \to \mathbb{R}^2$ to be the coordinate projection to $\mathbb{R}^2$.



Example.

Let $X \subset \mathbb{R}^3$ be a point cloud of a torus.

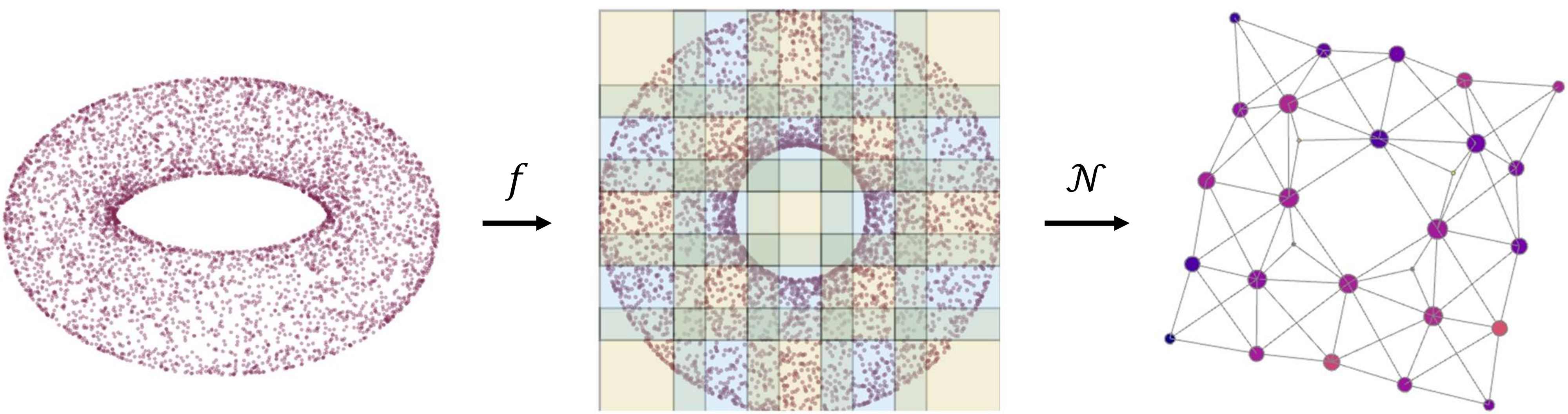
- Choose $f: X \to \mathbb{R}^2$ to be the coordinate projection to $\mathbb{R}^2$.
- Cover $f(X)$ with overlapping squares.



Example.

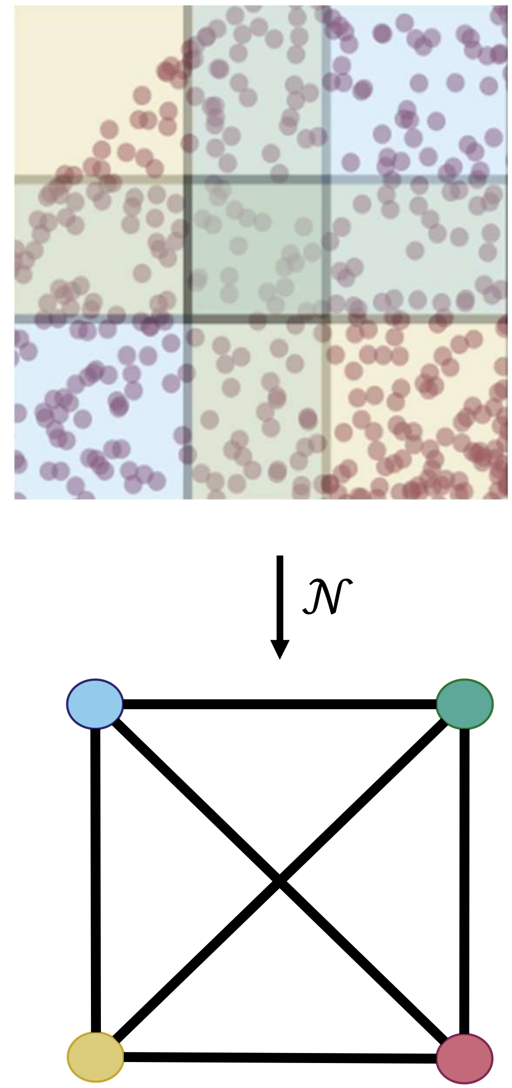
Let $X \subset \mathbb{R}^3$ be a point cloud of a torus.

- Choose $f: X \to \mathbb{R}^2$ to be the coordinate projection to $\mathbb{R}^2$.
- Cover $f(X)$ with overlapping squares.
- Choose DBSCAN as the clustering algorithm.



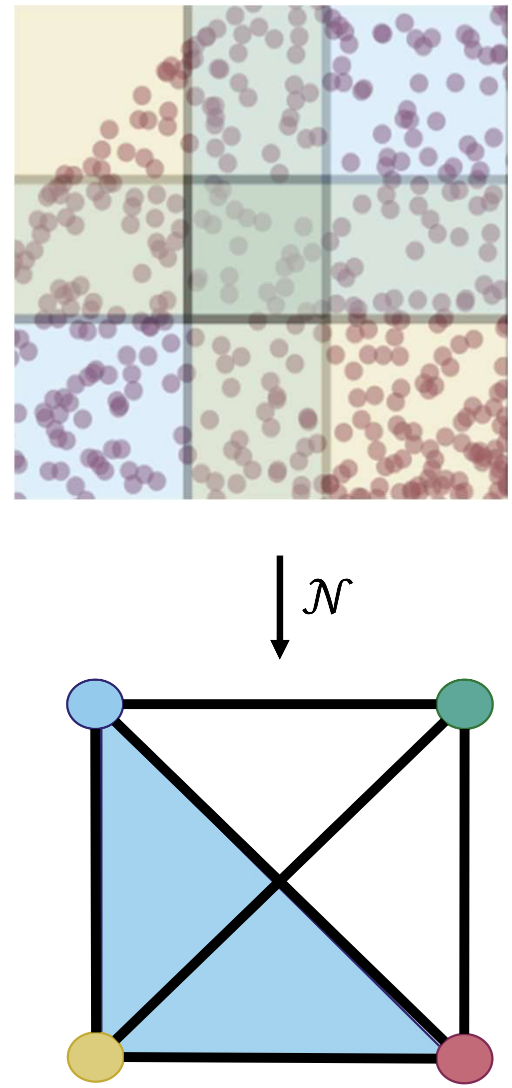
Mapper

The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.



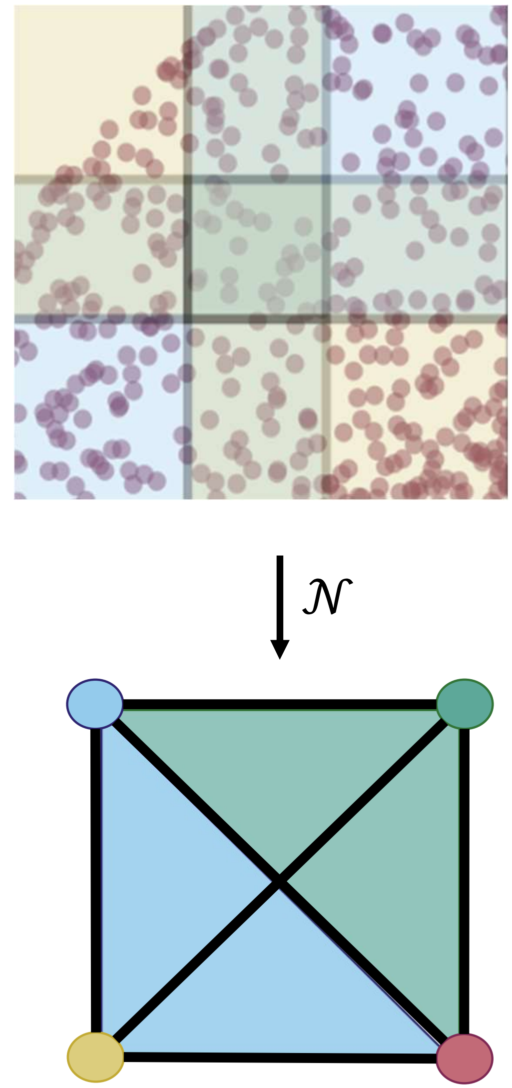
Mapper

The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.



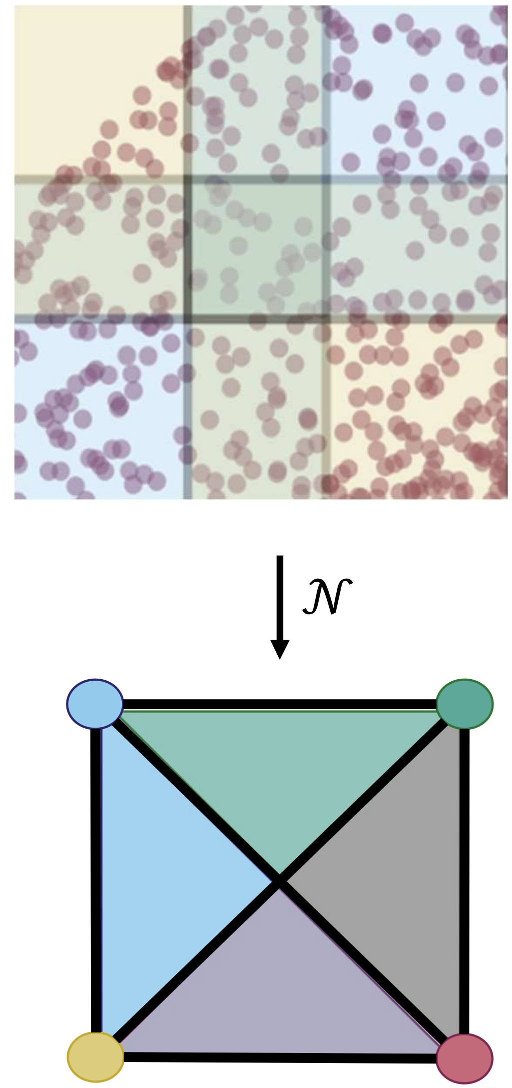
Mapper

The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.



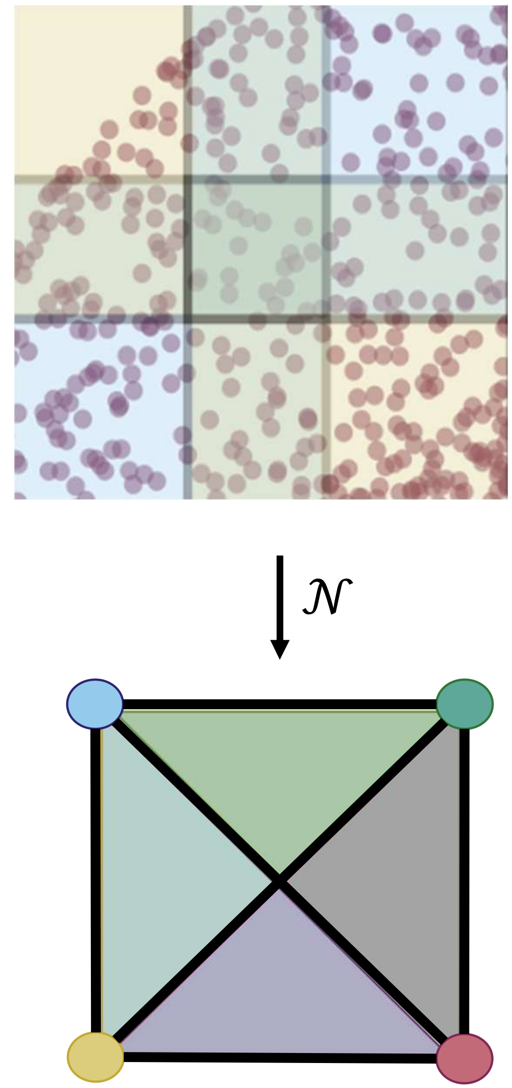
Mapper

The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.



Mapper

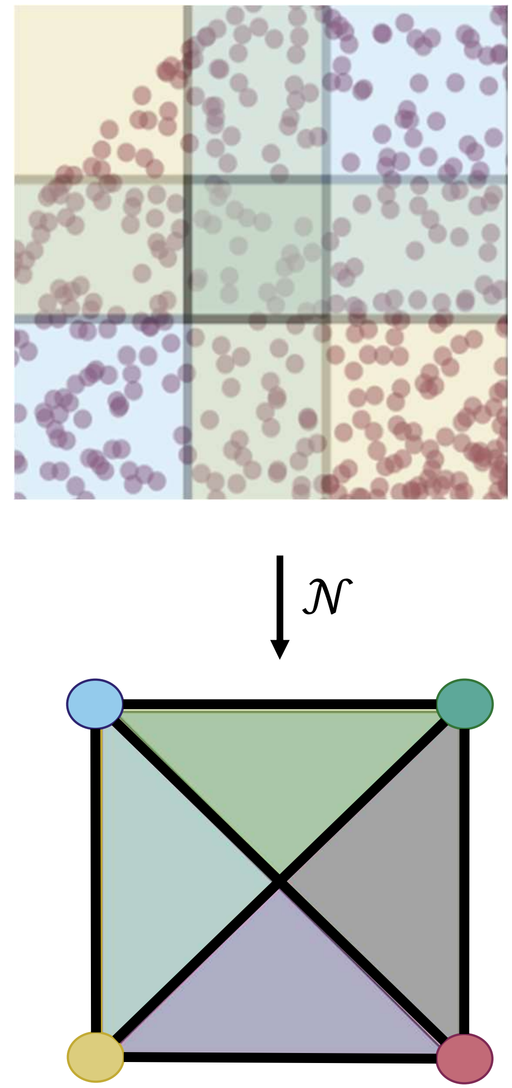
The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.



Mapper

The intersection structure of a cover can yield higher dimensional nerve when $f(X) \subset \mathbb{R}^2$.

Taking this into account could yield better manifold approximations of point clouds.



2-Mapper

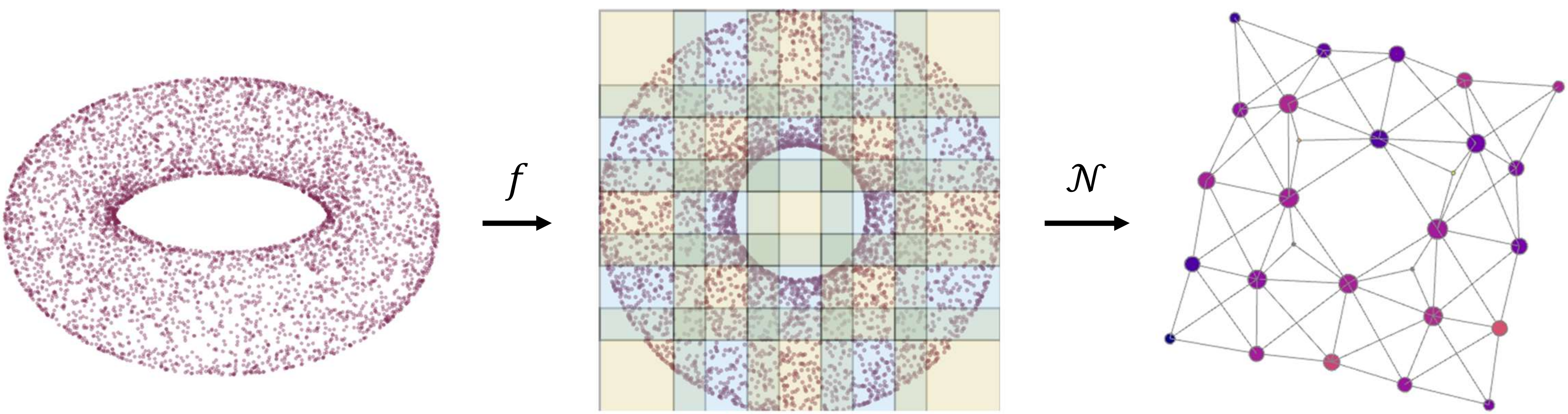
Definition. Given a point cloud X, continuous filter $f: X \to \mathbb{R}^n$, cover $\mathcal{U}$ and clustering algorithm, we define the *2-Mapper complex of* X as

$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^2\left(f^{-1}(\mathcal{U})\right)$$

2-Mapper

Definition. Given a point cloud X, continuous filter $f: X \to \mathbb{R}^n$, cover $\mathcal{U}$ and clustering algorithm, we define the *2-Mapper complex of* X as

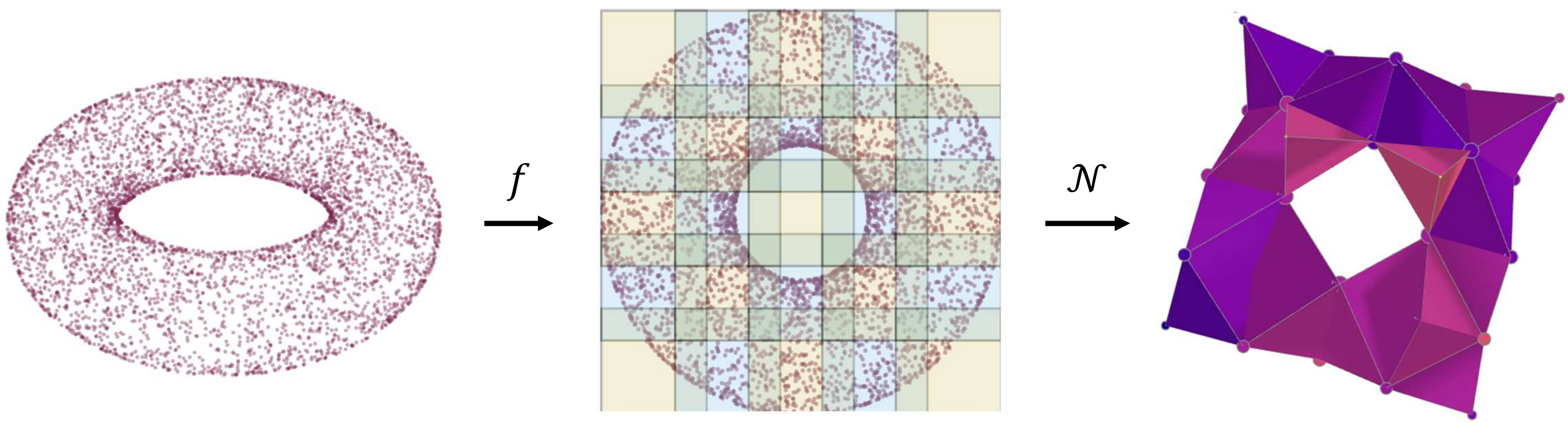
$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^2\left(f^{-1}(\mathcal{U})\right)$$



2-Mapper

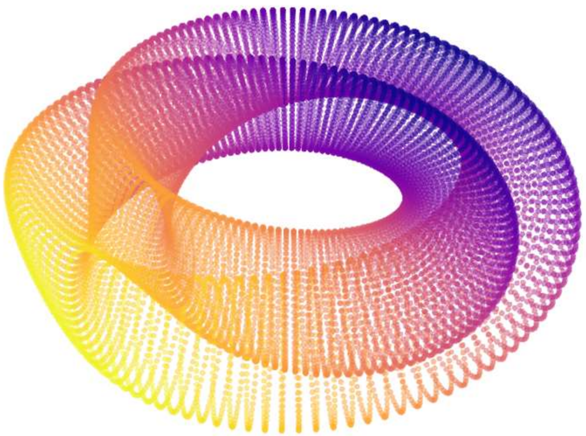
Definition. Given a point cloud X, continuous filter $f: X \to \mathbb{R}^n$, cover $\mathcal{U}$ and clustering algorithm, we define the *2-Mapper complex of X* as

$$M(f, \mathcal{U}) \coloneqq \mathcal{N}^2\left(f^{-1}(\mathcal{U})\right)$$



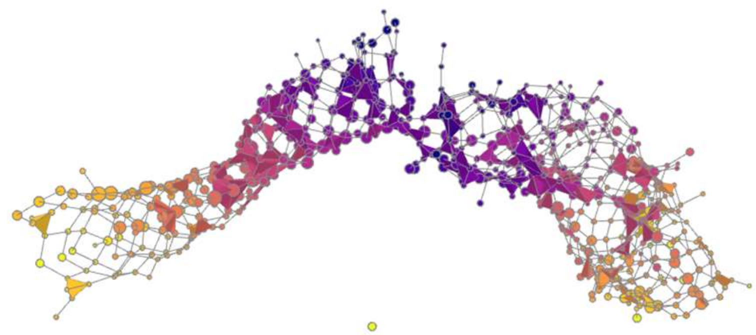
2-Mapper

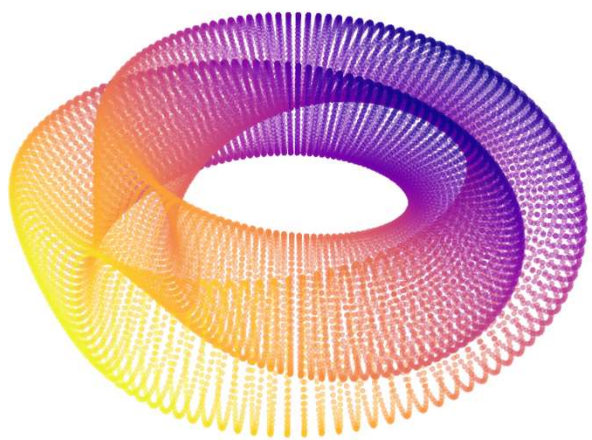
Question. How do we determine parameter choices for 2-Mapper?

