



AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE



Topological data and how to analyze them

Session 2 — Topological and Geometric Deep Learning: Theory, Methods and Applications

Universidad Complutense de Madrid

Session 2 — Topological data and how to analyze them

Outline

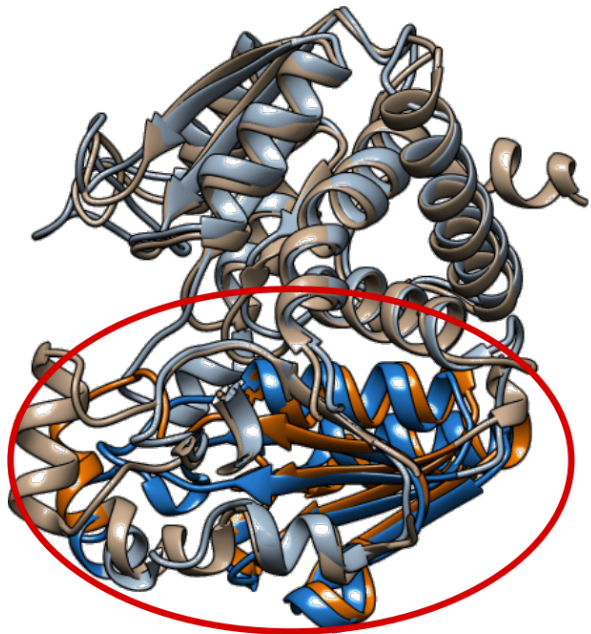
- Graphs
 - Matrix representations of graphs
- Simplicial complexes
 - Building simplicial complexes out of point clouds
- Simplicial Homology
- Persistent Homology
 - Filtrations
 - Barcodes
 - Stability
 - Other invariants
- Some success stories of PH
- (Discrete) Hodge theory

Motivation

Session 2 — Topological data and how to analyze them

Data has shape, shape has meaning, and meaning begets understanding

Gunnar Carlsson, paraphrased

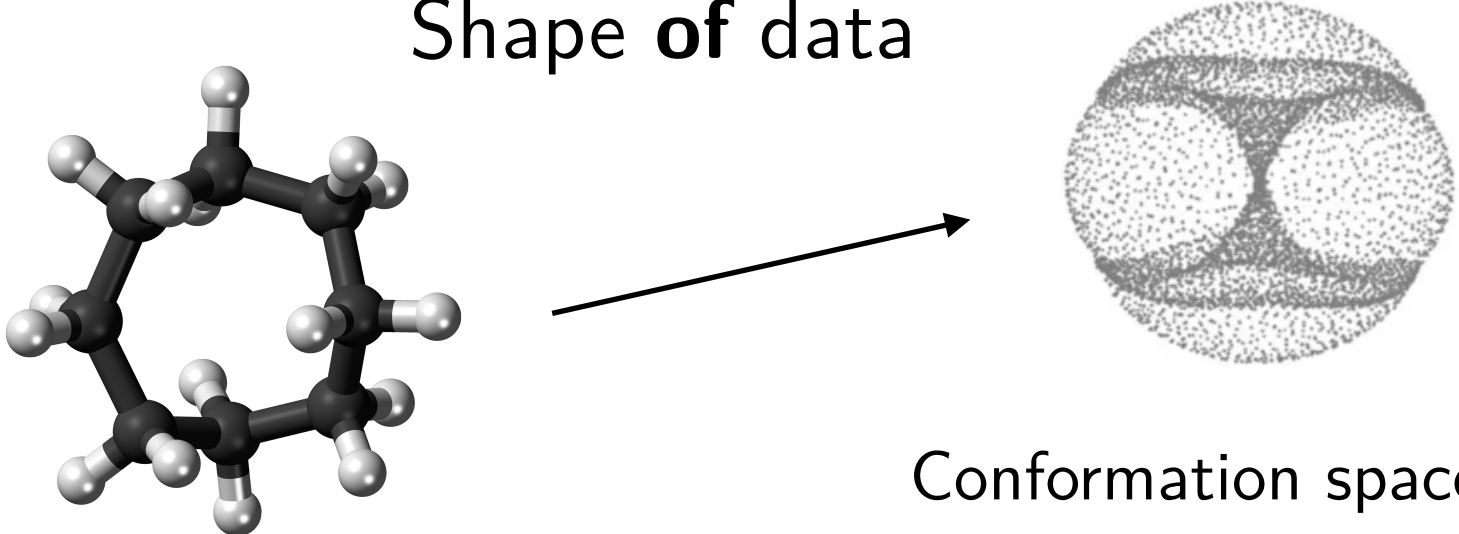


Shape **in data**

Distinguishing knotted/ unknotted proteins

Benjamin et al., (2023)

How do we *measure* shape?



Shape **of data**

1,031,644 coordinate vectors in \mathbb{R}^{72}

Conformation space of cyclo-octane

Martin and Watson, (2011)

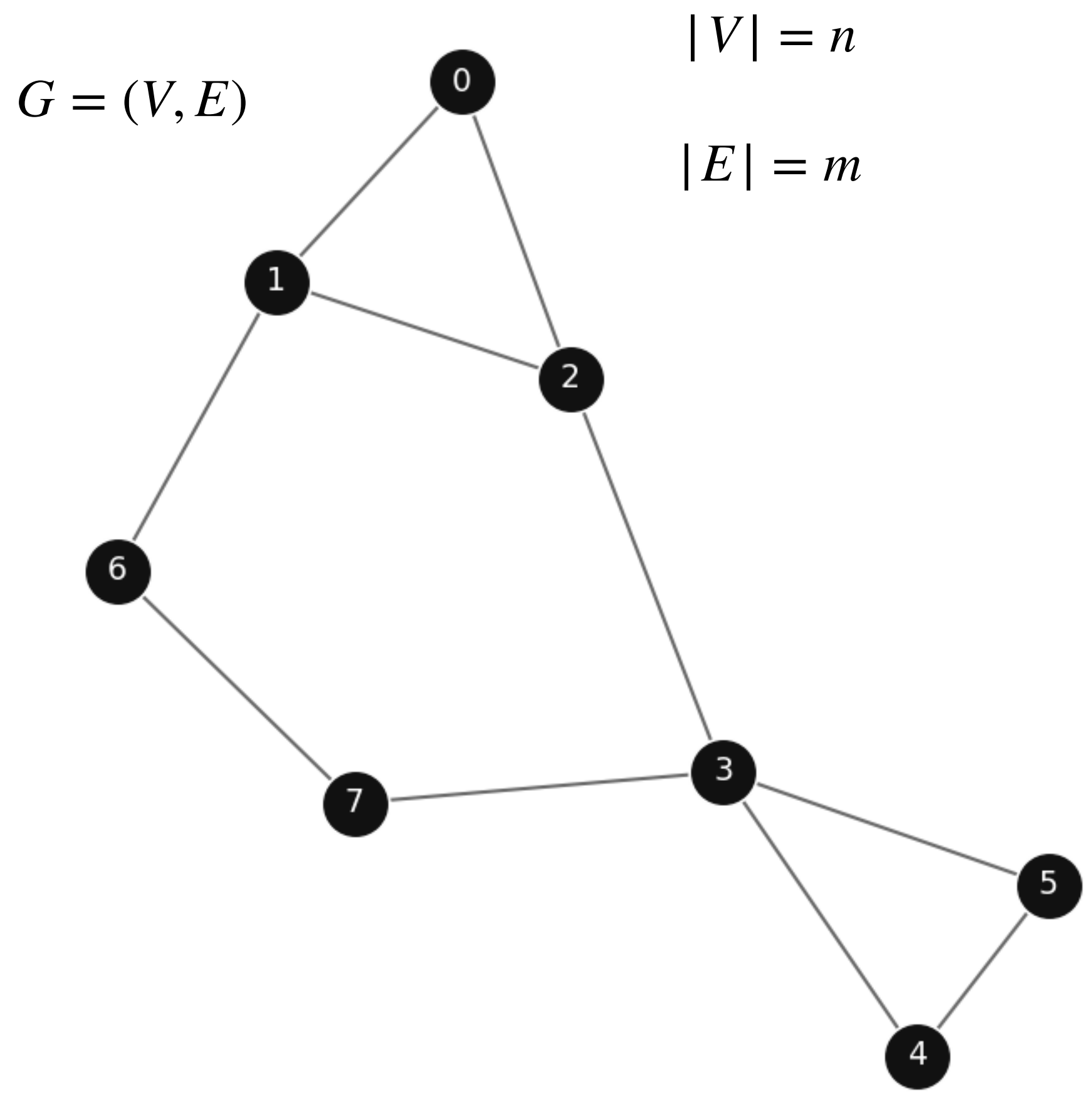
Geometry	=	fine detail (local properties)	+	quantitative answers
Topology	=	fundamental structure (global properties)	+	qualitative answers

Graphs

Session 2 — Topological data and how to analyze them

Graphs

Adjacency matrix



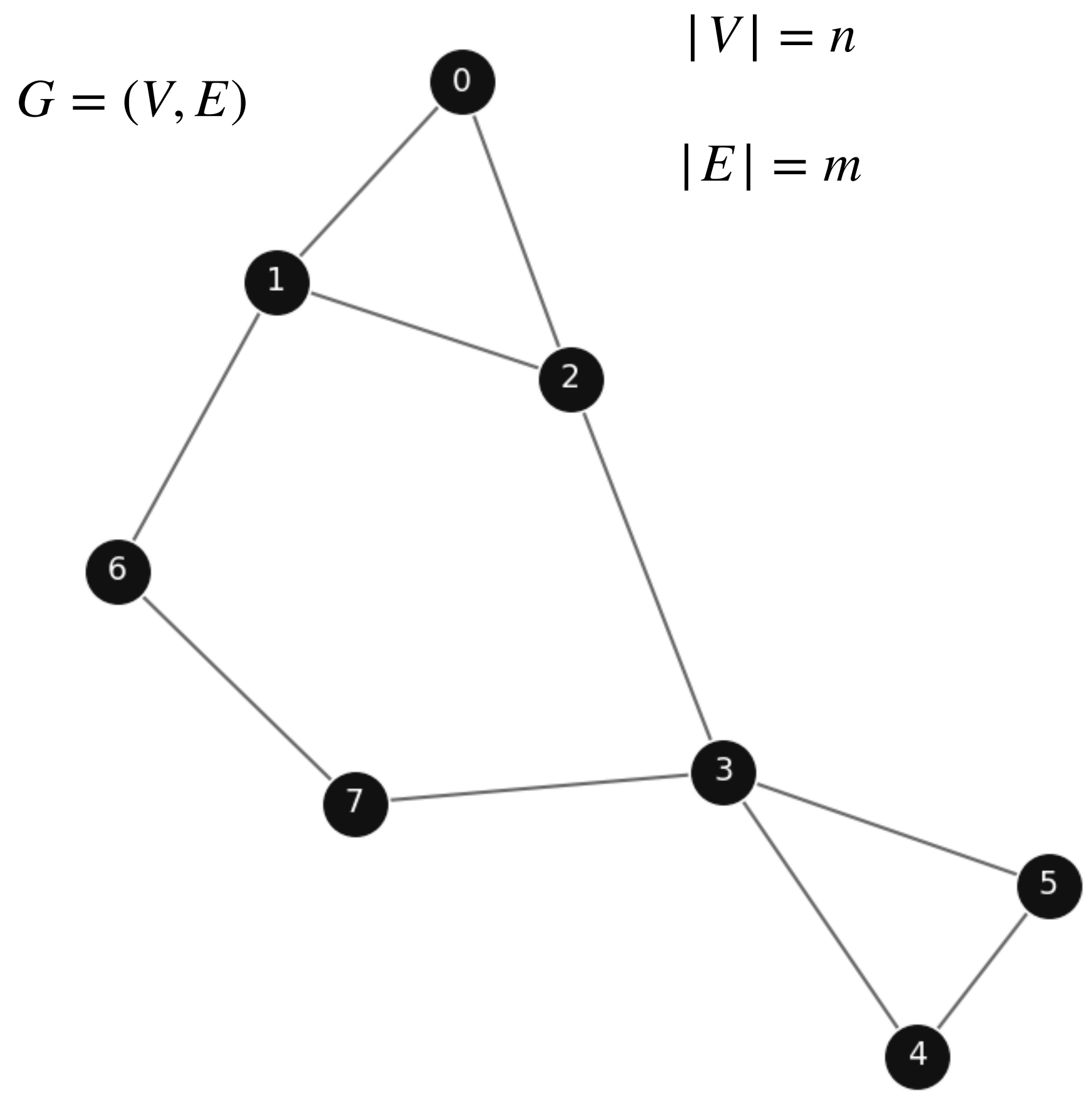
Adjacency matrix

$$A \in \mathbb{R}^{n \times n}, \quad A_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases}$$

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Graphs

Degree matrix



Neighbors

$$N(i) = \{j \in V \mid (i, j) \in E\}$$

$$= \{j \in V \mid A_{ij} = 1\}$$

Degree

$$d_i = |N(i)| = \sum_j A_{ij}$$

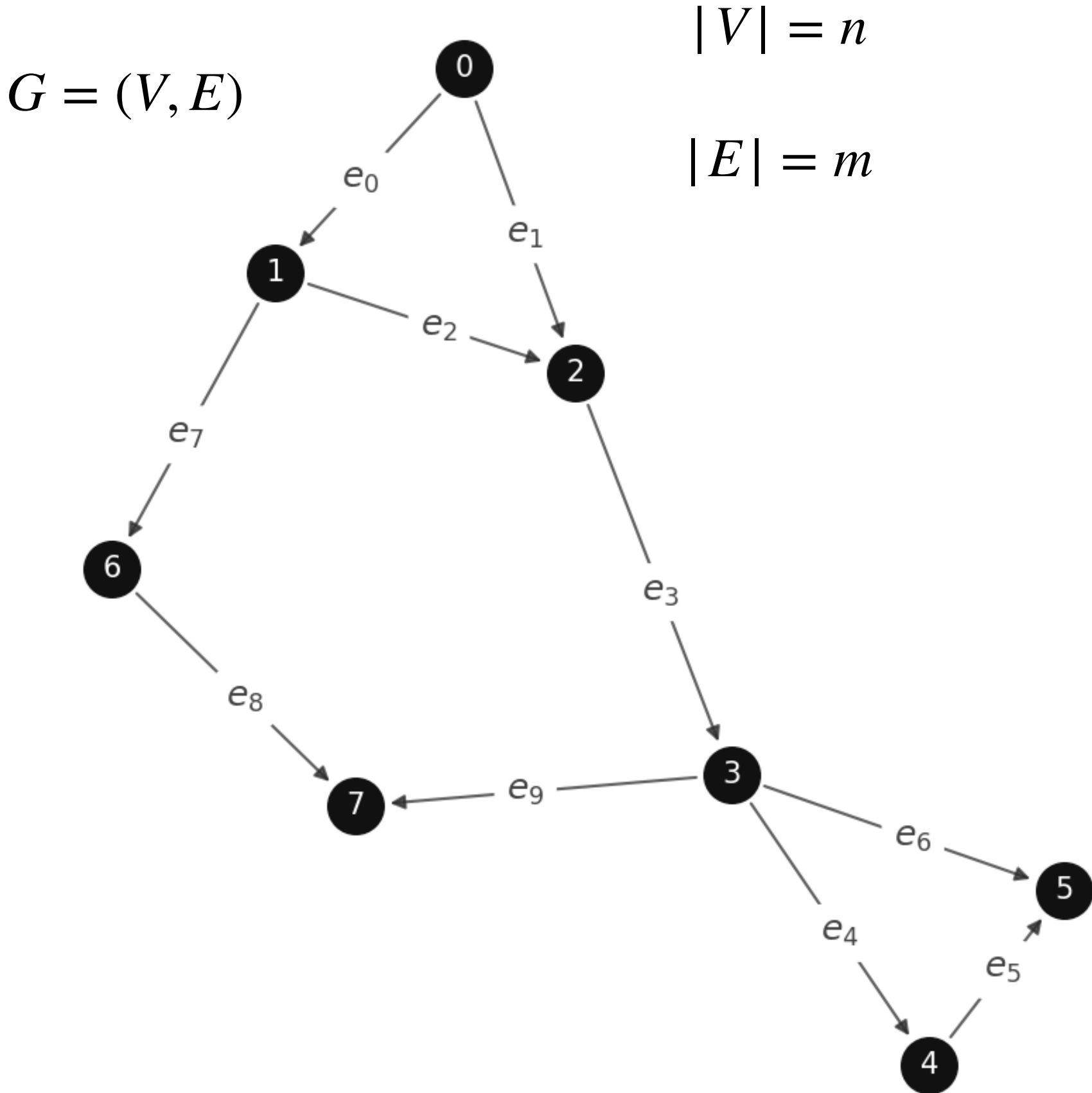
Degree matrix

Diagonal matrix $D \in \mathbb{R}^{n \times n}$ storing the degrees of the vertices

$$D = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Graphs

Incidence matrix



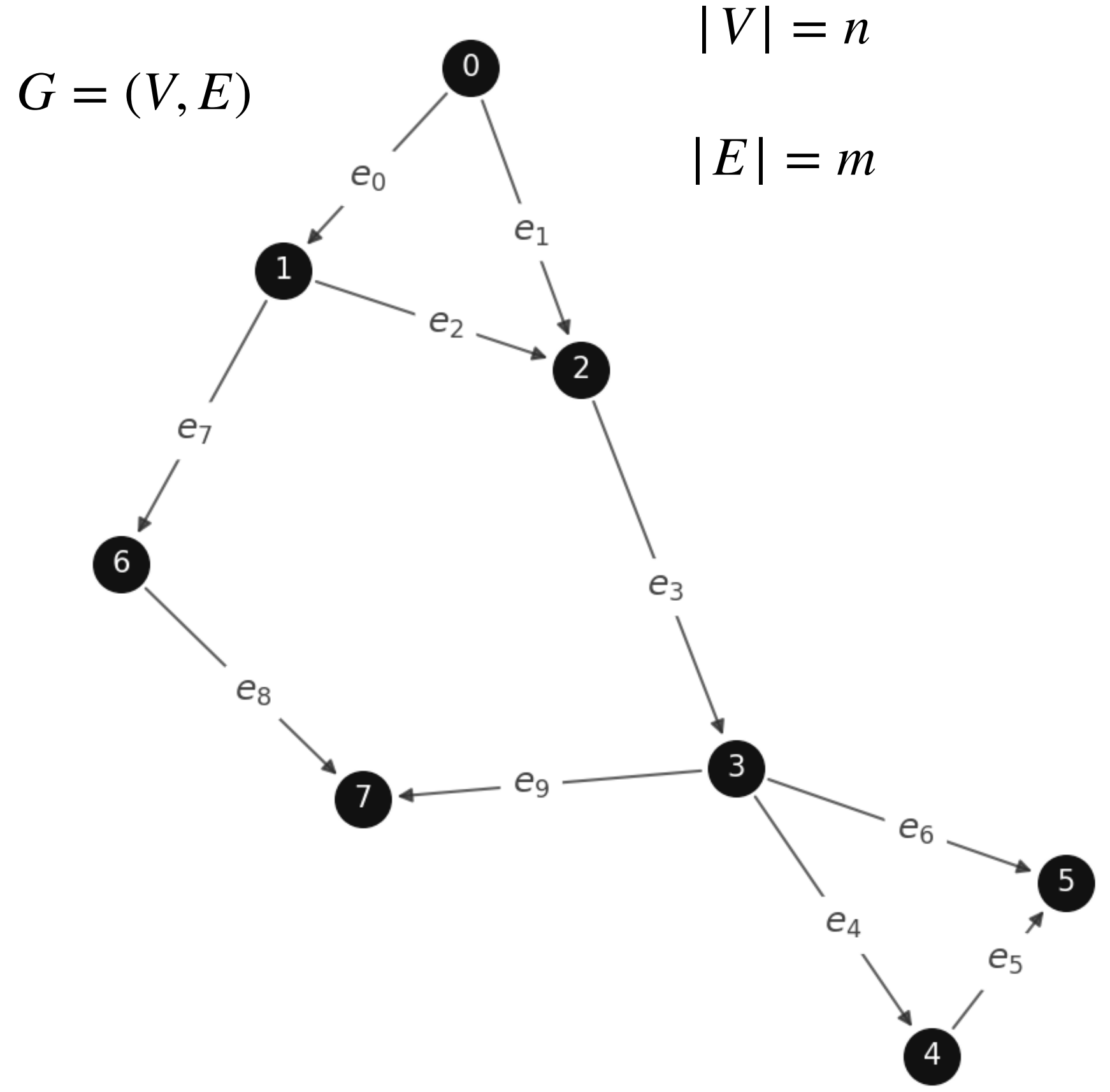
(Signed) Incidence matrix

$B \in \mathbb{R}^{n \times m}, \quad B_{ve} = \begin{cases} +1 & \text{if vertex } v \text{ is the tail of edge } e, \\ -1 & \text{if vertex } v \text{ is the head of edge } e, \\ 0 & \text{otherwise.} \end{cases}$

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}$$

Graphs

Graph Laplacian



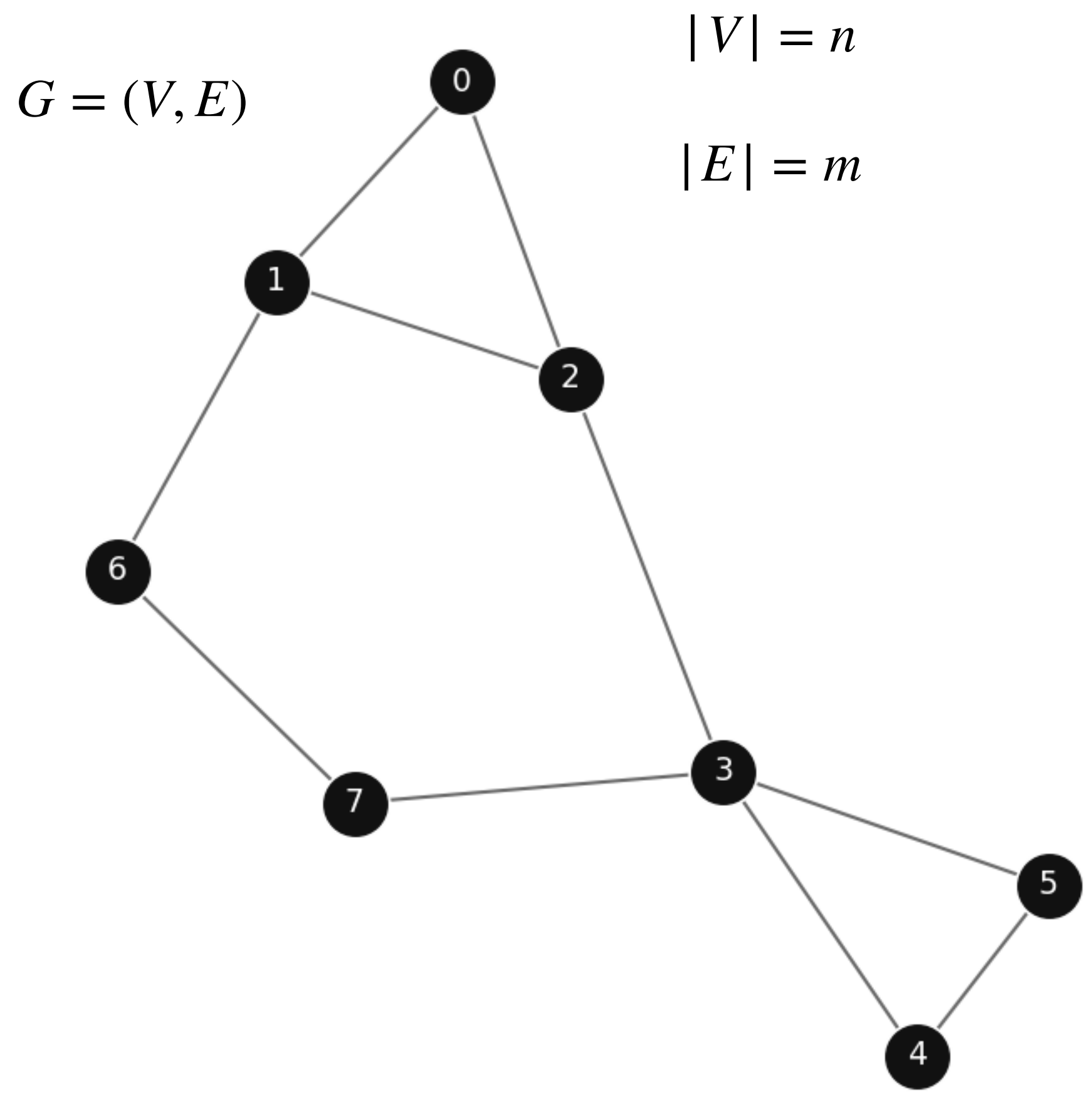
Graph Laplacian

$$L = BB^T$$

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}$$

Graphs

Graph Laplacian



Graph Laplacian

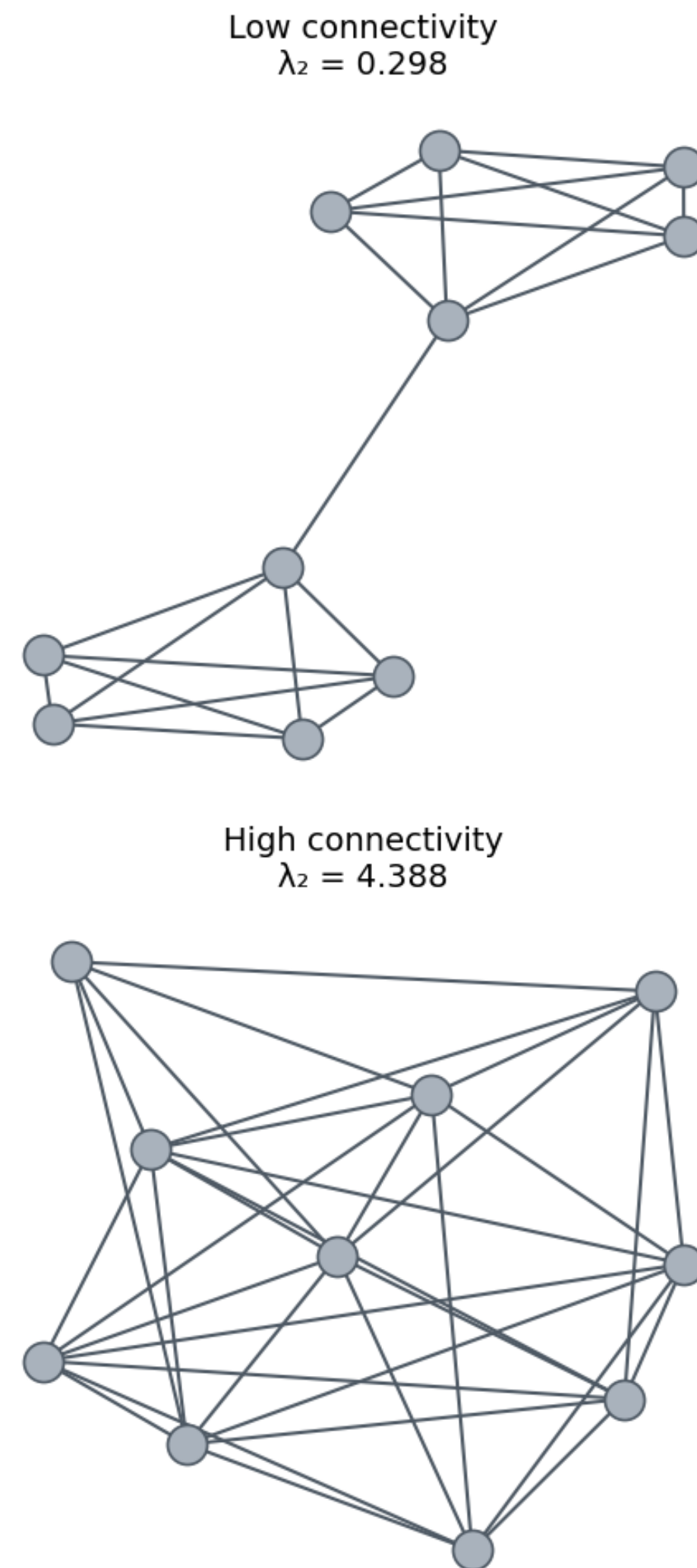
$$L = BB^T$$

$$L = D - A$$

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & 0 & -1 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 & -1 & -1 & 0 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 2 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & 2 \end{bmatrix}$$

