Application of topological data analysis in biomechanics

2020. 09. 04

Sangman Jung , Kyungsoo Kim
Kyung Hee University

Contents

- Introduction
- Topological data analysis (TDA)
- Method
- Application
- Summary

Introduction

- **What is** the topological data analysis (**TDA**):
  - The recently proposed **data analysis method in computational topology**.
  - It requires **algebraic topology** and **programming skills**

- **Why we use** this:
  - This method **finds topological structure** and extracts some topological features.
  - It can be **effective in finding hidden features of raw data** that cannot be found with conventional data analysis methods.

- **Purpose**:
  - To obtain topological features of biomechanical data **using TDA**.
  - **To quantify similarities or dissimilarities between subjects**.

Topological data analysis (TDA)

- Basic flow

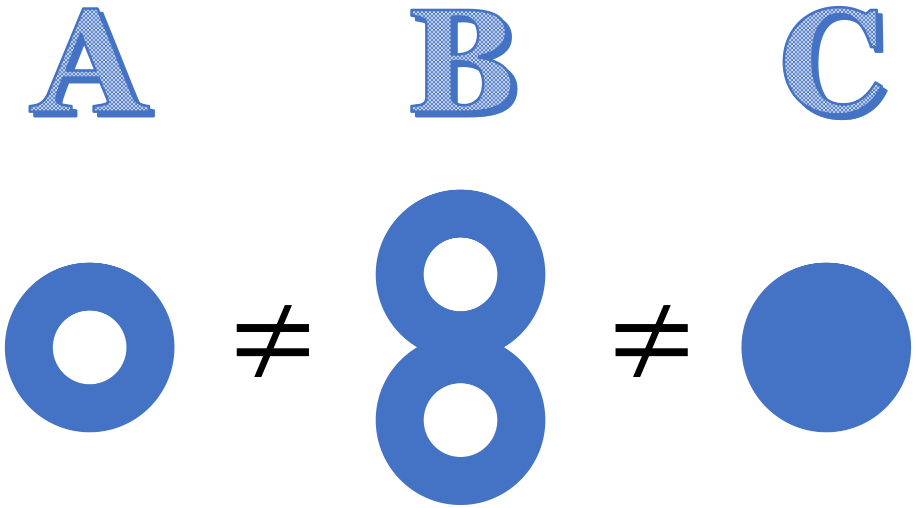
Raw dataset

↓

Find topological structures

↓

Extract topological features



“ Two shapes can be distinguished by examining their **holes**. ”

reference : Walker, B. (2008, May). Using persistent homology to recover spatial information from encounter traces. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing* (pp. 371-380).

Topological data analysis (TDA)

- Basic flow

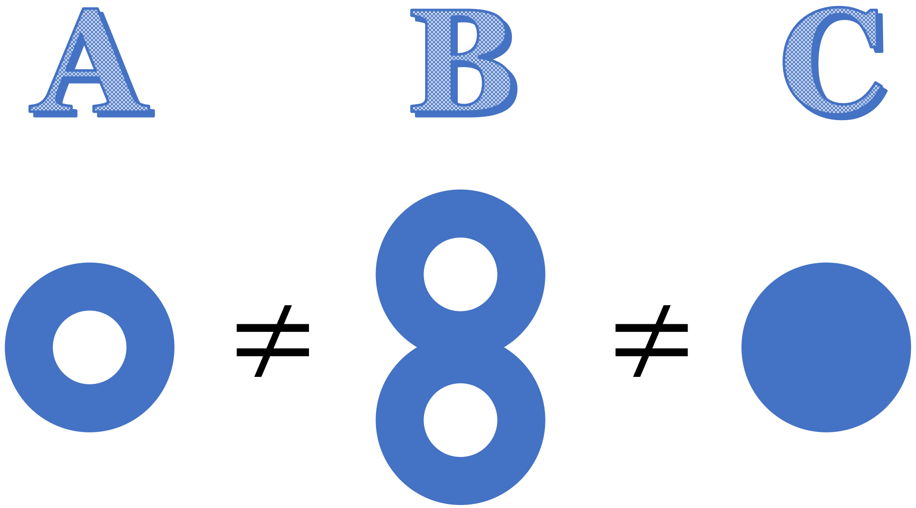
Raw dataset

↓

Find topological structures

↓

Extract topological features



“ Two shapes can be distinguished by examining their **holes**. ”

“ The number of k-dimensional holes in the space. ”

Betti Number

$\boldsymbol{\beta_k}$

reference : Walker, B. (2008, May). Using persistent homology to recover spatial information from encounter traces. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing* (pp. 371-380).

Topological data analysis (TDA)

- Betti Number $\boldsymbol{\beta_k}$:

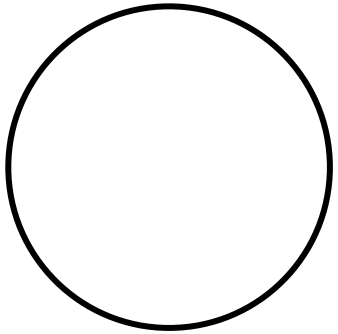
  (intuitively)

$\boldsymbol{\beta_0}$: the number of **connected components** (connectivity)

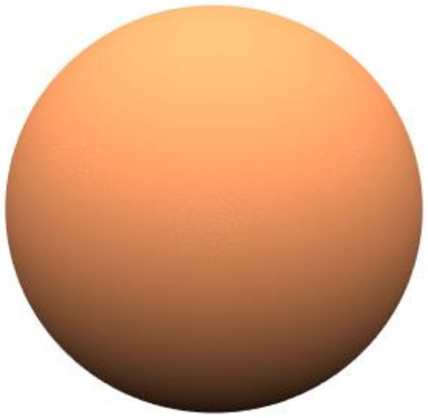
$\boldsymbol{\beta_1}$: the number of 1-dimensional **holes** or **loops**