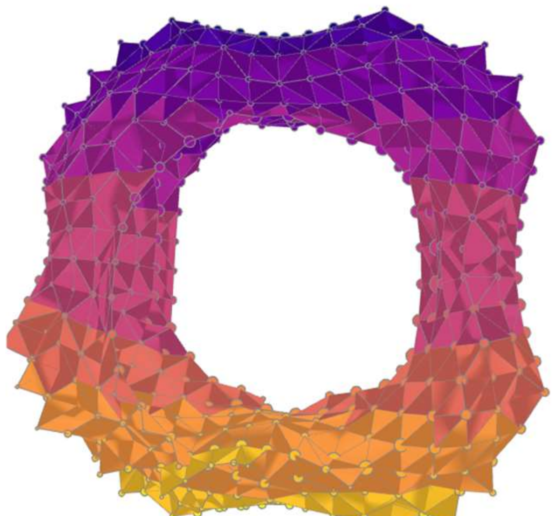


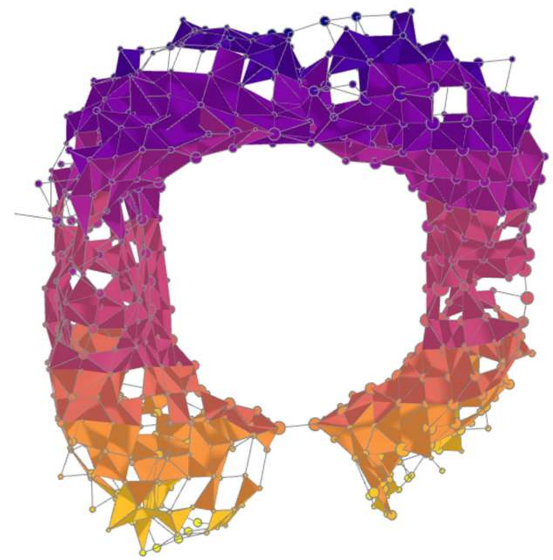
2-Mapper

<u>Question.</u> How do we determine parameter choices for 2-Mapper?









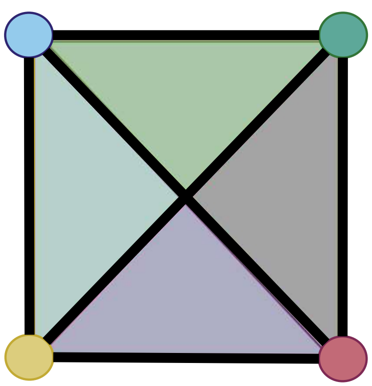
Optimal Covers

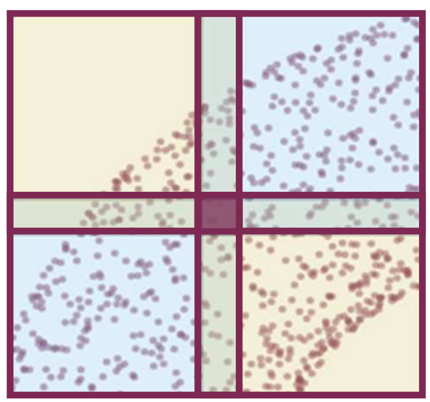
2-simplices only require a non-empty triple intersection of cover sets.

Optimal Covers

2-simplices only require a non-empty triple intersection of cover sets.

Cubical Cover





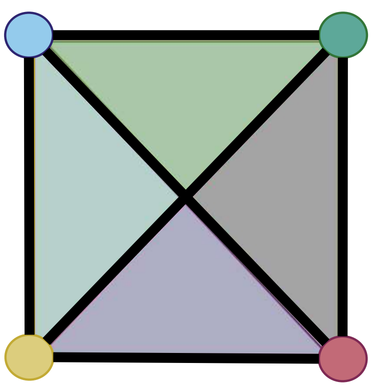
Four 2-Simplices

Optimal Covers

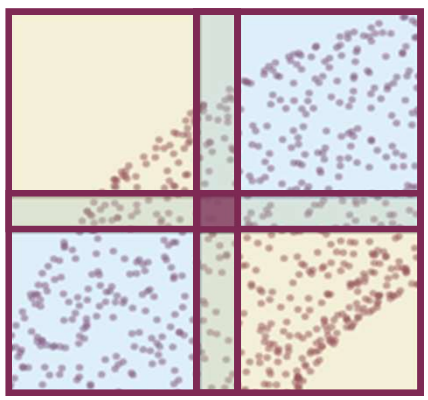
2-simplices only require a non-empty triple intersection of cover sets.

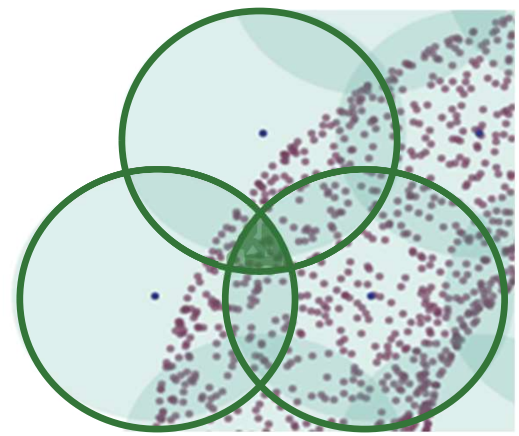
Cubical Cover

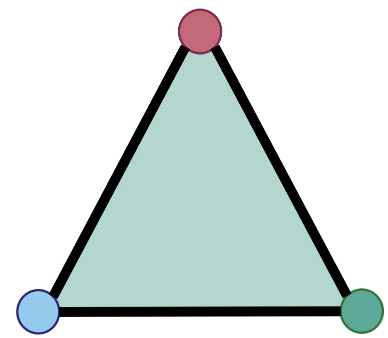
Lattice Cover



Four 2-Simplices







One 2-Simplex

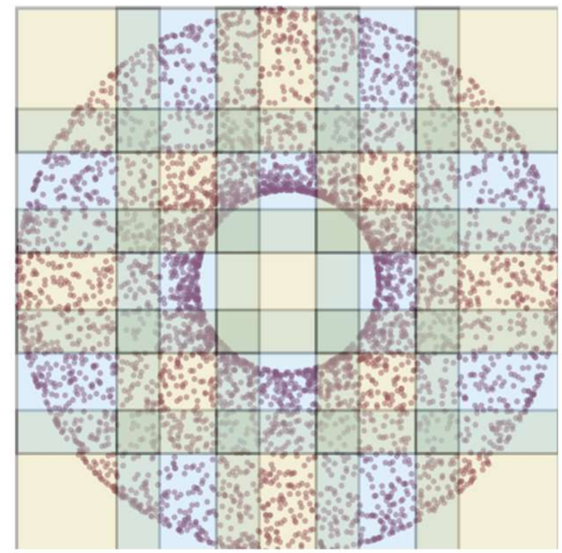
Defining Covers

Definition. Let $X \subset \mathbb{R}^d$ be a data set. The *bounding box* on X is the space $B = \prod_{i=1}^{d} [m_i, M_i]$, where $m_i = \min\limits_{x \in X} \pi_i(x)$ and $M_i = \max\limits_{x \in X} \pi_i(x)$ for coordinate projections π_i.

We define each cover with parameters k and g representing the scale for the number of cover sets and their overlap proportion, respectively.

Cubical Cover

Definition. Let $\mathrm{Z} \subset \mathbb{R}^n$ be a compact topological space with bounding box B. The cubical cover $\mathcal{U}$ on Z constructed with $k \geq 1$-intervals and overlap fraction $0 < g < 1$ is a cover of boxes $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ so that each cover set is of the form



$$U_{\alpha,s} = \prod_{i=1}^{n} \left[c_{\alpha,i} - \frac{l_i}{2}, c_{\alpha,i} + \frac{l_i}{2}\right]$$

Where $c_{\alpha,i} = m_1 + (\alpha_i - 1)(1 - g)l_i + \frac{1}{2} l_i$
and $l_i = \frac{M_i - m_i}{k-(k-1)g}$.

Cubical Lattice Cover

Definition. Let $Z \subset \mathbb{R}^n$ be a compact topological space with bounding box $B = \prod_{i=1}^{n}[m_i, M_i]$. We define the *cubical lattice cover* $\mathcal{U}$ over Z constructed with k-intervals and overlap fraction g as the cover $\mathcal{U} = \{U_\xi\}_{\xi \in B \cap \mathbb{Z}^n}$ whose cover sets are hypercubes defined

$$U_\xi = \prod_{i=1}^{n}\left[c\left(\xi_i - \frac{1}{2(1-g)}\right), c\left(\xi_i + \frac{1}{2(1-g)}\right)\right],$$

with $c = \max\limits_{1 \le i \le n} \frac{M_i - m_i}{k-(k-1)g}$

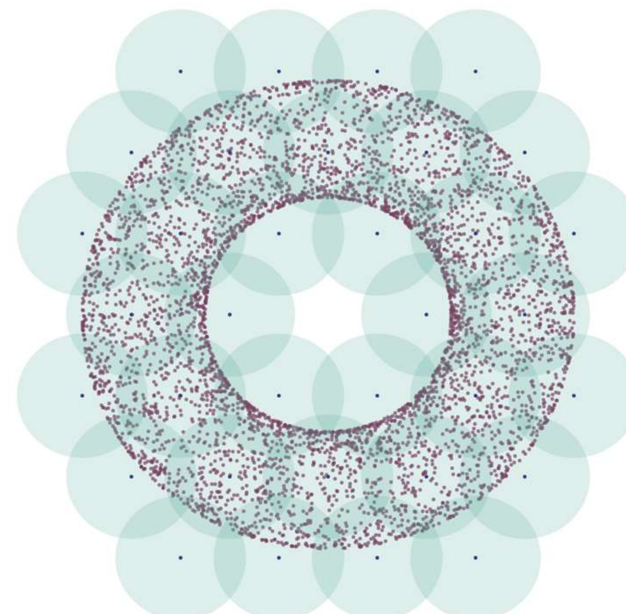
A_2^*-Lattice Cover

<u>Definition.</u> Let $Z \subset \mathbb{R}^n$ be a compact topological space with bounding box $B = \prod_{i=1}^{n}[m_i, M_i]$. Let A_2^* denote the root lattice generated with matrix

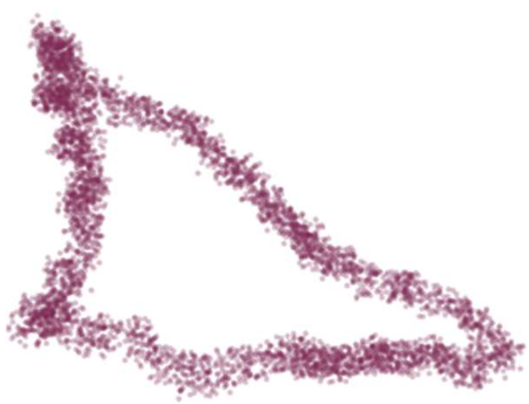
$$M_{A_2^*} = \begin{pmatrix} 1 & 0 \\ \dfrac{-1}{2} & \dfrac{\sqrt{3}}{2} \end{pmatrix}$$

The A_2^*-*lattice cover* defined with k intervals and overlap fraction g is a cover $\mathcal{L} = \left\{B_\varepsilon\left(c\xi M_{A_2^*}\right)\right\}_{c\xi \in B \cap \mathbb{Z}^2}$ with

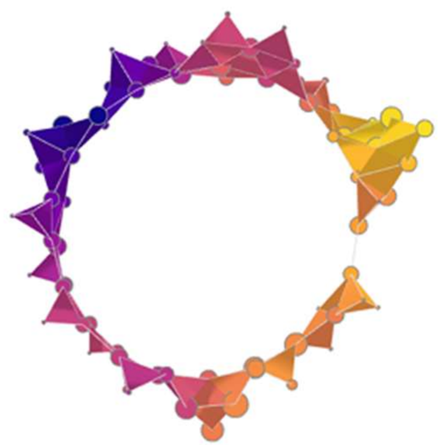
$$\varepsilon = \frac{1+g}{\sqrt{3}} \max_{i=1,2} \frac{M_i - m_i}{k}.$$



Analysis of 2-Mapper through Cover Choice



Stochastic Triangle in $\mathbb{R}^8$



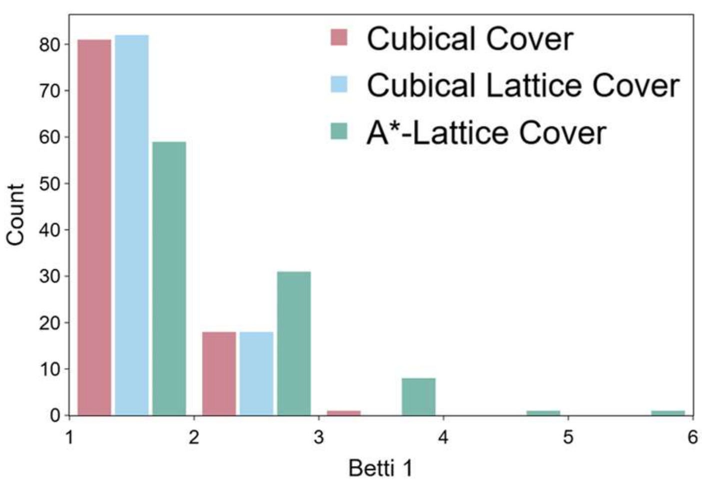
Cubical Cover

Cubical Lattice Cover

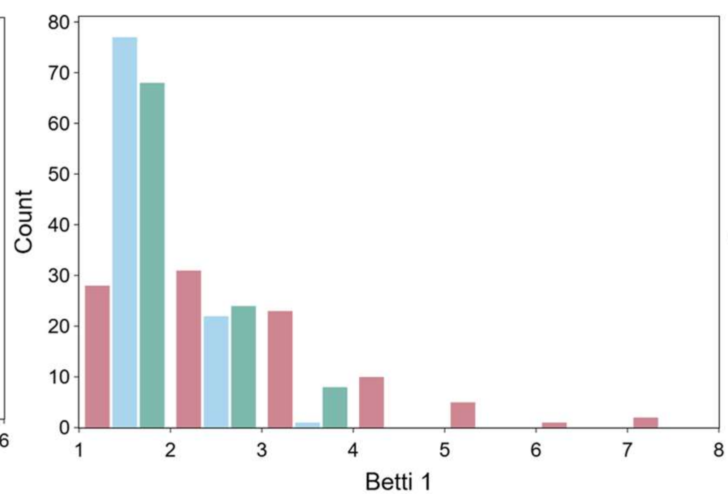
A_2^*-Lattice Cover

Analysis of 2-Mapper through Cover Choice

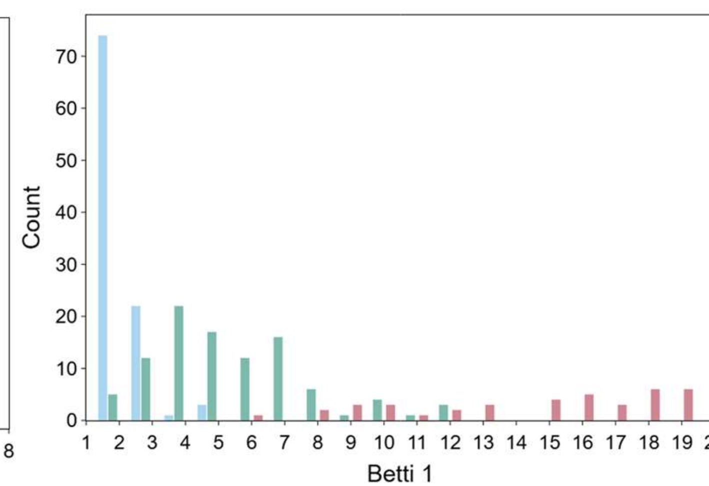
- 100 stochastic triangles generated in $\mathbb{R}^8$
- 2-Mapper complexes constructed with $k = 10$ and $g = 0.3$



2 Principal Components



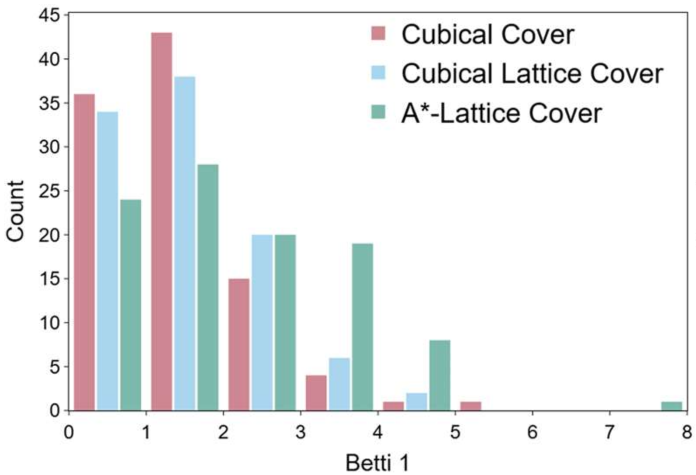
3 Principal Components



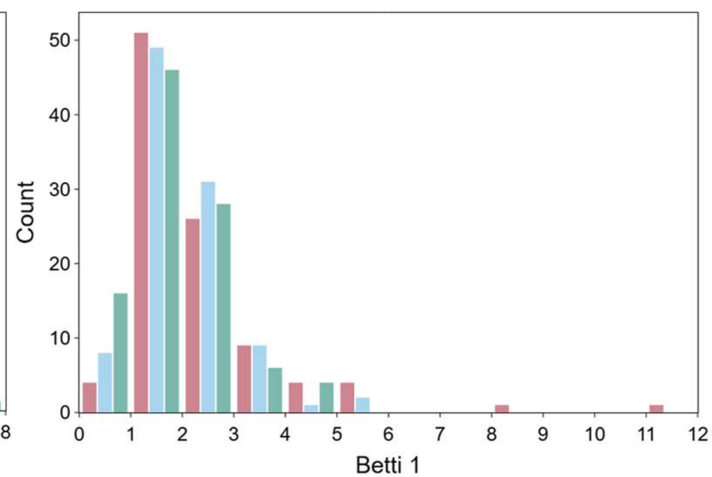
4 Principal Components

Analysis of 2-Mapper through Cover Choice

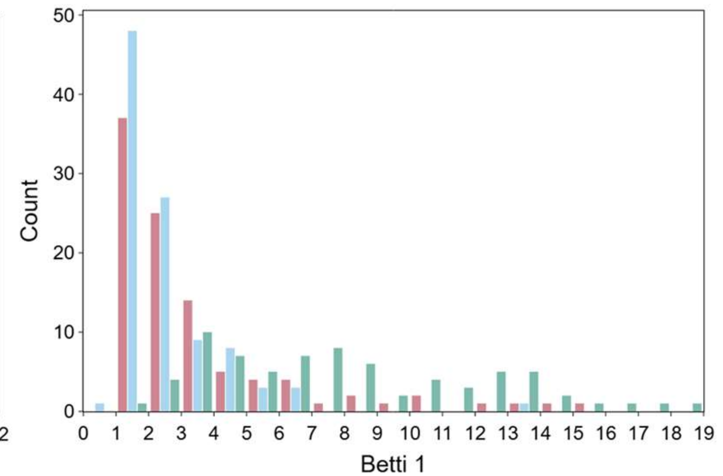
- 100 stochastic triangles generated in $\mathbb{R}^8$
- 2-Mapper complexes constructed with $k = 10$ and $g = 0.3$