Graphs

Graph Laplacian properties



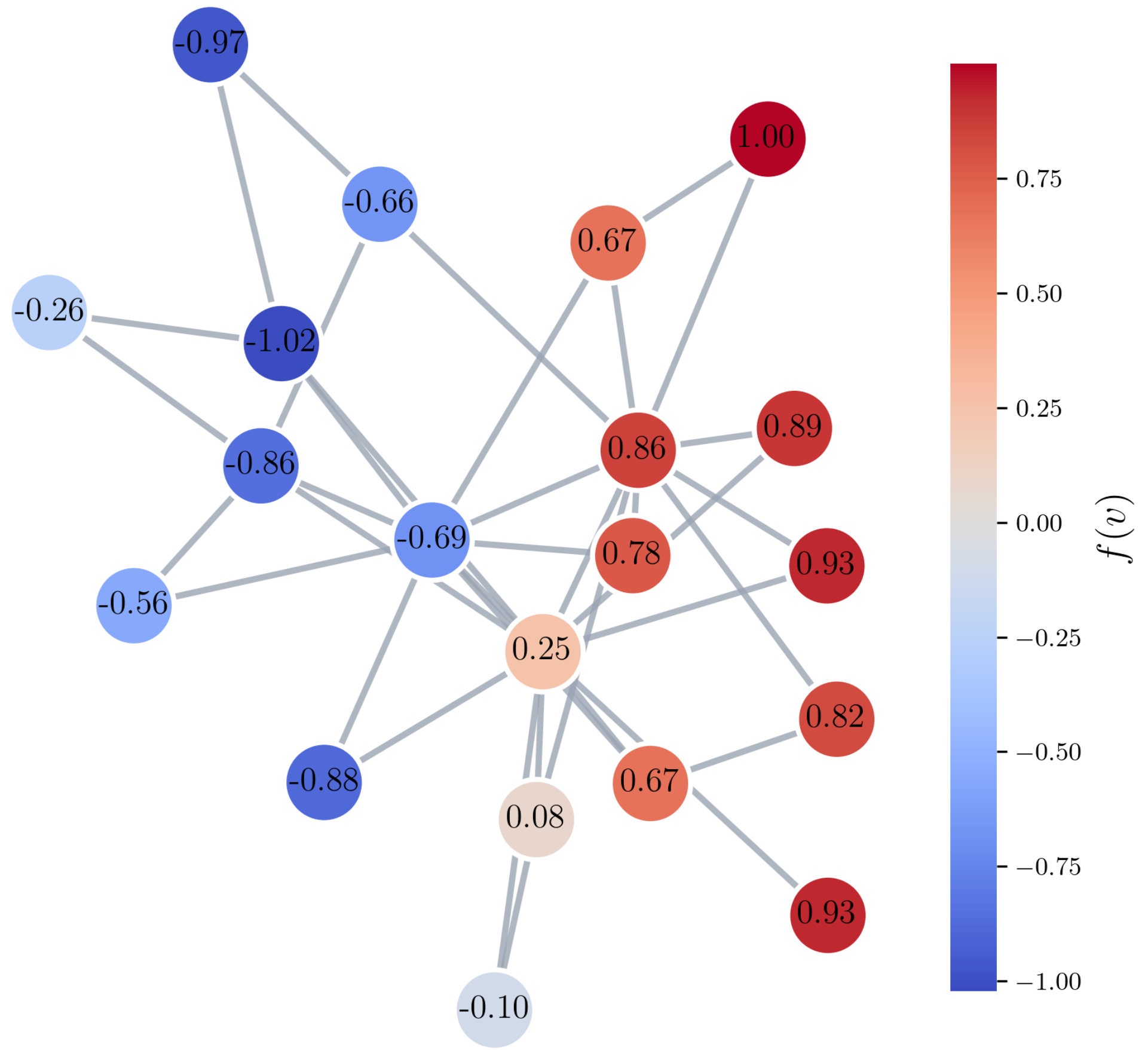
Graph Laplacian - Properties

1. L is **symmetric** and **positive semi-definite** $\implies L$ is diagonalizable and all its eigenvalues are non-negative
2. 0 is always an eigenvalue: $L\mathbf{1} = D\mathbf{1} - A\mathbf{1} = 0$
3. $\dim \ker L = \#\{\text{connected components } G\}$
4. **Fiedler value** (λ_2): smallest non-zero eigenvalue, tells us how “well connected” the graph is
 1. Small $\lambda_2 \implies$ graph almost disconnected
 2. Large $\lambda_2 \implies$ graph well knit together

Graphs

Graph Laplacian for signal processing

L is usually understood as acting over *signals* on the graph, i.e., functions $f: V \rightarrow \mathbb{R}$



$$\mathbf{f} = (f_1, \dots, f_n) \in \mathbb{R}^n$$

$$(L\mathbf{f})_i = d_i f_i - \sum_{j \sim i} f_j$$

How much the signal differs on average with respect to the neighbors

$$\mathbf{f}^\top L\mathbf{f} = \sum_{(i,j) \in E} (f_i - f_j)^2$$

How smooth the signal is

Graph Fourier Coefficients

$$\mathbf{f} = \sum_{i=1}^n a_i u_i \quad Lu_i = \lambda_i u_i$$

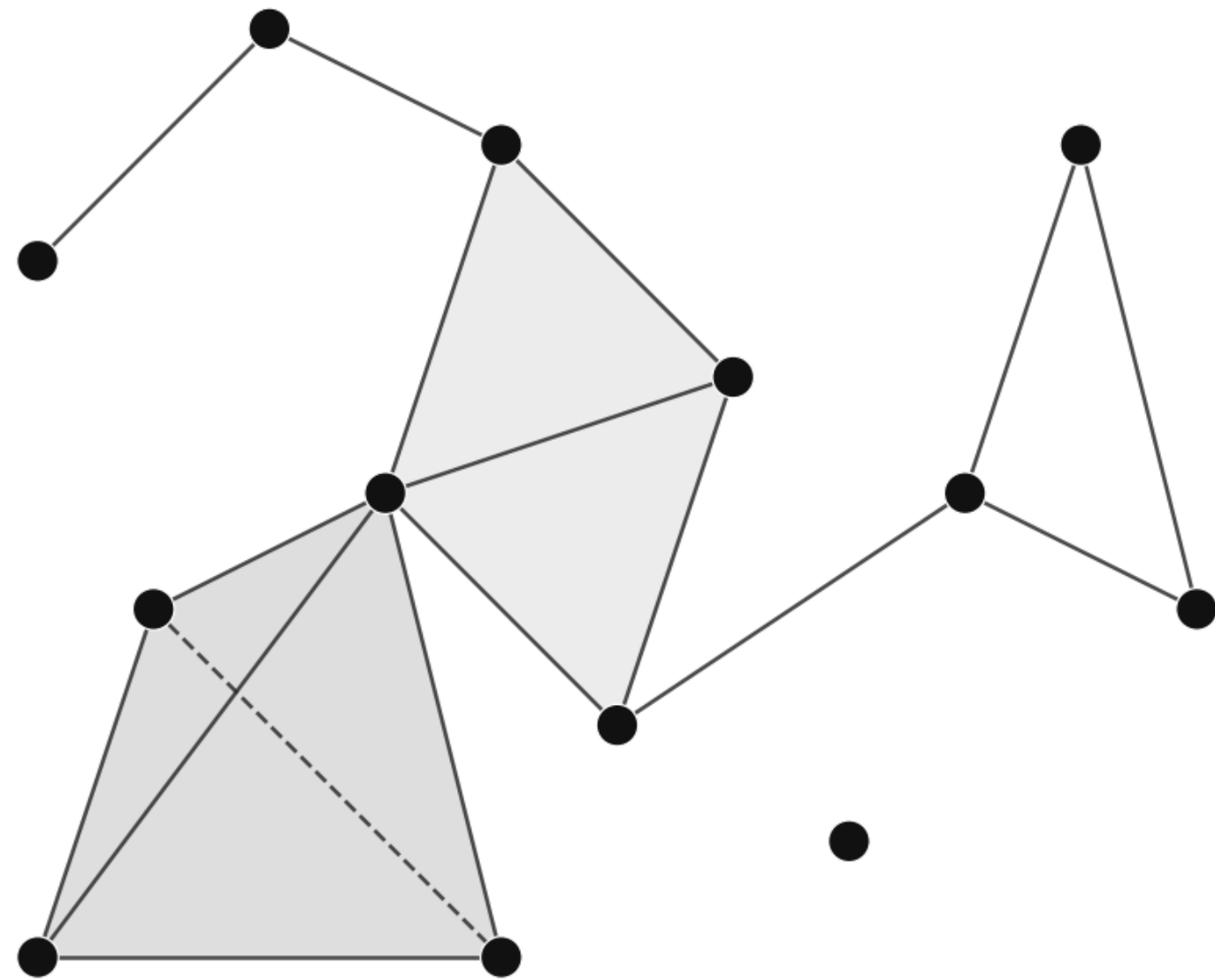
- Low frequency modes correspond to smoother signals
- High frequency modes correspond to signal that fluctuate a lot
- You can low-pass filter

Simplicial Complexes

Session 2 — Topological data and how to analyze them

Simplicial complexes

Basic definitions



Def. (Abstract) Simplicial Complex

Given a discrete set $S = \{s_1, \dots, s_n\}$, a simplicial complex X is a set of subsets $\sigma \subset S$ closed under inclusion ($\forall \tau \subseteq \sigma \implies \tau \in X$)

- **k -dimensional simplex:** $\sigma = \{s_{i_0}, \dots, s_{i_k}\}$
- X_k : set of k -dimensional simplices
- X has dimension $d =$ the dimension of its highest dimensional simplex
- **k -skeleton** ($X_{\leq k}$): all simplices with dimension $\leq k$

Example: a graph is a 1-dimensional simplicial complex

Simplicial complexes

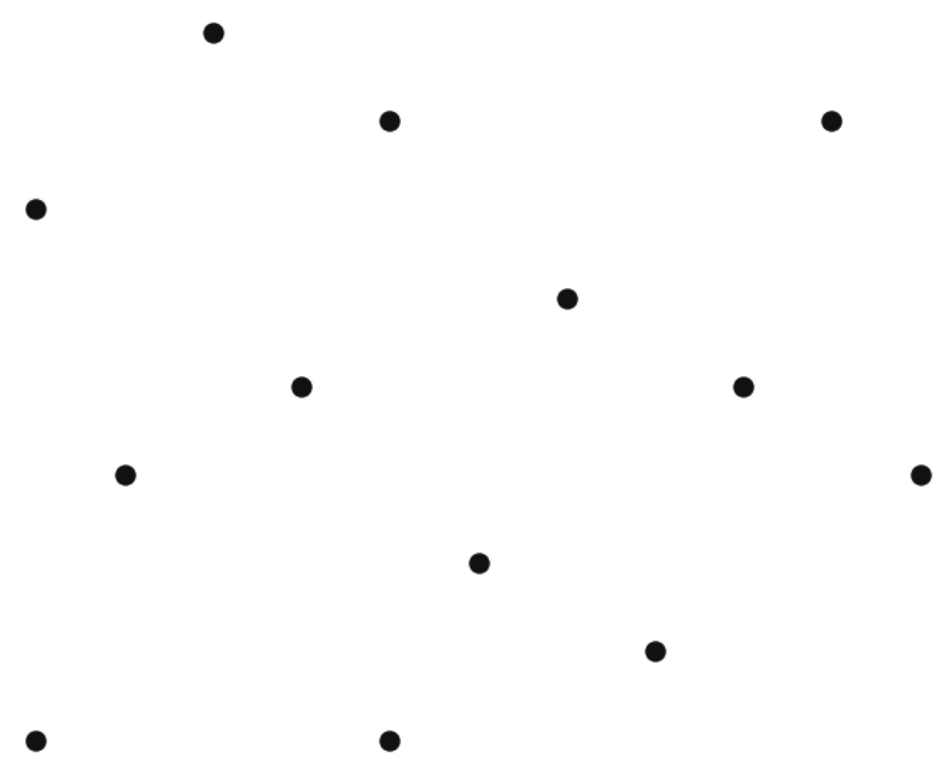
Building simplicial complexes from point clouds

Finite metric space

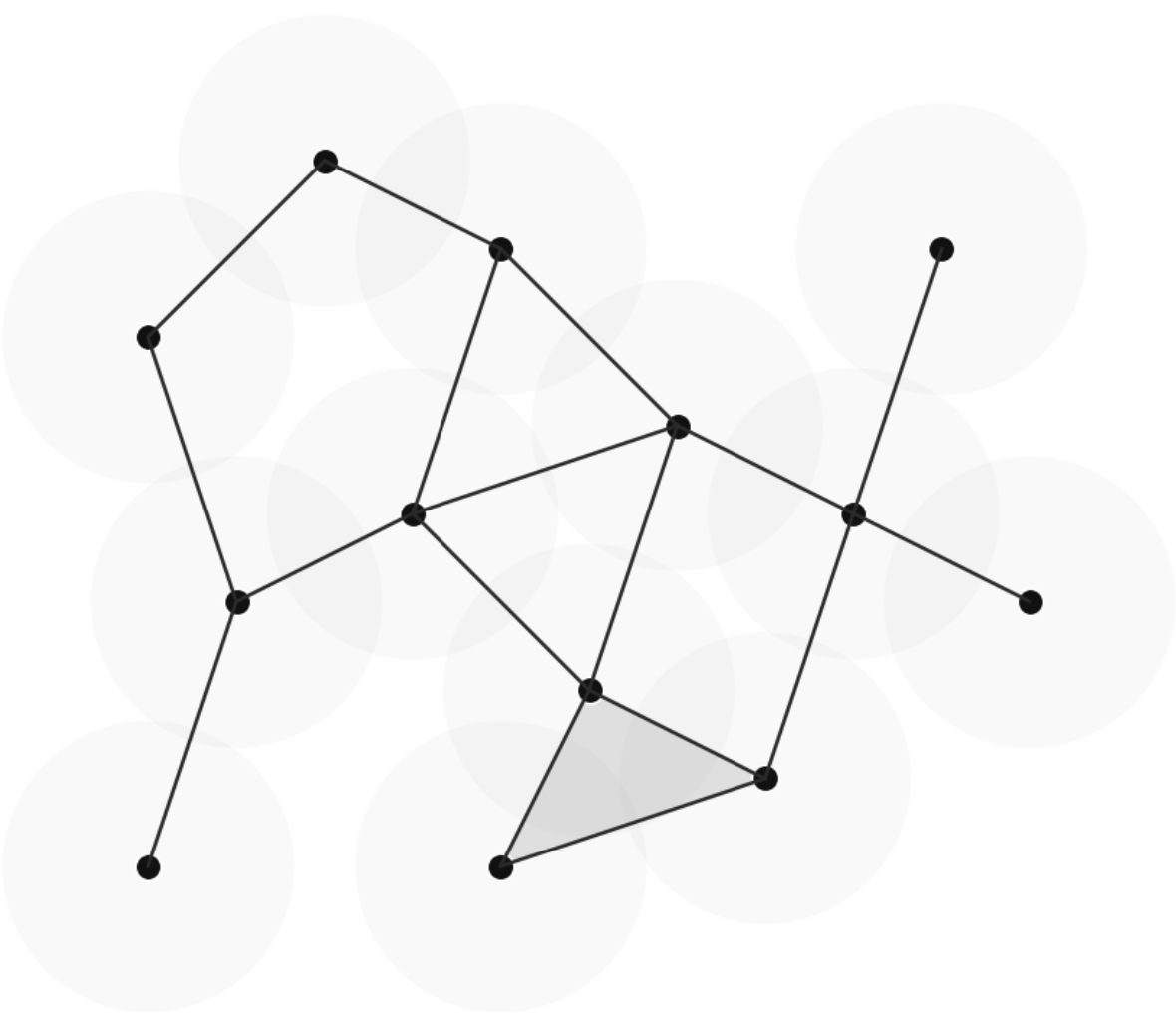
Given (S, d) and $r \geq 0$, the Čech complex (or Nerve complex) at scale r is:

$$C_r(S, d) = \left\{ \{s_{i_0}, \dots, s_{i_k}\} : \bigcap_{j=0}^k B(s_{i_j}, r) \neq \emptyset \right\}$$

Monotonicity
For $r \leq s$: $C_r(S) \subseteq C_s(S)$



Point cloud



Čech complex

Simplicial complexes

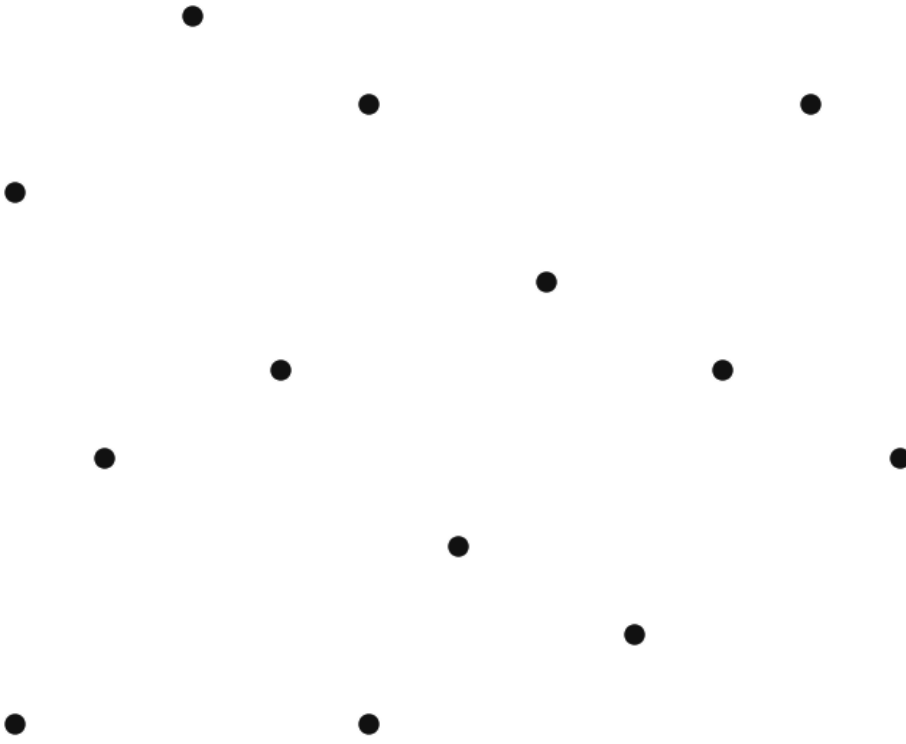
Building simplicial complexes from point clouds

Given (S, d) and $r \geq 0$, the **Vietoris-Rips complex** at scale r is:

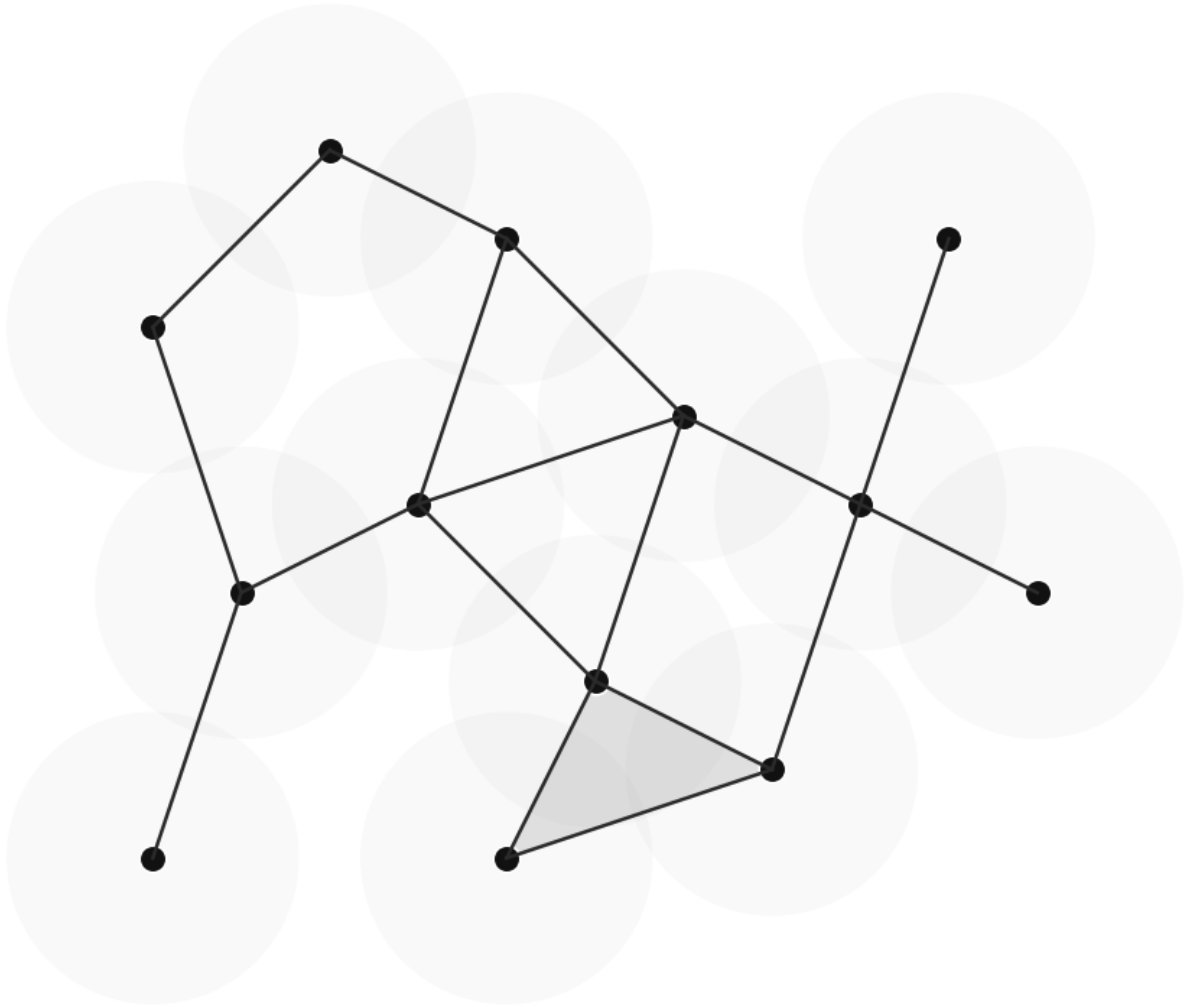
$$VR_r(S, d) = \left\{ \{s_{i_0}, \dots, s_{i_k}\} : d(s_i, s_j) \leq r, \forall i_0 \leq i < j \leq i_k \right\}$$

Monotonicity

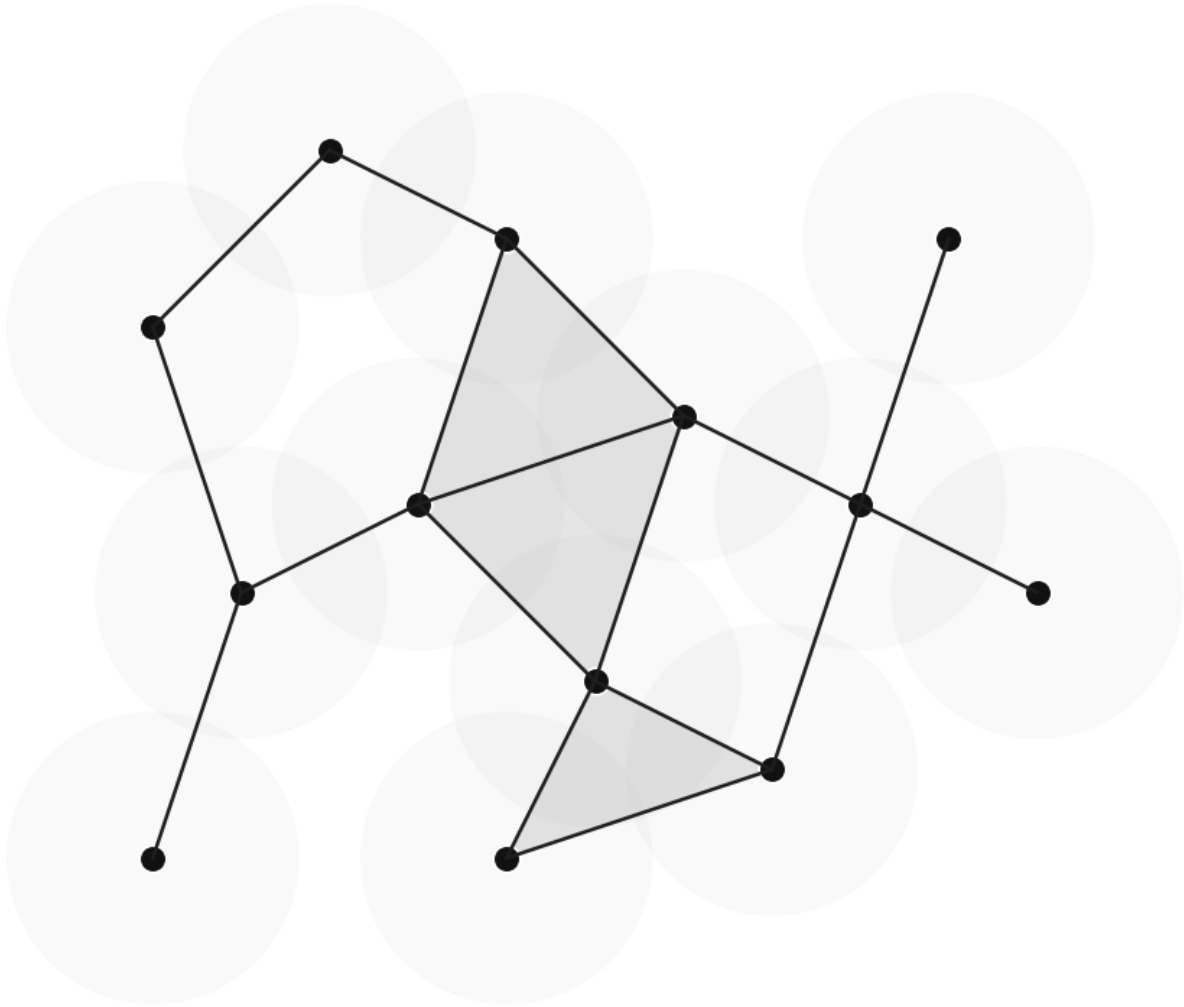
For $r \leq s$: $VR_r(S) \subseteq VR_s(S)$