$\boldsymbol{\beta_2}$: the number of enclosed solid **voids** (2-dimensional voids)

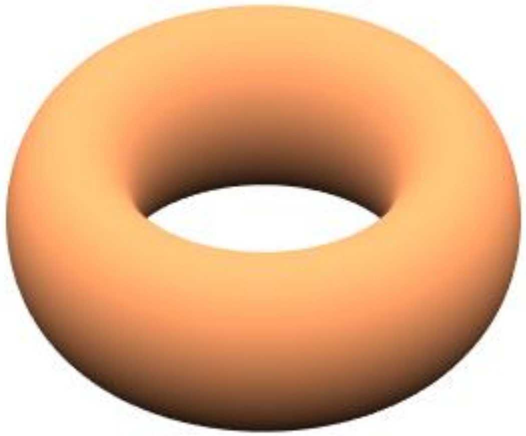
$(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\mathbf{1}, \mathbf{0}, \mathbf{0})$



$(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\mathbf{1}, \mathbf{1}, \mathbf{0})$



$(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\mathbf{1}, \mathbf{1}, \mathbf{1})$

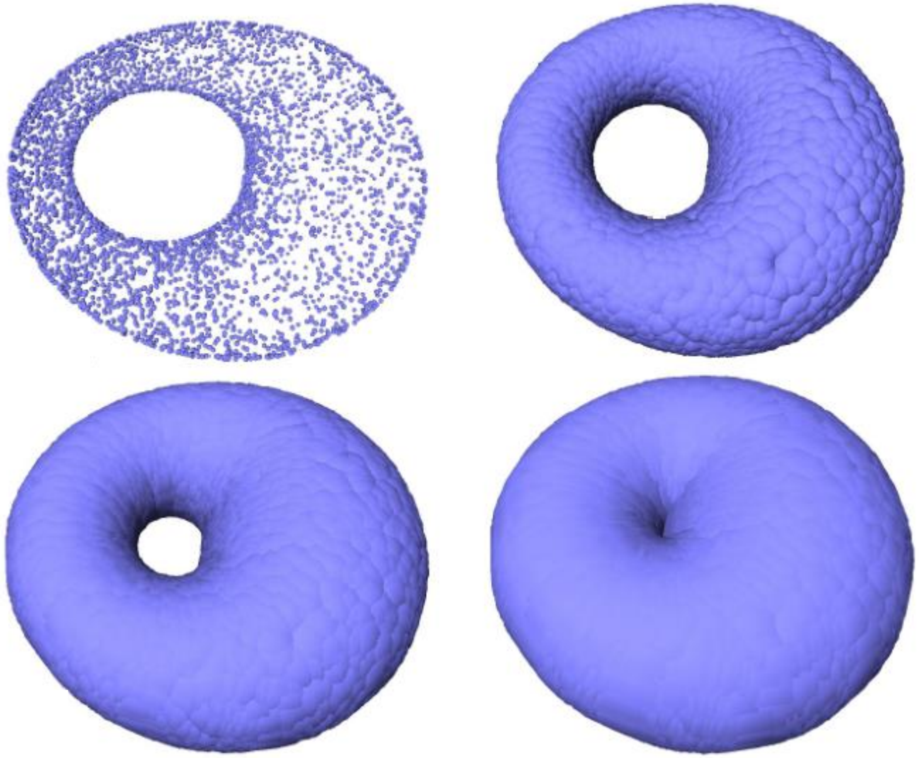


$(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\mathbf{1}, \mathbf{2}, \mathbf{1})$

reference : Moon, C., Giansiracusa, N., & Lazar, N. A. (2018). Persistence terrace for topological inference of point cloud data. *Journal of Computational and Graphical Statistics*, *27*(3), 576-586.

Topological data analysis (TDA)