$\mathbb{R}^2$ Coordinate Projection



$\mathbb{R}^3$ Coordinate Projection

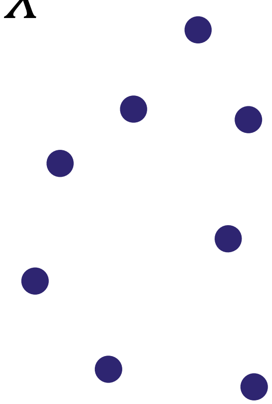


$\mathbb{R}^4$ Coordinate Projection

Persistent Homology

Construct a filtration of simplicial complexes for $r \geq 0$,

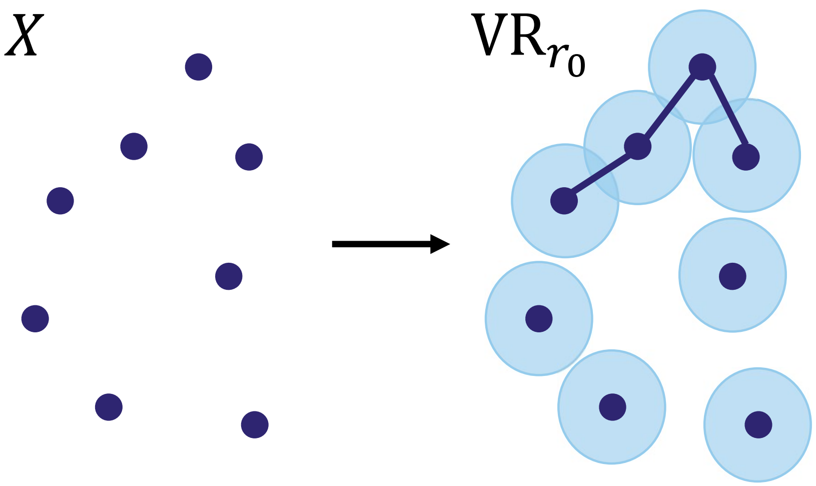
X



Persistent Homology

Construct a filtration of simplicial complexes for $r \geq 0$,

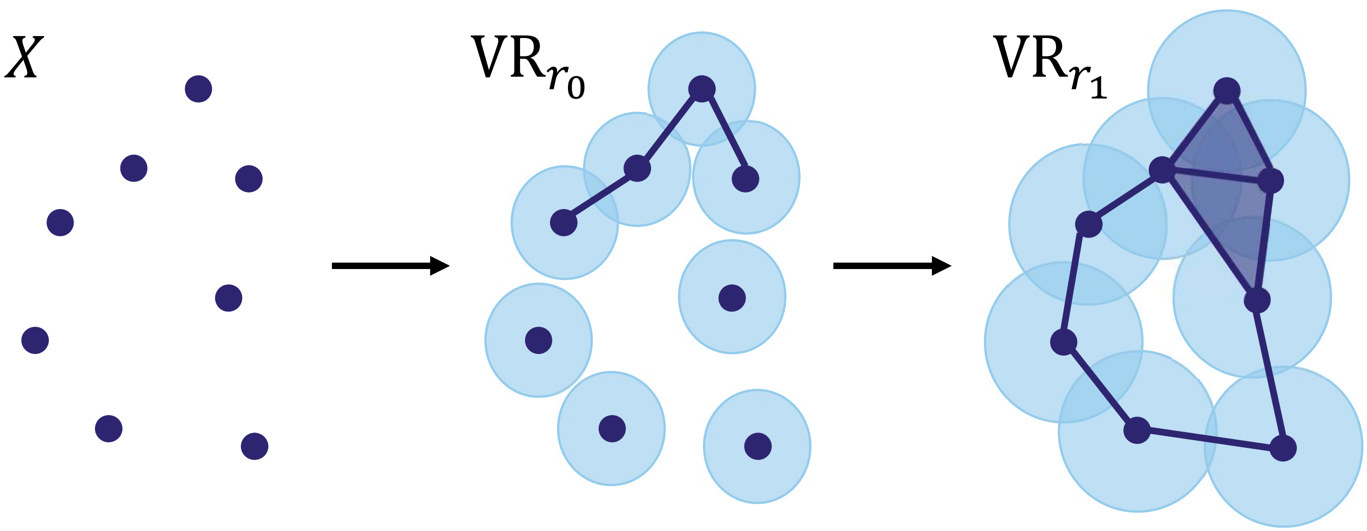
$$\mathrm{VR}_{r_0}(X)$$



Persistent Homology

Construct a filtration of simplicial complexes for $r \geq 0$,

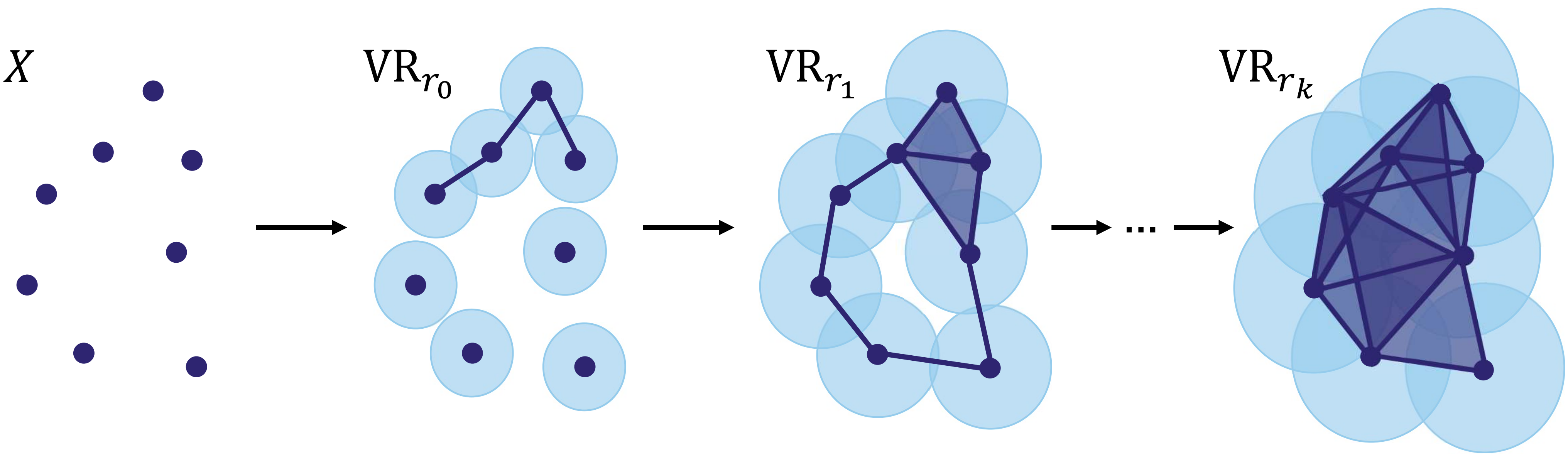
$$\mathrm{VR}_{r_0}(X) \subseteq \mathrm{VR}_{r_1}(X)$$



Persistent Homology

Construct a filtration of simplicial complexes for $r \geq 0$,

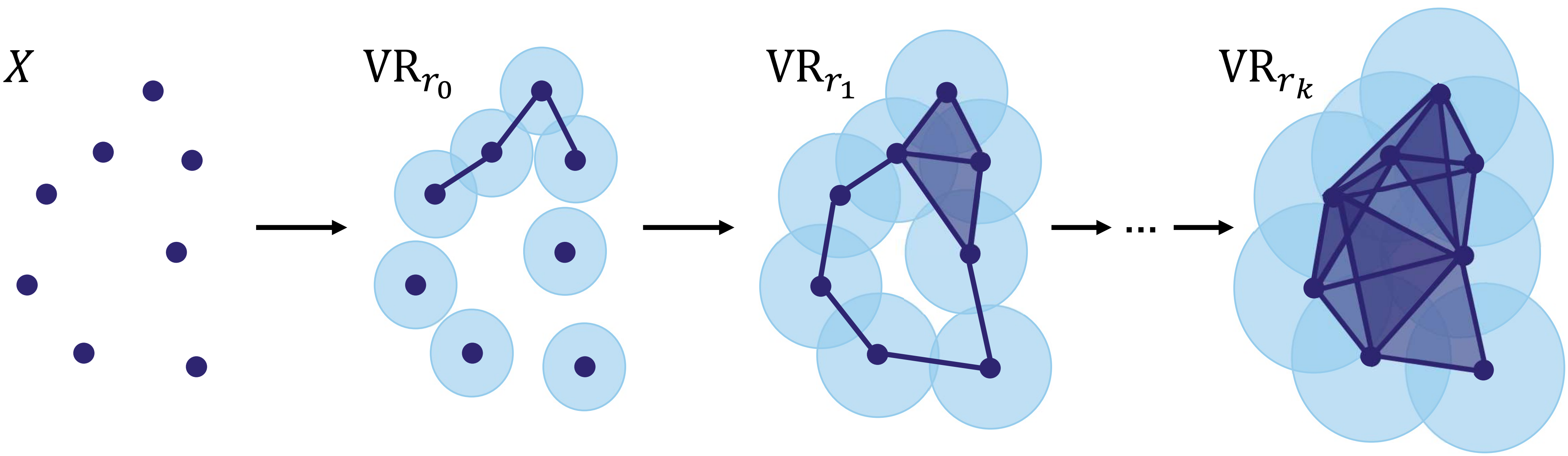
$$\mathrm{VR}_{r_0}(X) \subseteq \mathrm{VR}_{r_1}(X) \subseteq \cdots \subseteq \mathrm{VR}_{r_k}(X)$$



Persistent Homology

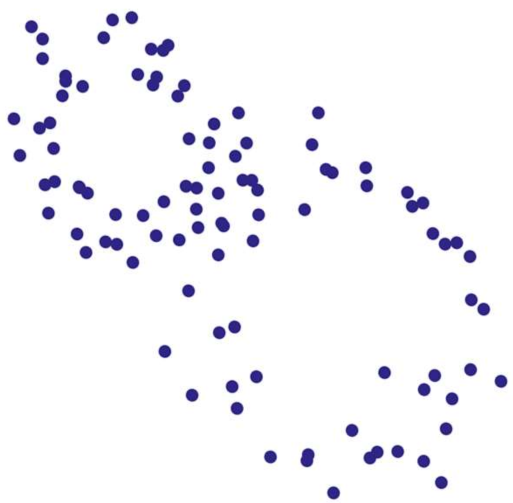
Applying homology yields a *persistence vector space:*

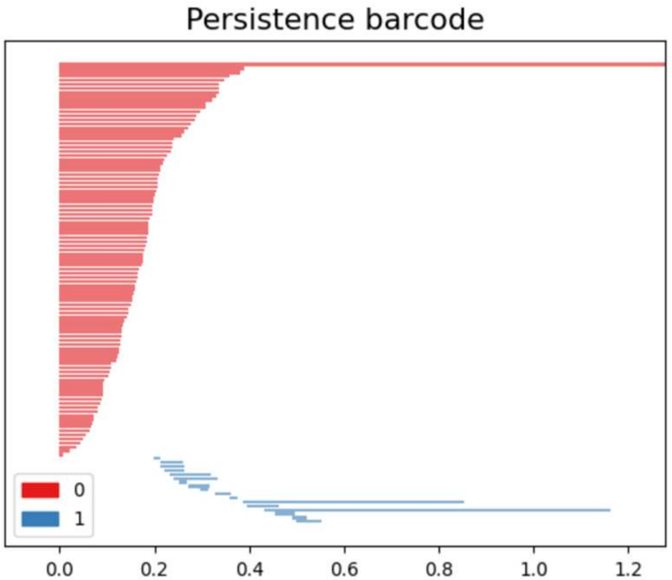
$$H_*\big(\mathrm{VR}_{r_0}(X)\,;\mathbb{Z}_2\big) \to H_*\big(\mathrm{VR}_{r_1}(X)\,;\mathbb{Z}_2\big) \to \cdots \to H_*\big(\mathrm{VR}_{r_k}(X)\,;\mathbb{Z}_2\big)$$

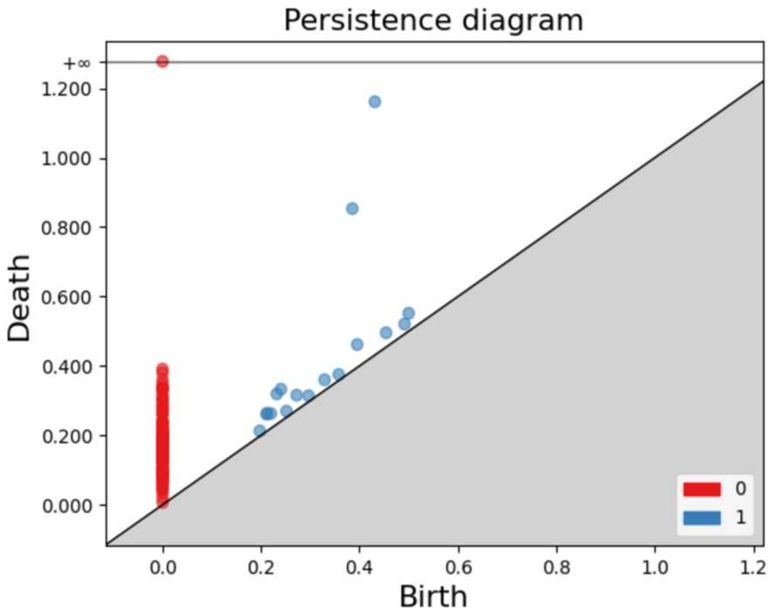


Barcodes and Persistent Diagrams

We track the birth and death times for each cycle as a persistent barcode or persistence diagram.







Distances between diagrams

Definition. Let $D_k(V)$ and $D_k(W)$ be two degree-k persistence diagrams for persistence vector spaces V and W.

Let $\Pi = \{\pi : D_k(V) \to D_k(W)\}$ be the set of bijections between their points to each other or to the diagonal $\{(x, x) : x \in \mathbb{R}_+\}$.

The *bottleneck distance* is defined as

$$d_B\big(D_k(V), D_k(W)\big) = \inf_{\pi \in \Pi} \sup_{x \in D_k(V)} \|x - \pi(x)\|_\infty$$

Multiscale Mapper

<u>Definition.</u> Let X and Z be topological spaces. For a well-behaved continuous map $f : X \to Z$ and finite open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of Z the *generalized Mapper complex* is

$$\mathrm{M}(f, \mathcal{U}) \coloneqq \mathcal{N}\big(f^*(\mathcal{U})\big)$$

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale Mapper

<u>**Definition.**</u> A *tower of covers* with *resolution* $s \in \mathbb{R}_{\geq 0}$ is a collection of covers $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon \geq s}$ with maps $u_{\varepsilon,\delta}: \mathcal{U}_\varepsilon \to \mathcal{U}_\delta$ so that $u_{\varepsilon,\varepsilon} = \mathrm{Id}$ and $u_{\varepsilon,\gamma} = u_{\delta,\gamma} \circ u_{\varepsilon,\delta}$ for all $s \leq \varepsilon \leq \delta \leq \gamma$.

We write $\mathrm{res}(\mathfrak{U}) = s$ for the resolution of tower $\mathfrak{U}$.

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale Mapper

<u>Definition.</u> Let X and Z be topological spaces and $f\colon X \to Z$ be a well-behaved continuous map. Let $\mathfrak{U}$ be a tower of covers of Z.

The *Multiscale Mapper* is the tower of simplicial complexes defined

$$MM(\mathfrak{U}, f) \coloneqq \mathcal{N}(f^*(\mathfrak{U})).$$

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale Mapper

Definition. Let X and Z be topological spaces and $f: X \to Z$ be a well-behaved continuous map. Let $\mathfrak{U}$ be a tower of covers of Z.

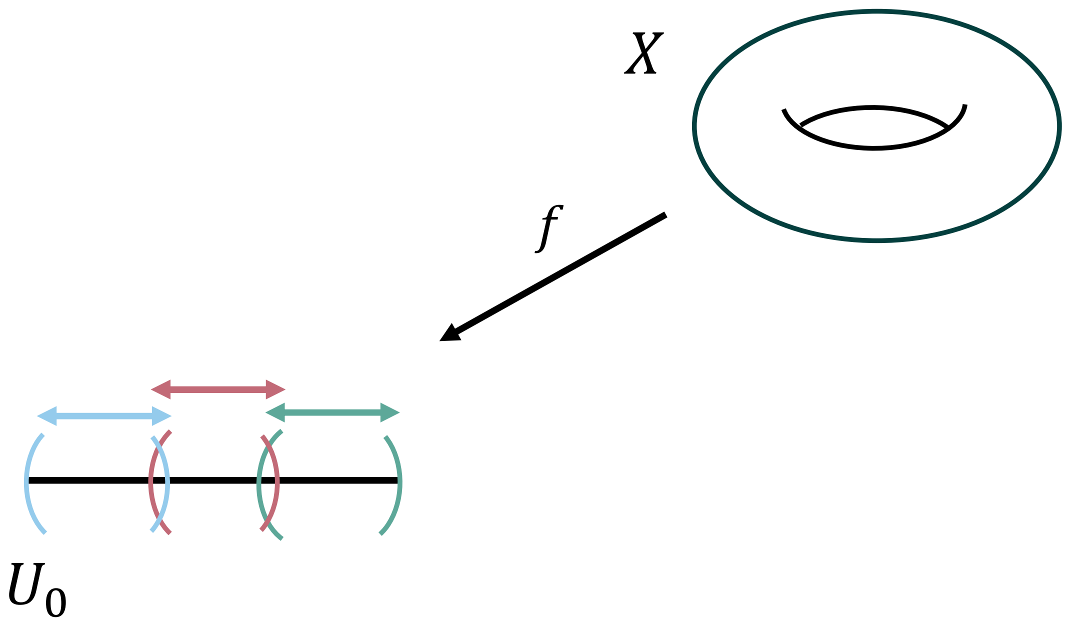
The *Multiscale Mapper* is the tower of simplicial complexes defined

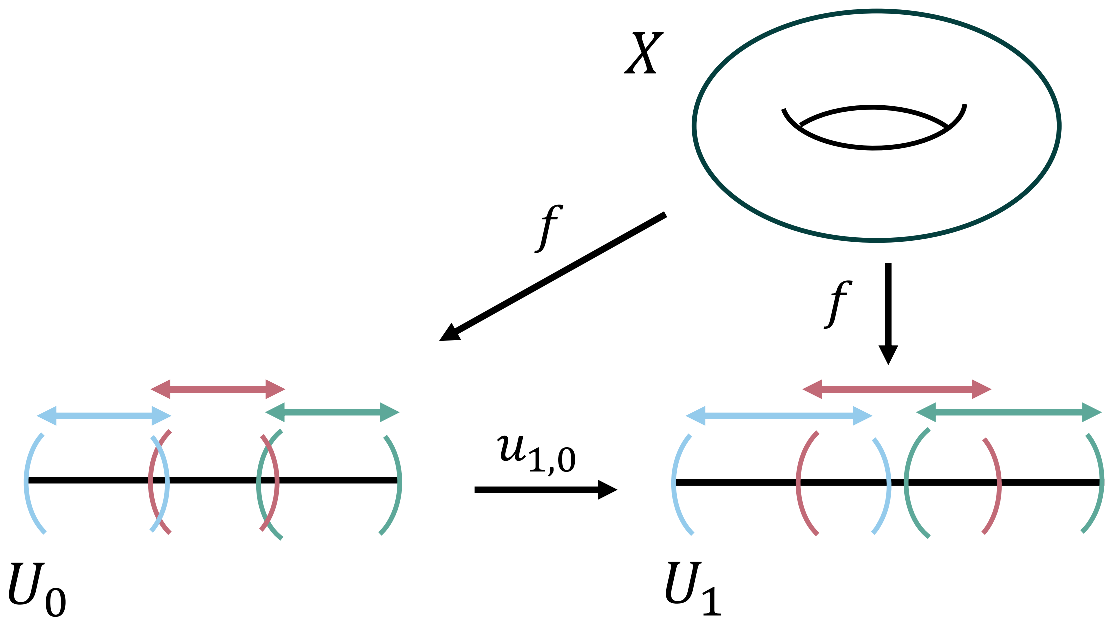
$$MM(\mathfrak{U}, f) \coloneqq \mathcal{N}(f^*(\mathfrak{U}).$$