Point cloud



Čech complex

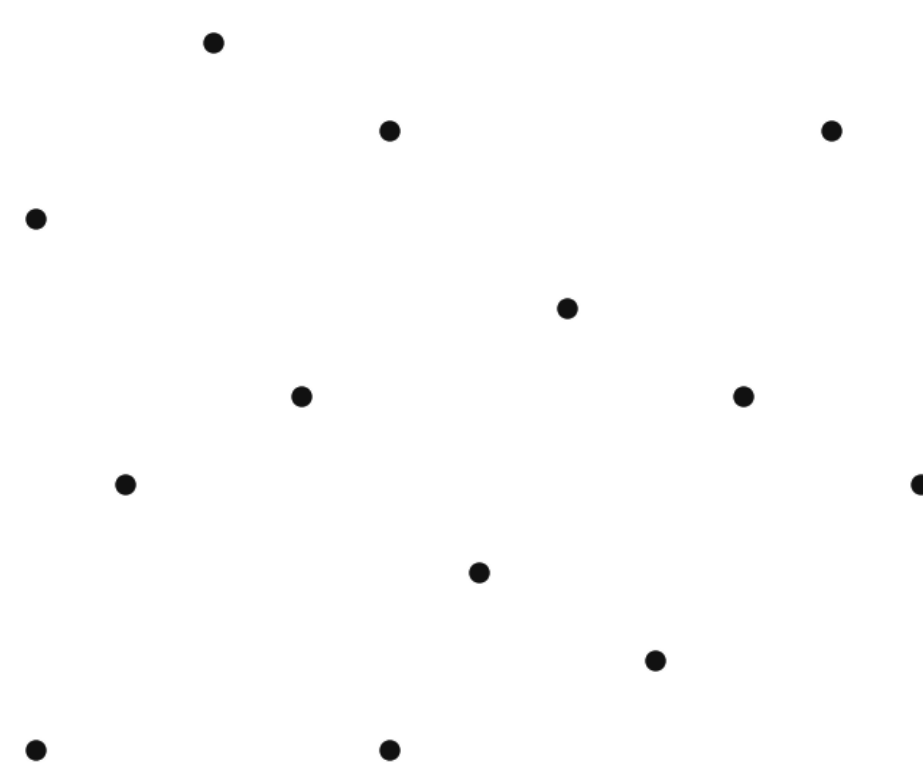


VR complex

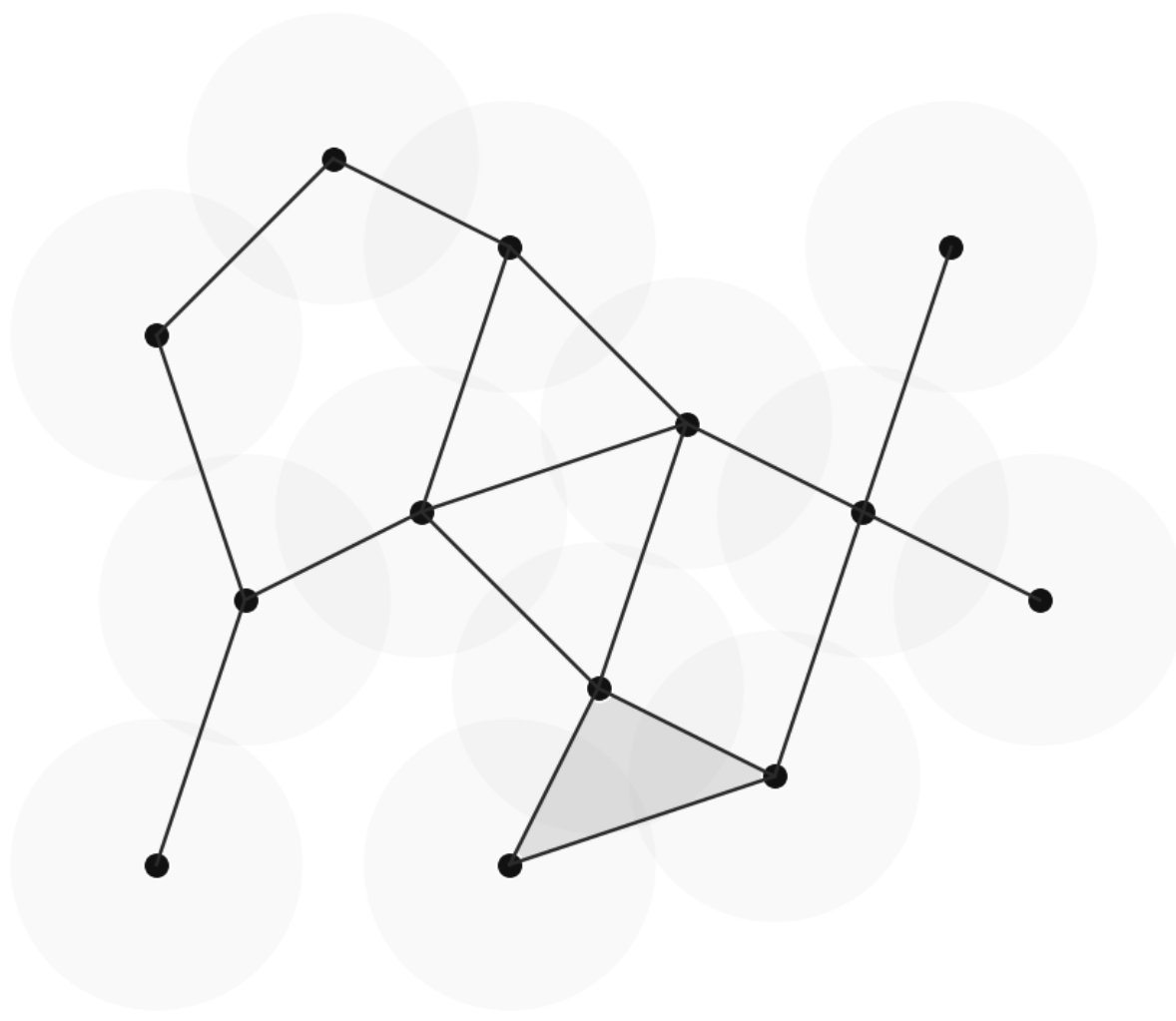
Simplicial complexes

Building simplicial complexes from point clouds

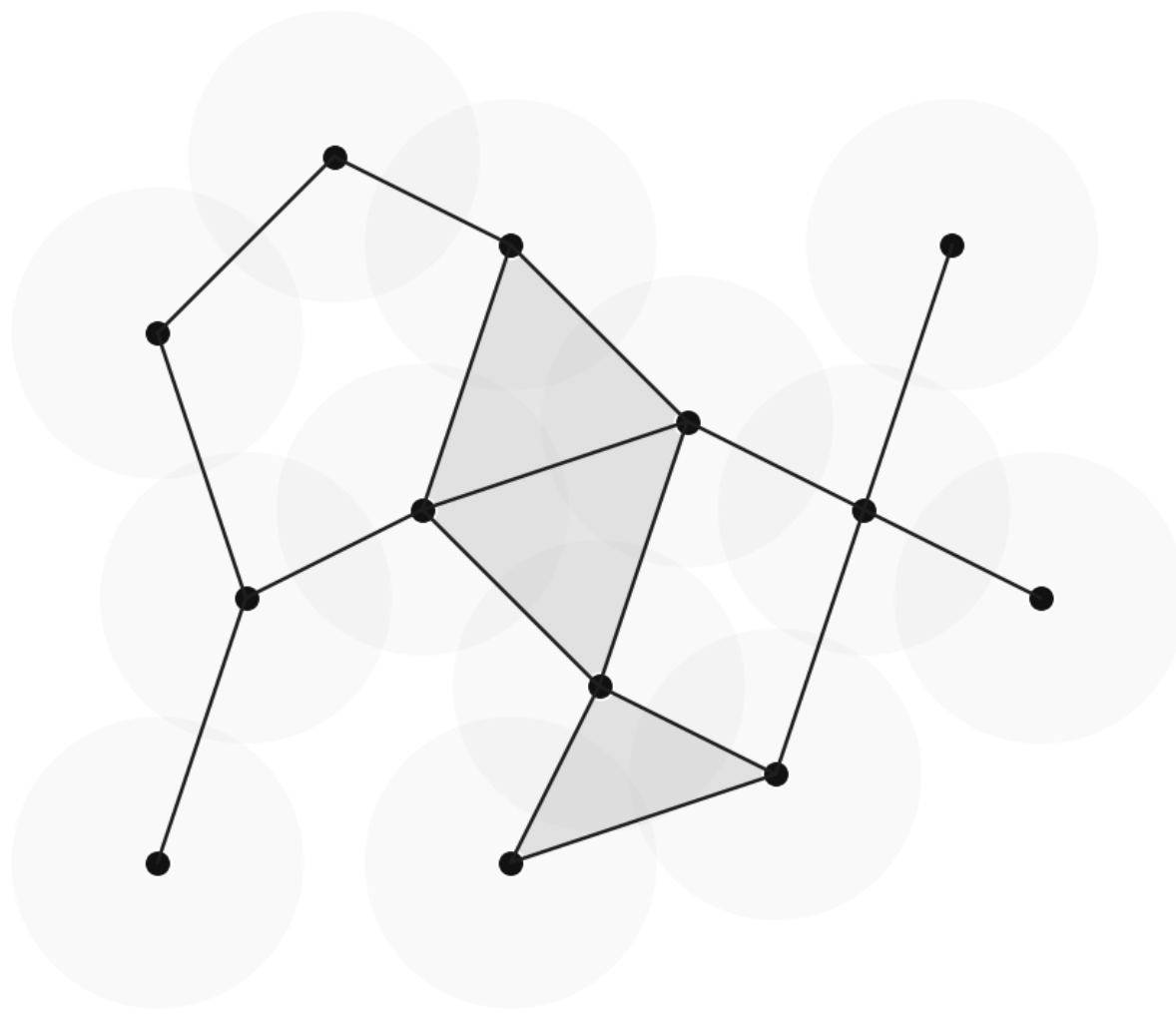
For $r \geq 0$: $C_r(S) \subseteq VR_r(S) \subseteq C_{\sqrt{2}r}(S)$



Point cloud



Čech complex

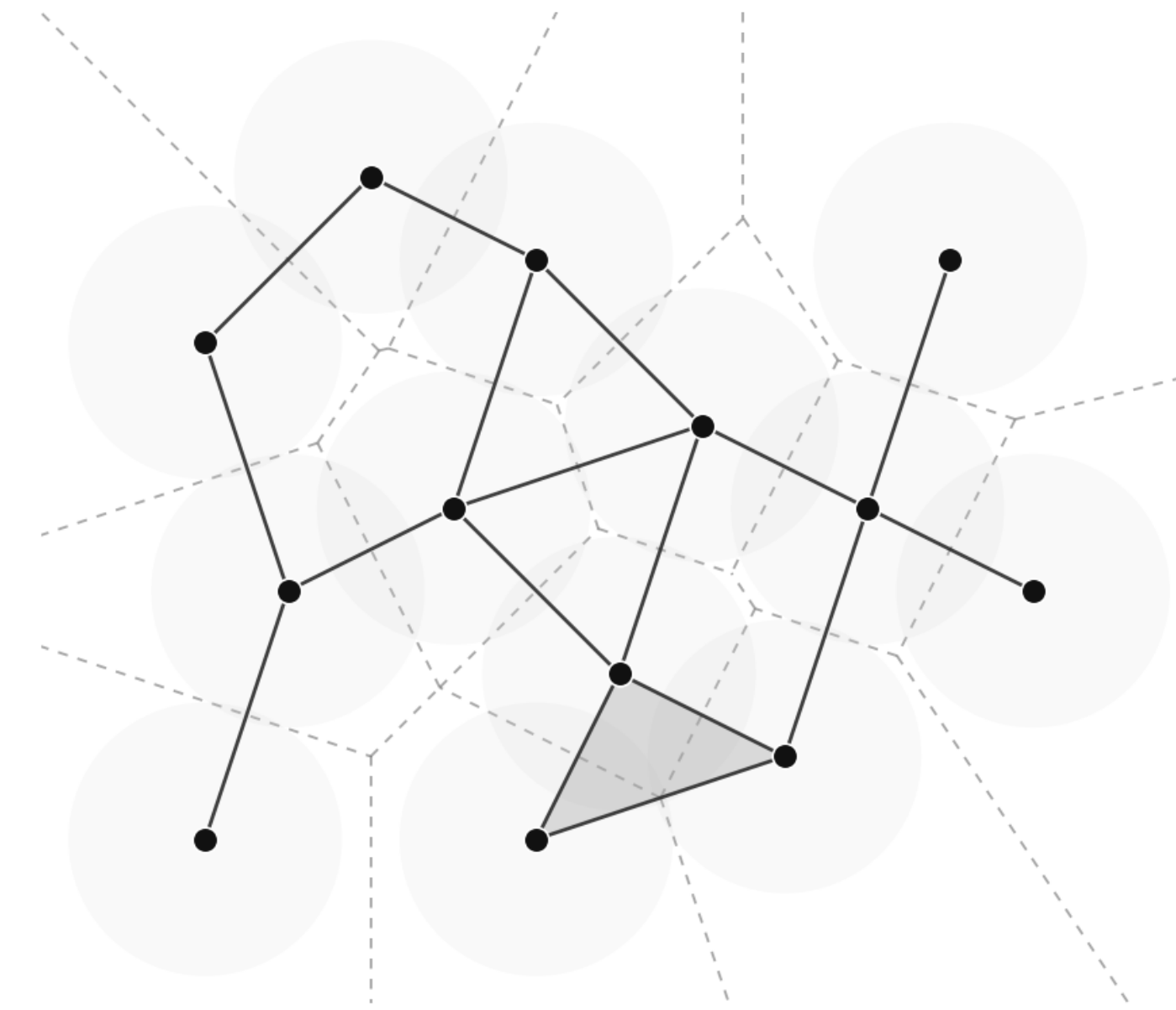


VR complex

Simplicial complexes

Other simplicial complexes

- Alpha shapes and Alpha complexes
- Dowker complexes
- k -NN complexes



H. Edelsbrunner, D. Kirkpatrick and R. Seidel, "On the shape of a set of points in the plane", *IEEE Transactions on Information Theory* 29.4, 1983, pp. 551–559

C. H. Dowker, "Homology Groups of Relations", *Annals of Mathematics* 56.1, 1952, pp. 84–95

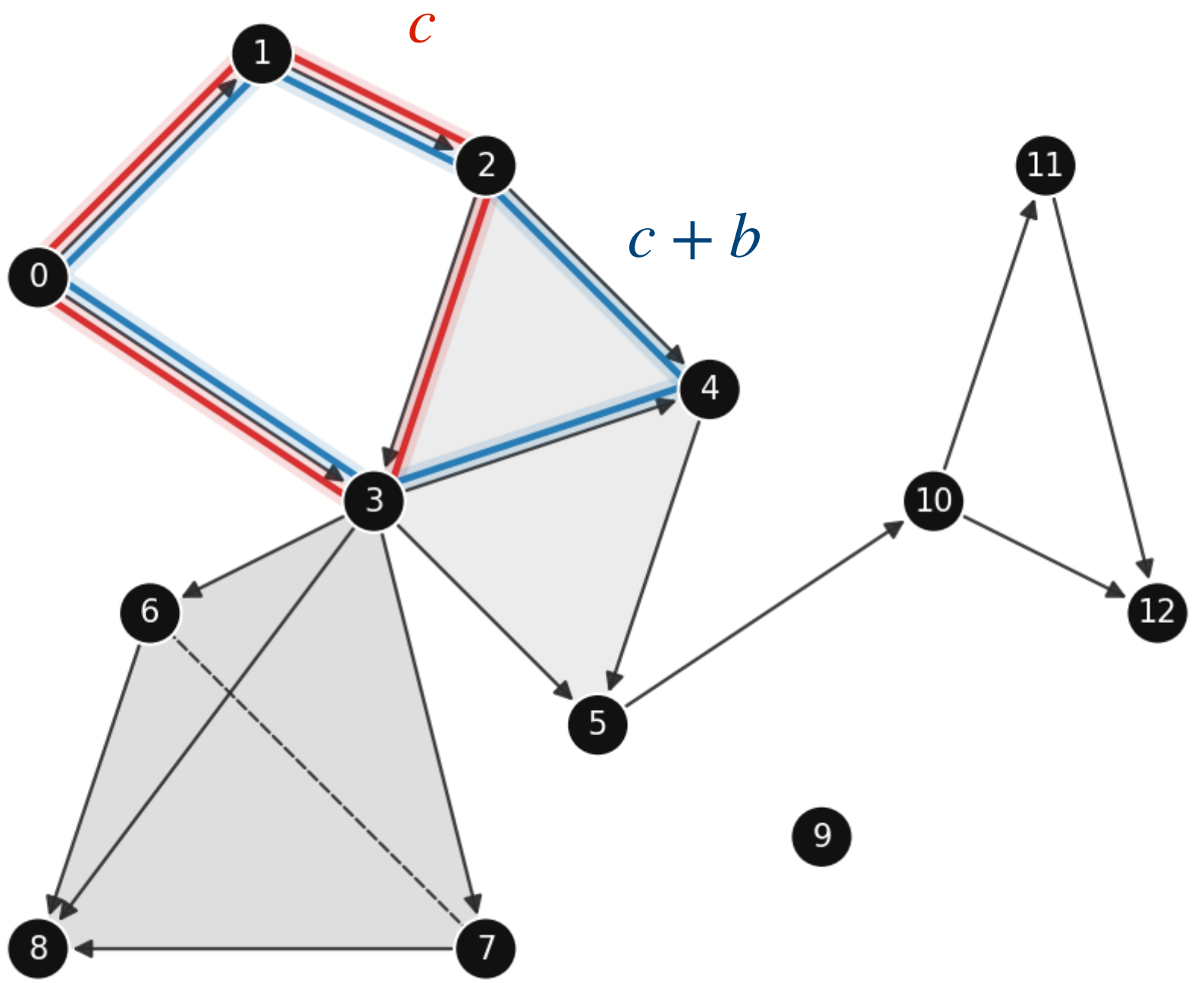
M. Q. Le and D. Taylor, "Persistent Homology with k -nearest-neighbor Filtrations reveals Topological Convergence of PageRank", 2022, arXiv: 2206.04725

Simplicial homology

Session 2 — Topological data and how to analyze them

Simplicial homology

Chain groups



Homology group:

$$H_k(X, \mathbb{k}) = \frac{\ker \partial_k}{\text{im } \partial_{k+1}}$$

Cycles

Boundaries

Example:

$$c = (0,1) + (1,2) + (2,3) - (0,3) \in \ker \partial_1$$

$$\partial_1(c) = (0) - (1) + (1) - (2) + (2) - (3) - (0) + (3) = 0$$

$$b = (2,3) + (3,4) - (2,4) = \partial_2(2,3,4) \in \text{im } \partial_2$$

$$c - b = (0,1) + (1,2) + (2,4) - (3,4) - (0,3) \cong c$$

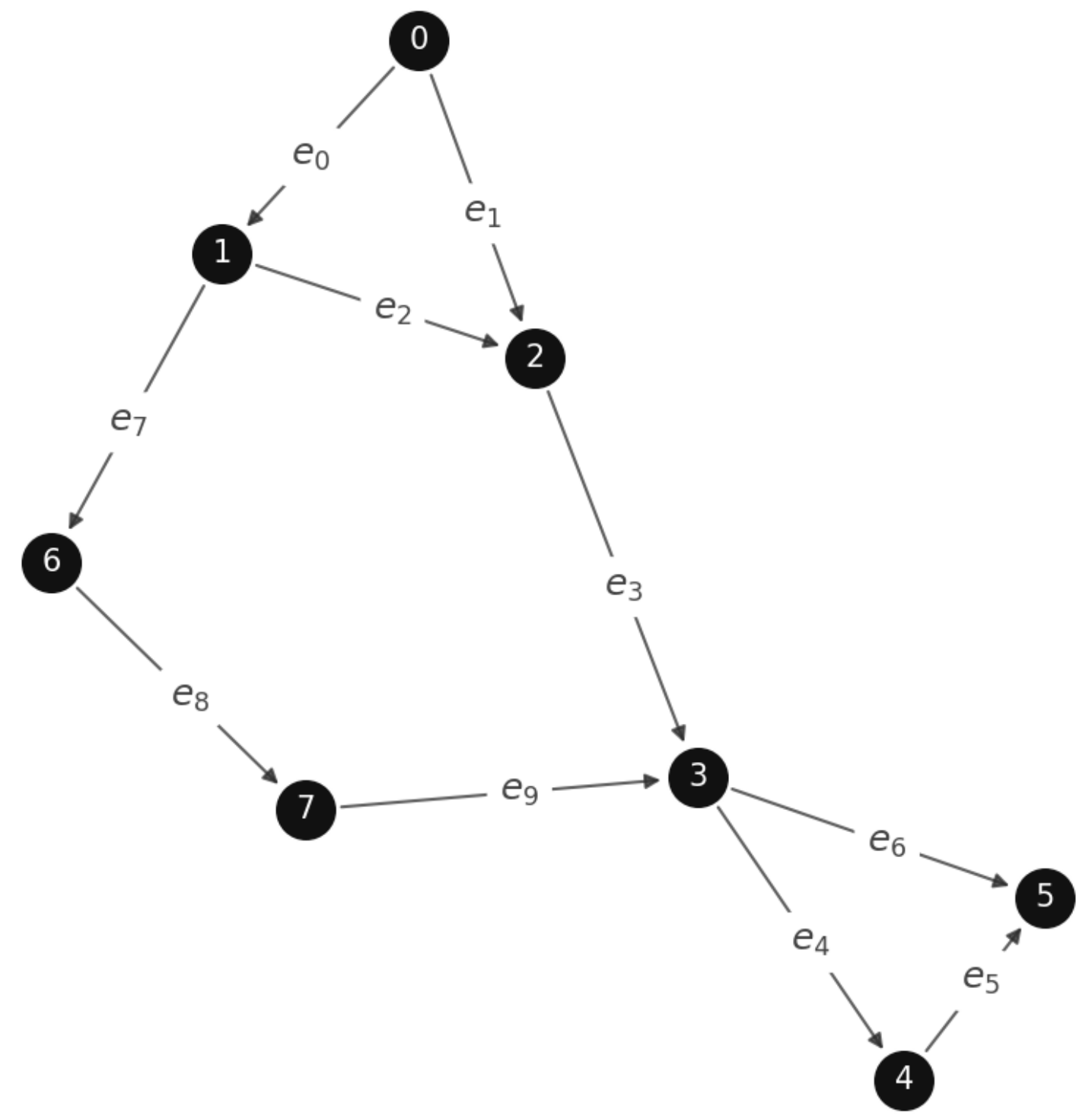
Betti number:

$$\beta_k(X, \mathbb{k}) = \dim H_k(X, \mathbb{k})$$

“Number of k -dimensional topological features in X ”

Simplicial homology

Incidence matrices revisited



$$C_1(G) \xrightarrow{\partial_1} C_0(G)$$

$$C_1(G) \cong \mathbb{R}^{|E|} \quad C_0(G) \cong \mathbb{R}^{|V|}$$

$$B \in \mathbb{R}^{n \times m}, \quad B_{ve} = \begin{cases} +1 & \text{if vertex } v \text{ is the tail of edge } e, \\ -1 & \text{if vertex } v \text{ is the head of edge } e, \\ 0 & \text{otherwise.} \end{cases}$$

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} = [\partial_1]_{\mathcal{B}_E}^{\mathcal{B}_V}$$