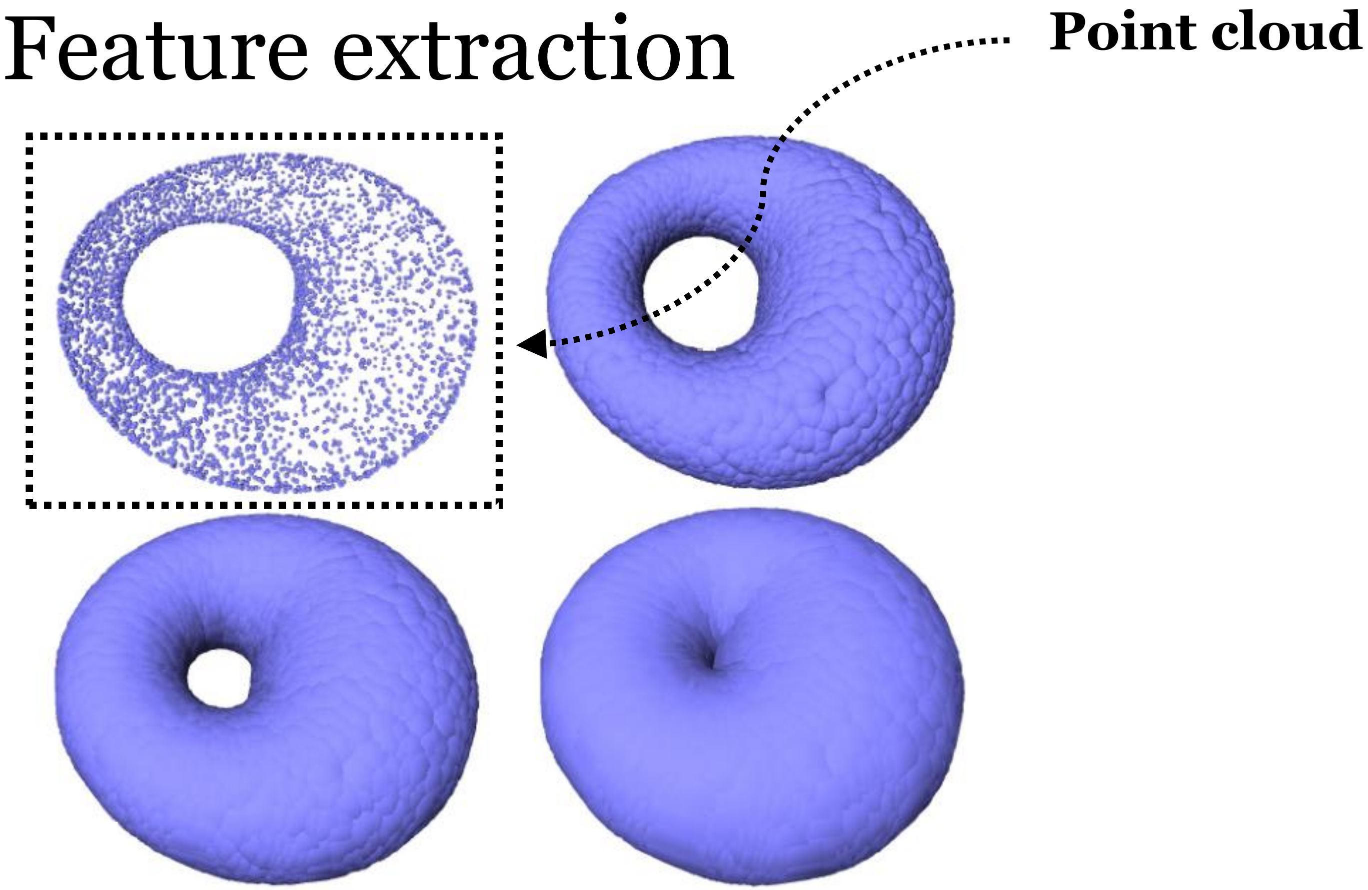
- Feature extraction



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

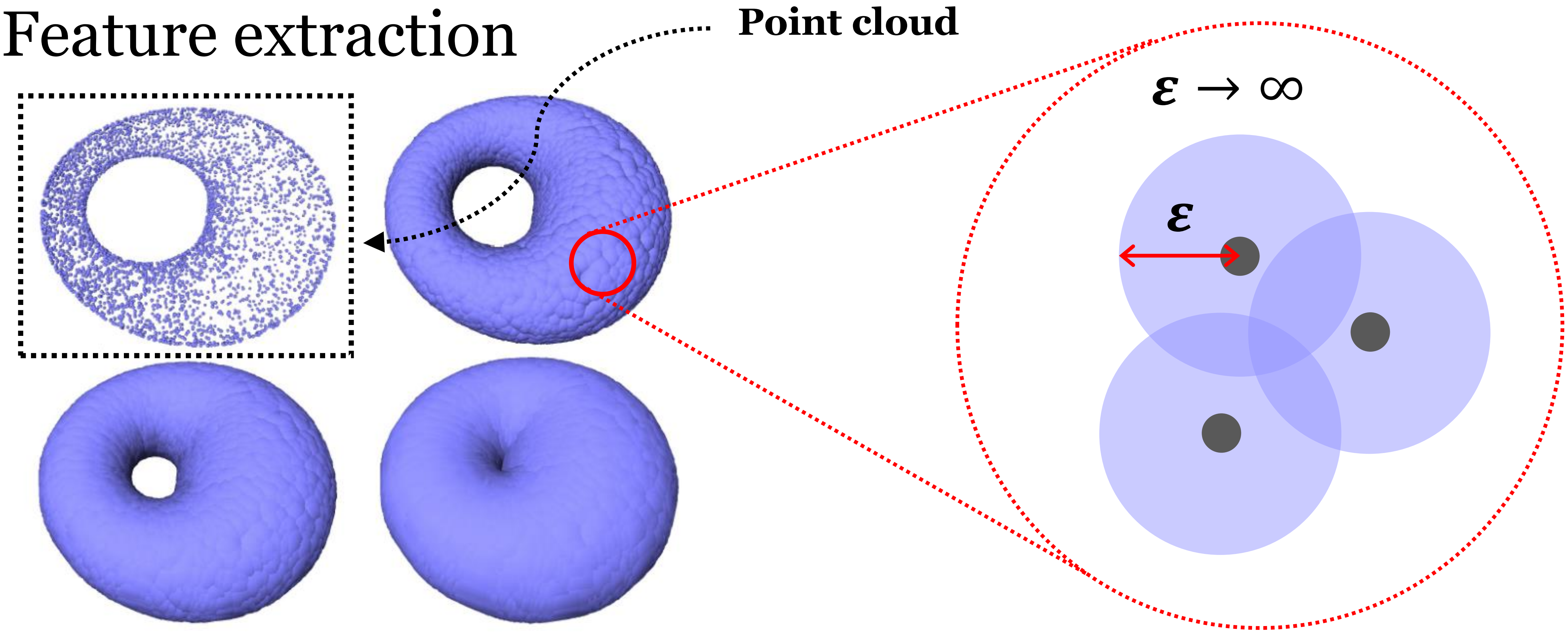
- Feature extraction