For a finite sequence $\mathrm{res}(\mathfrak{U}) \leq \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_n$,

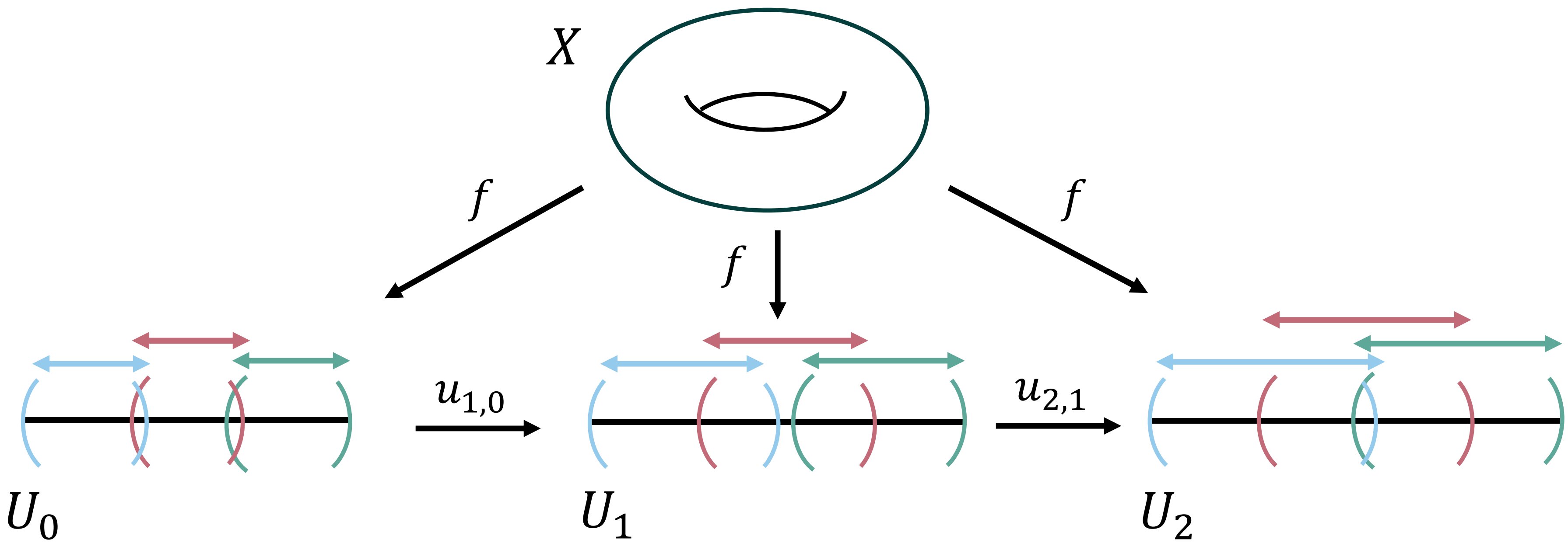
$$H_*\left(\mathcal{N}\left(f^*(\mathcal{U}_{\varepsilon_1})\right)\right) \to H_*\left(\mathcal{N}\left(f^*(\mathcal{U}_{\varepsilon_2})\right)\right) \to \cdots \to H_*\left(\mathcal{N}\left(f^*(\mathcal{U}_{\varepsilon_n})\right)\right).$$

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

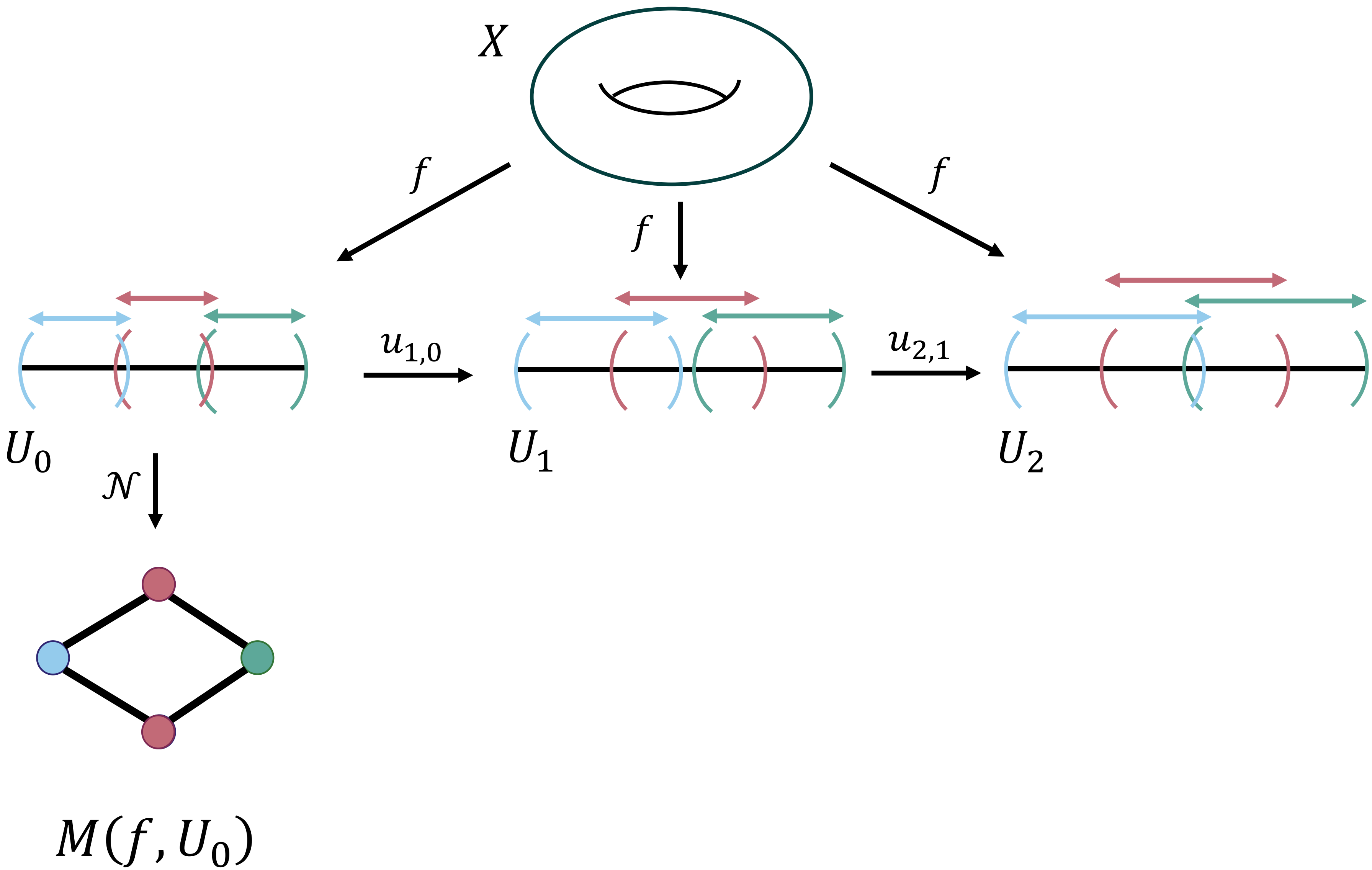

X
f
U0



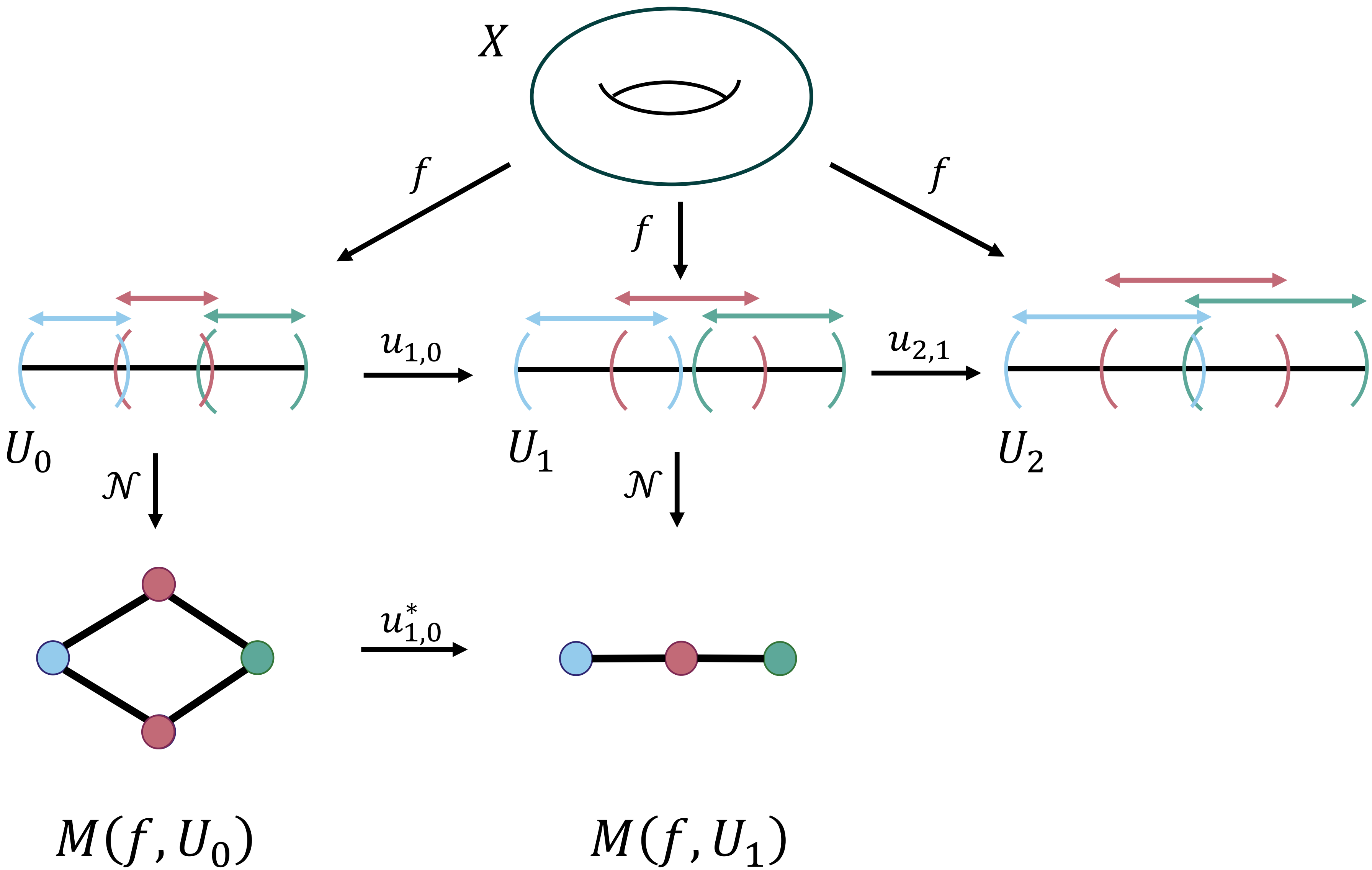
X
f
f
$u_{1,0}$
U_0
U_1

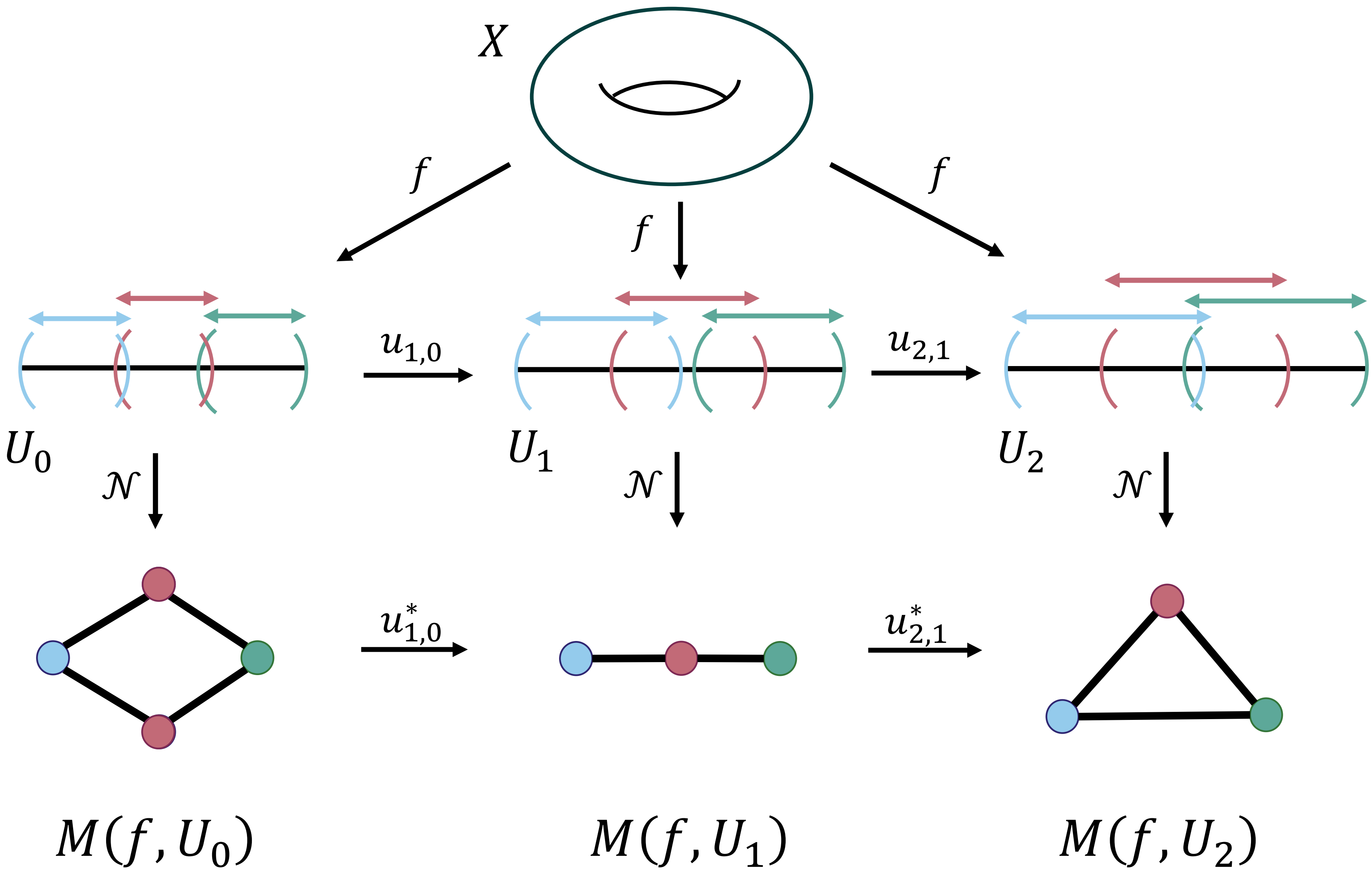


X
f
f
f
$u_{1,0}$
$u_{2,1}$
U_0
U_1
U_2



X
f
f
f
$u_{1,0}$
$u_{2,1}$
U_0
U_1
U_2
$\mathcal{N}$
$M(f, U_0)$


X
f
f
f
$u_{1,0}$
$u_{2,1}$
U_0
U_1
U_2
$\mathcal{N}$
$\mathcal{N}$
$u^*_{1,0}$
$M(f, U_0)$
$M(f, U_1)$



X
f
f
f
U_0
$u_{1,0}$
U_1
$u_{2,1}$
U_2
$\mathcal{N}$
$\mathcal{N}$
$\mathcal{N}$
$u^*_{1,0}$
$u^*_{2,1}$
$M(f, U_0)$
$M(f, U_1)$
$M(f, U_2)$

Multiscale Mapper

Definition. Let $c \geq 1$ and $s > 0$. A finite tower of covers $\mathfrak{W} = \{\mathcal{W}_\varepsilon\}$ over a compact metric space (Z, d_Z) is (c, s)-*good* if

1. $\mathrm{res}(\mathfrak{W}) = s$, and $s \leq \mathrm{diam}(Z)$
2. $\mathrm{diam}\left(W_{\varepsilon,\alpha}\right) \leq \varepsilon$ for all $W_{\varepsilon,\alpha} \in \mathcal{W}_\varepsilon$, $\varepsilon \geq s$
3. $\forall\ O \subset Z$ with $\mathrm{diam}(O) \geq s$, there exists $W \in \mathcal{W}_{c\cdot\mathrm{diam}(O)}$ such that $O \subset W$.

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale Mapper

Definition. Let $c \geq 1$ and $s > 0$. A finite tower of covers $\mathfrak{W} = \{\mathcal{W}_\varepsilon\}$ over a compact metric space (Z, d_Z) is (c, s)-*good* if

1. $\mathrm{res}(\mathfrak{W}) = s$, and $s \leq \mathrm{diam}(Z)$
2. $\mathrm{diam}\big(W_{\varepsilon,\alpha}\big) \leq \varepsilon$ for all $W_{\varepsilon,\alpha} \in \mathcal{W}_\varepsilon$, $\varepsilon \geq s$
3. $\forall\ O \subset Z$ with $\mathrm{diam}(O) \geq s$, there exists $W \in \mathcal{W}_{c \cdot \mathrm{diam}(O)}$ such that $O \subset W$.

Example. A tower of covers $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon \geq s}$ with $\mathcal{U}_\varepsilon = \big\{B_{\varepsilon/2}(z) \mid z \in Z\big\}$ is a $(2, s)$-good tower of covers of compact metric space Z.

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale Mapper

Theorem (Dey, M., W.). For two (c, s)-good tower of covers $\mathfrak{U}, \mathfrak{V}$, the bottleneck distance between two diagrams produced from their multiscale mappers is bounded,

$$d_B\left(D_k\big(MM(f, \mathfrak{U})\big), D_k\big(MM(f, \mathfrak{V})\big)\right) \leq c.$$

Tamal K. Dey, Fecundo Mèmoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms,* pages 997-1013. 2016.

Multiscale 2-Mapper

Definition. Let $Z \subset \mathbb{R}^n$ be a compact topological space. Then for lens function $f: Z \to \mathbb{R}^m$ and tower of covers $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon \geq s}$ over $f(Z)$, the *Multiscale 2-Mapper*, denoted $\mathrm{MM}_2(f, \mathfrak{U})$, for a finite sequence $s \leq \varepsilon_1 \leq \cdots \leq \varepsilon_k$ is of filtration of 2-Mapper complexes

$$\mathcal{N}^2\big(f^*(\mathcal{U}_s)\big) \to \mathcal{N}^2\left(f^*\left(\mathcal{U}_{\varepsilon_1}\right)\right) \to \cdots \to \mathcal{N}^2\left(f^*\left(\mathcal{U}_{\varepsilon_k}\right)\right).$$

Tower of Cubical Covers

Definition. Let $X \subset \mathbb{R}^n$ be a compact topological space with bounding box B and cubical cover $\mathcal{U}_s$ constructed with k-intervals and overlap fraction g so that each cover set is of the form

$$U_{\alpha,s} = \prod_{i=1}^{n} \left[c_{\alpha,i} - \frac{l_i}{2}, c_{\alpha,i} + \frac{l_i}{2} \right]$$

Where $c_{\alpha,i} = m_1 + (\alpha_i - 1)(1 - g)l_i + \frac{1}{2} l_i$ and $l_i = \frac{M_i - m_i}{k - (k-1)g}$.