Simplicial homology

Nerve theorem

Nerve theorem

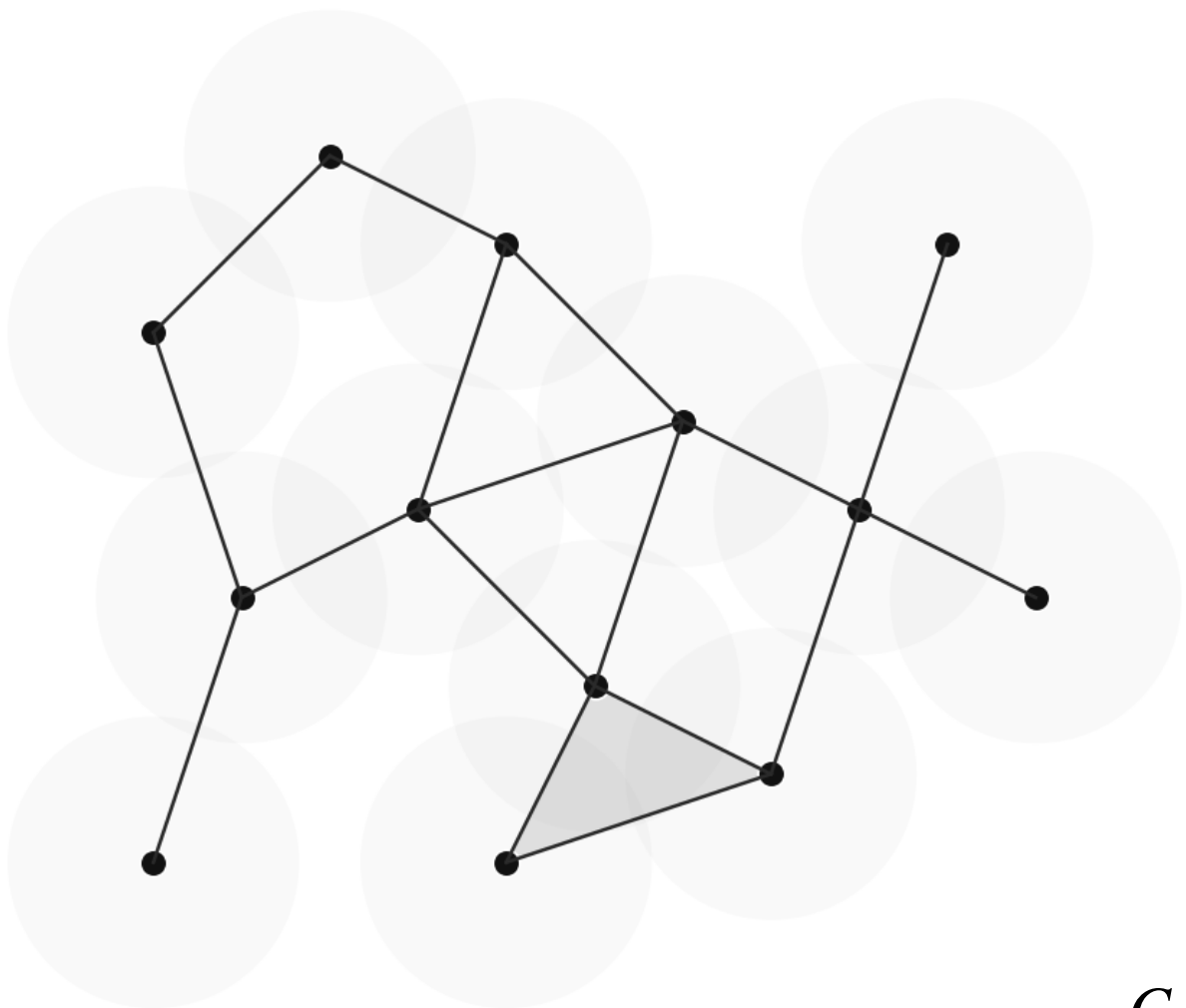
X — topological space

$\mathcal{U} = (U_i)_{i \in I}$ — cover of X

$$\text{Nerve}(\mathcal{U}) = \left\{ J \subseteq I : \bigcap_{j \in J} U_j \neq \emptyset \right\}$$

If any $\bigcap_{j \in J \subseteq I} U_j \neq \emptyset$ is contractible, with some *tameness conditions* on X and \mathcal{U} , $\text{Nerve}(\mathcal{U}) \cong X$

(homotopy equivalent)



Čech complex is a nerve

$$C_r(S, d) = \left\{ \{s_{i_0}, \dots, s_{i_k}\} : \bigcap_{j=0}^k B(s_{i_j}, r) \neq \emptyset \right\}$$

The Vietoris-Rips complex, although not a nerve, is isomorphic to one¹

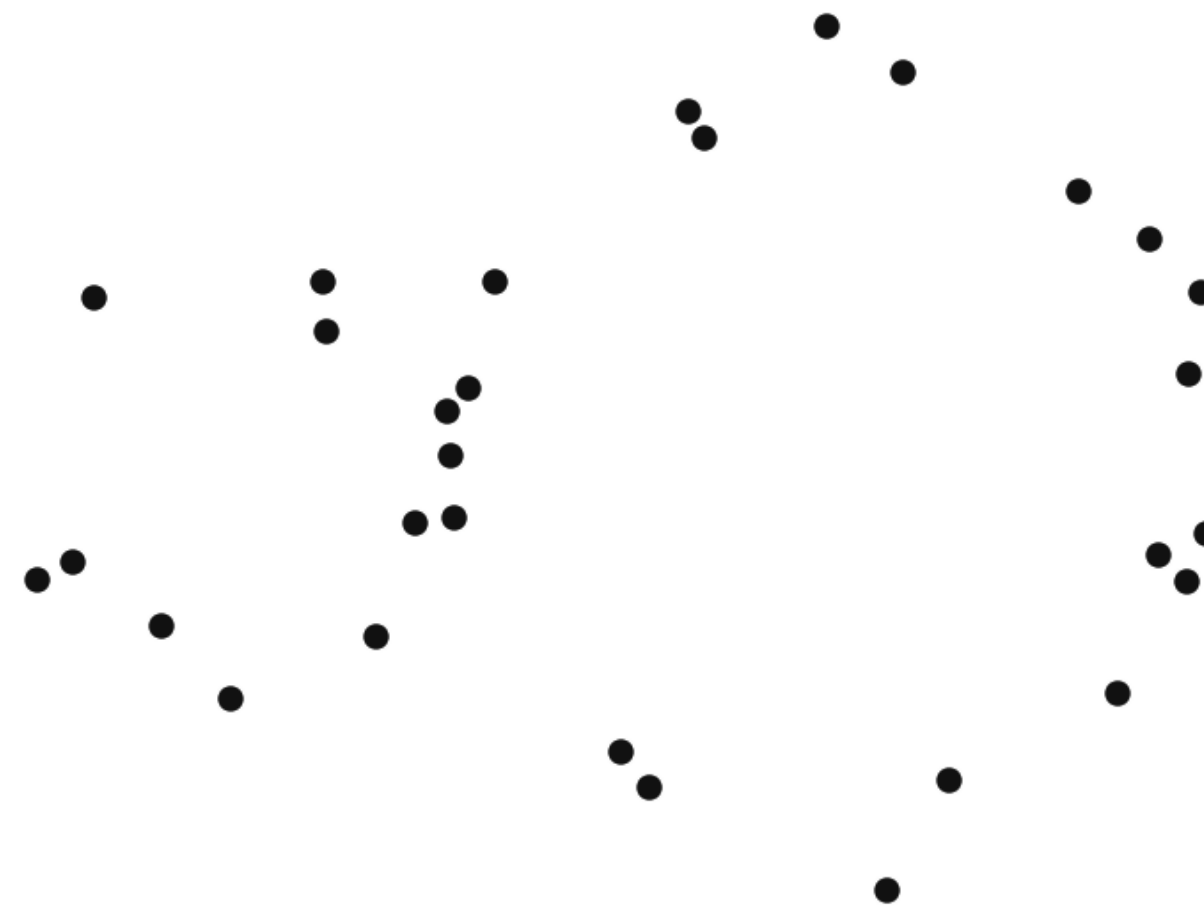
¹Ghrist, R., and A. Muhammad. "Coverage and Hole-Detection in Sensor Networks via Homology." *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks*, 2005., April 2005, 254–60.

¹Bauer, Ulrich, Michael Kerber, Fabian Roll, and Alexander Rolle. "A Unified View on the Functorial Nerve Theorem and Its Variations." *Expositiones Mathematicae* 41, no. 4 (2023): 125503.

Simplicial homology

Nerve theorem — the practical consequences

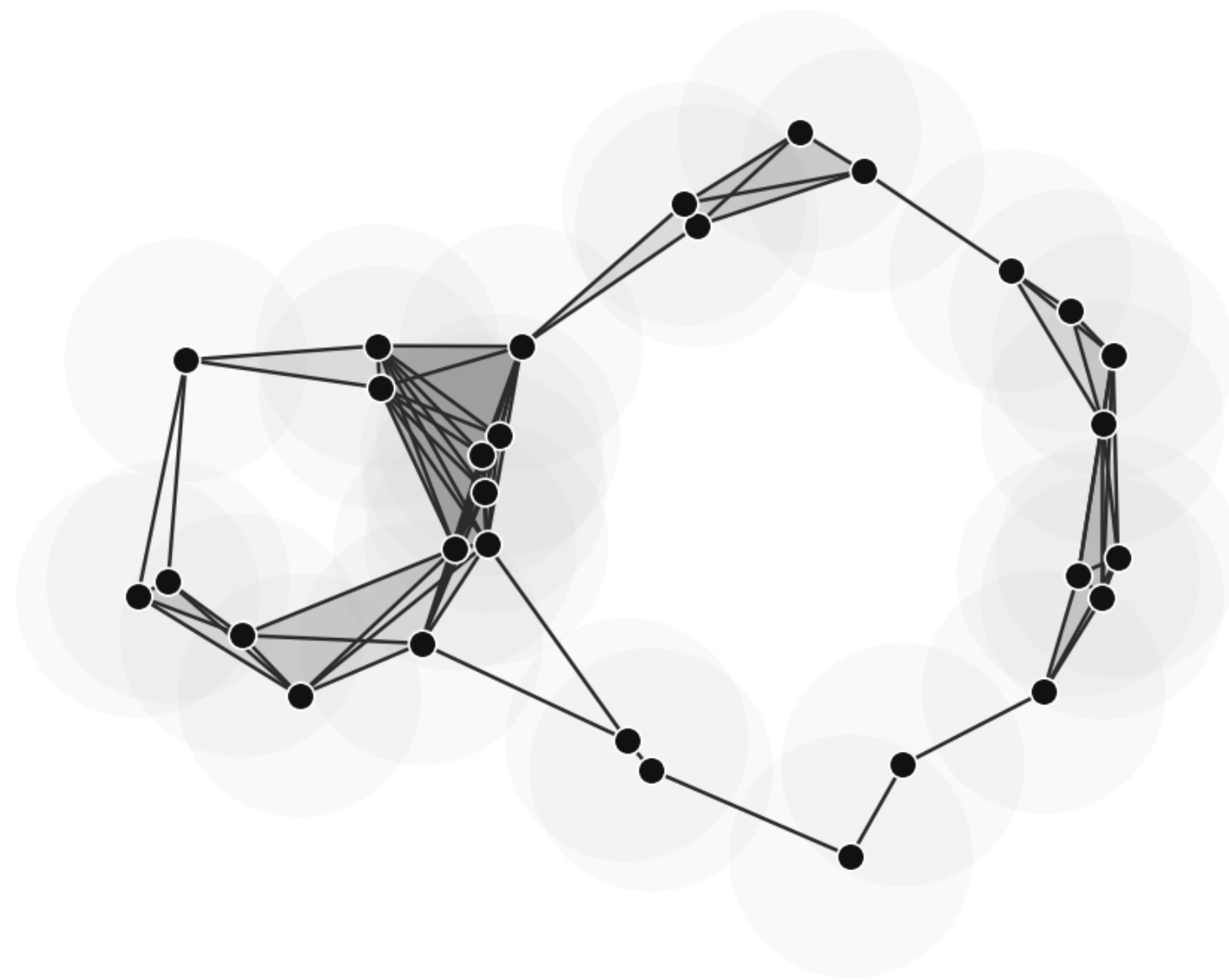
Suppose you are given the following point cloud, how can you design a test to conclude that this sample comes from a distribution over two circles?



Simplicial homology

Nerve theorem — the practical consequences

1. Approximate S with a Čech complex for a some radius that captures the holes: $C_r(S)$
2. Compute homology in degree 1: $H_1(C_r(S))$
3. If $\beta_1(C_r(S)) = 2$, then you are observing two holes



How do you find the radius r ?

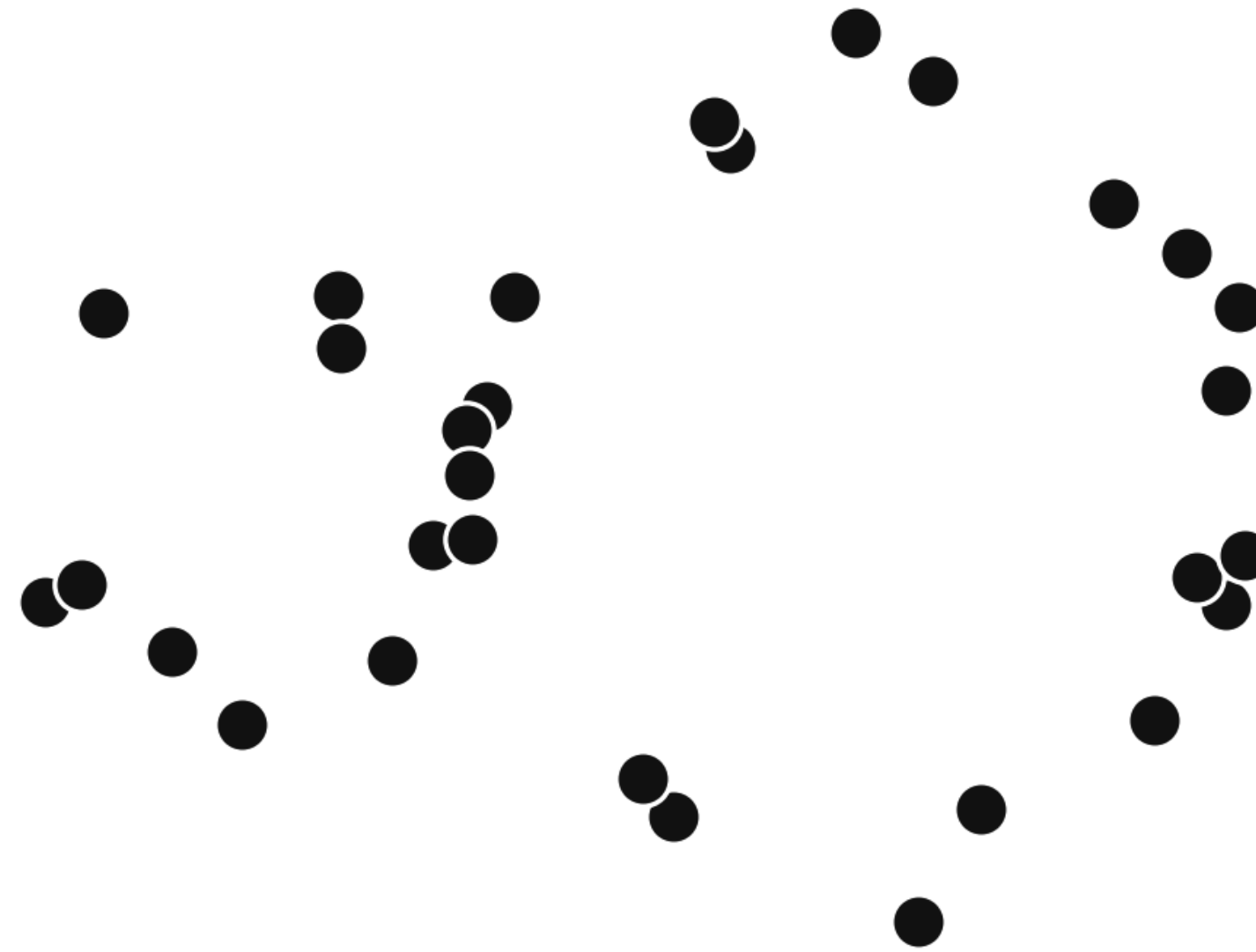
Persistent homology

Session 2 — Topological data and how to analyze them

Persistent homology

Filtrations

Consider all possible radii!



Filtration: $\{X_t : t \in \mathbb{R}\}$ such that for $s \leq t$, $X_s \subseteq X_t$

Persistent homology

Persistence modules

Functoriality of homology:

Given a simplicial map $f : X \rightarrow Y$, we can define a map at the level of homology:

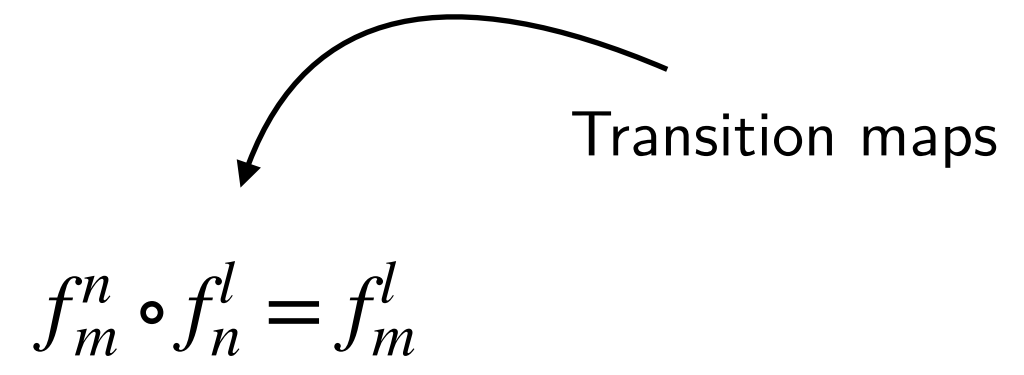
- For chains: $f_{\#} : C_k(X) \rightarrow C_k(Y)$ by sending $\sigma \mapsto f(\sigma)$ if $f(\sigma) \in Y_k$, and 0 otherwise; and extend linearly
- This assignment respects boundaries: $\partial_k \circ f_{\#} = f_{\#} \circ \partial_k$

This induces a map $f_* : H_k(X) \rightarrow H_k(Y)$ that satisfies:

$$(f \circ g)_* = f_* \circ g_* \qquad (\text{id}_X)_* = \text{id}_{H_k(X)}$$

Applying this to a filtration:

$$X_{\bullet} : X_1 \hookrightarrow X_2 \hookrightarrow \dots \hookrightarrow X_N \quad \rightsquigarrow \quad H_k(X_{\bullet}) : H_k(X_1) \xrightarrow{f_2^1} H_k(X_2) \xrightarrow{f_3^2} \dots \xrightarrow{f_N^{N-1}} H_k(X_N)$$



Persistent homology

Persistence modules

Functoriality of homology:

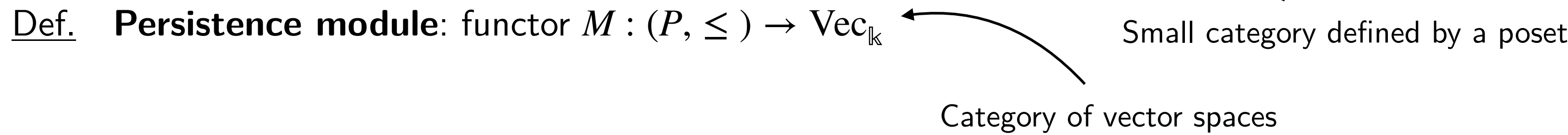
Given a simplicial map $f : X \rightarrow Y$, we can define a map at the level of homology:

- For chains: $f_{\#} : C_k(X) \rightarrow C_k(Y)$ by sending $\sigma \mapsto f(\sigma)$ if $f(\sigma) \in Y_k$, and 0 otherwise; and extend linearly
- This assignment respects boundaries: $\partial_k \circ f_{\#} = f_{\#} \circ \partial_k$

This induces a map $f_* : H_k(X) \rightarrow H_k(Y)$ that satisfies:

$$(f \circ g)_* = f_* \circ g_* \qquad (\text{id}_X)_* = \text{id}_{H_k(X)}$$

Applying this to a filtration:



- For every $t \in P$, we have a vector space $M_t = M(t)$
- For every pair of comparable points $s \leq t$ we have a map $M(s \leq t) : M_s \rightarrow M_t$ (**transition map**)

Persistent homology

Structure theorem and persistence barcodes and diagrams

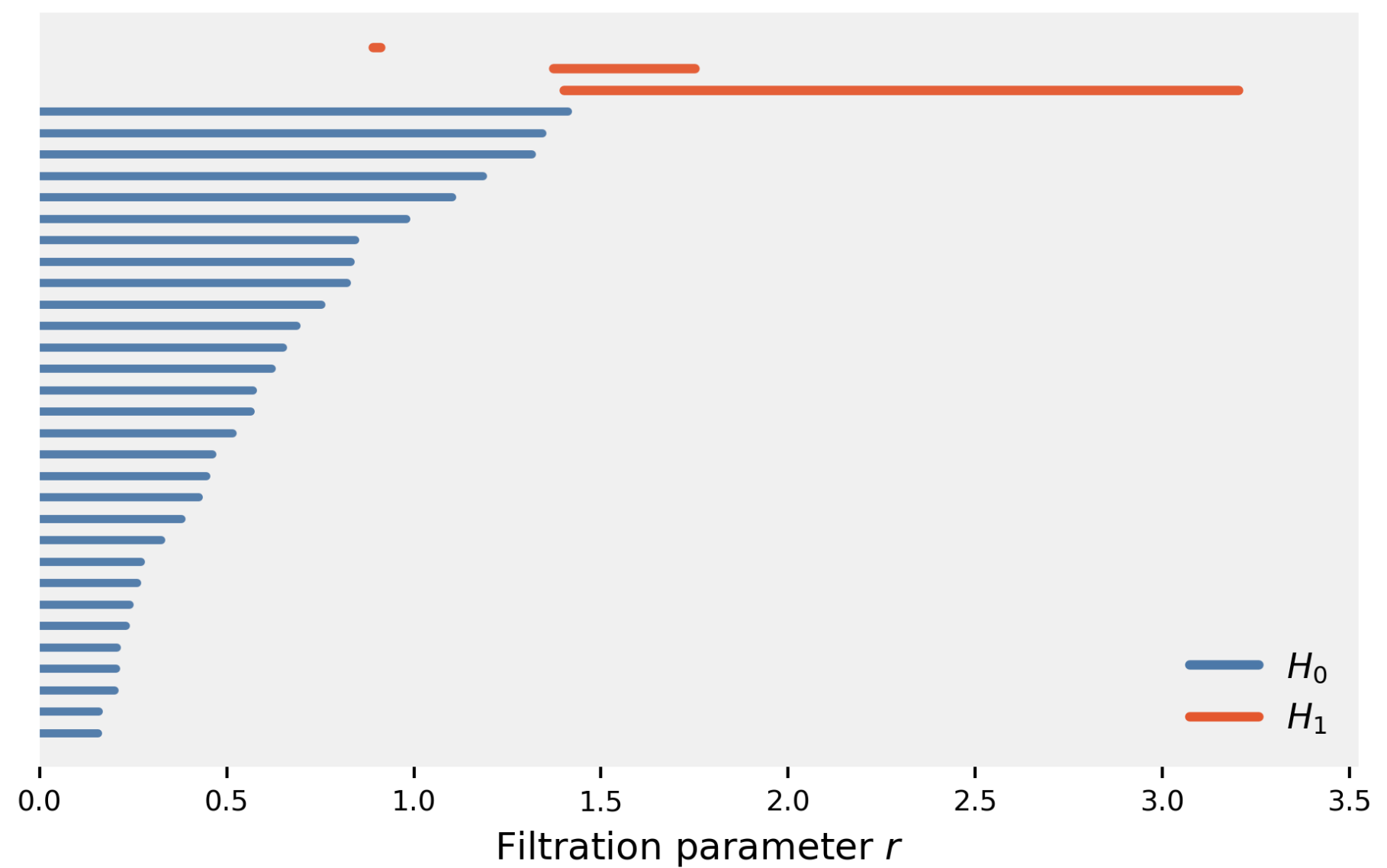
Structure theorem:

$$H_k(X_\bullet) \cong \bigoplus_{i \in I} \mathbb{k}[b_i, d_i]$$

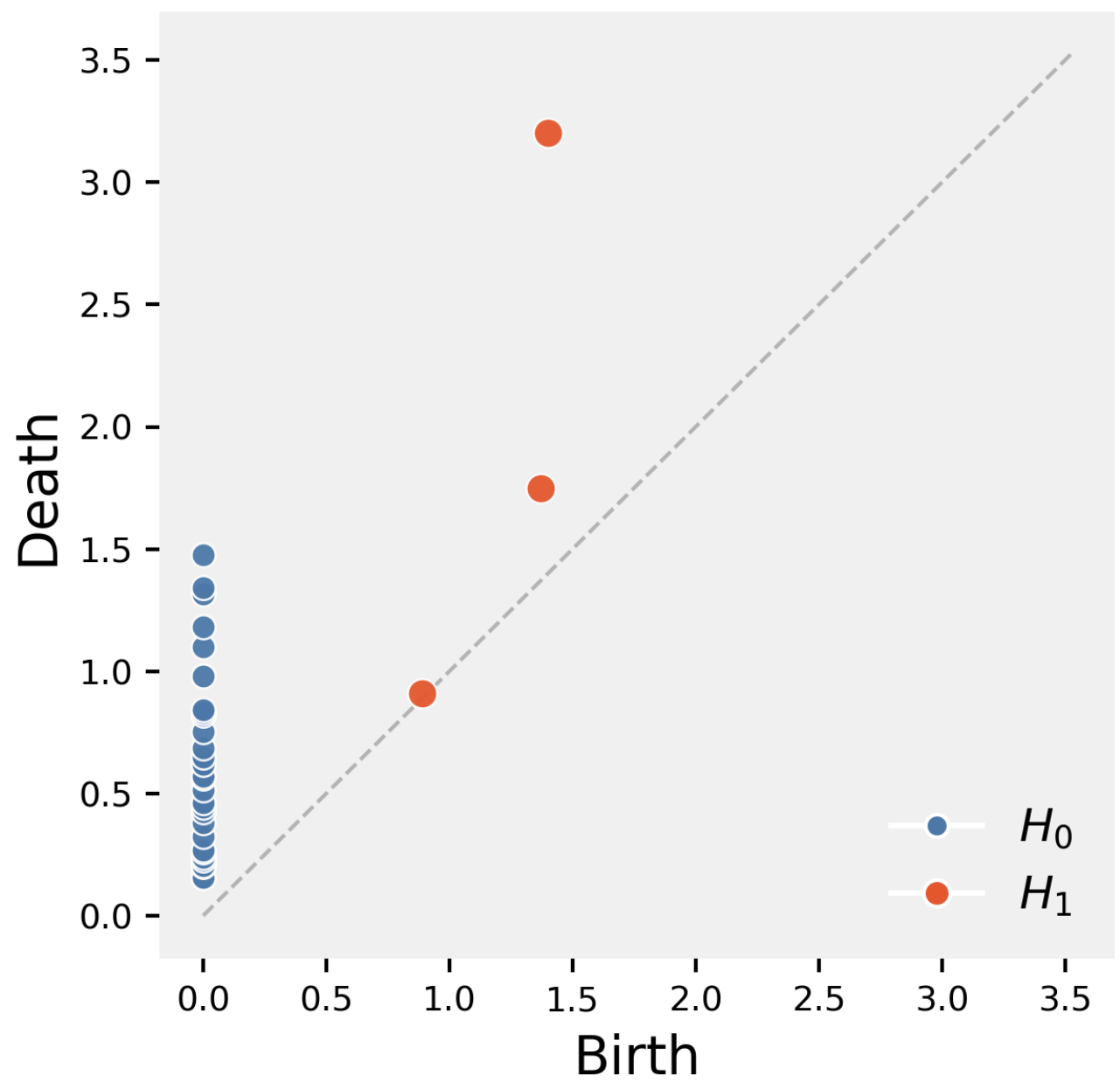
$$\mathbb{k}[b, d)_t = \begin{cases} \mathbb{k} & \text{if } d \leq t < d \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{k}[b, d)(s \leq t) = \begin{cases} \text{id}_{\mathbb{k}} & \text{if } d \leq s \leq t < d \\ 0 & \text{otherwise} \end{cases}$$

Persistence barcode



Persistence diagram



Azumaya, Gorô. "Corrections and Supplementaries to My Paper Concerning Krull-Remak-Schmidt's Theorem." *Nagoya Mathematical Journal* 1 (June 1950): 117–24.
 Crawley-Boevey, William. "Decomposition of Pointwise Finite-Dimensional Persistence Modules." *Journal of Algebra and Its Applications* 14, no. 05 (2015): 1550066.