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

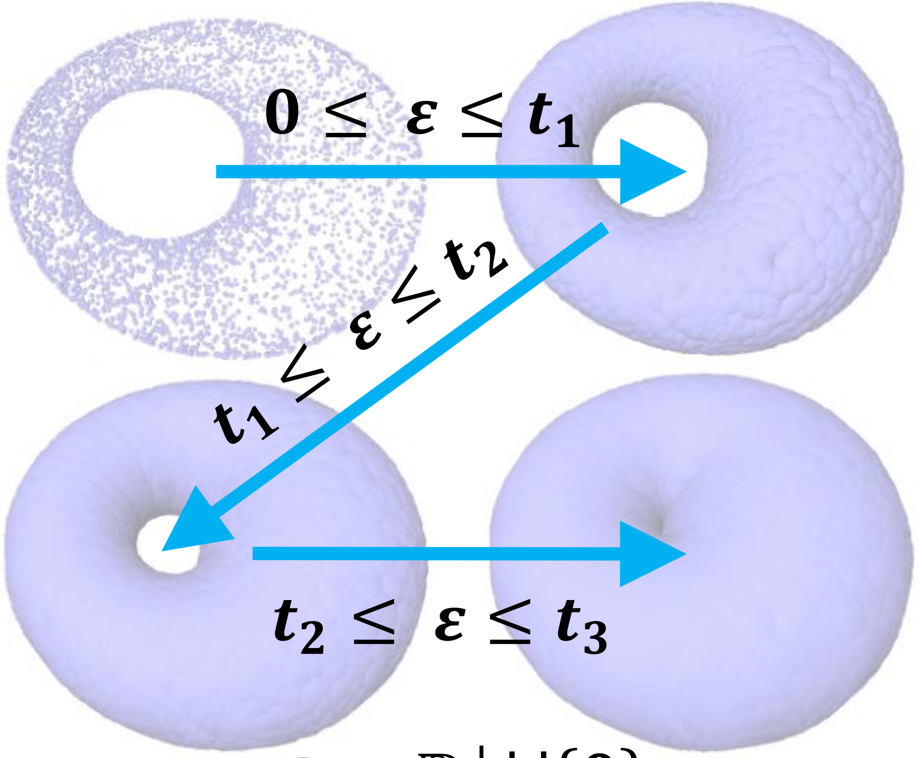
- Feature extraction



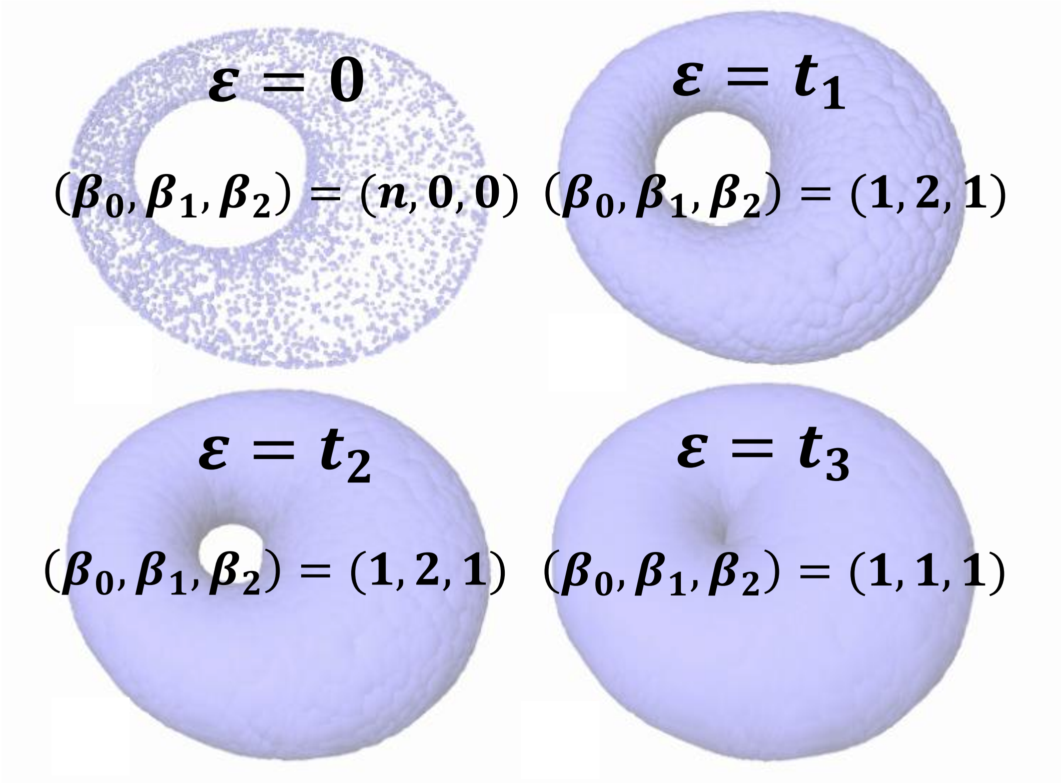
Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