Tower of Cubical Covers

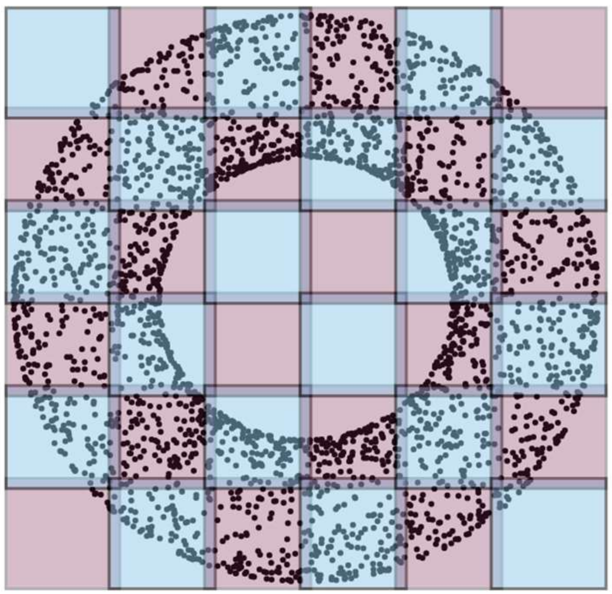
Definition. The *tower of cubical covers* $\mathfrak{U}$ on X is a tower $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon \geq s}$ with resolution $\mathrm{res}(\mathfrak{U}) = s = \|(l_1, \ldots, l_n)\|_2$.
For $\varepsilon \geq s$, each cover $\mathcal{U}_\varepsilon = \left\{U_{\alpha,\varepsilon}\right\}_{\alpha \in A}$ such that for each $\alpha \in A$,

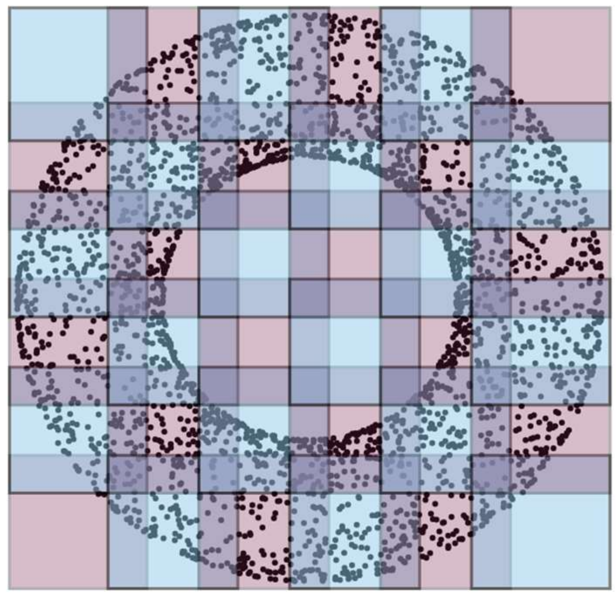
$$U_{\alpha,\varepsilon} = \prod_{i=1}^{n} \left[c_{\alpha,i} - \frac{1}{2}(l_i - \varepsilon'), c_{\alpha,i} + \frac{1}{2}(l_i + \varepsilon')\right],$$

For some $\varepsilon' \geq 0$ so that $\mathrm{diam}\left(U_{\alpha,\varepsilon}\right) = \varepsilon$.

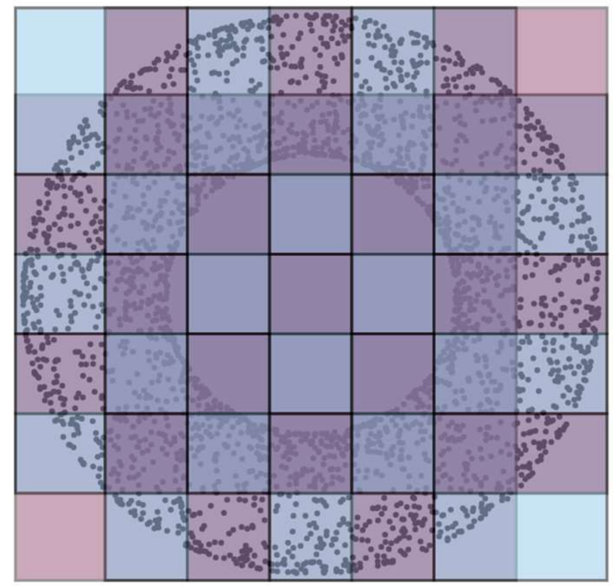
This is a filtration over the parameter g.

Tower of Cubical Covers constructed with $0.1 \leq g \leq 0.5$ and $k = 6$.





$g = 0.1$



$g = 0.3$

$g = 0.5$

Tower of Lattice Covers

Definition. For $X \subset \mathbb{R}^n$ with bounding box B, let $\mathcal{U}_s$ be a cubical lattice cover constructed with k-intervals and overlap fraction g.

The *tower of cubical lattice covers* $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon \geq s}$ is a tower of covers with resolution $\mathrm{res}(\mathfrak{U}) = s = \frac{c\sqrt{n}}{1-g}$.
For each $\varepsilon \geq s$ we define each cover $\mathcal{U}_\varepsilon = \left\{U_{\alpha,\varepsilon}\right\}_{\alpha \in A}$ so that

$$U_{\alpha,\varepsilon} = \prod_{i=1}^{n} \left[c\xi_{\alpha,i} - \frac{1}{2}\left(\frac{c}{1-g} - \varepsilon'\right), c\xi_{\alpha,i} + \frac{1}{2}\left(\frac{c}{1-g} + \varepsilon'\right)\right],$$

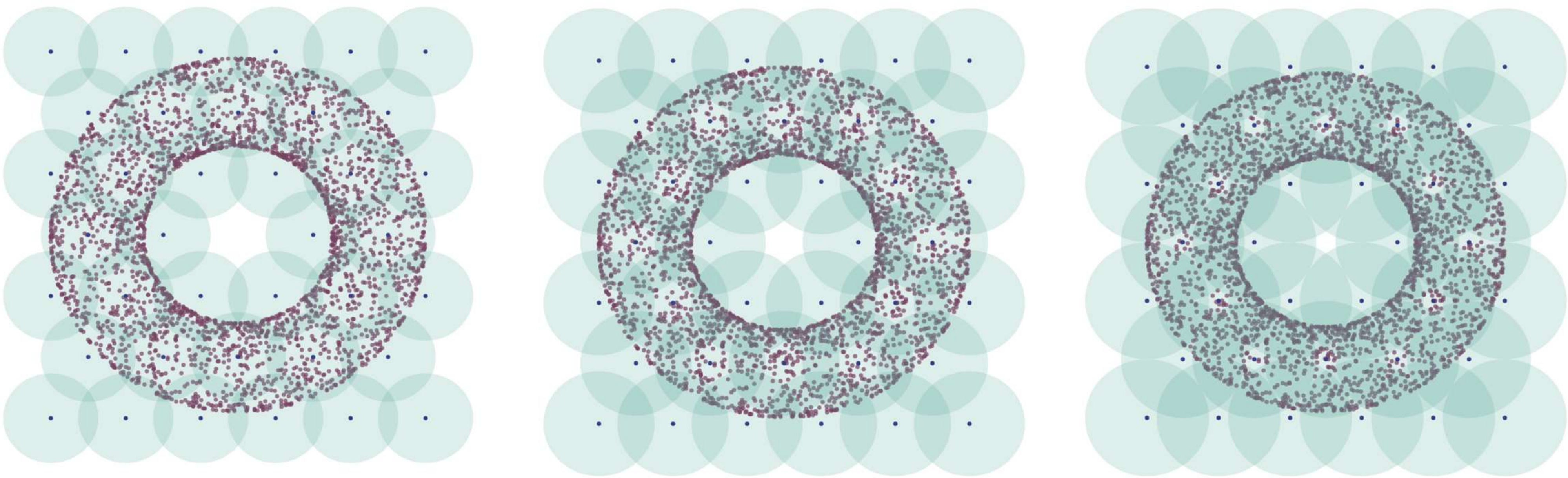
For some $\varepsilon' \geq 0$ so that $\mathrm{diam}\left(U_{\alpha,\varepsilon}\right) = \varepsilon$.

Tower of A_2^*-Lattice Covers

Definition. For $X \subset \mathbb{R}^n$ with bounding box B, let $\mathcal{L}_s$ be an A_2^*-lattice cover constructed with k-intervals and overlap fraction g.

The *tower of* A_2^**-lattice covers* $\mathfrak{L} = \{\mathcal{L}_\varepsilon\}_{\varepsilon \geq s}$ is a tower of covers with resolution $\mathrm{res}(\mathfrak{L}) = s = \frac{2c(1+g)}{\sqrt{3}}$.
For each $\varepsilon \geq s$ we define each cover $\mathcal{L}_\varepsilon = \left\{B\left(c\xi_\alpha M_{A_2^*}, \frac{1}{2}\varepsilon\right)\right\}_{\alpha \in A}$ with $\varepsilon = s + \varepsilon'$ for some $\varepsilon' \geq 0$.

Tower of A_2^*-Lattice Covers constructed with $0.1 \leq g \leq 0.5$ and $k = 6$.



$g = 0.1$ $g = 0.3$ $g = 0.5$

Stability for Covers

Theorems (F.). A tower of cubical (lattice) covers with resolution s constructed with k-intervals and overlap fraction g is $(3, s)$-good for $k \geq \sqrt{n}$.

Theorem (F.). A tower of A_2^*-lattice covers with resolution s, constructed with k-intervals and overlap fraction g is $(3, s)$-good for $k \geq \frac{2}{\sqrt{3}}(1 + g)$.

Stability for Covers

Theorems (F.). A tower of cubical (lattice) covers with resolution s constructed with k-intervals and overlap fraction g is $(3, s)$-good for $k \geq \sqrt{n}$.

Theorem (F.). A tower of A_2^*-lattice covers with resolution s, constructed with k-intervals and overlap fraction g is $(3, s)$-good for $k \geq \frac{2}{\sqrt{3}}(1 + g)$.

- Two Multiscale 2-Mapper objects with these towers are close with respect to the bottleneck distance!

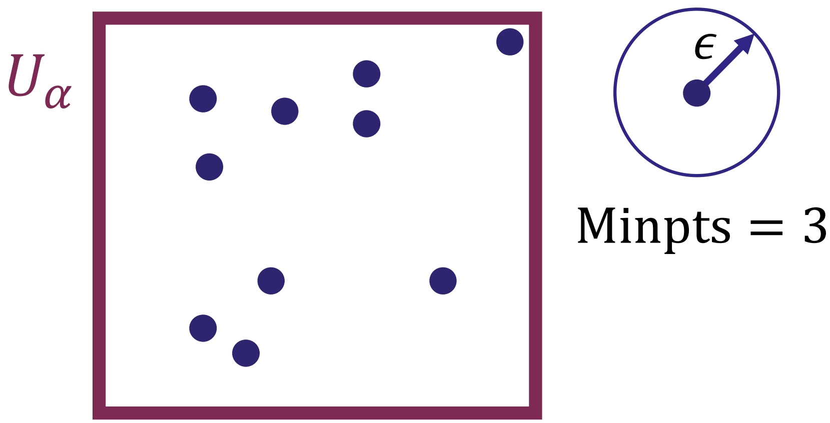
Multiscale Mapper with DBSCAN

<u>**Definition.**</u> *DBSCAN* is a non-parametric density-based clustering algorithm constructed with parameters ϵ and Minpts representing the search radius and minimum number of points per cluster, respectively.

W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

Multiscale Mapper with DBSCAN

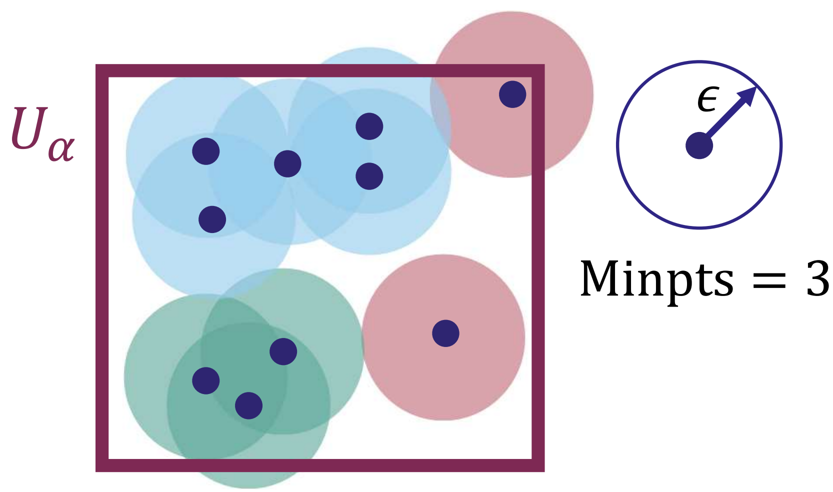
Definition. *DBSCAN* is a non-parametric density-based clustering algorithm constructed with parameters ϵ and Minpts representing the search radius and minimum number of points per cluster, respectively.



W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

Multiscale Mapper with DBSCAN

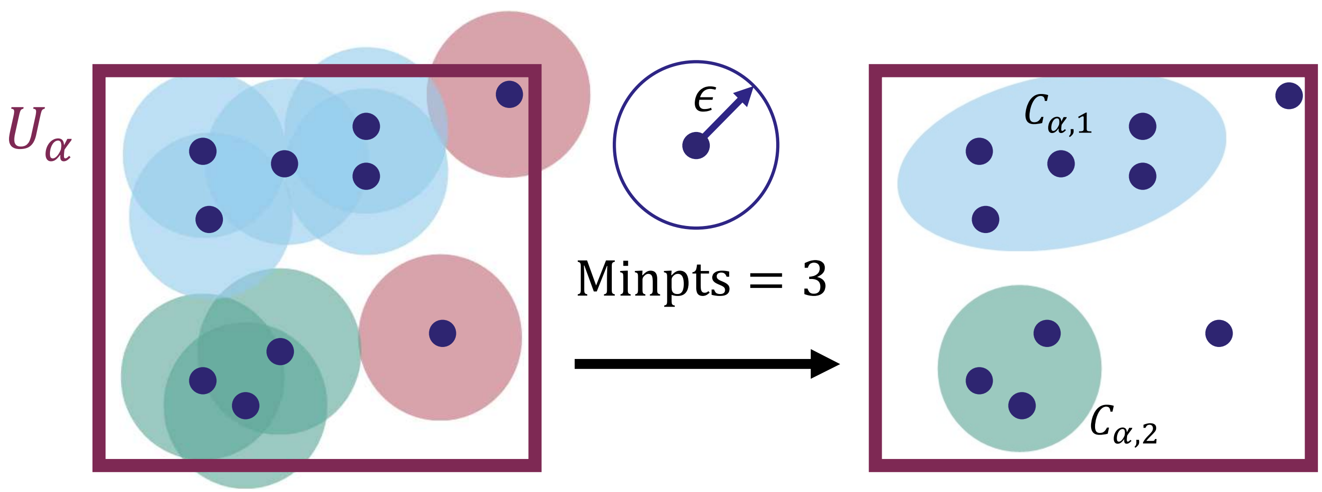
Definition. *DBSCAN* is a non-parametric density-based clustering algorithm constructed with parameters ϵ and Minpts representing the search radius and minimum number of points per cluster, respectively.



W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

Multiscale Mapper with DBSCAN

Definition. *DBSCAN* is a non-parametric density-based clustering algorithm constructed with parameters ϵ and Minpts representing the search radius and minimum number of points per cluster, respectively.



$$U_\alpha = C_{1,\alpha} \sqcup C_{2,\alpha} \sqcup N_\alpha$$

W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

Multiscale Mapper with DBSCAN

Theorem (Bungula, Darcy). Let X be a data set with tower of covers $\mathcal{U}$ and use DBSCAN to cluster X with fixed parameters ϵ and $\mathrm{Minpts} \leq 2$. Define $\mathcal{C}_\varepsilon = \left\{C_{p,\varepsilon} : p \in U_\varepsilon \; is \; a \; core \; point\right\}$. Then $\mathfrak{C} = \{\mathcal{C}_\varepsilon\}_{\varepsilon \geq s}$ is a tower of covers with maps $c_{\varepsilon,\delta} : \mathcal{C}_\varepsilon \to \mathcal{C}_\delta$ for all $\mathcal{U}_\varepsilon \subseteq \mathcal{U}_\delta$ which are closed under composition.

W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

Multiscale Mapper with DBSCAN

<u>**Theorem (Bungula, Darcy).**</u> Let X be a data set and assume DBSCAN is used to cluster X with fixed hyperparameters ϵ and $\mathrm{Minpts} \leq 2$. Then for all $\mathcal{U}_\varepsilon \subseteq \mathcal{U}_\delta$,

- There is a filtration of simplicial complexes $\phi_{\varepsilon,\delta} : \mathcal{N}(\mathcal{C}_\varepsilon) \to \mathcal{N}(\mathcal{C}_\delta)$.
- There is a filtration of homology groups $f_{\varepsilon,\delta} : H_k\big(\mathcal{N}(\mathcal{C}_\varepsilon)\big) \to H_k\big(\mathcal{N}(\mathcal{C}_\delta)\big)$ for each degree k.

W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

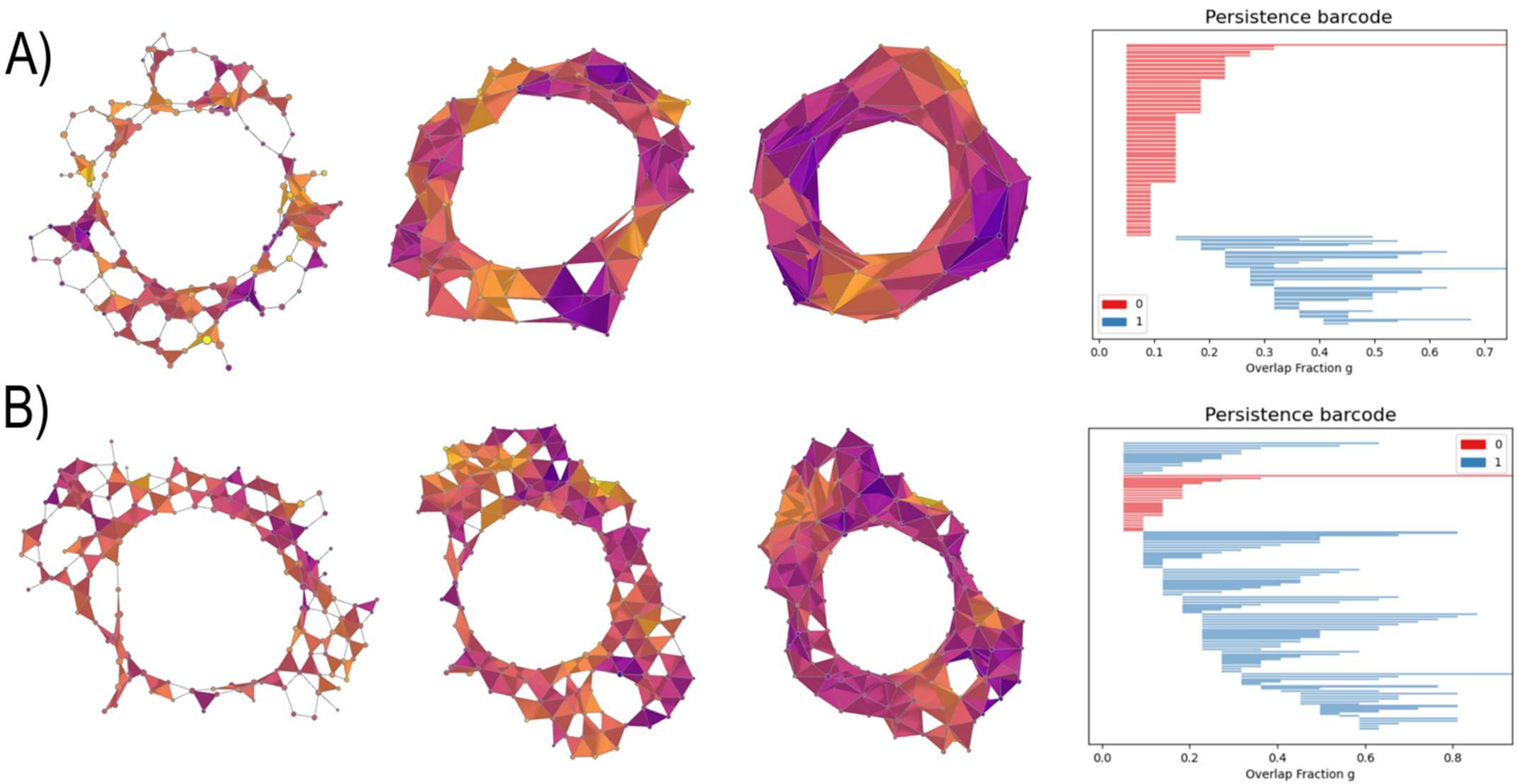
Stability for Covers with DBSCAN

Theorem (F.). Let $\mathfrak{U}$ be a tower of cubical, cubical lattice, or A_2^*-lattice covers with resolution s over a data set X. Define $\mathfrak{C} = \{\mathcal{C}_\varepsilon\}_{\varepsilon \geq s'}$ to be the cluster cover subordinate to $\mathfrak{U}$ using DBSCAN with fixed parameters ϵ and $Minpts \leq 2$. Then $\mathfrak{C}$ is a $(4, s')$-good cover where $s' \leq s$.