Persistent homology

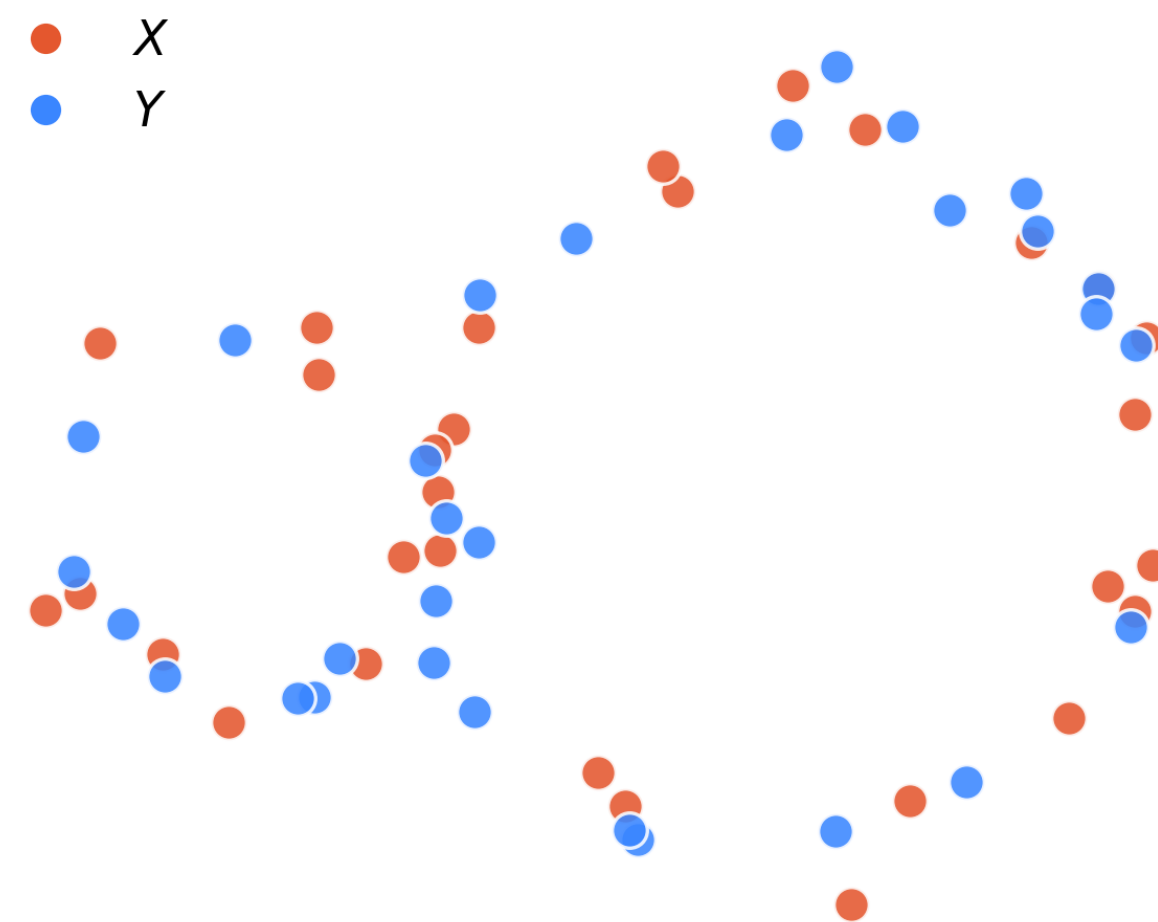
Stability

Recipe for a stability result in TDA:

A given choice of invariant $\text{inv} : \mathcal{D} \rightarrow \mathcal{F}$ is *stable* if it is Lipschitz

$$d_{\mathcal{F}}(\text{inv}(X), \text{inv}(Y)) \leq C \cdot d_{\mathcal{D}}(X, Y), \quad \forall X, Y \in \mathcal{D}$$

Gromov-Hausdorff distance:



For (X, d_X) , (Y, d_Y) metric spaces and $C \subset X \times Y$ a correspondence

Distortion: $\text{dis}(C) = \sup\{ |d_X(x, x') - d_Y(y, y')| : (x, y), (x', y') \in C\}$

$$d_{GH}(X, Y) = \frac{1}{2} \inf_{C \subset X \times Y} \max\{\text{dis}(C)\}$$

Persistent homology

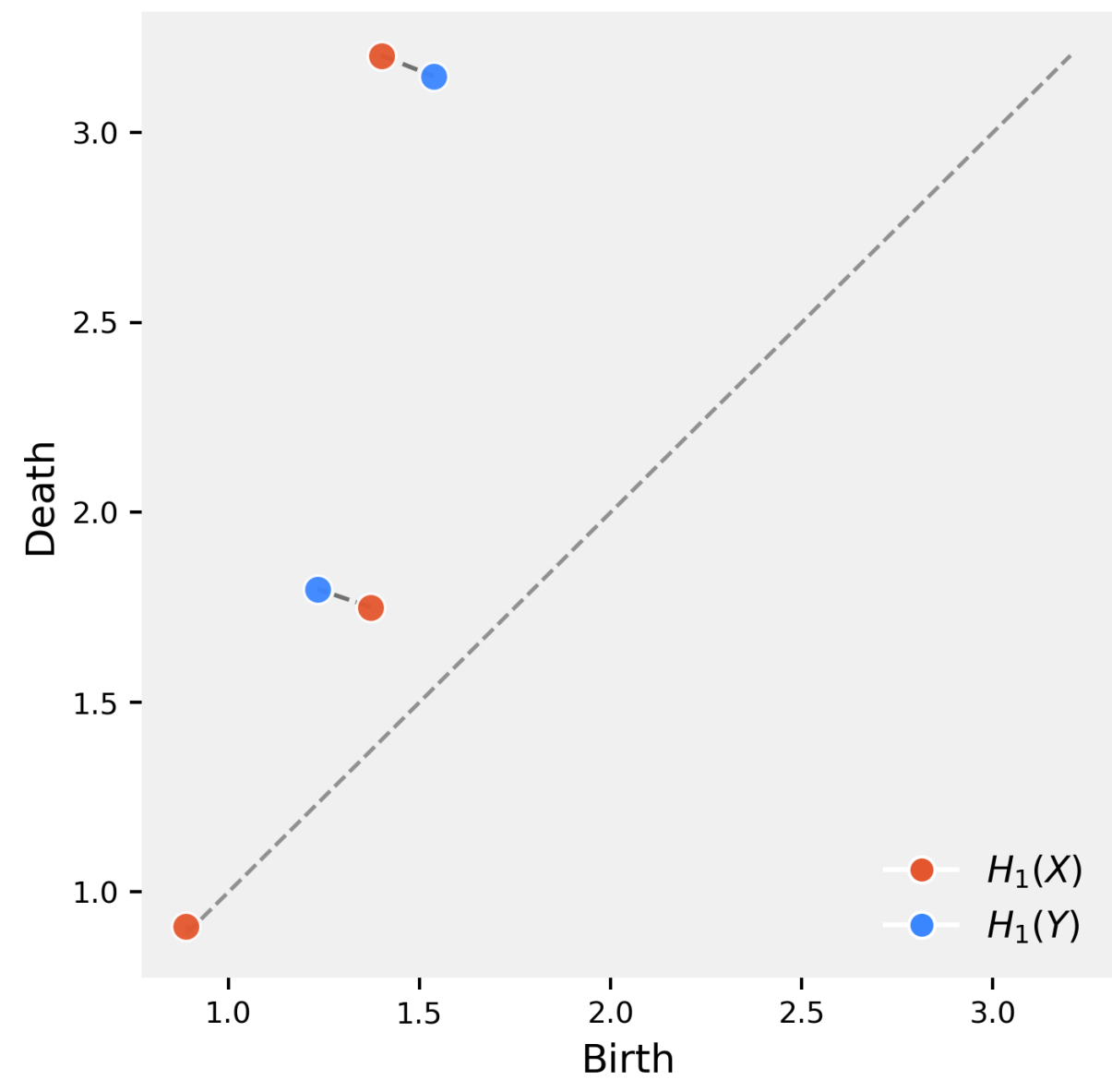
Stability

Recipe for a stability result in TDA:

A given choice of invariant $\text{inv} : \mathcal{D} \rightarrow \mathcal{F}$ is *stable* if it is Lipschitz

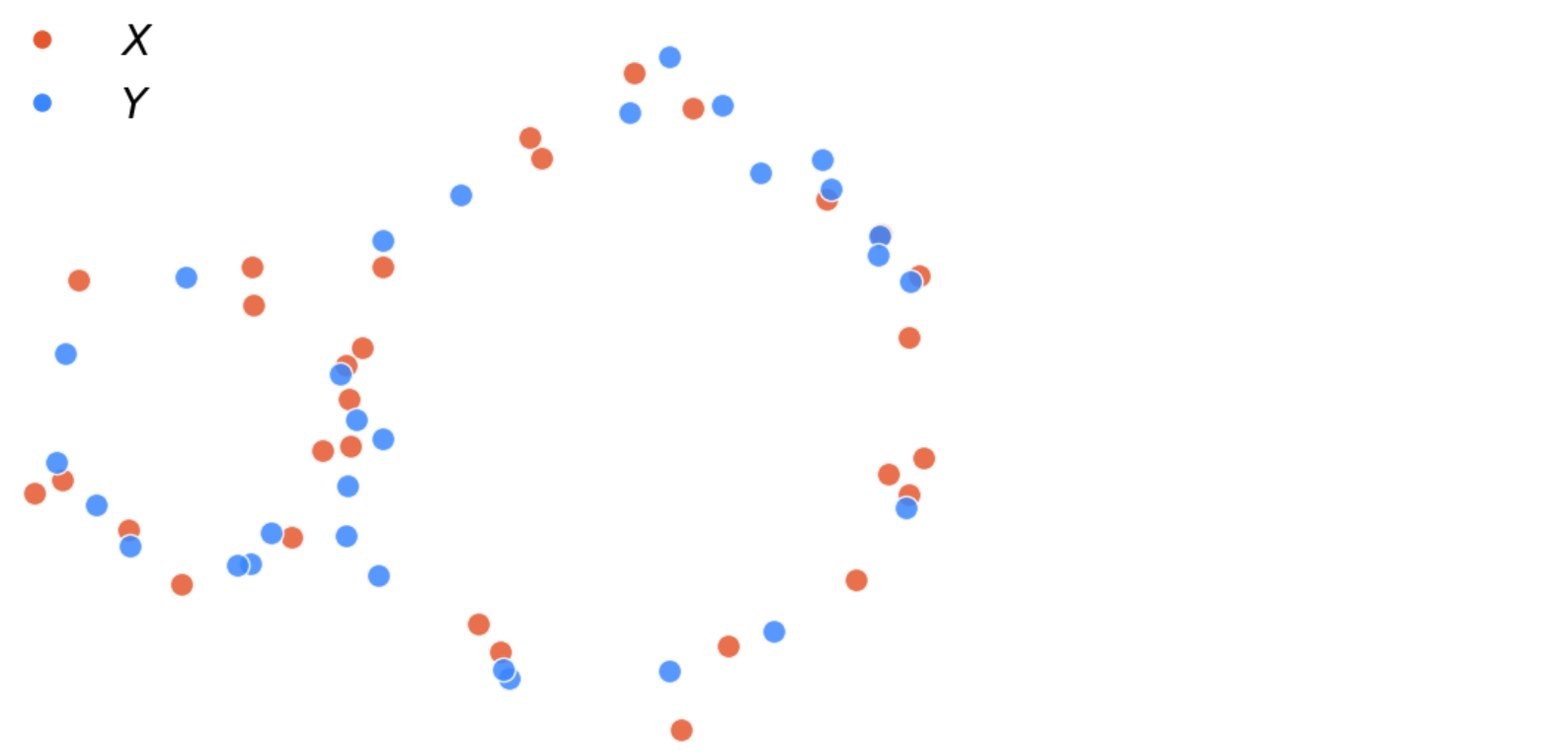
$$d_{\mathcal{F}}(\text{inv}(X), \text{inv}(Y)) \leq C \cdot d_{\mathcal{D}}(X, Y), \quad \forall X, Y \in \mathcal{D}$$

Bottleneck distance:



$$d_B(\text{PD}(X), \text{PD}(Y)) = \inf_{\psi: \text{PD}(X) \rightarrow \text{PD}(Y)} \max\{\|x - \psi(x)\|_{\infty} : \forall x \in \text{PD}(X)\}$$

• X
• Y



$$\text{PD}(X) := \text{PD}(H_1(X))$$

$$\text{PD}(Y) := \text{PD}(H_1(Y))$$

Persistent homology

Stability

Recipe for a stability result in TDA:

A given choice of invariant $\text{inv} : \mathcal{D} \rightarrow \mathcal{F}$ is *stable* if it is Lipschitz

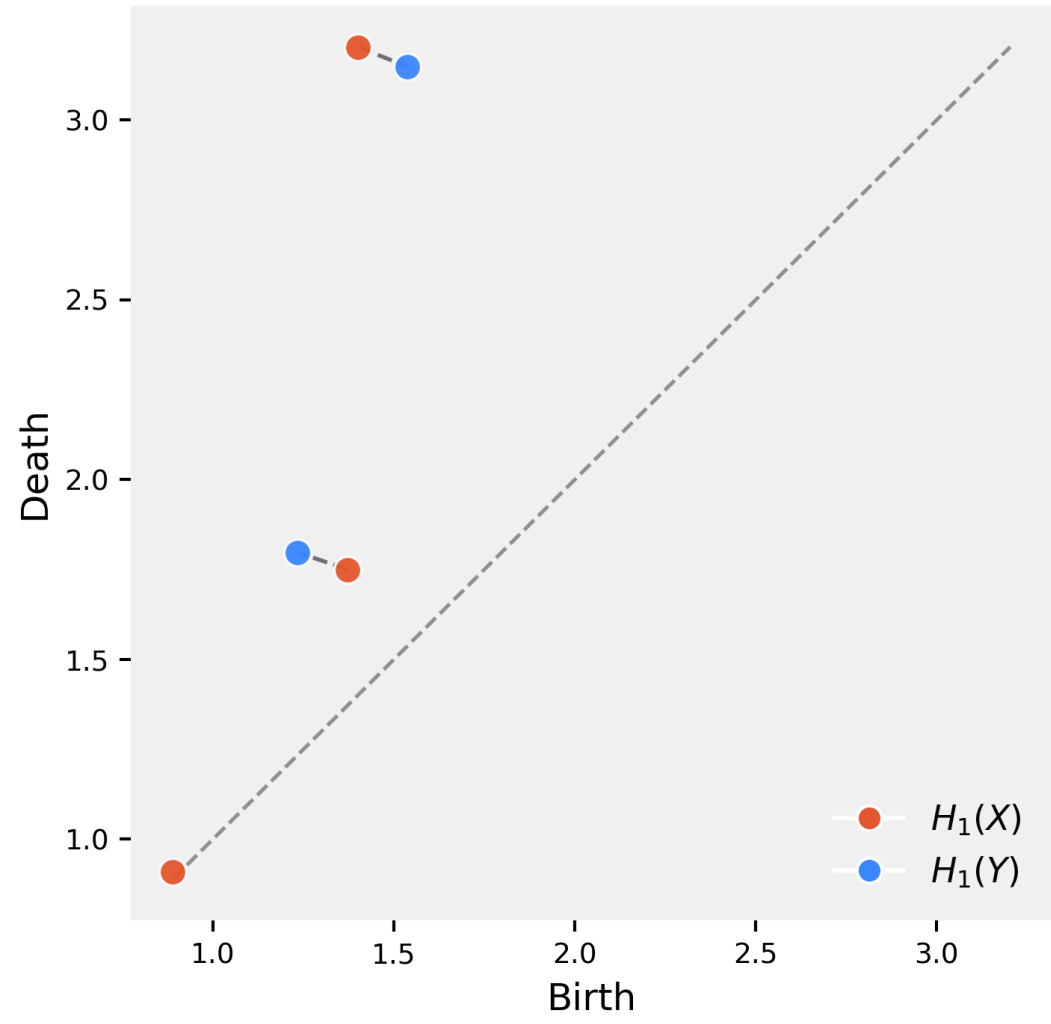
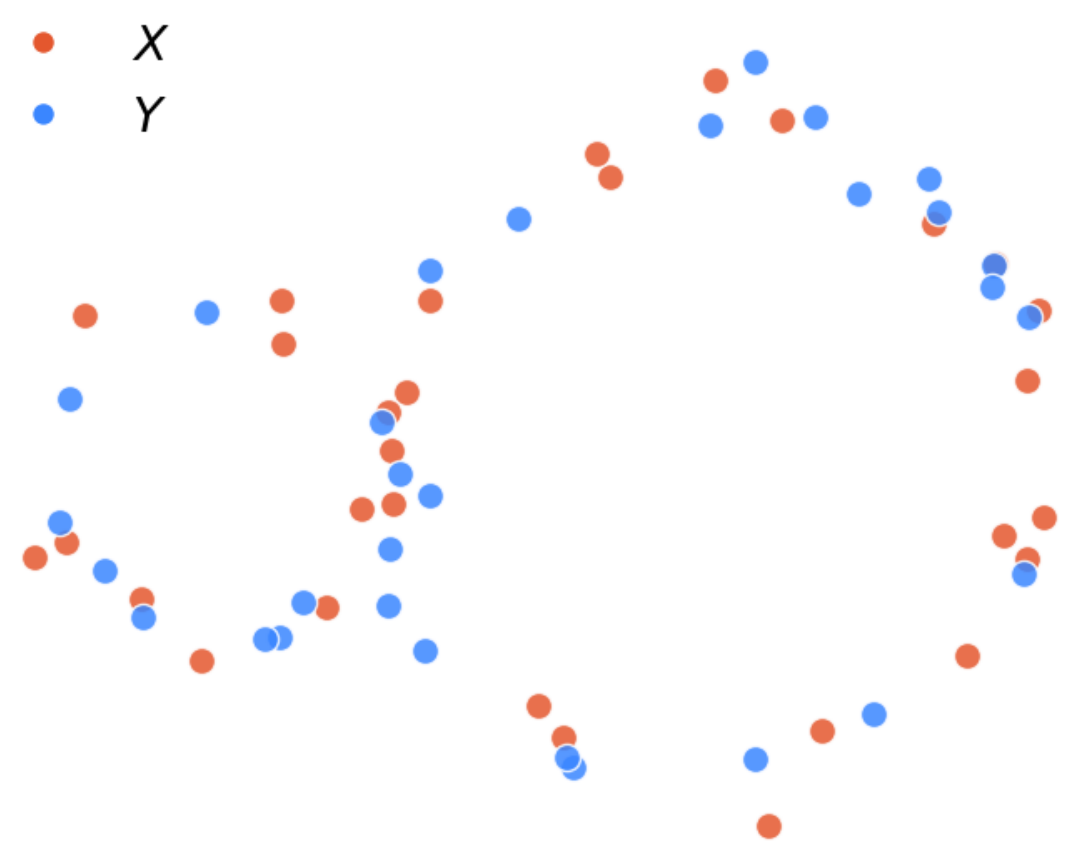
$$d_{\mathcal{F}}(\text{inv}(X), \text{inv}(Y)) \leq C \cdot d_{\mathcal{D}}(X, Y), \quad \forall X, Y \in \mathcal{D}$$

Bottleneck stability:

$$d_B(\text{PD}(X), \text{PD}(Y)) \leq 2 \cdot d_{GH}(X, Y)$$

$$\text{PD}(X) := \text{PD}(H_1(X))$$

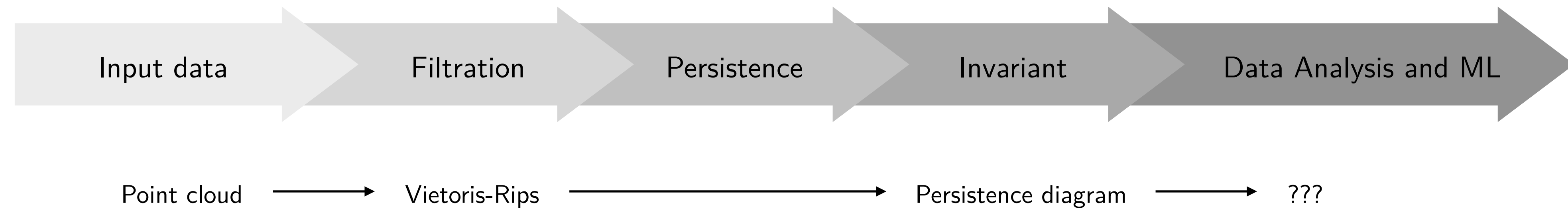
$$\text{PD}(Y) := \text{PD}(H_1(Y))$$



Chazal, Frédéric, Vin de Silva, and Steve Oudot. "Persistence Stability for Geometric Complexes." *Geometriae Dedicata* 173, no. 1 (2014): 193–214.

Persistent homology

The general pipeline



The space of diagrams* is a Polish space, amenable to statistics and probability¹

However, it is an Alexandrov space with negative curvature bounded from below, which makes (Fréchet) means and geodesics non-unique, and hinders the development of statistical theory²

In addition, most ML/DL methods expect input data coming from an Euclidean space

A possible solution: *vectorize*³ persistence diagrams, i.e., embed them in a Hilbert space where we can develop statistical theory

*with a different choice of metric between diagrams called the 2-Wasserstein distance

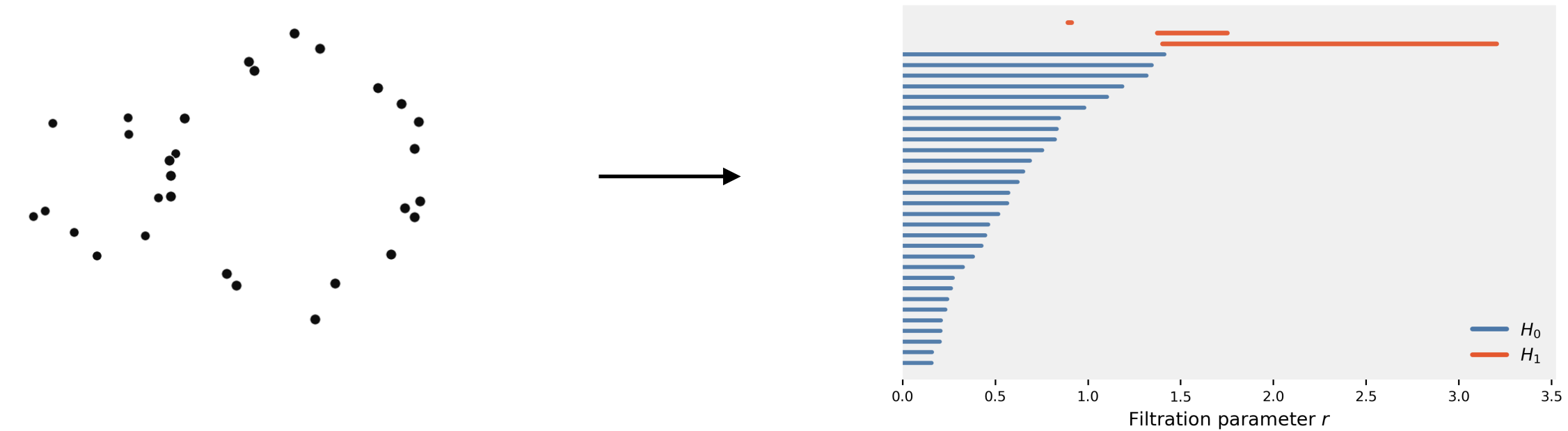
¹Mileyko, Yuriy, Sayan Mukherjee, and John Harer. "Probability Measures on the Space of Persistence Diagrams." *Inverse Problems* 27, no. 12 (2011): 124007.

²Turner, Katharine, Yuriy Mileyko, Sayan Mukherjee, and John Harer. "Fréchet Means for Distributions of Persistence Diagrams." *Discrete & Computational Geometry* 52, no. 1 (2014): 44–70.

³Ali, Dashti, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. "A Survey of Vectorization Methods in Topological Data Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, no. 12 (2023): 14069–80.

Persistent homology

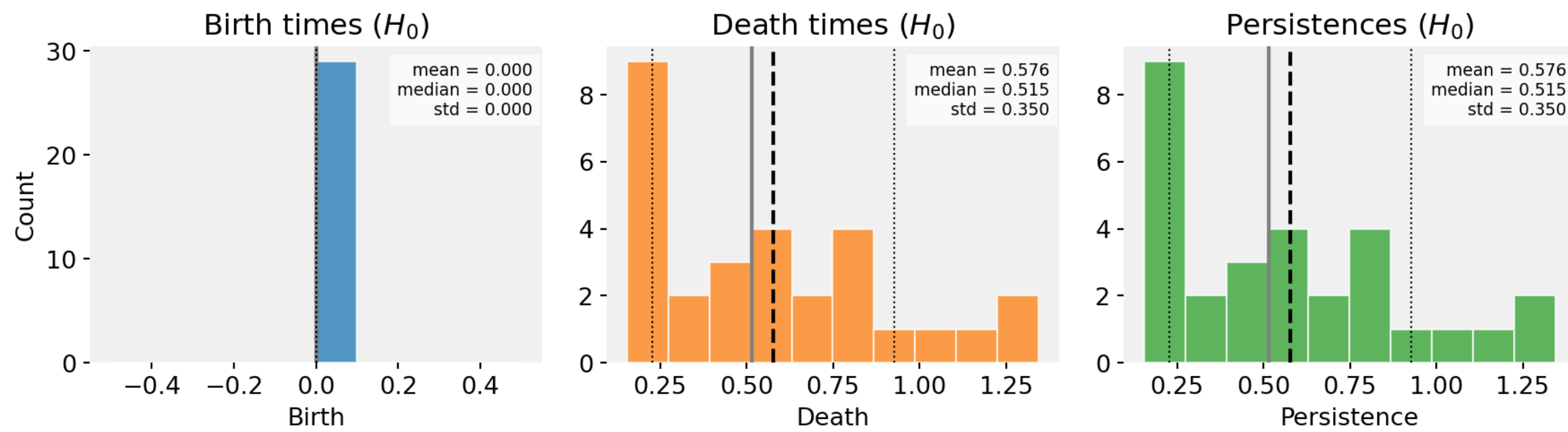
Statistics summary of the barcode



Define numerical descriptors for your barcode:

- Births (vector)
- Deaths (vector)
- Persistences (vector)
- Ratios birth/death (vector)
- Number of bars in the diagram (scalar)
- Total persistence (scalar)
- Persistent entropy¹
- ...