- Feature extraction



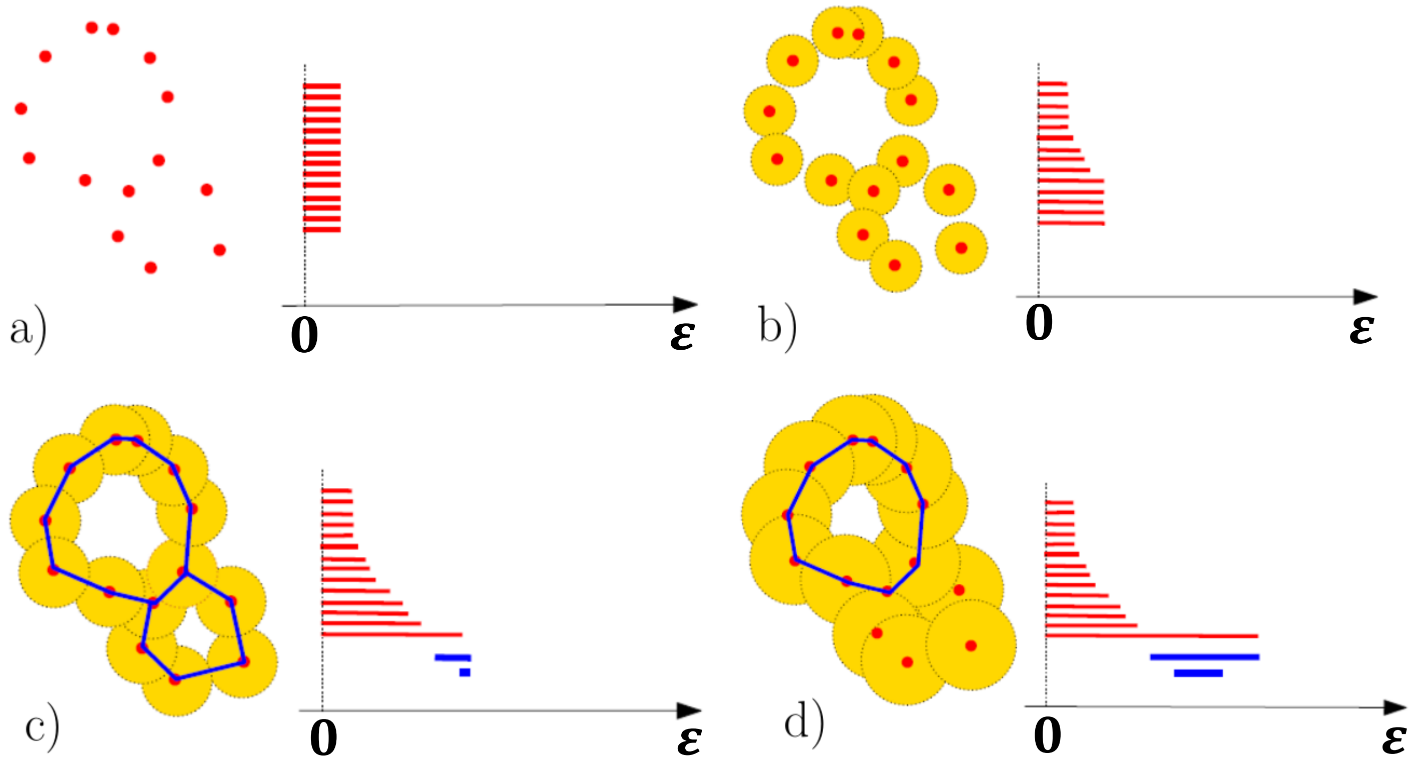
$t_i \in \mathbb{R}^+ \cup \{0\}$



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

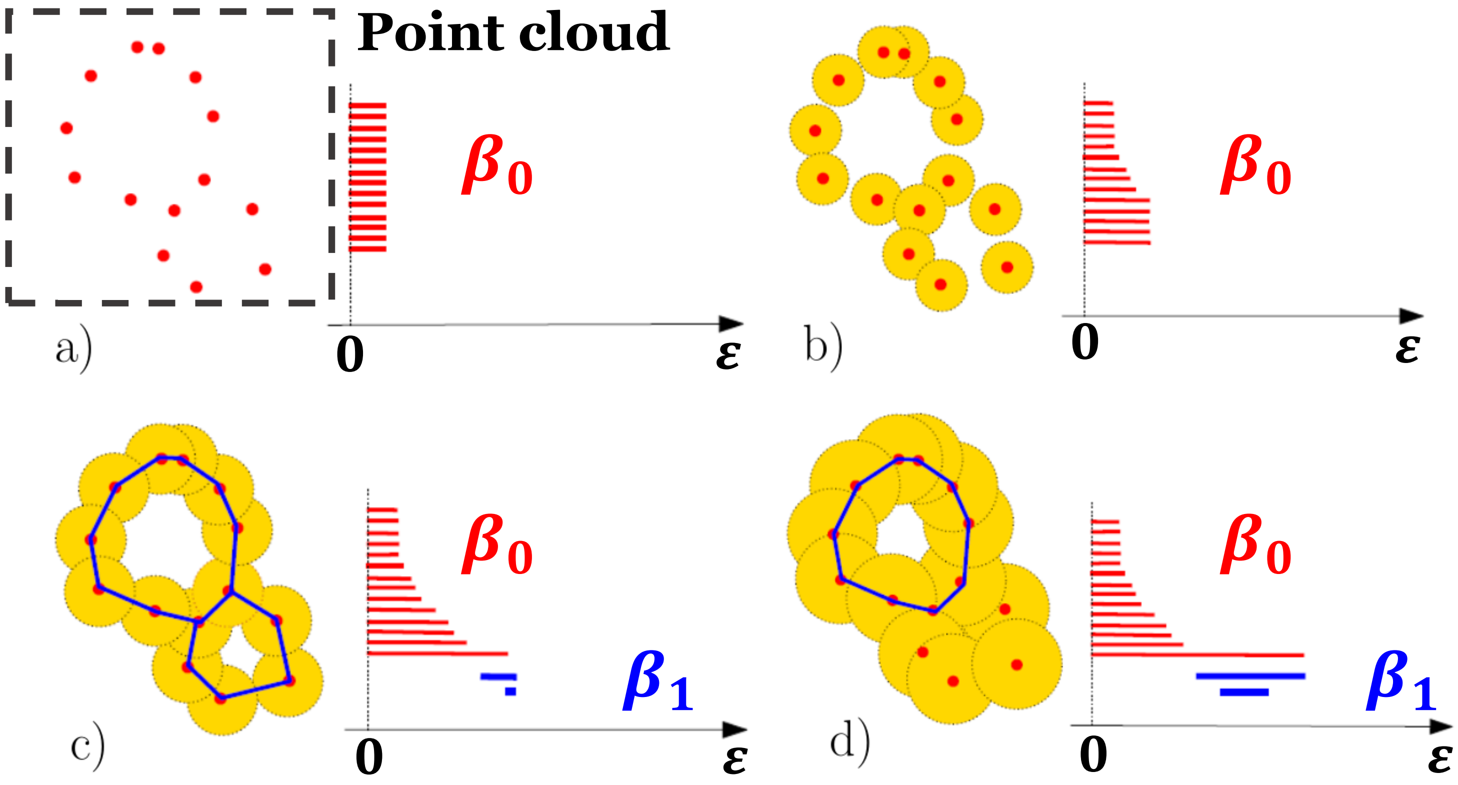
- Persistent barcode



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

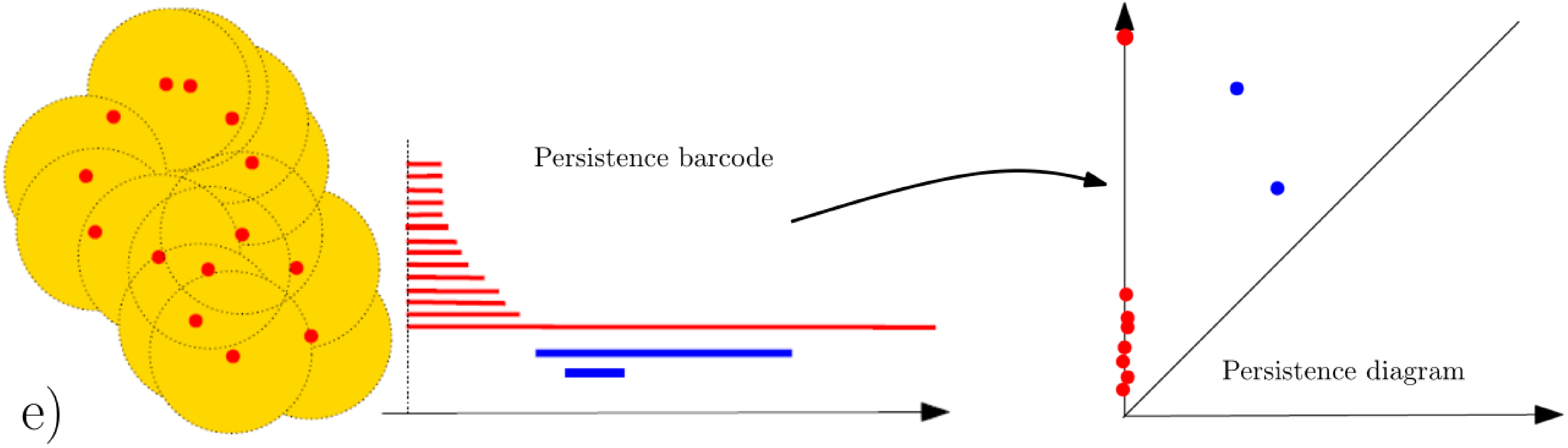
- Persistent barcode



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

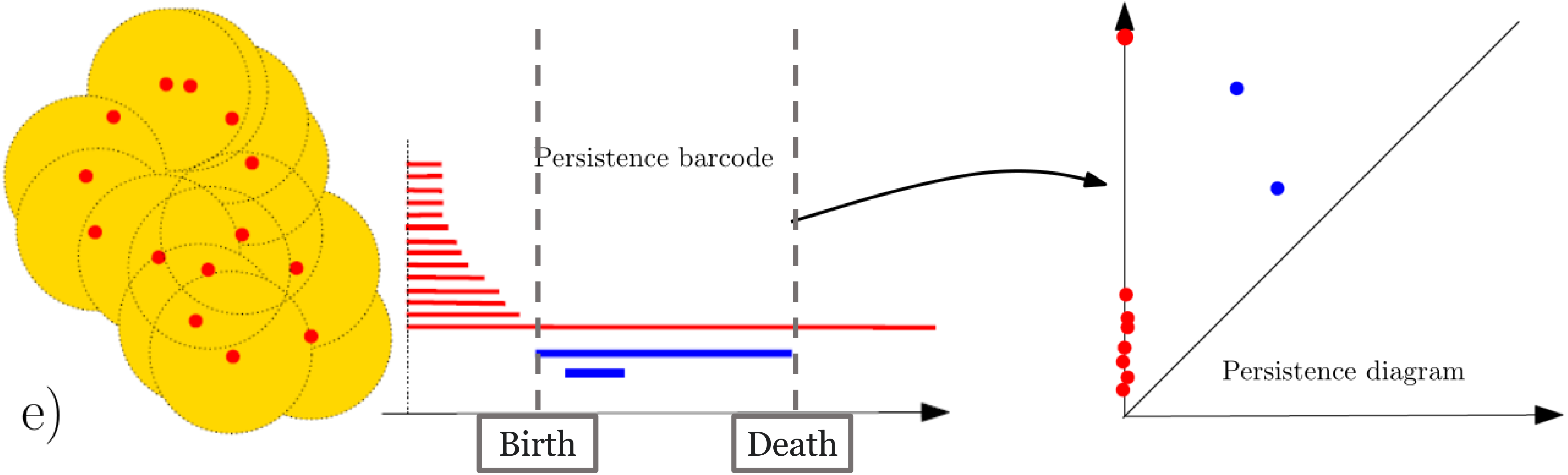
- Persistence diagram



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

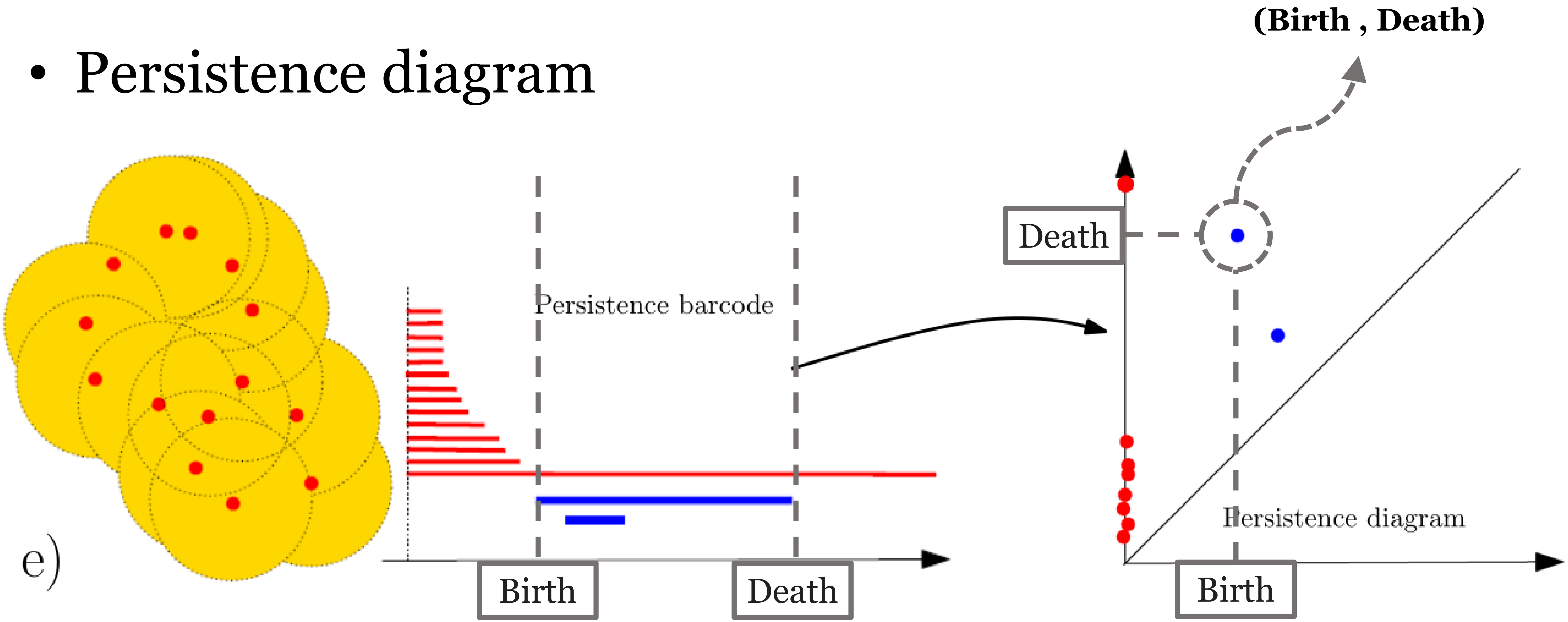
- Persistence diagram