Stability for Covers with DBSCAN

Theorem (F.). Let $X \subset \mathbb{R}^m$ be a data set, $f : X \to \mathbb{R}^n$ be a continuous lens, and $\mathfrak{U}, \mathfrak{W}$ be two towers of covers (cubical, cubical lattice, or A_2^*-lattice). Cluster X using DBSCAN with fixed parameters ϵ and $\mathrm{Minpts} \leq 2$. Then for each $k \geq 0$,

$$d_B\big(D_k(\mathrm{MM}_2(f, \mathfrak{C}_{\mathfrak{U}})), D_k(\mathrm{MM}_2(f, \mathfrak{C}_{\mathfrak{W}}))\big) \leq 4$$



The bottleneck distances for these two barcodes is 0.0894 in degree zero and 0.1789 in degree one.

Building an Algorithm

Let $\mathrm{MM}_2(f, \mathfrak{C})$ be the Multiscale 2-Mapper computed over data set X with finite tower of covers $\mathfrak{U}$ and clustered with DBSCAN using parameters ϵ and Minpts.

Building an Algorithm

Let $\mathrm{MM}_2(f, \mathfrak{C})$ be the Multiscale 2-Mapper computed over data set X with finite tower of covers $\mathfrak{U}$ and clustered with DBSCAN using parameters ϵ and Minpts.

$$\mathrm{M}_2(\mathrm{f}, \mathcal{C}_{\varepsilon_0}) \overset{c^*_{0,1}}{\rightarrow} \mathrm{M}_2(\mathrm{f}, \mathcal{C}_{\varepsilon_1}) \overset{c^*_{1,2}}{\rightarrow} \cdots \overset{c^*_{t-1,t}}{\rightarrow} \mathrm{M}_2(f, \mathcal{C}_{\varepsilon_t})$$

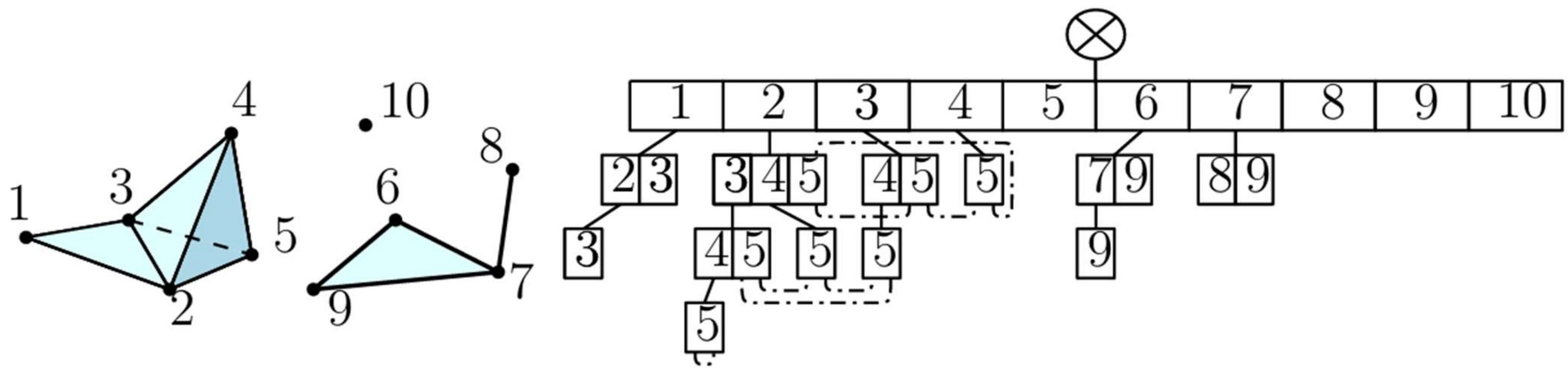
Write $M_i := \mathrm{M}_2(f, \mathcal{C}_{\varepsilon_i})$ where M_i has vertex set V_i.

For each node $n \in V_i$ we write $n = (X_n, U_{\alpha,\varepsilon_i}, C_{p_\alpha,\varepsilon_i})$

Data Cover set Cluster

Building an Algorithm

Simplex trees are used for computing with simplicial complexes and filtrations of simplicial complexes.



Each simplex σ is stored with filtration time t_σ.

J.-D. Boissonnat and C. Maria, The Simplex Tree: An Efficient Data Structure for General Simplicial Complexes, Algorithmica 70 (2014), no. 3, 406–427, doi: 10.1007/s00453-014-9887-3.

Building an Algorithm

Dream Construction. For data set X, lens $f : X \to \mathbb{R}^n$, tower of covers $\mathfrak{U}$, and DBSCAN parameters ϵ and Minpts.

- Create 2-Mapper complexes M_i for each cover in cluster cover $\mathfrak{C}$.
- Start with an empty simplex tree.
- Insert each M_i sequentially.
- Done!

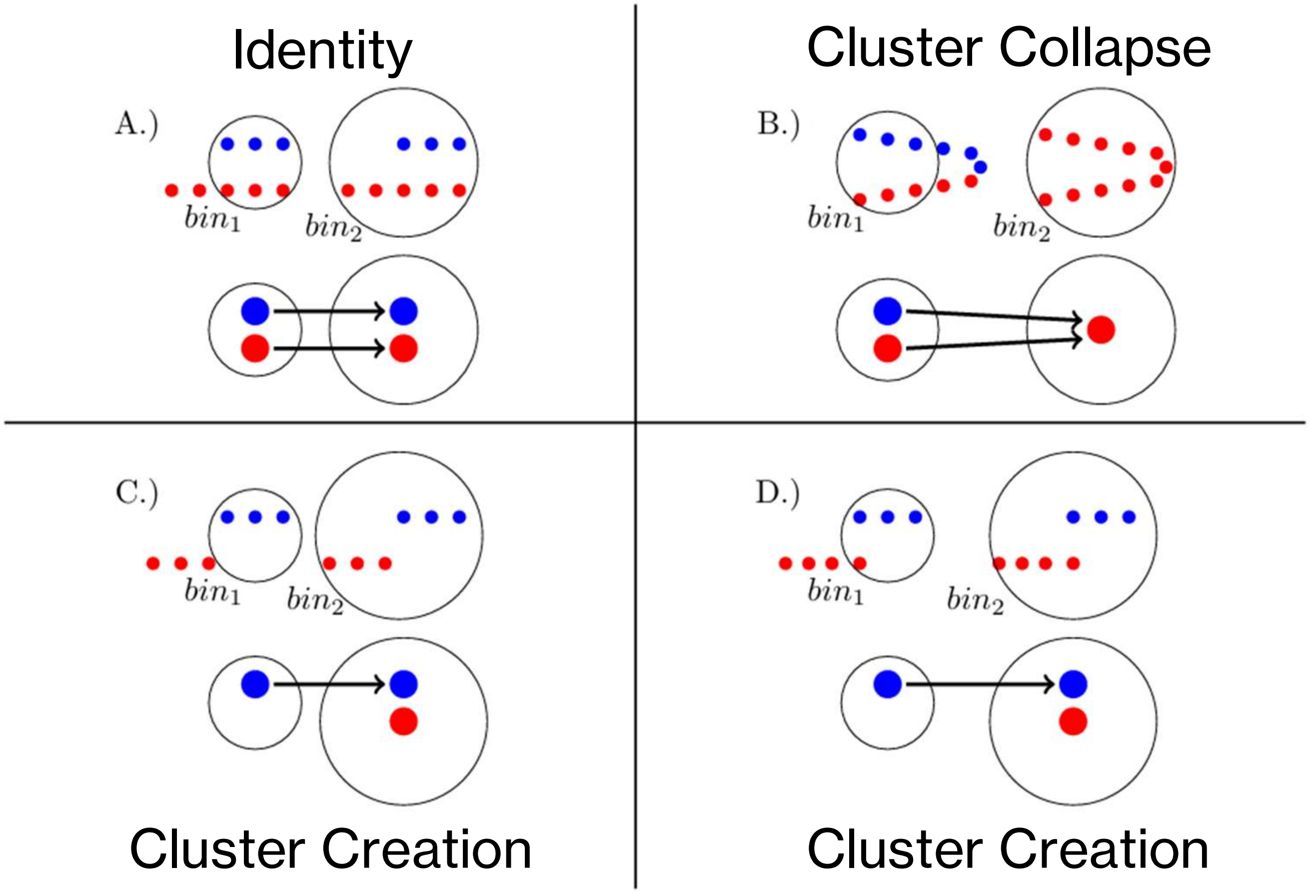
Building an Algorithm

Dream Construction. For data set X, lens $f : X \to \mathbb{R}^n$, tower of covers $\mathfrak{U}$, and DBSCAN parameters ϵ and Minpts.

- Create 2-Mapper complexes M_i for each cover in cluster cover $\mathfrak{C}$.
- Start with an empty simplex tree.
- Insert each M_i sequentially.
- ~~Done!~~

Clustering complicates our algorithm.

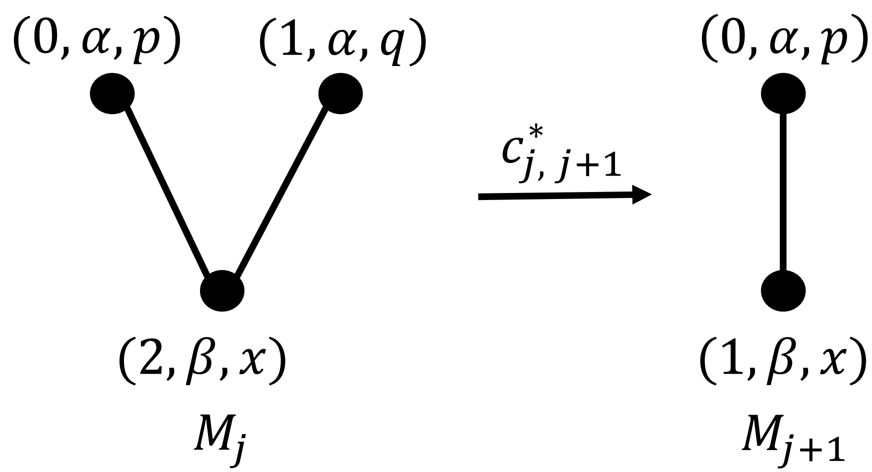
Building an Algorithm



W. Bungula and I. Darcy, Bi-Filtration and Stability of TDA Mapper for Point Cloud Data, arXiv: 2409.17360 [math.AT], 2024.