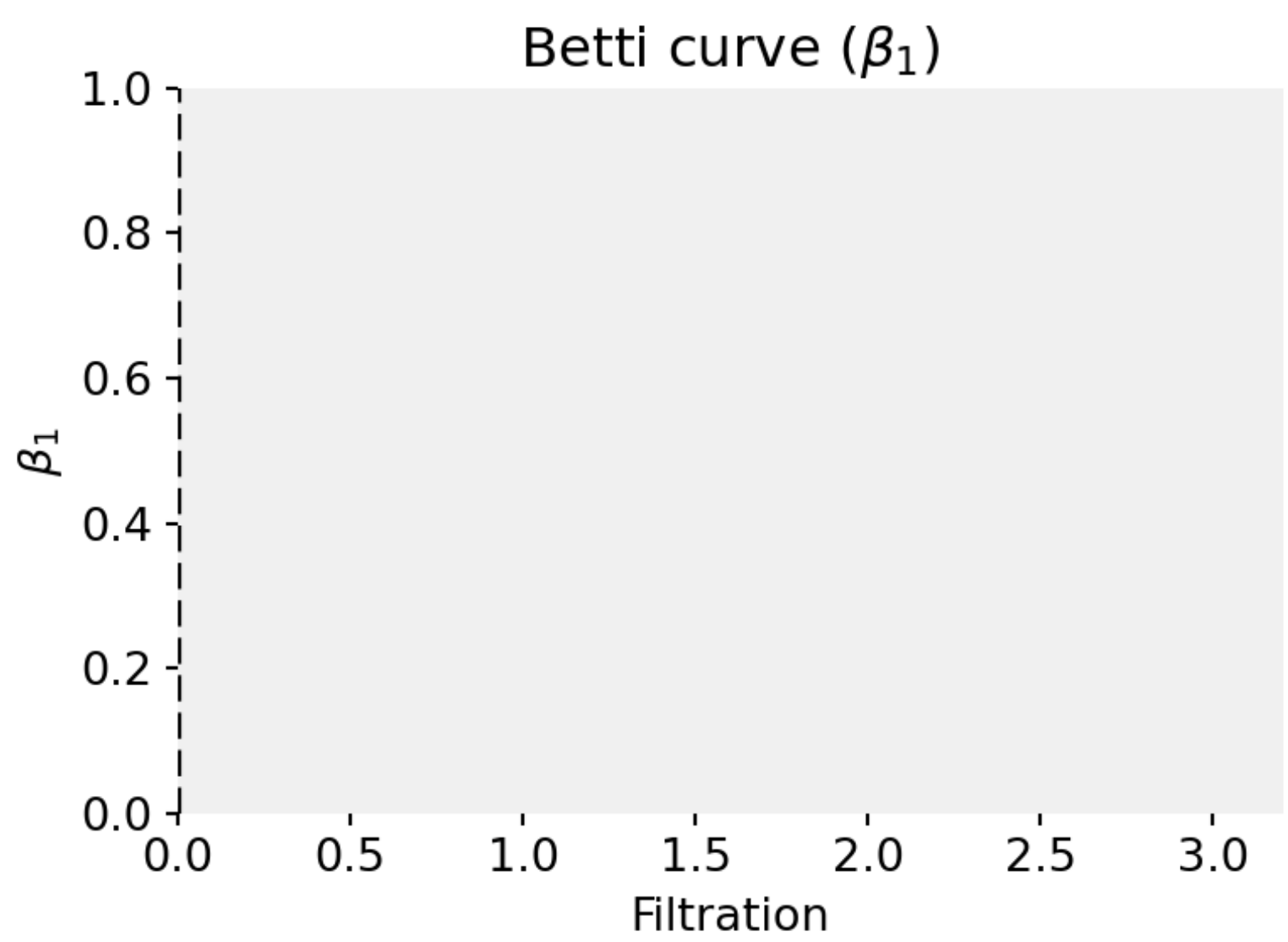
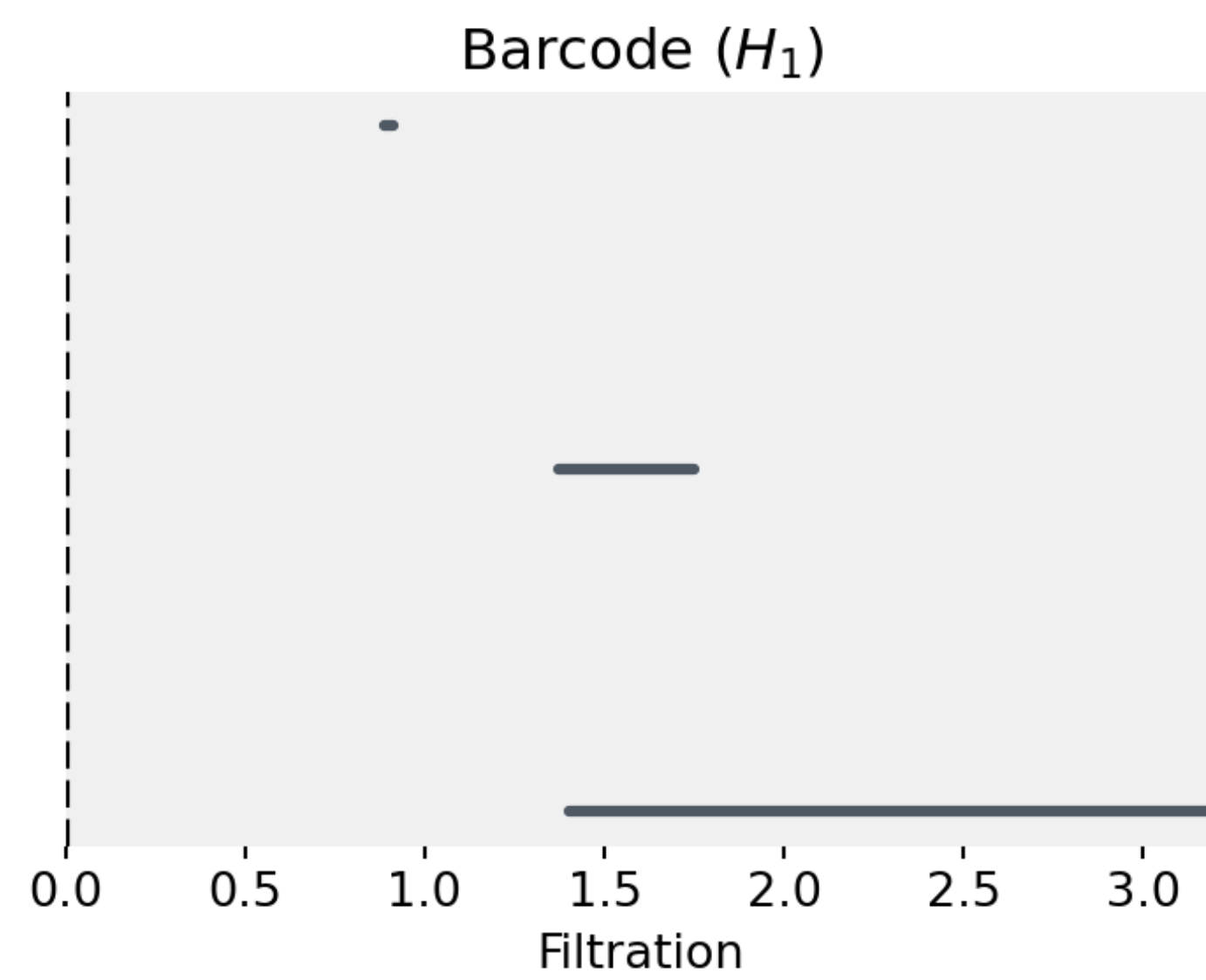
For the vectorial measures, pick some statistics (mean, standard deviation, median, quartiles...) and build a vector collecting their values for the empirical distributions of these descriptors of the barcode



¹Chintakunta, Harish, Thanos Gentimis, Rocio Gonzalez-Diaz, Maria-Jose Jimenez, and Hamid Krim. "An Entropy-Based Persistence Barcode." *Pattern Recognition* 48, no. 2 (2015): 391–401.

Persistent homology

Betti curves



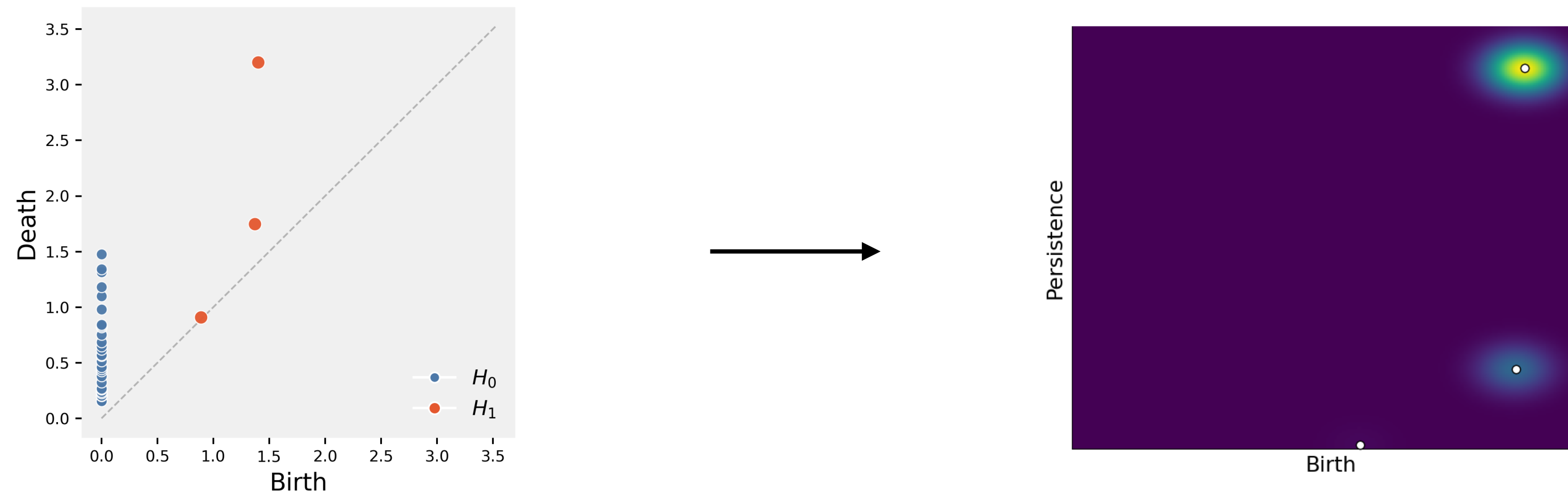
$$\beta_k : \mathbb{R} \rightarrow \mathbb{N}$$

$$\beta_k(t) = \#\{\text{alive bars in } H_k(X_t)\}$$

Giusti, Chad, Eva Pastalkova, Carina Curto, and Vladimir Itskov. "Clique Topology Reveals Intrinsic Geometric Structure in Neural Correlations." *Proceedings of the National Academy of Sciences* 112, no. 44 (2015): 13455–60.

Persistent homology

Persistence images



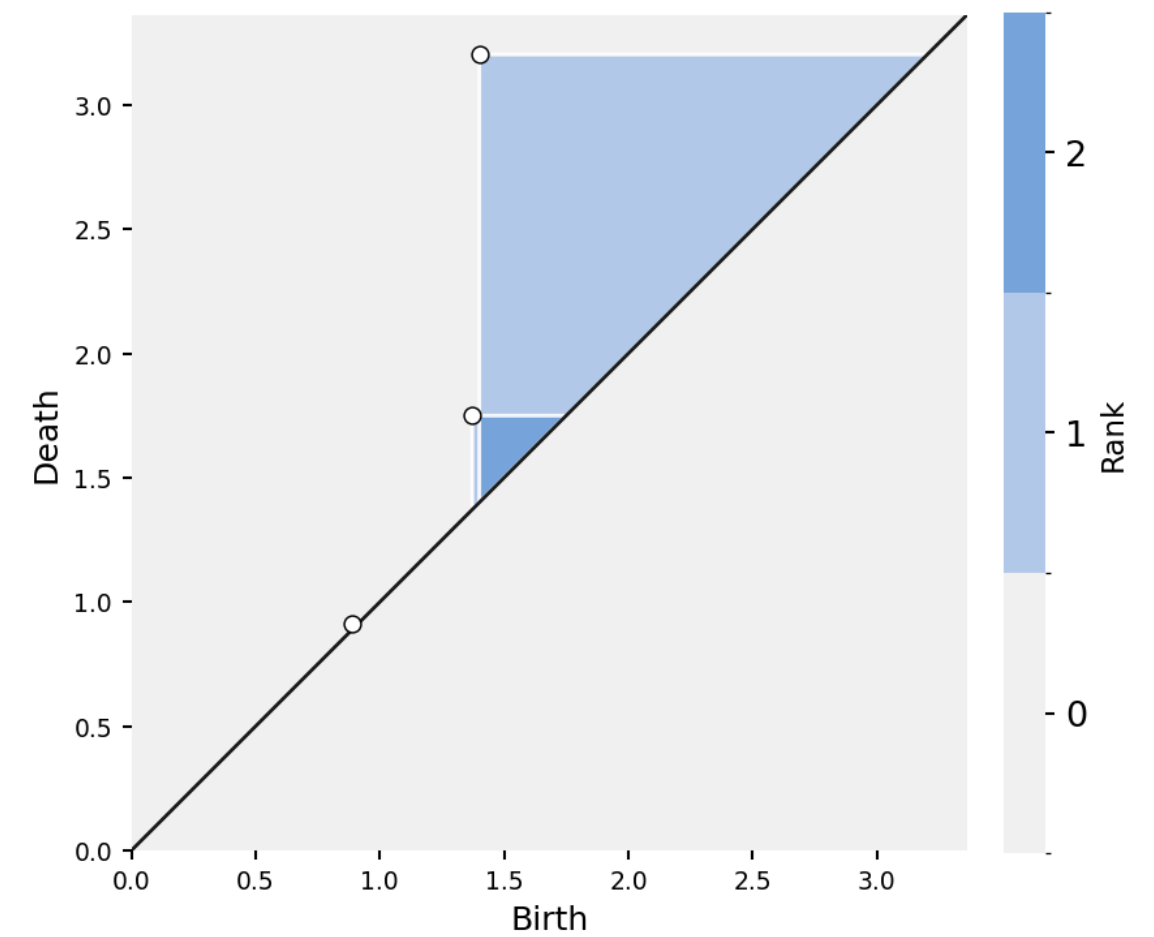
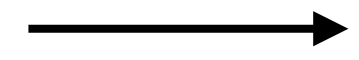
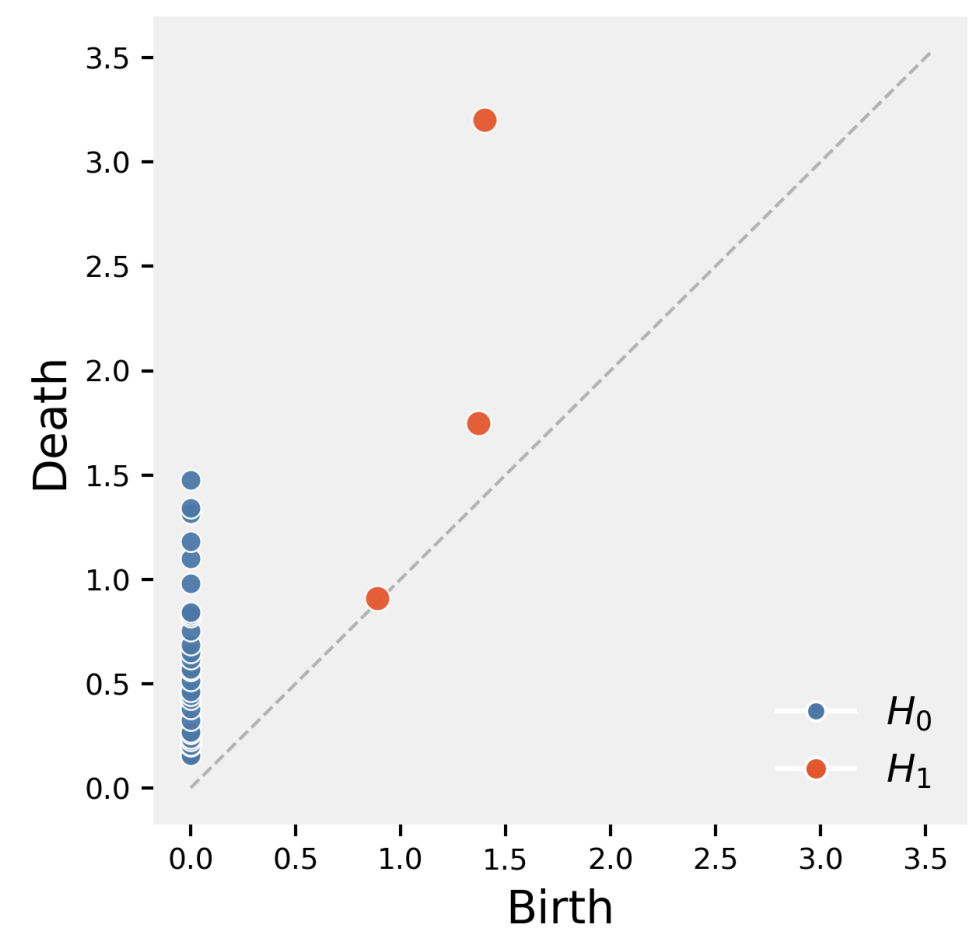
1. Change to *persistence coordinates* in the persistence diagram $(x, y) \mapsto (x, y - x)$
2. Let $\phi_u : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a differentiable probability distribution with mean $u = (u_x, u_y) \in \mathbb{R}^2$
3. Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a non-negative weighting function, that is 0 in the horizontal axis, continuous and piece-wise differentiable

$$\rho(x, y) = \sum_{u \in \text{PD}(X)} f(u) \phi_u(x, y)$$

Adams, Henry, Tegan Emerson, Michael Kirby, et al. "Persistence Images: A Stable Vector Representation of Persistent Homology." *Journal of Machine Learning Research* 18, no. 8 (2017): 1–35.

Persistent homology

Rank function



$$\text{rk}_{H_k(X_\bullet)} : \mathbb{R}^2 \rightarrow \mathbb{R}$$

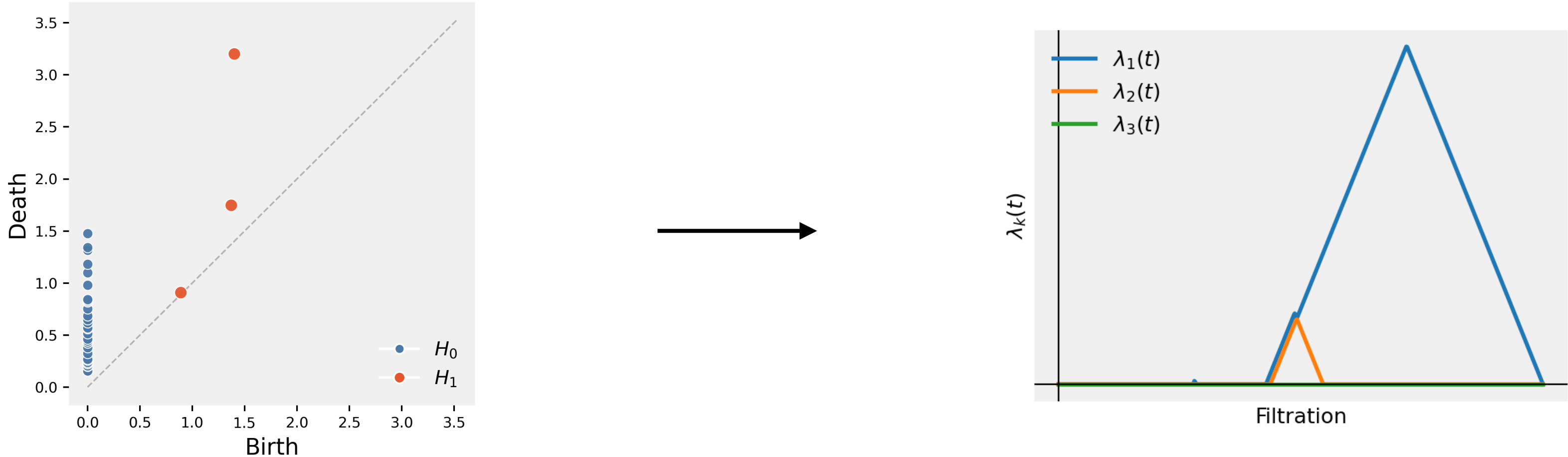
$$\text{rk}_{H_k(X_\bullet)}(x, y) = \begin{cases} \dim \text{im } H_k(X_\bullet)(x \leq y) & \text{if } x \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

This is a complete invariant!
Equivalent to barcodes and diagrams

Frosini, Patrizio. "A Distance for Similarity Classes of Submanifolds of a Euclidean Space." *Bulletin of the Australian Mathematical Society* 42, no. 3 (1990): 407–15.
 Frosini, Patrizio. "Discrete Computation of Size Functions." *Journal of Combinatorics, Information & System Sciences* 17, nos. 3–4 (1992): 232–50.

Persistent homology

Persistence landscapes



Change coordinates in the persistence diagram $(x, y) \mapsto (m, h) = \left(\frac{x + y}{2}, \frac{y - x}{2} \right)$