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Topological data analysis (TDA)

- Persistence diagram



Figures in this slide: Chazal, F., & Michel, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

Method: point cloud

Given a dataset $X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pn} \end{pmatrix} \in \mathbb{R}^{p \times n}$, let $\boldsymbol{x_j} = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{pj} \end{pmatrix}$, $j = 1, \cdots, n$.

Assume that $\boldsymbol{x_j}$ are centered and normalized:

$$\sum_{i=1}^{p} x_{ij} = 0 \qquad \|\boldsymbol{x_j}\|^2 = 1$$

Distance metric: (**point cloud**)

$$c_X(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sqrt{1 - |corr(\boldsymbol{x_i}, \boldsymbol{x_j})|} \in \mathbb{R}^{n \times n} \text{ where}$$

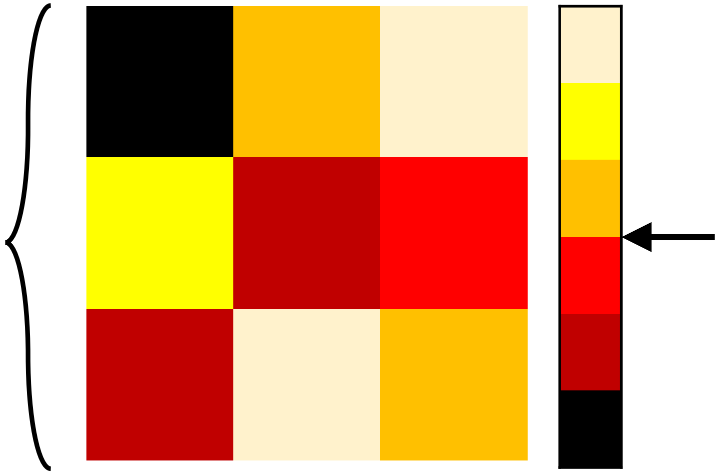
$corr(\cdot, \cdot)$ is the Pearson correlation coefficient between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$

Method: point cloud

Heatmap :
Data visualization technique with colored representation based on matrix element value.

$$M = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{p1} & \cdots & w_{pn} \end{pmatrix} \in \mathbb{R}^{p \times n}$$

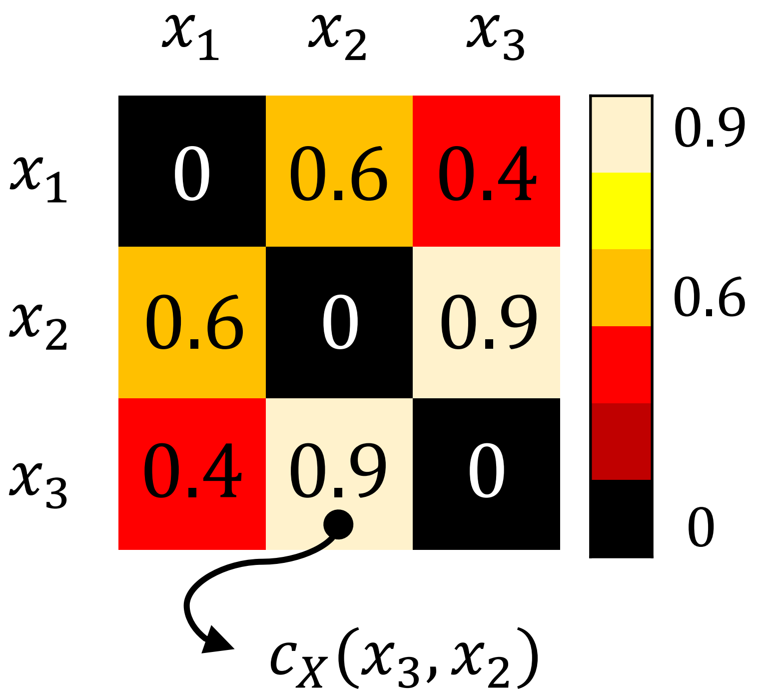
Heatmap



$\boldsymbol{w_{ij}}$
(sorted)