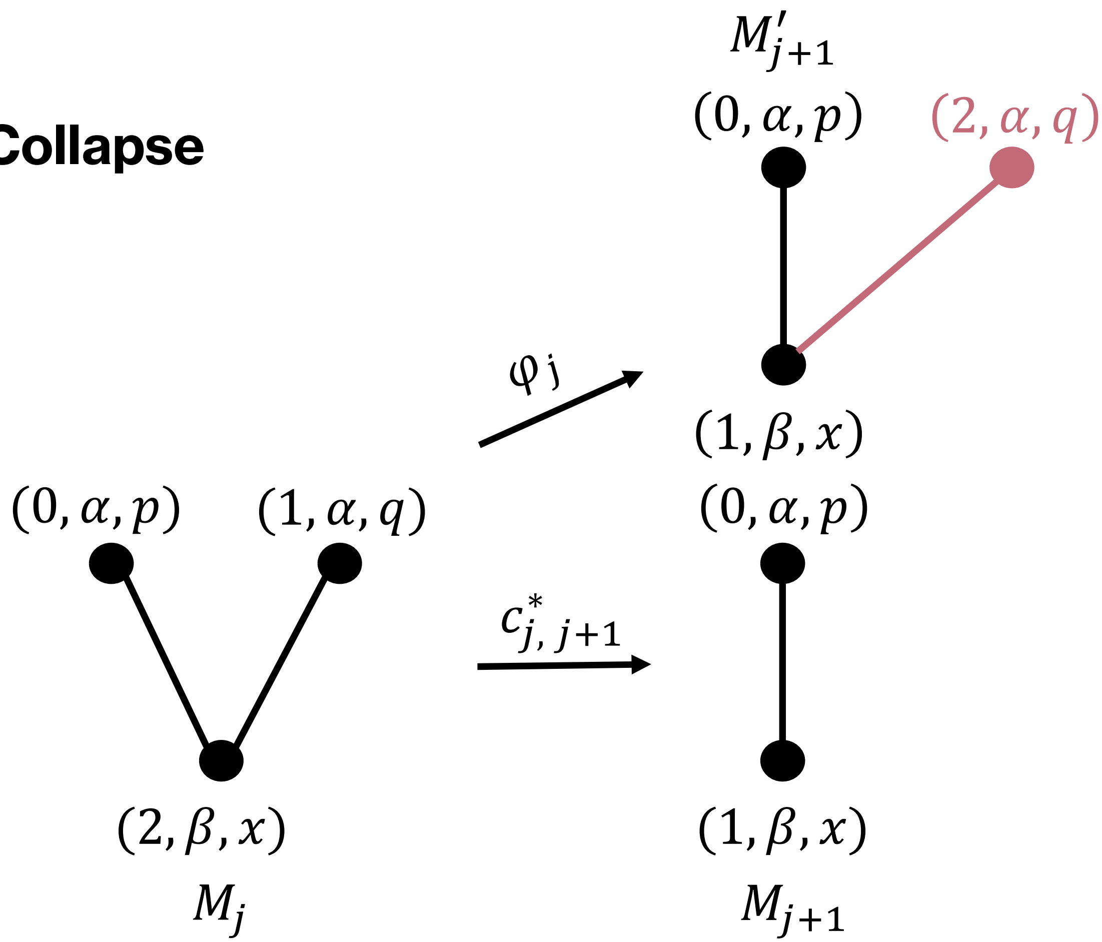
Building an Algorithm

Single Cluster Collapse



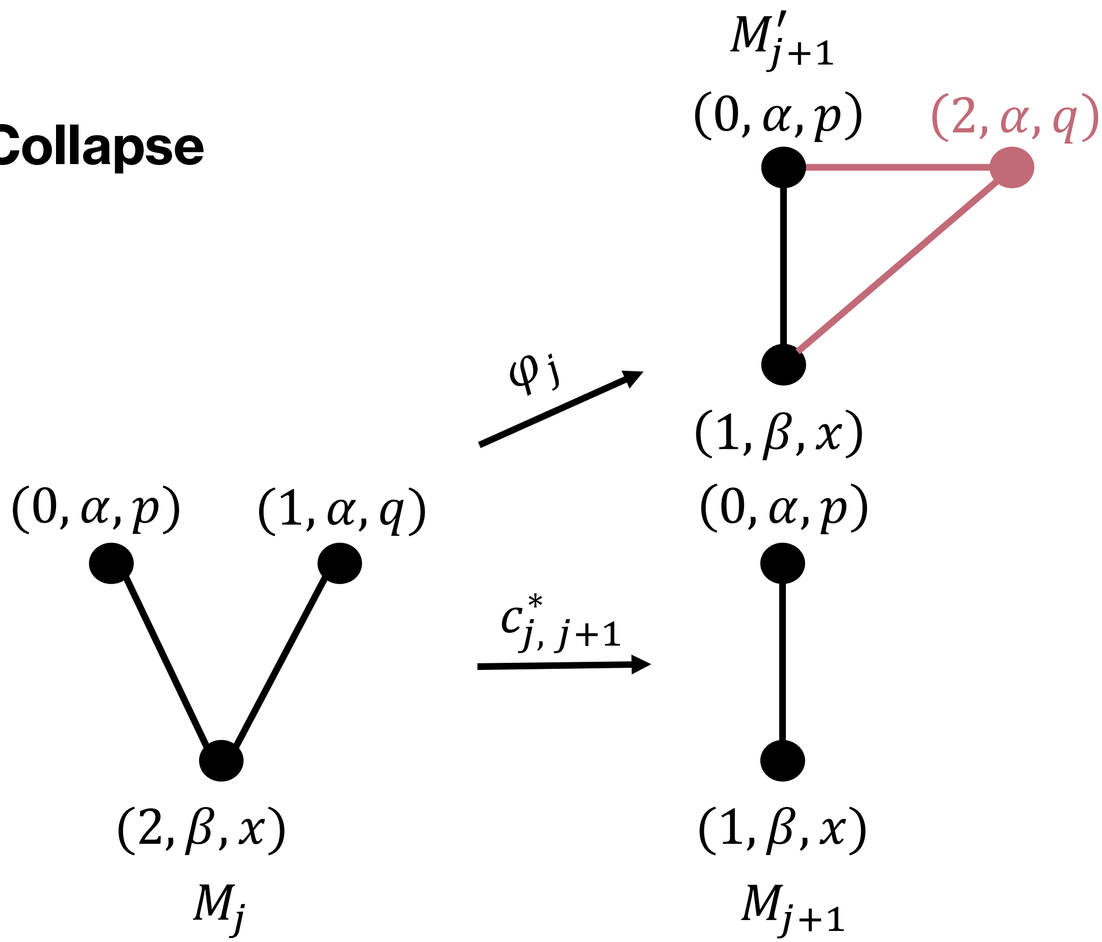
Building an Algorithm

Single Cluster Collapse



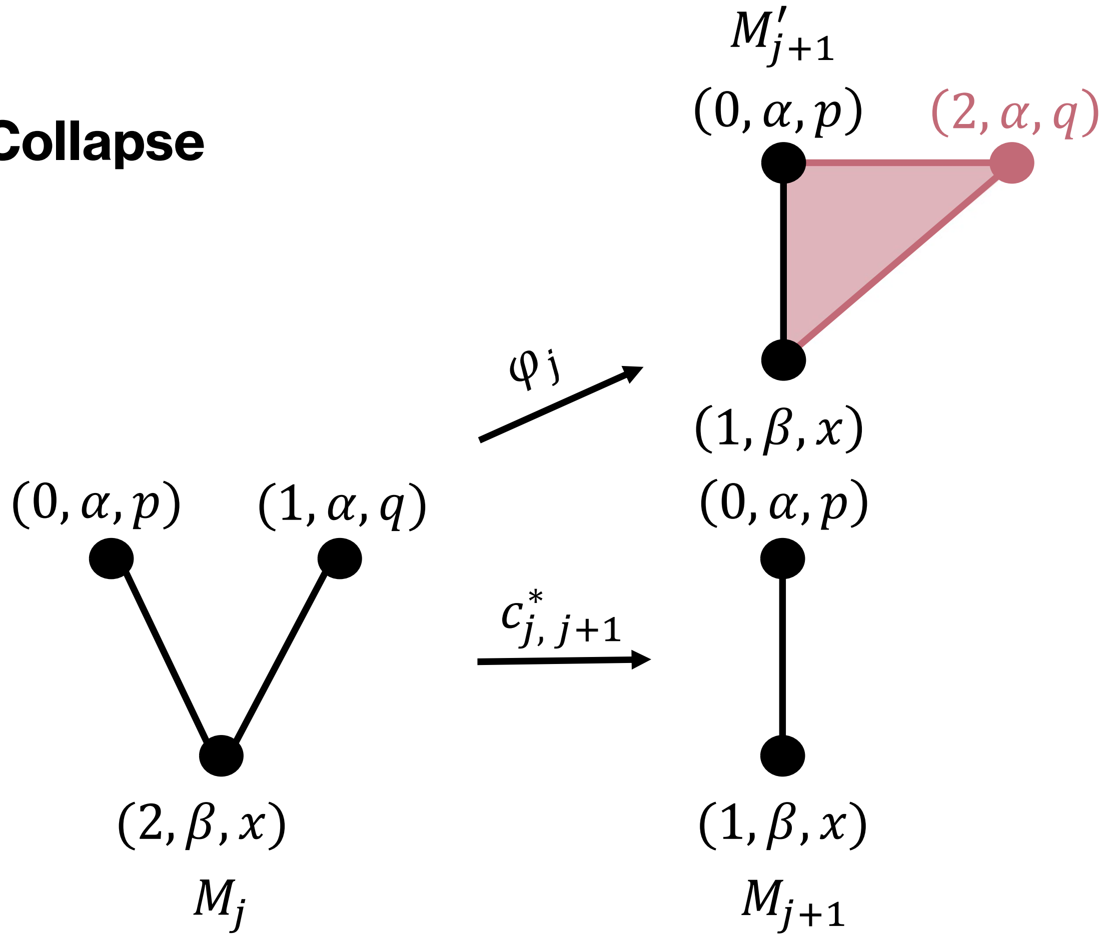
Building an Algorithm

Single Cluster Collapse



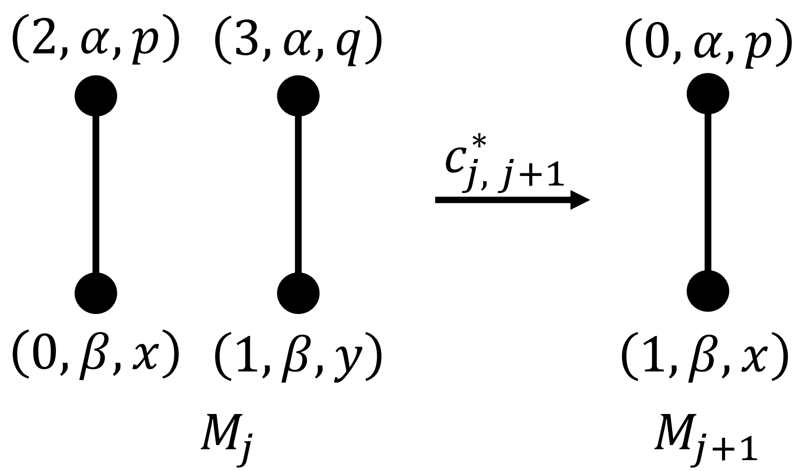
Building an Algorithm

Single Cluster Collapse



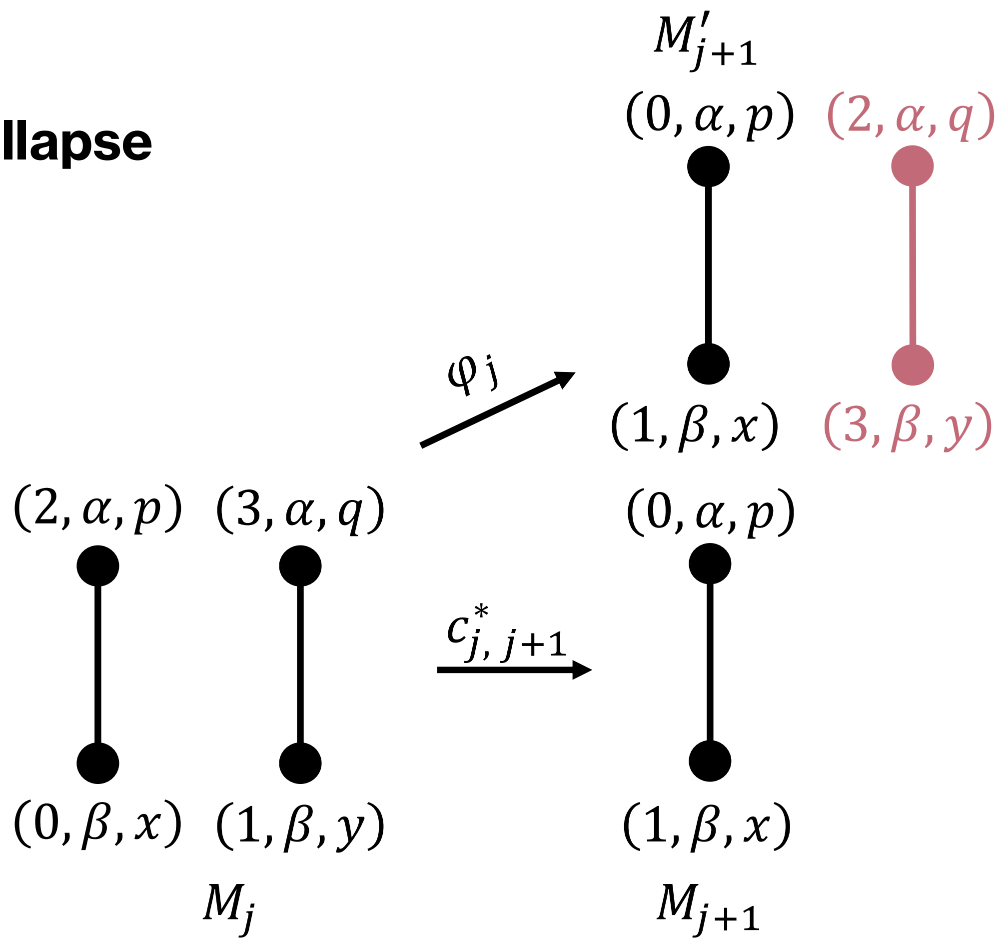
Building an Algorithm

Double Cluster Collapse



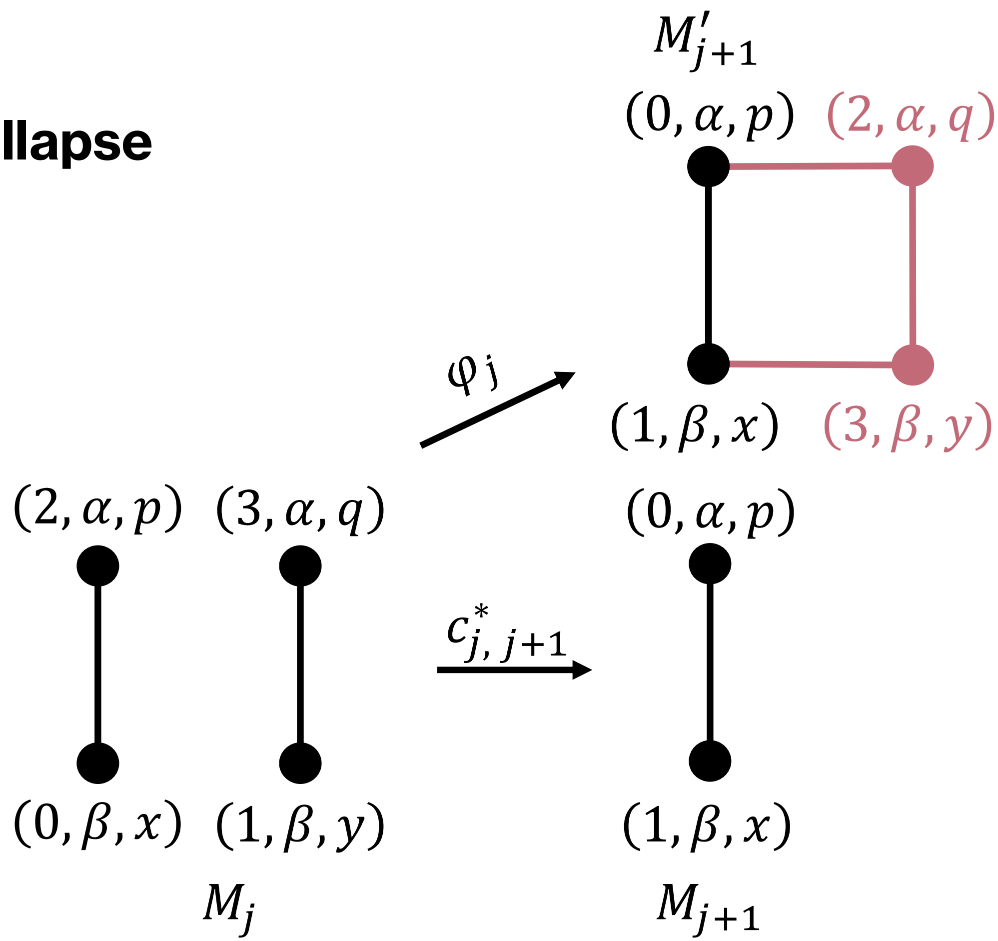
Building an Algorithm

Double Cluster Collapse



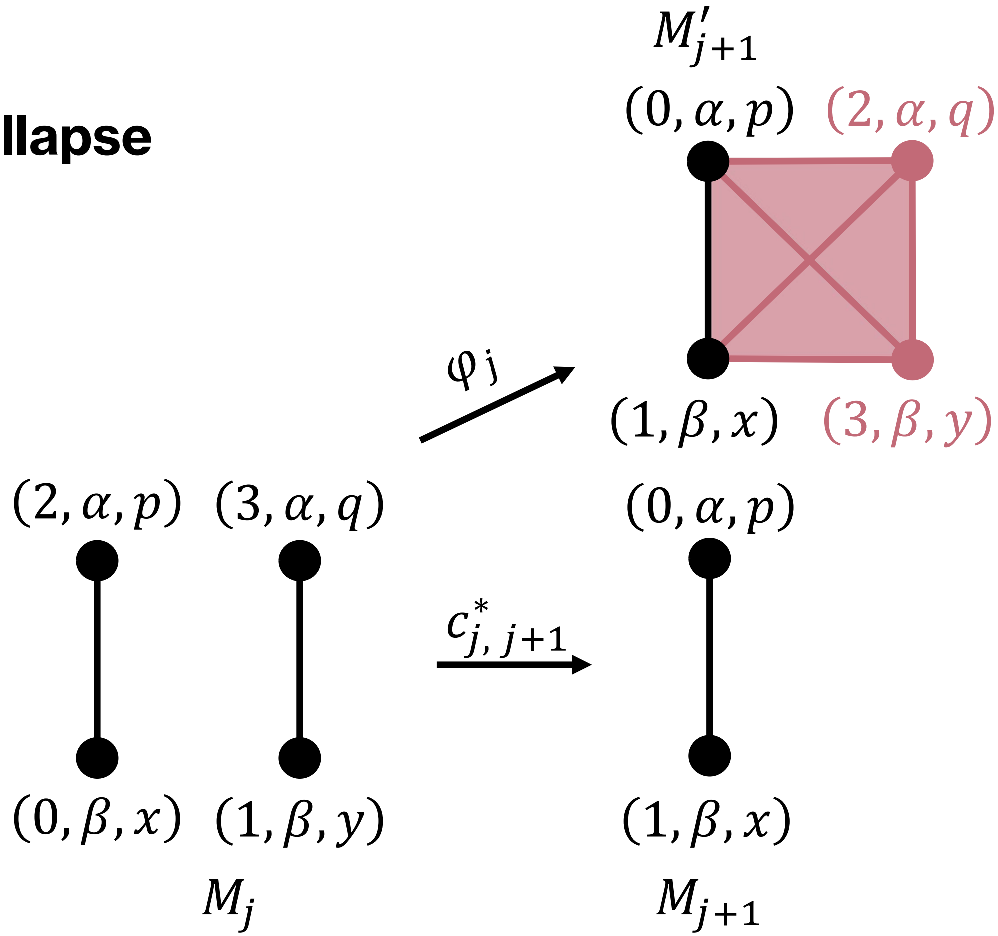
Building an Algorithm

Double Cluster Collapse



Building an Algorithm

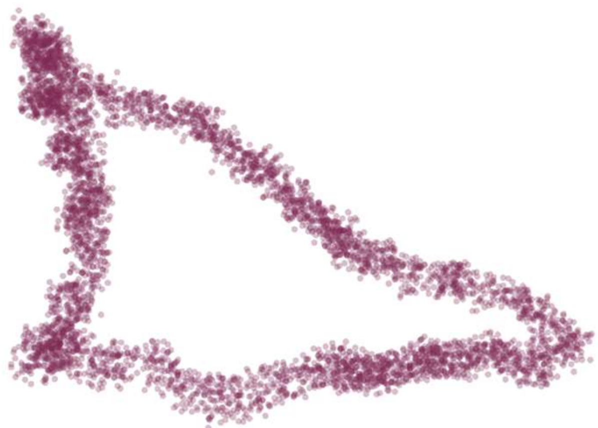
Double Cluster Collapse

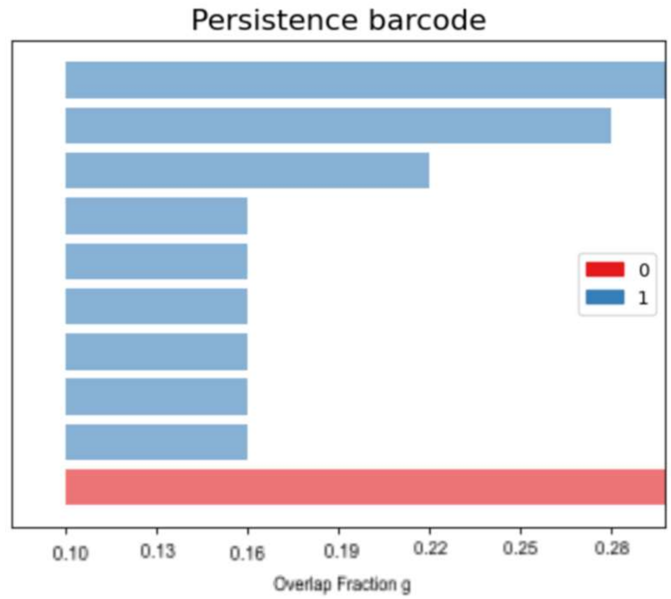


Building an Algorithm

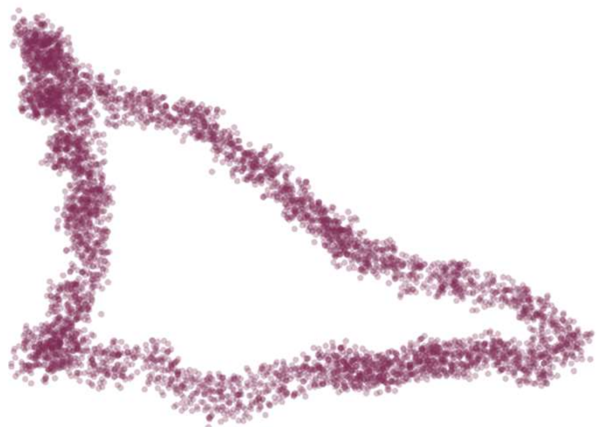
Sketch: For data set X, lens $f : X \to \mathbb{R}^n$, tower of covers $\mathfrak{U}$, and DBSCAN parameters ϵ and Minpts.

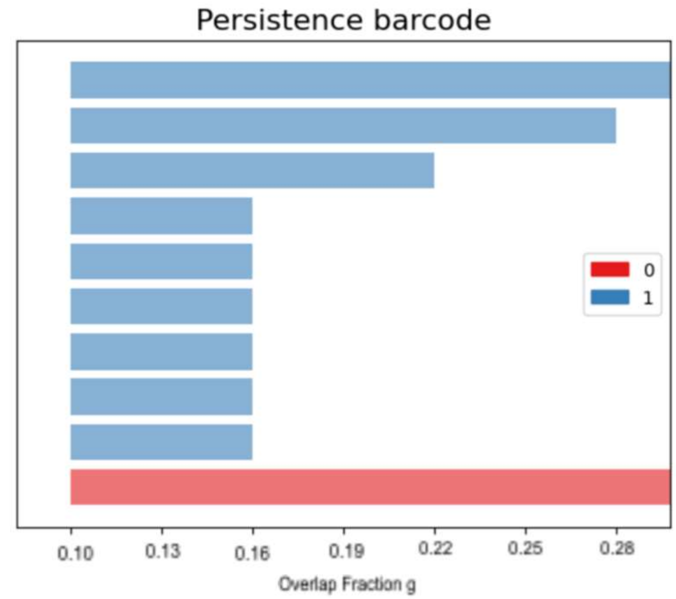
- Create 2-Mapper complexes M_i for each cover in cluster cover $\mathfrak{C}$.
- For pairs (M_i, M_{i+1}) create maps $\varphi_i : M_i \to M'_{i+1}$ which deal with single and double collapses of clusters.
- Create Simplex Tree with 2-Mapper complexes $M_0, M'_1, \dots, M'_t$

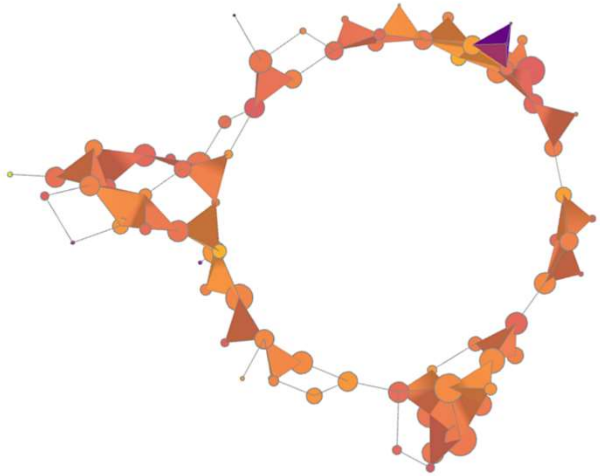




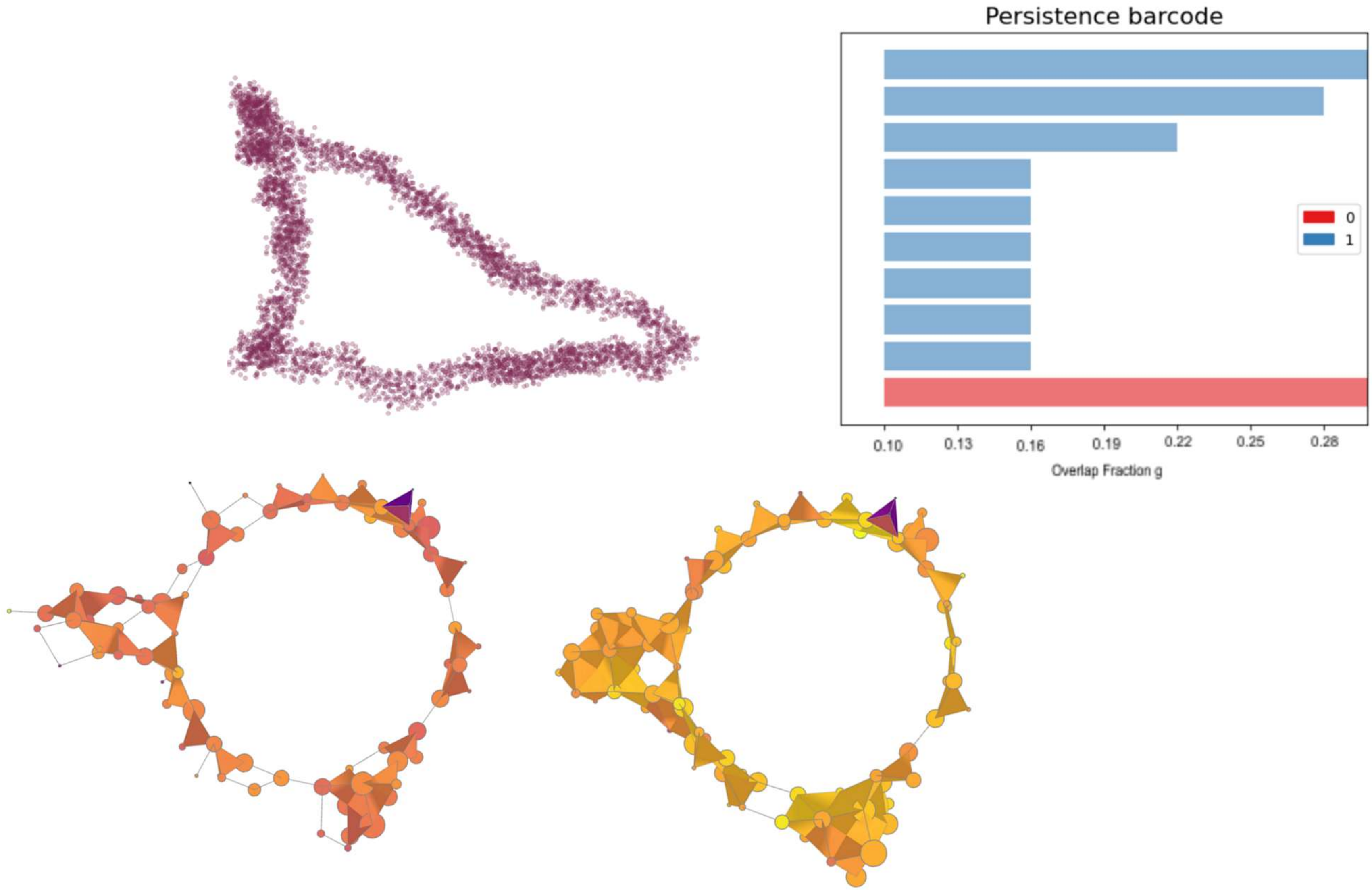
Persistence barcode
0
1
0.10
0.13
0.16
0.19
0.22
0.25
0.28
Overlap Fraction g