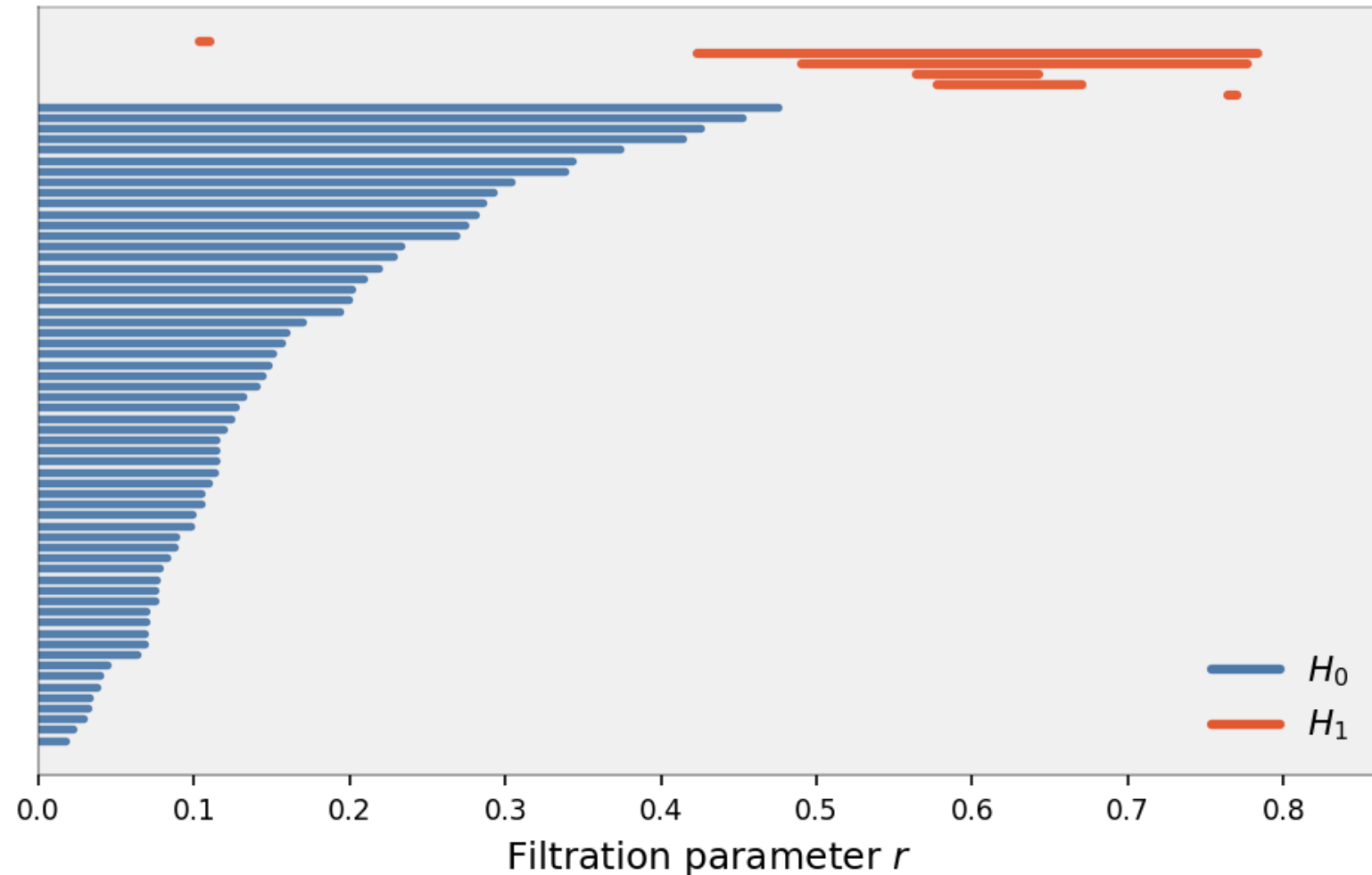
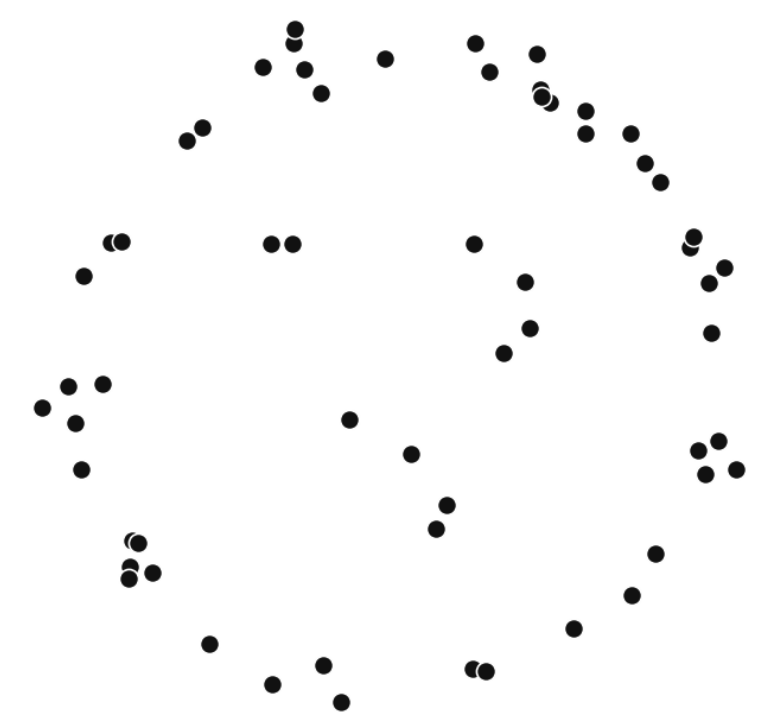
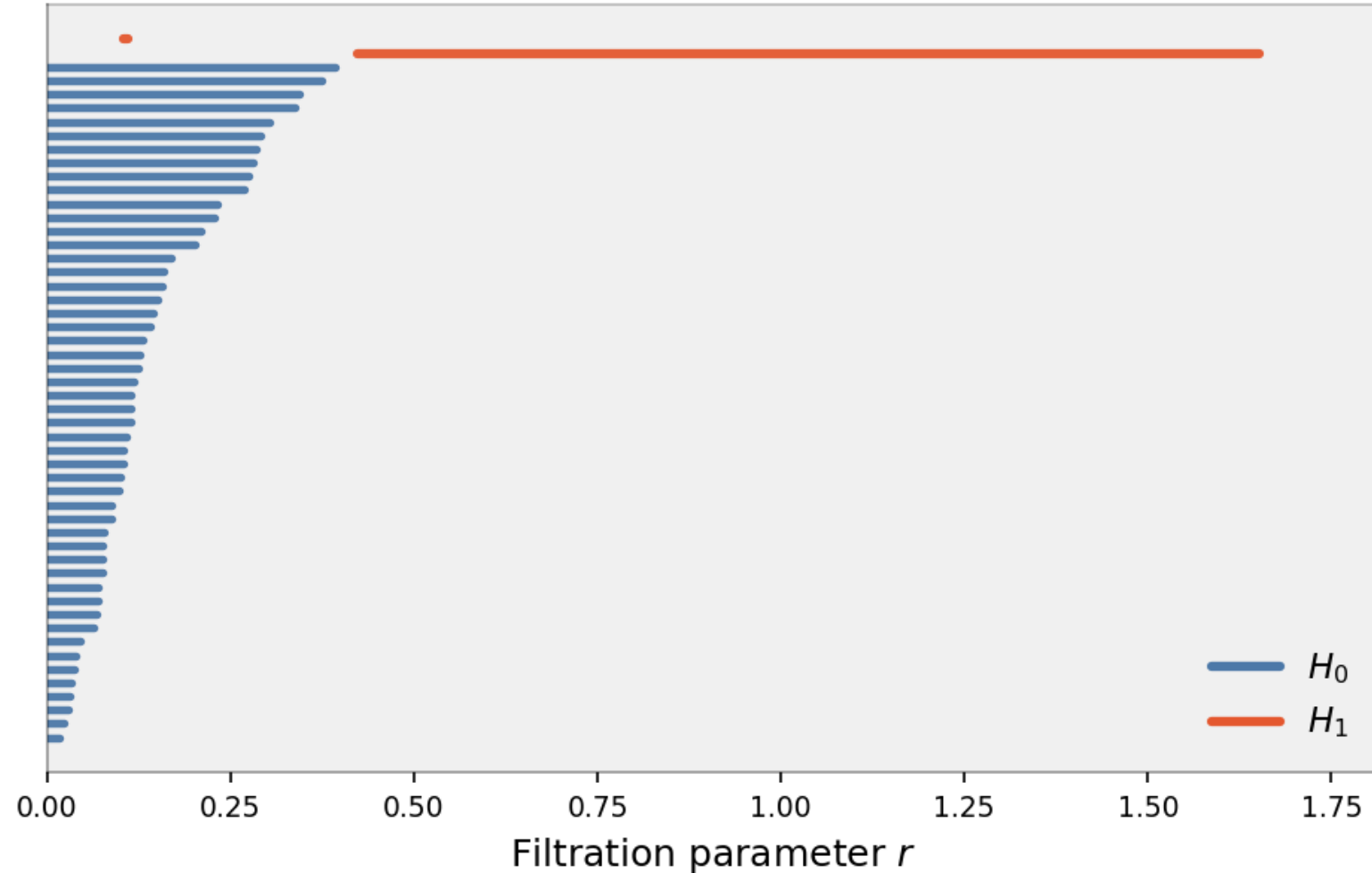
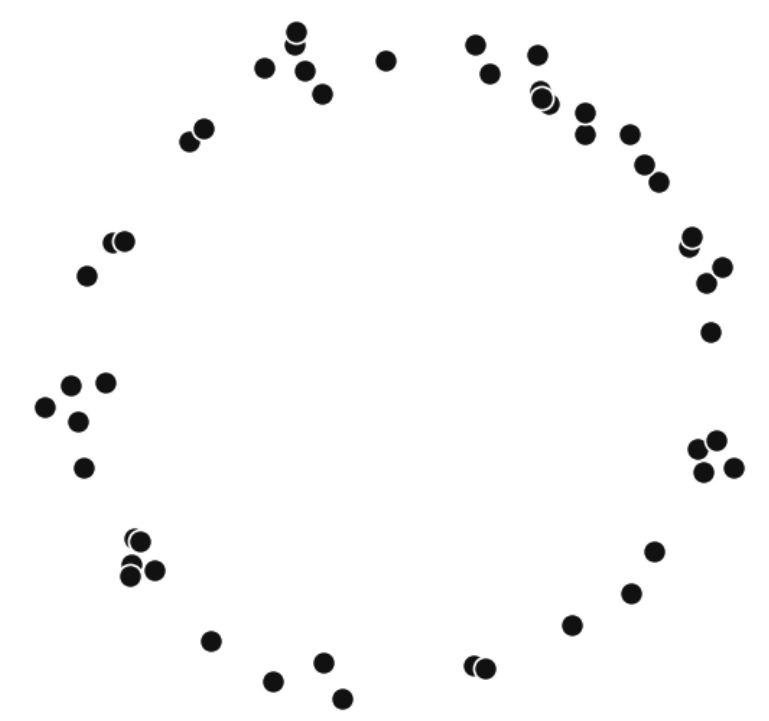
$$\lambda_k : \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$$

$$\lambda_k(x) = \sup\{m \geq 0 : \text{rk}(x - m, x + m) \geq k\}$$

Bubenik, Peter. "Statistical Topological Data Analysis Using Persistence Landscapes." *Journal of Machine Learning Research* 16, no. 3 (2015): 77–102.

Persistent homology

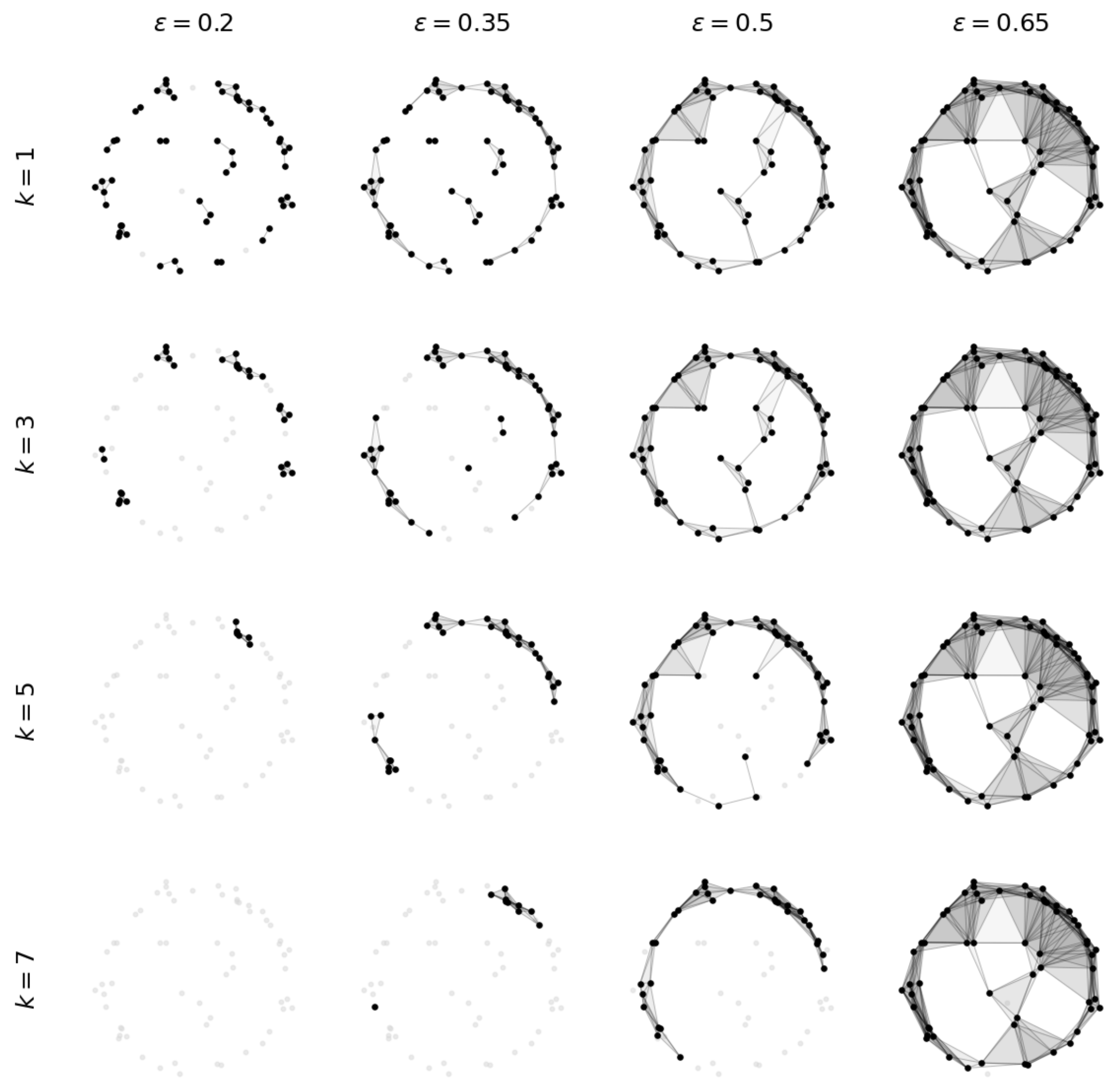
Not robust to outliers



Persistent homology

Not robust to outliers

Degree-Rips



Persistent homology

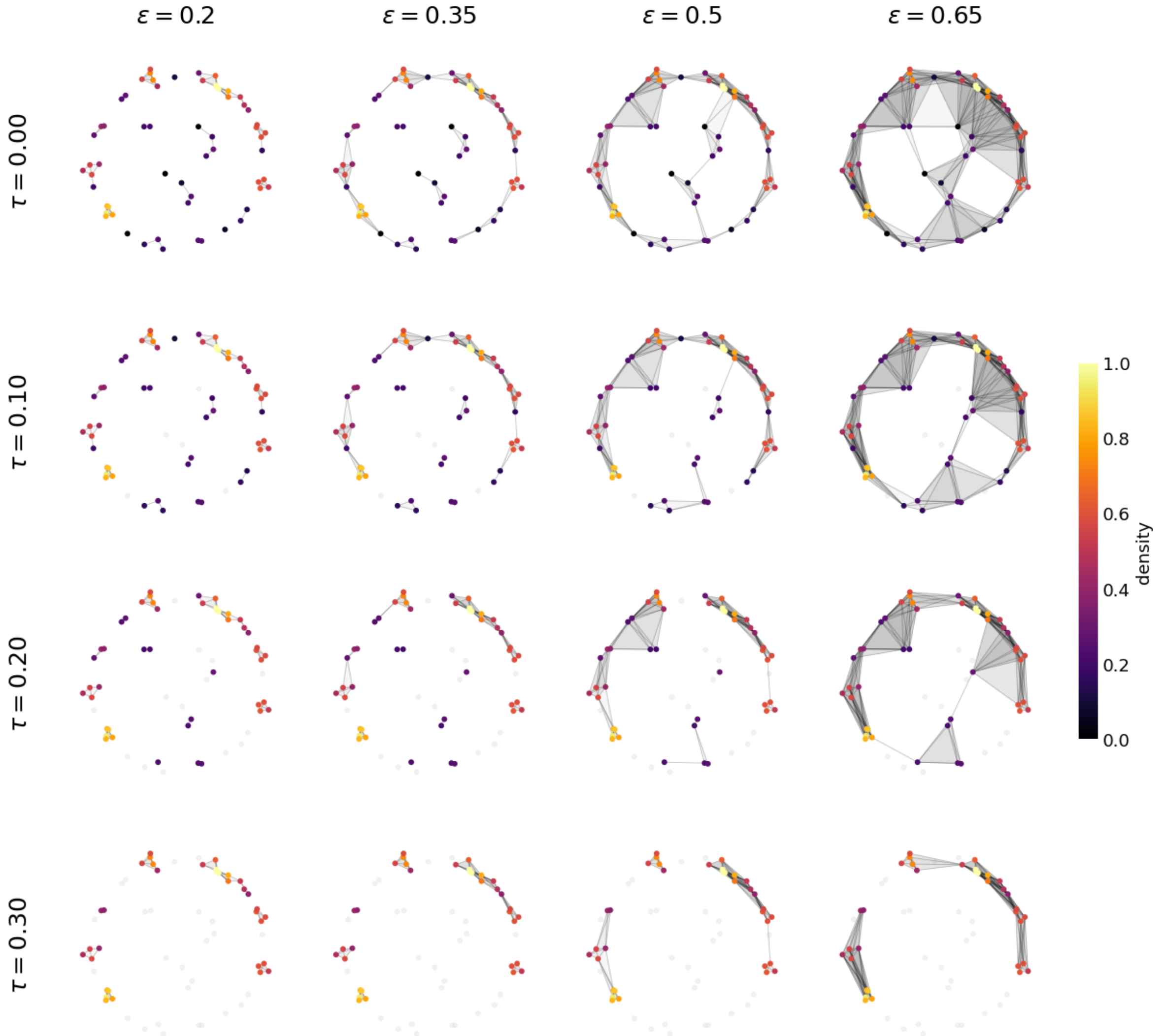
Not robust to outliers

Structure theorem:

$$H_k(X_{\bullet, \bullet}) \cong \bigoplus_{i \in I} M_i$$

Not necessarily intervals

Function-Rips



Some success stories

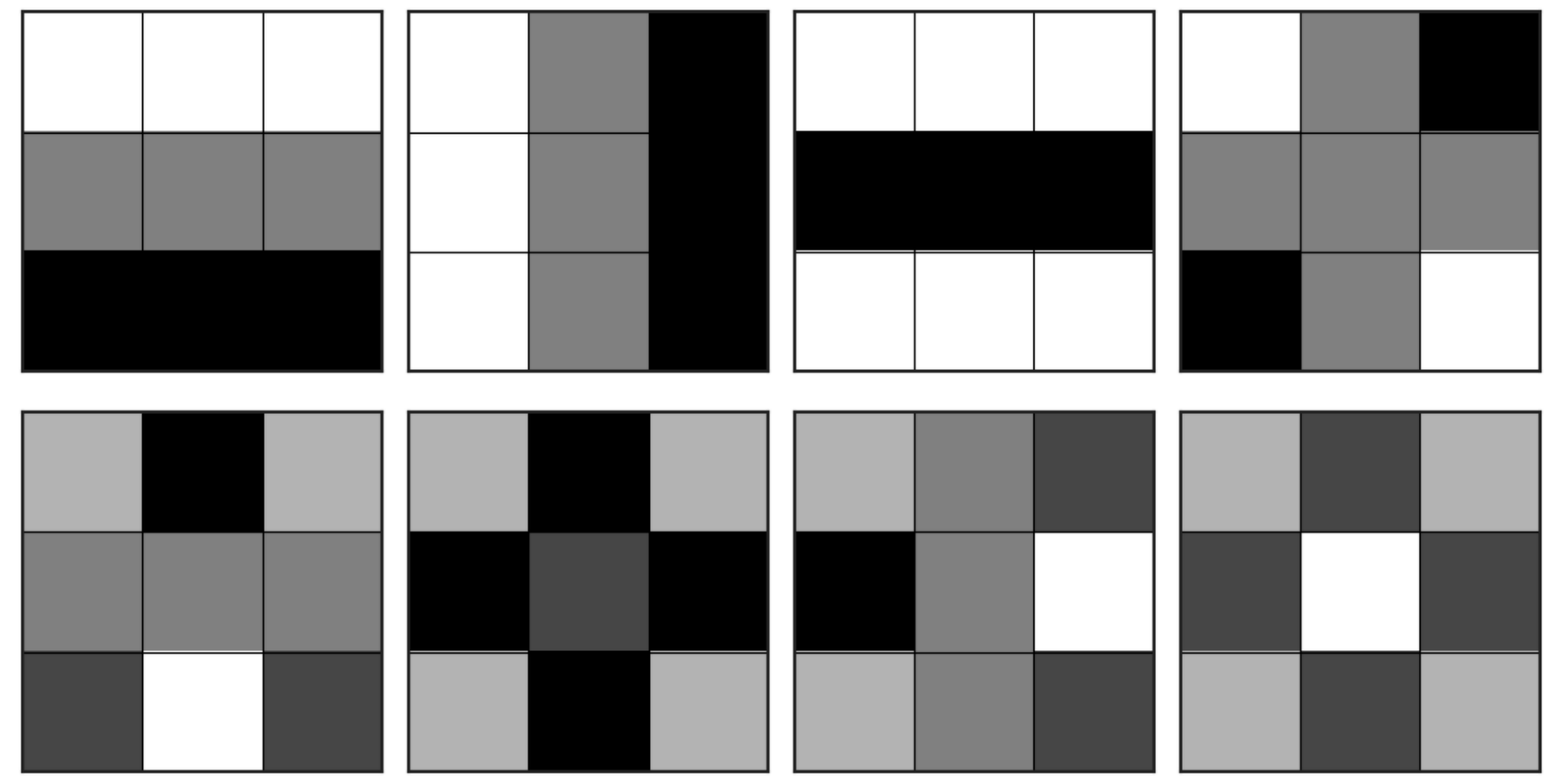
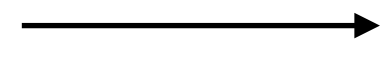
Session 2 — Topological data and how to analyze them

The shape of local image patches

The data



Hateren and van der Schaaf (1998)



Obtain 3×3 *high contrast* patches (points in \mathbb{R}^9)

Throw out patches with the same color everywhere

Some other preprocessing steps

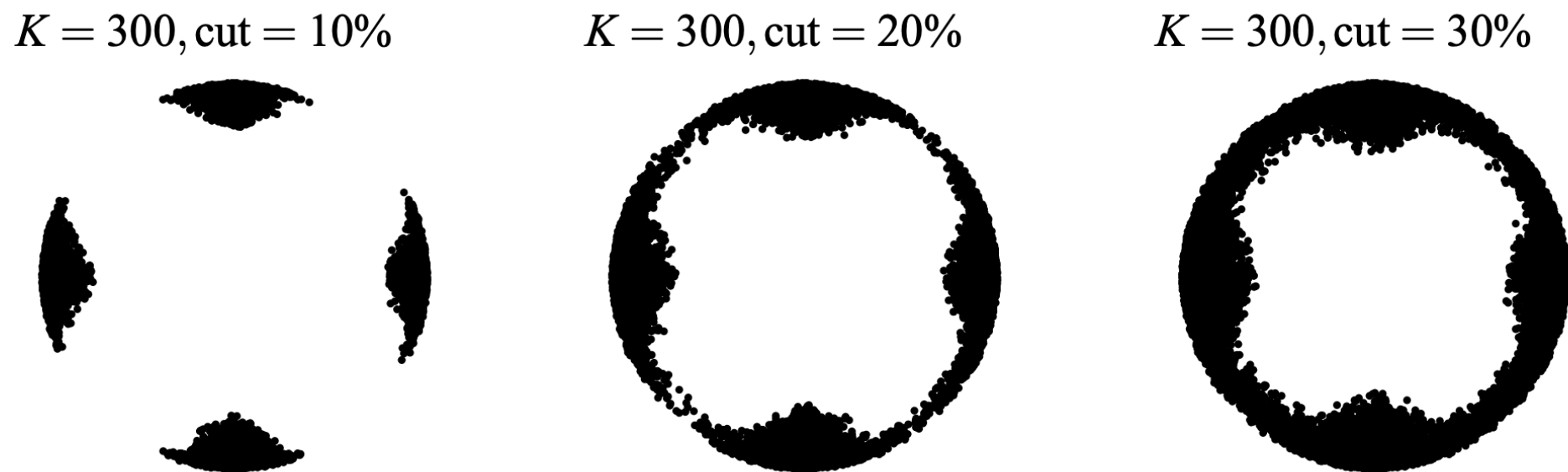
Select the *densest points*: the distance from x to its k **th nearest neighbor** ($\rho_k(x)$) is inversely proportional fo the density at x
 $X(k, p)$: the top p percent densest points according to kNN

Carlsson, Gunnar, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. "On the Local Behavior of Spaces of Natural Images." *International Journal of Computer Vision* 76, no. 1 (2008): 1–12.

The shape of local image patches

The densest patches

Projection onto the first two coordinates



H_1 persistent homology barcodes

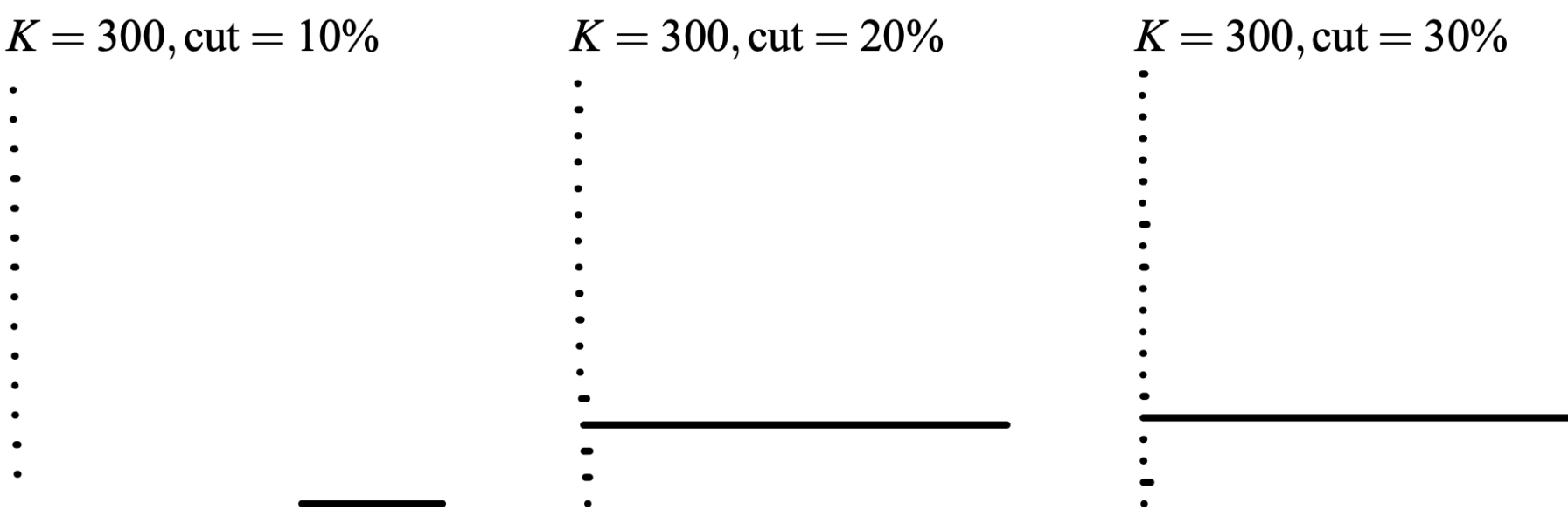
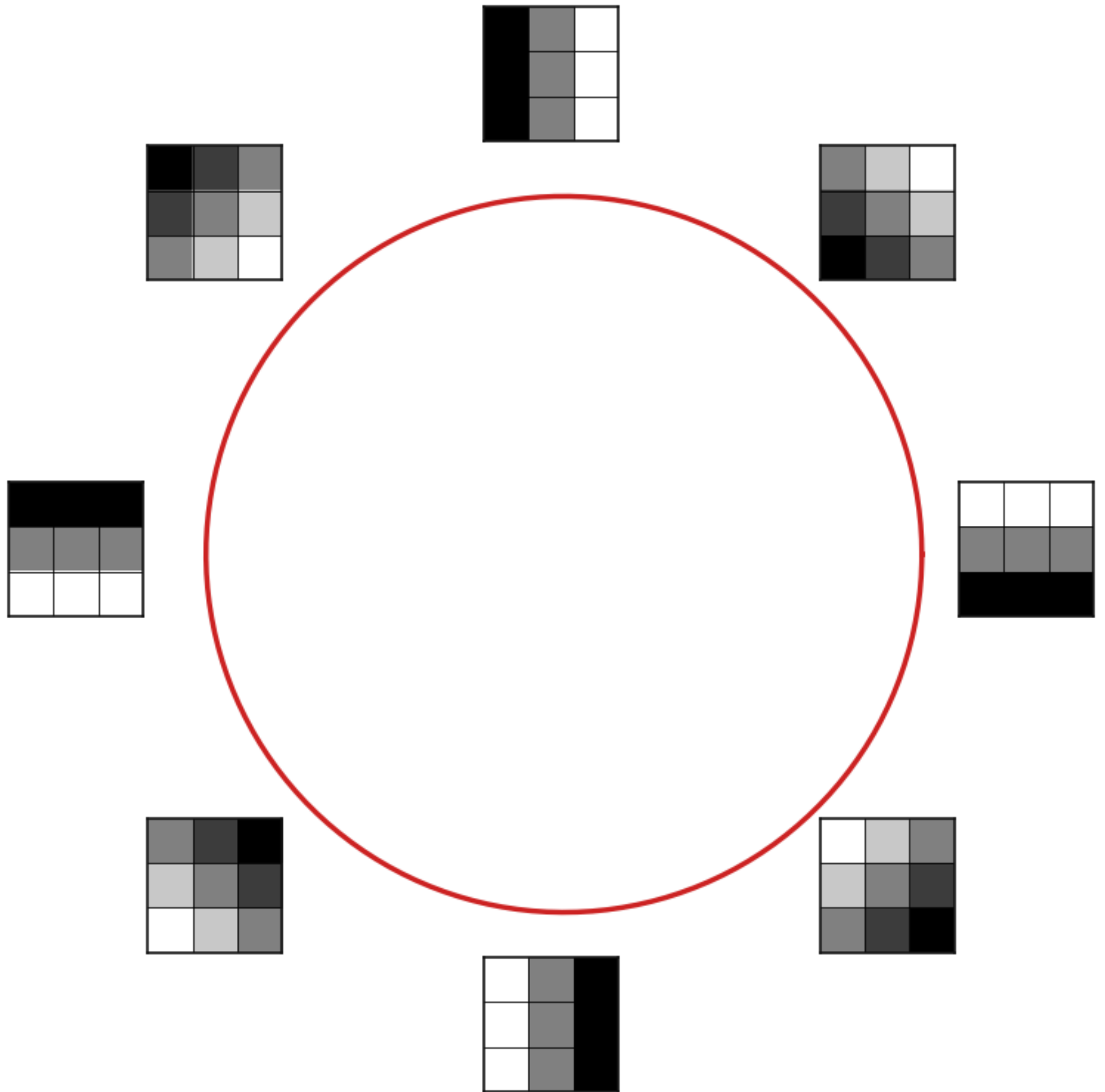


Figure from: de Silva and Carlsson (2004)



De Silva, Vin, and Gunnar Carlsson. "Topological Estimation Using Witness Complexes." *Proceedings of the First Eurographics Conference on Point-Based Graphics* (Goslar, DEU), SPBG'04, June 2, 2004, 157–66.

The shape of local image patches

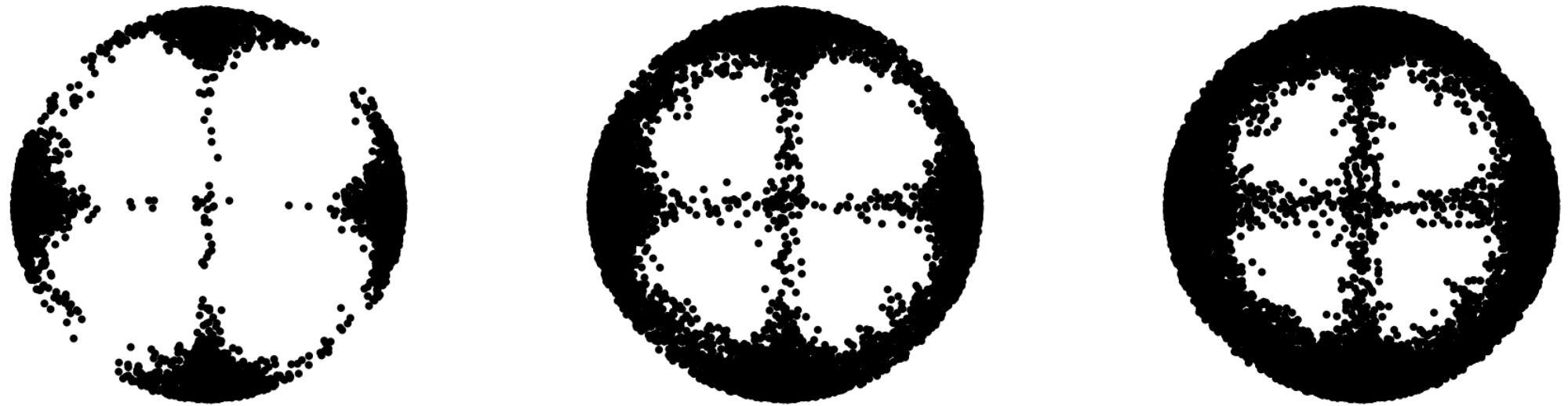
Intermediate estimate

Projection onto the first two coordinates

$K = 15, \text{cut} = 10\%$

$K = 15, \text{cut} = 20\%$

$K = 15, \text{cut} = 30\%$



H_1 persistent homology barcodes

$K = 15, \text{cut} = 10\%$

$K = 15, \text{cut} = 20\%$

$K = 15, \text{cut} = 30\%$

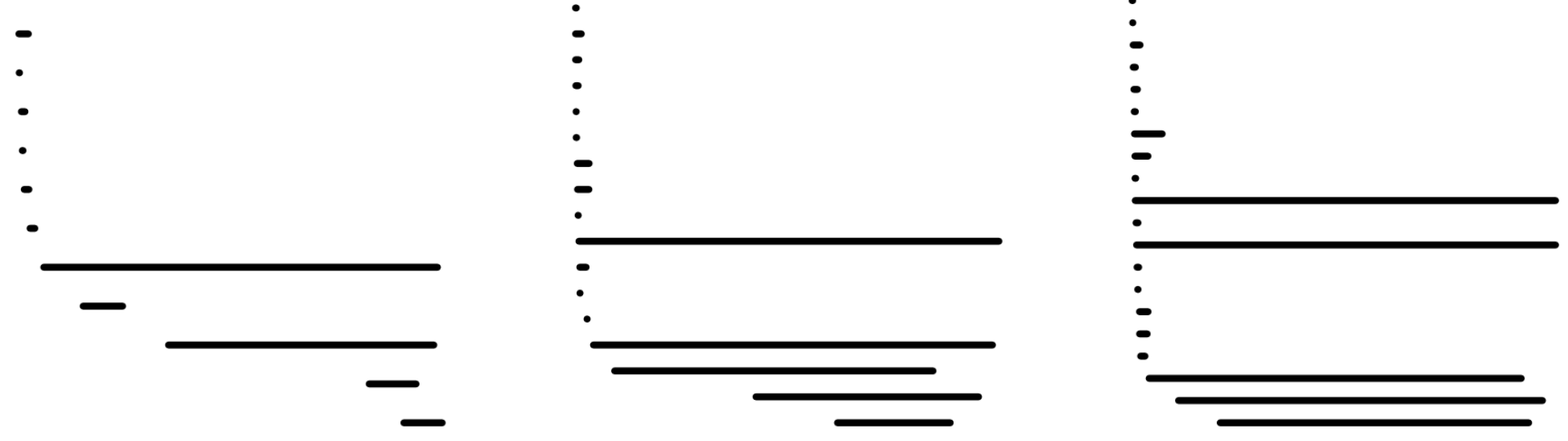
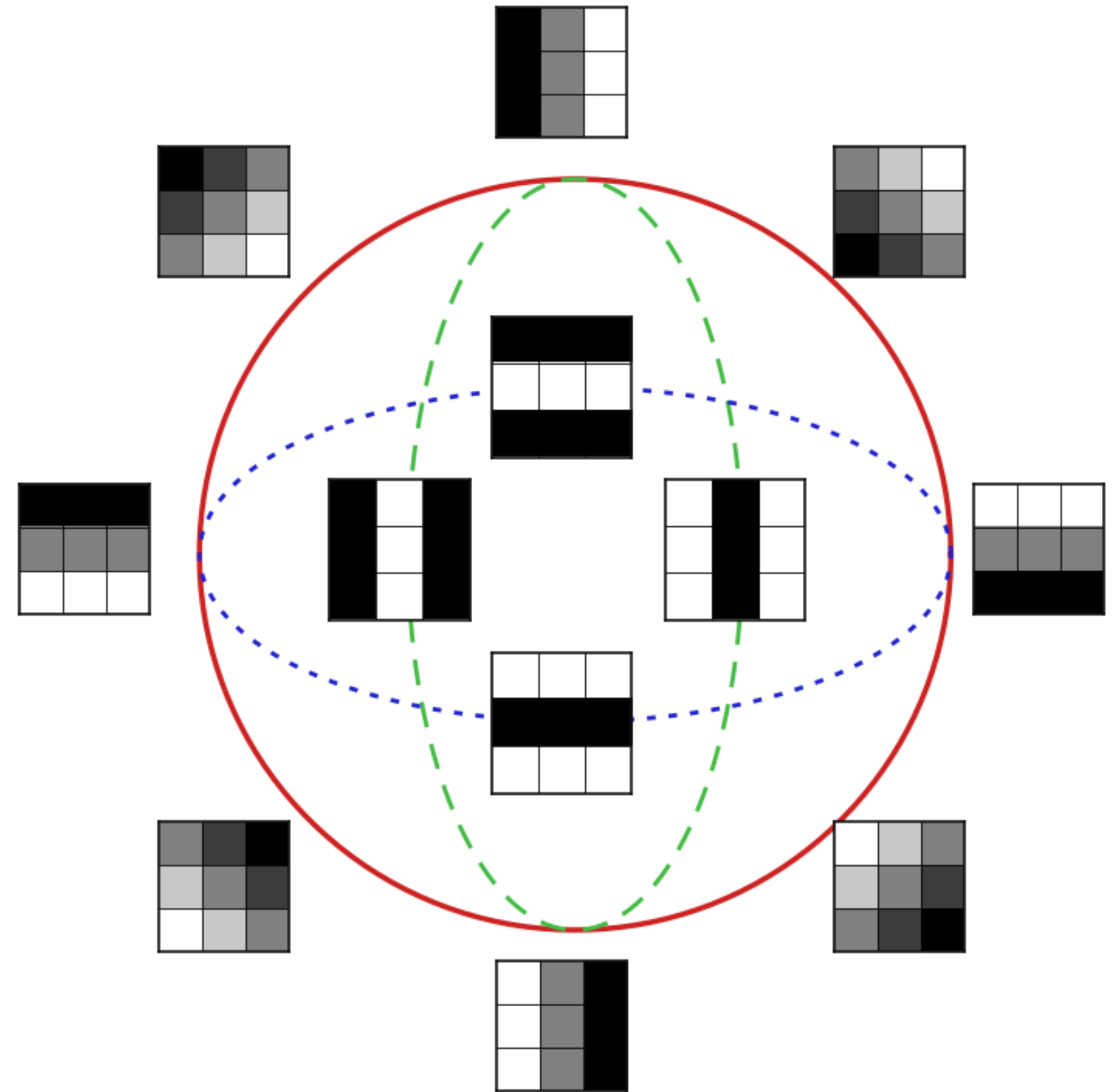


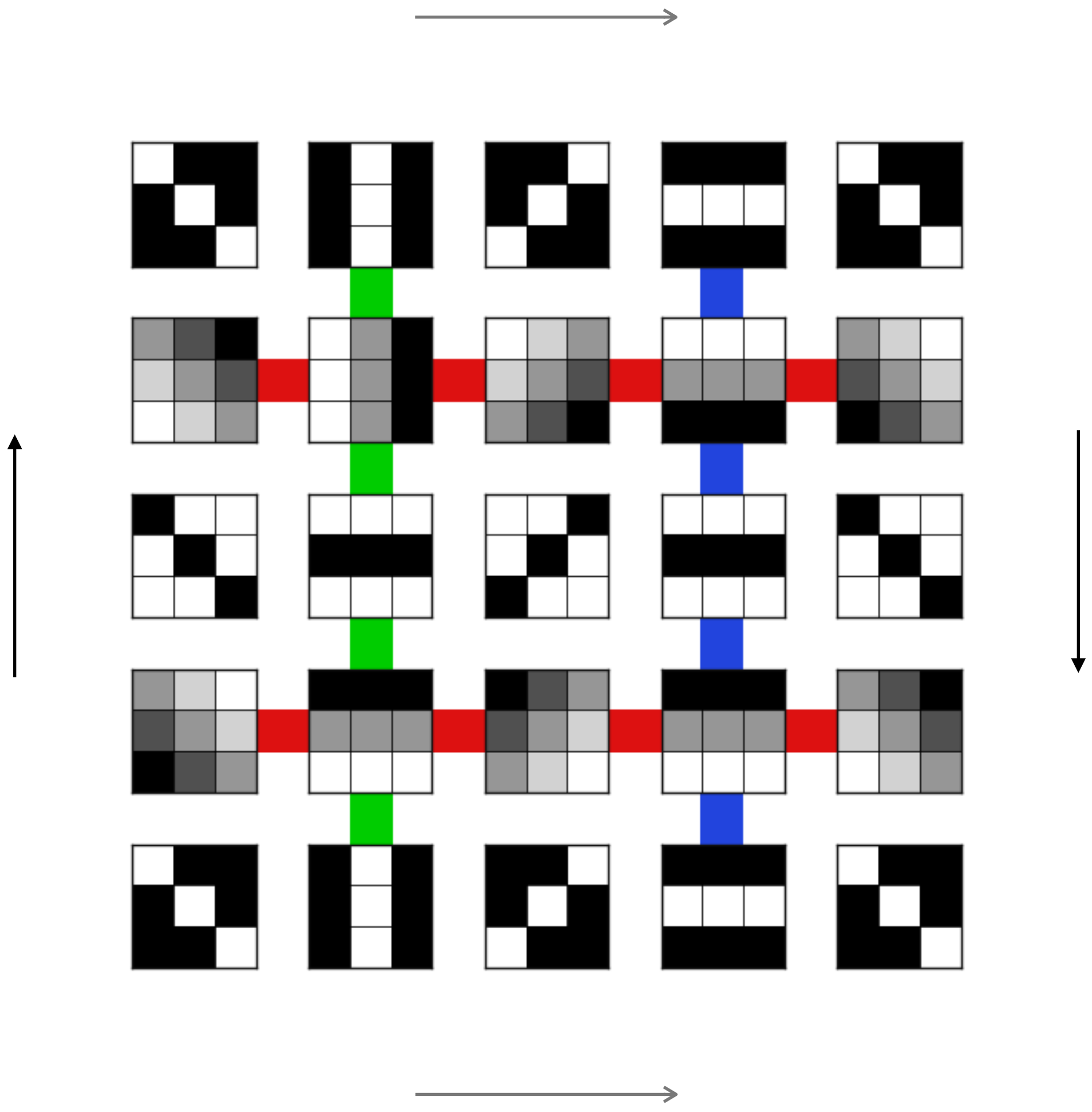
Figure from: de Silva and Carlsson (2004)



Carlsson, Gunnar, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. "On the Local Behavior of Spaces of Natural Images." *International Journal of Computer Vision* 76, no. 1 (2008): 1–12.

The shape of local image patches

Including more patches

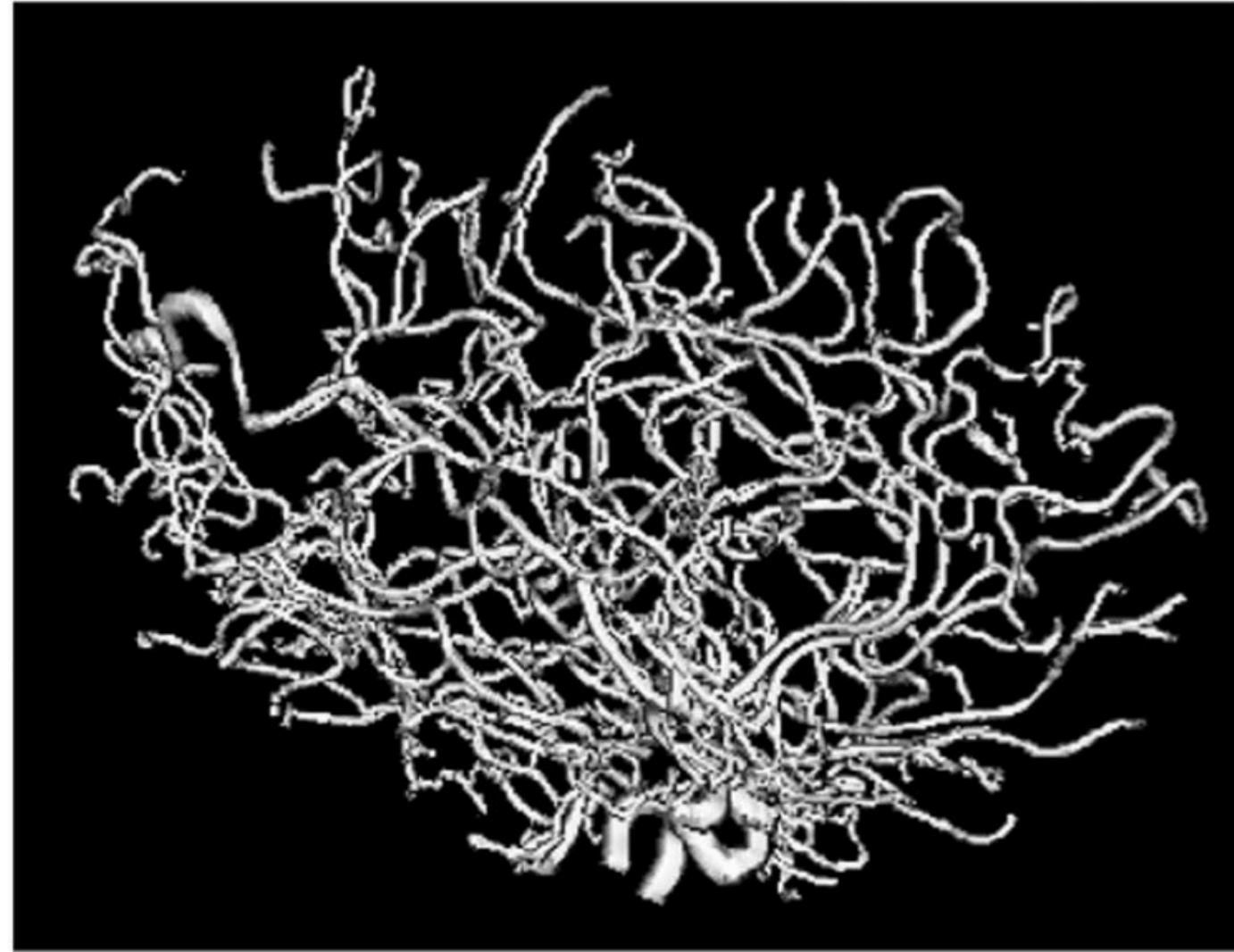


The patches form a Klein bottle!

De Silva, Vin, and Gunnar Carlsson. "Topological Estimation Using Witness Complexes." *Proceedings of the First Eurographics Conference on Point-Based Graphics (Goslar, DEU), SPBG'04, June 2, 2004, 157-66.*

Brain Artery Trees

The data



Tube-tracking vessel segmentation algorithm applied to 3-dimensional Magnetic Resonance Angiography (MRA)

98 trees, subject ages form 18 to 72

Compute persistence barcodes directly on these 3D structures

Select the 100 highest persistences in dimensions 0 and 1

$$(p_1, \dots, p_{100}), \quad (q_1, \dots, q_{100})$$



Figures from: Bendich, Paul, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. "Persistent Homology Analysis of Brain Artery Trees." *The Annals of Applied Statistics* 10, no. 1 (2016): 198–218.

Brain Artery Trees

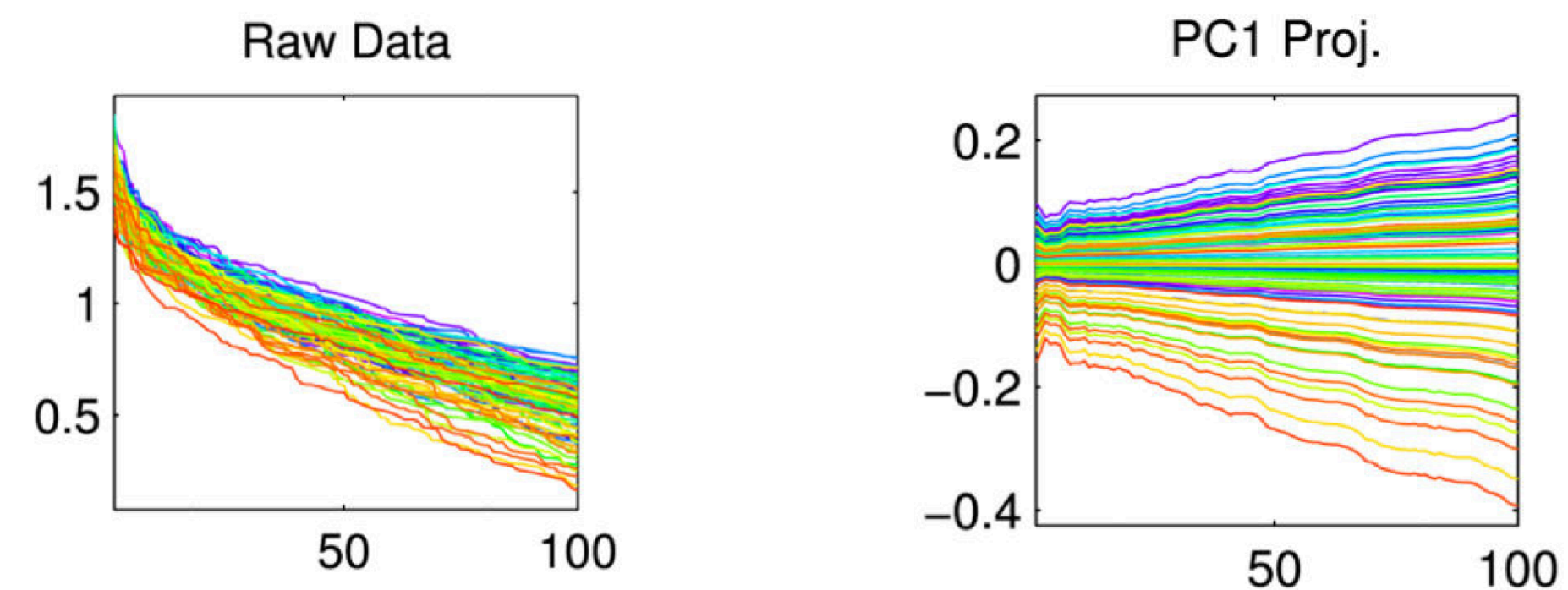
The data

- **Age effects:** strong correlations between these features and the ages of the subject

Checked that these correlations were not due solely to the total artery length, already proved to decrease with age

- **Sex effects:**

- Compute the means for the 1-dimensional homology vectors (q_1, \dots, q_{100}) of men and women
- Compute MSE between the means
- Perform a random permutation test: p -value of 0.03 to reject the null hypothesis that these vectors have the same distribution



Figures from: Bendich, Paul, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. "Persistent Homology Analysis of Brain Artery Trees." *The Annals of Applied Statistics* 10, no. 1 (2016): 198–218.

(A bit of) Hodge Theory

Session 2 — Topological data and how to analyze them

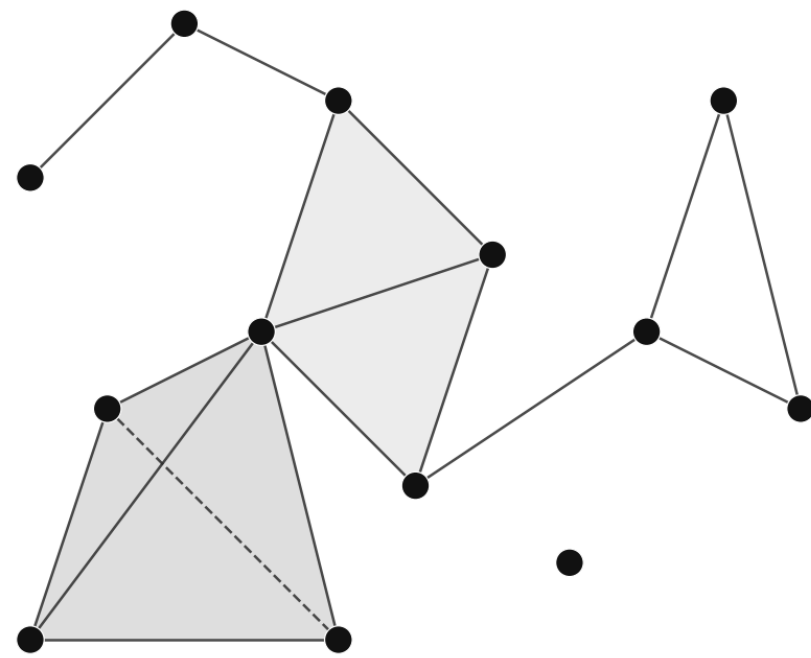
Hodge Theory

Combinatorial Hodge Laplacians

So far we have only used the *vector space* structure of the chain spaces

We can also endow these spaces with *inner products*

Example: weighted simplicial complexes (X, w) with $w : X \rightarrow \mathbb{R}_{>0}$



$$C_k(X, \mathbb{R}) = \text{span}\{(i_0, \dots, i_k) : i_0 < \dots < i_k\}$$

$$\langle \sigma, \tau \rangle_{C_k} = \begin{cases} w(\sigma) & \text{if } \sigma = \tau \\ 0 & \text{otherwise} \end{cases}$$

The k -simplices form an orthogonal basis of $C_k(X)$ with norms prescribed by their weights

Adjoint boundary maps:

$$\langle \sigma, \partial_k \tau \rangle_{C_{k-1}} = \langle \partial_k^* \sigma, \tau \rangle_{C_k}, \quad \forall \sigma \in C_{k-1}(X), \tau \in C_k(X)$$

Hodge Laplacian operator:

$$C_{k+1}(X) \begin{matrix} \xrightarrow{\partial_{k+1}} \\ \xleftarrow{\partial_{k+1}^*} \end{matrix} C_k(X) \begin{matrix} \xrightarrow{\partial_k} \\ \xleftarrow{\partial_k^*} \end{matrix} C_{k-1}(X)$$

$$L_k = \partial_{k+1} \circ \partial_{k+1}^* + \partial_k^* \circ \partial_k = L_k^{\text{up}} + L_k^{\text{down}}$$

Hodge Theory

Hodge decomposition

Hodge Laplacian operator:

$$C_{k+1}(X) \begin{matrix} \xrightarrow{\partial_{k+1}} \\ \xleftarrow{\partial_{k+1}^*} \end{matrix} C_k(X) \begin{matrix} \xrightarrow{\partial_k} \\ \xleftarrow{\partial_k^*} \end{matrix} C_{k-1}(X)$$

$$L_k = \partial_{k+1} \circ \partial_{k+1}^* + \partial_k^* \circ \partial_k = L_k^{\text{up}} + L_k^{\text{down}}$$

Theorem: the chain space has the following orthogonal decomposition

$$C_k(X) = \text{im } \partial_{k+1} \oplus \text{ker } L_k \oplus \text{im } \partial_k^*$$

Hodge decomposition

Harmonic representatives

Theorem: $\text{ker } L_k = \text{ker } \partial_k \cap \text{ker } \partial_{k+1}^* \cong H_k(X, \mathbb{R})$

Example: graph case $G = (V, E)$

$$0 \begin{matrix} \xrightarrow{0} \\ \xleftarrow{0} \end{matrix} C_1(G) \begin{matrix} \xrightarrow{\partial_1} \\ \xleftarrow{\partial_1^*} \end{matrix} C_0(X)$$

$$L_0 = L_0^{\text{up}} = \partial_1 \circ \partial_1^*$$

$$L_1 = L_1^{\text{down}} = \partial_1^* \circ \partial_1$$

ℓ^2 inner product

$$\partial_1 = B_1, \quad \partial_1^* = B_1^\top$$

$$L_0 = B_1 B_1^\top$$

$$L_1 = B_1^\top B_1$$

Graph Laplacian

Lim, Lek-Heng. "Hodge Laplacians on Graphs." *SIAM Review* 62, no. 3 (2020): 685–715.

To sum up

Today we have learnt how to...

- How to describe graphs through matrices
- How to add higher-order interactions via simplicial complexes
- How to measure topological features using homology
- How to implement persistent homology in point cloud data
- Some case-studies of success for PH

We have not seen

- How to compute barcodes and other invariants
- Cohomology
- More Hodge Theory
- Euler transform, PH transform, magnitude...

Thank you for your attention!



AIDOS LAB
AI FOR DATA-ORIENTED SCIENCE