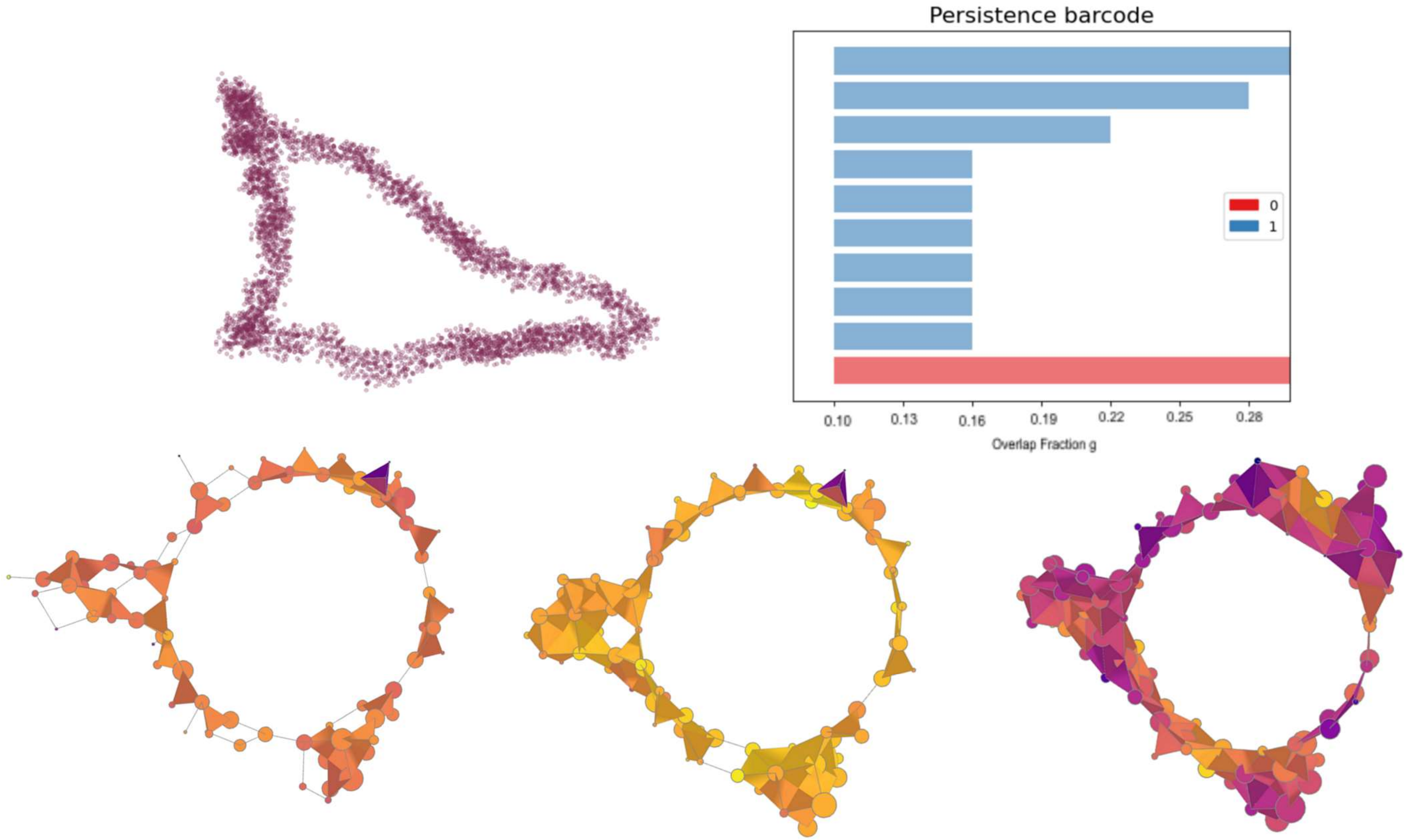

Persistence barcode
0
1
0.10
0.13
0.16
0.19
0.22
0.25
0.28
Overlap Fraction g



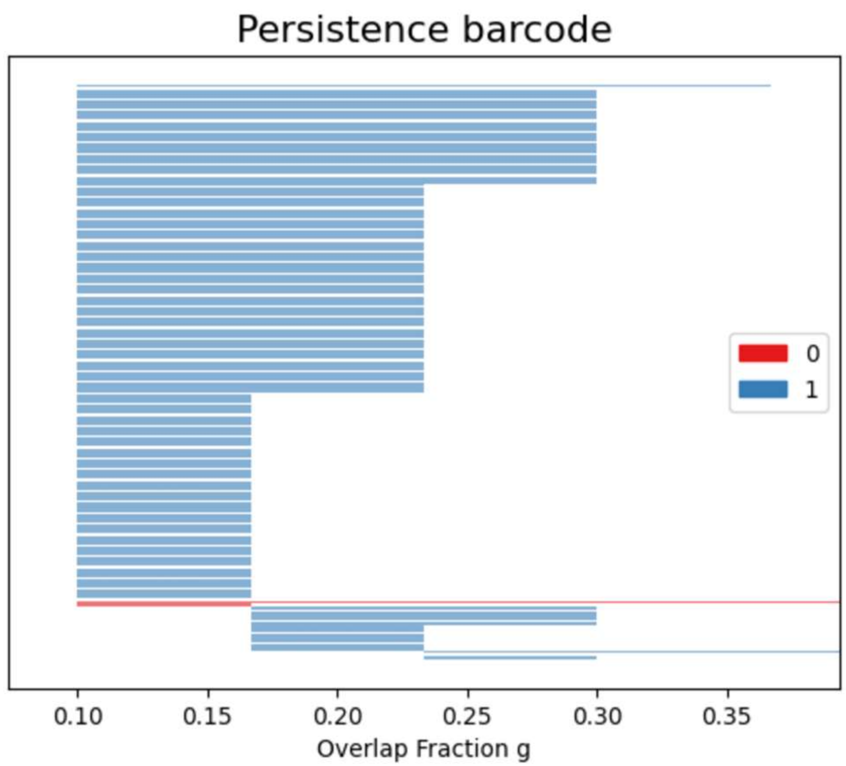
$g = 0.1$

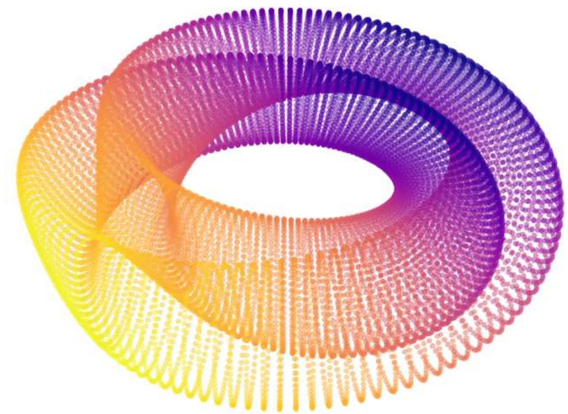


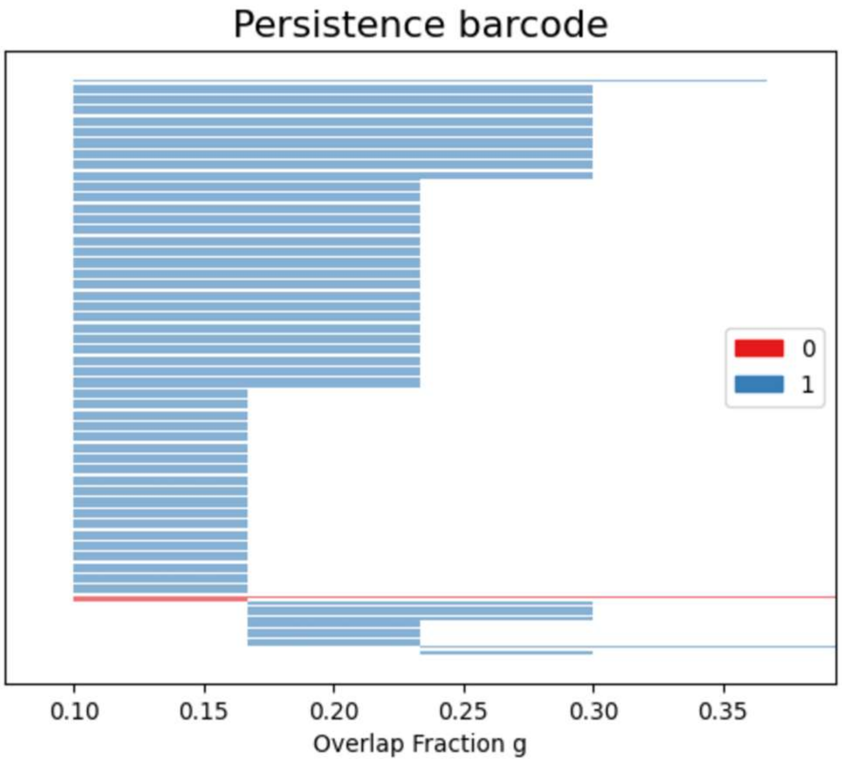
$g = 0.1$ $\qquad$ $g = 0.2$


Persistence barcode
0
1
0.10
0.13
0.16
0.19
0.22
0.25
0.28
Overlap Fraction g

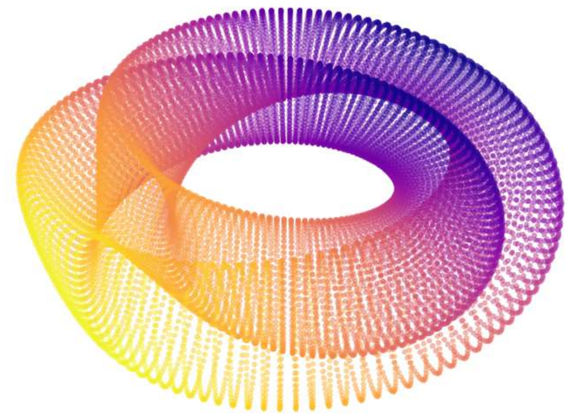
$g = 0.1$ $g = 0.2$ $g = 0.3$

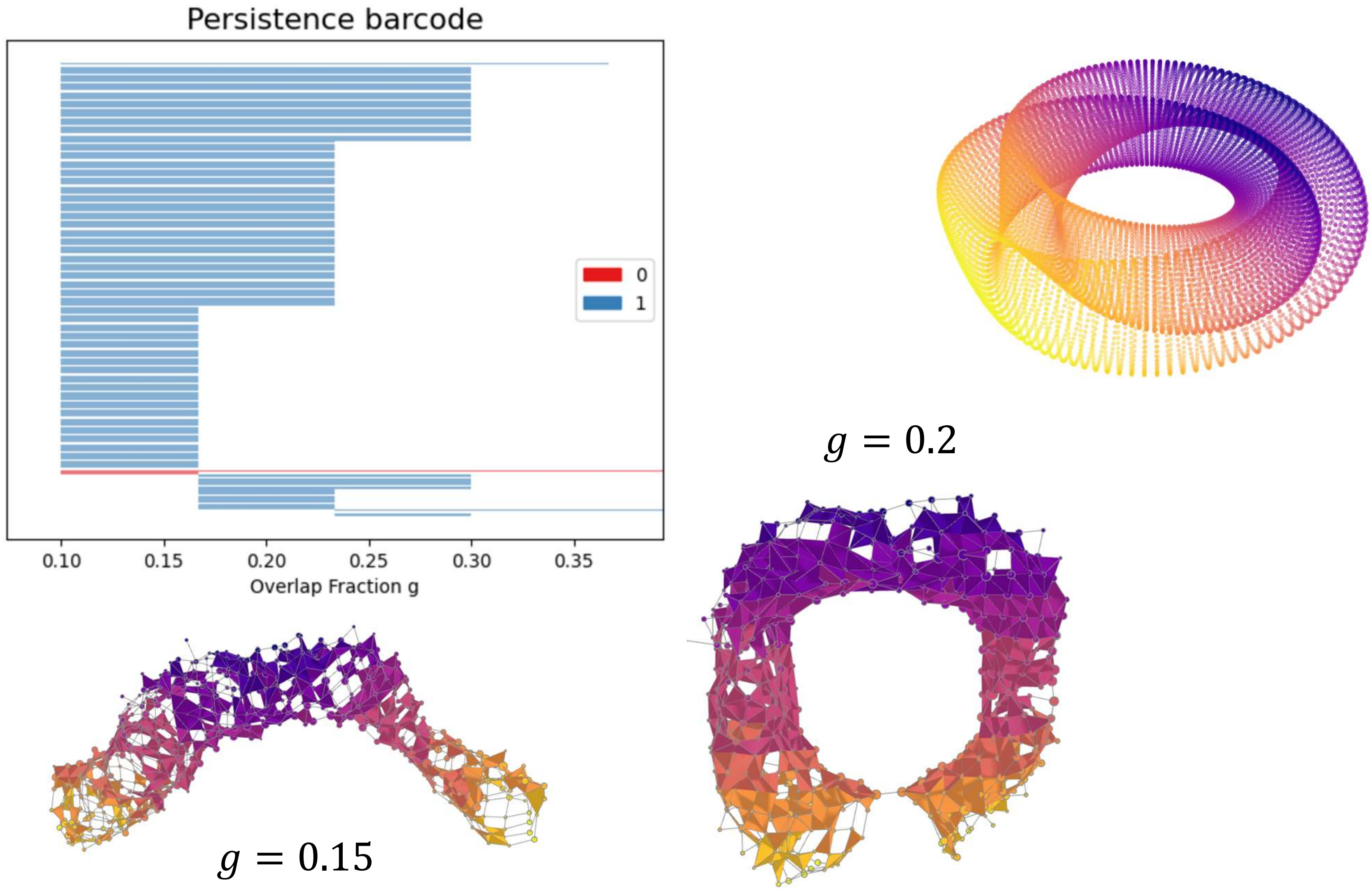

Persistence barcode
0
1
0.10
0.15
0.20
0.25
0.30
0.35
Overlap Fraction g

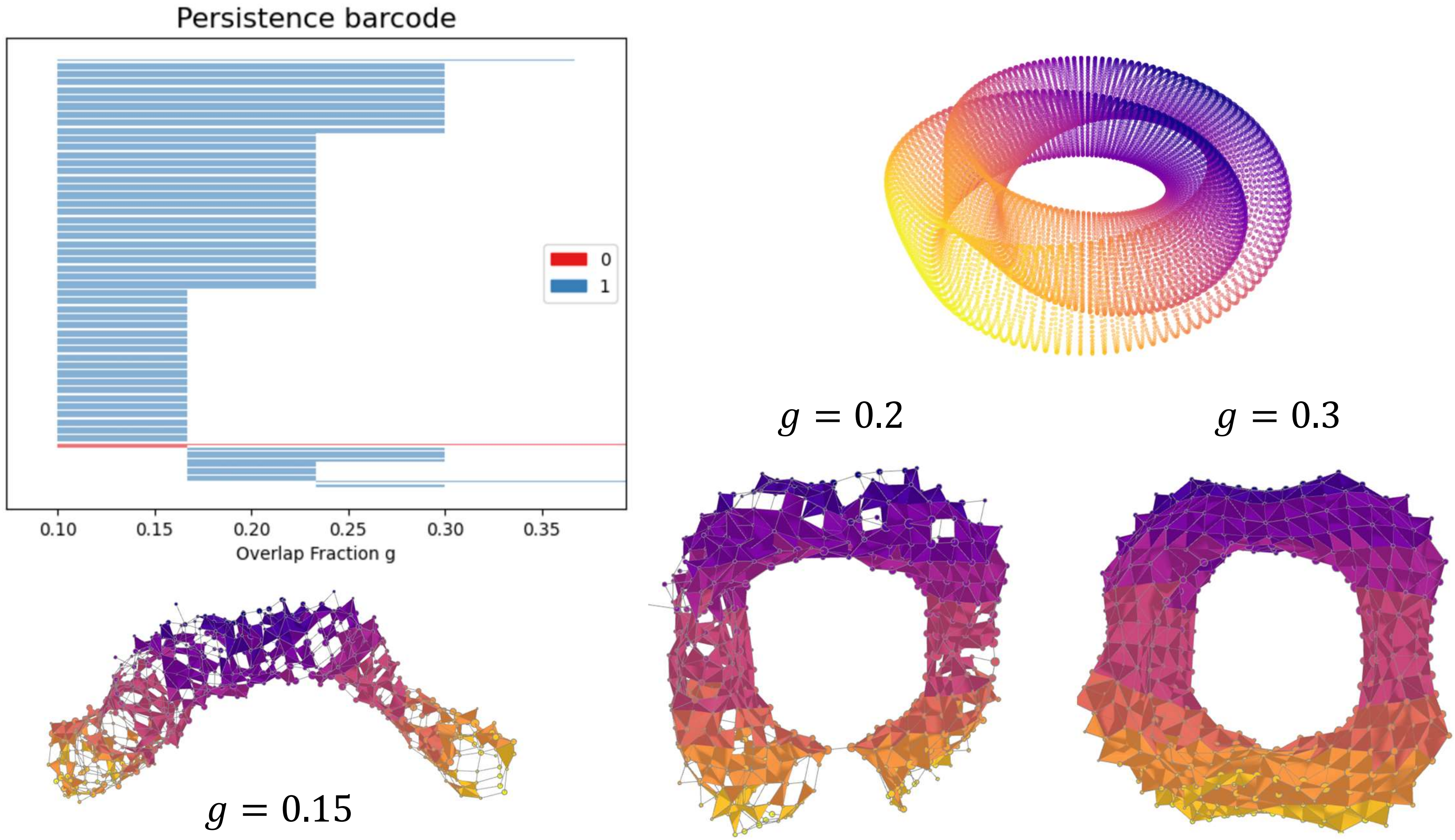



Persistence barcode
0
1
0.10
0.15
0.20
0.25
0.30
0.35
Overlap Fraction g

$g = 0.15$




Persistence barcode
0
1
0.10
0.15
0.20
0.25
0.30
0.35
Overlap Fraction g
$g = 0.2$
$g = 0.15$



Persistence barcode
0
1
0.10
0.15
0.20
0.25
0.30
0.35
Overlap Fraction g
$g = 0.2$
$g = 0.3$
$g = 0.15$

Future Directions

- Estimation for likely trajectories in dynamical systems: *Trajectory Mapper.*
- *Engineering Mapper Complexes:* Computing probabilities for 2-Mapper given certain parameters.
- Visualization of clustering in genetic ancestry for human populations