$$c_X(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{1 - |corr(\boldsymbol{x}_i, \boldsymbol{x}_j)|} \in \mathbb{R}^{3\times 3}$$

heatmap representation



Method: barcode

- Features in $\boldsymbol{\beta_0}$ are always shown but features in $\boldsymbol{\beta_1}$ or $\boldsymbol{\beta_2}$ are not.

- The result of barcode for $\boldsymbol{\beta_0}$ does not generate a noise in the process

  $\longrightarrow$ accurate topological information can be obtained.

- Thus, we only consider the zeroth betti number $\boldsymbol{\beta_0}$ in this study.

Method: barcode

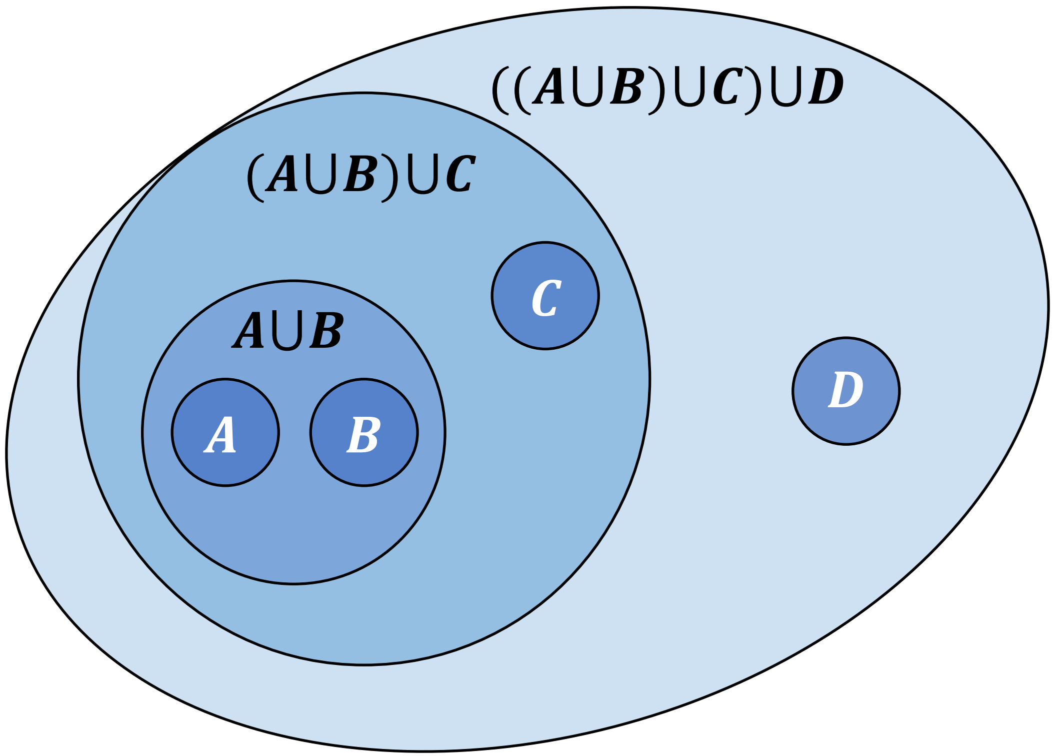
- In order to obtain the more accurate results for $\boldsymbol{\beta_0}$,
  we consider geometric information of $\boldsymbol{\beta_0}$, with a topological information.

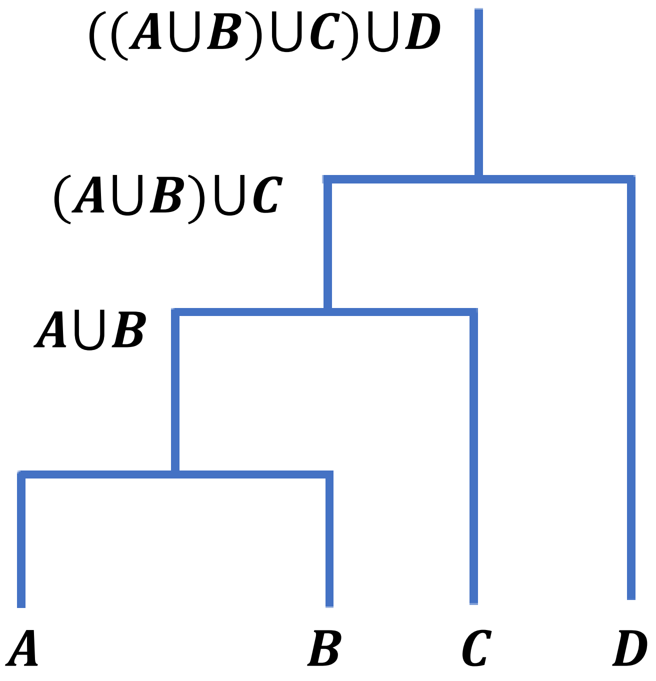
- We use the dendrogram in this study as geometric information of $\boldsymbol{\beta_0}$ as

Lee et al., Persistent brain network homology from the perspective of dendrogram, IEEE transactions on medical imaging, 2012.

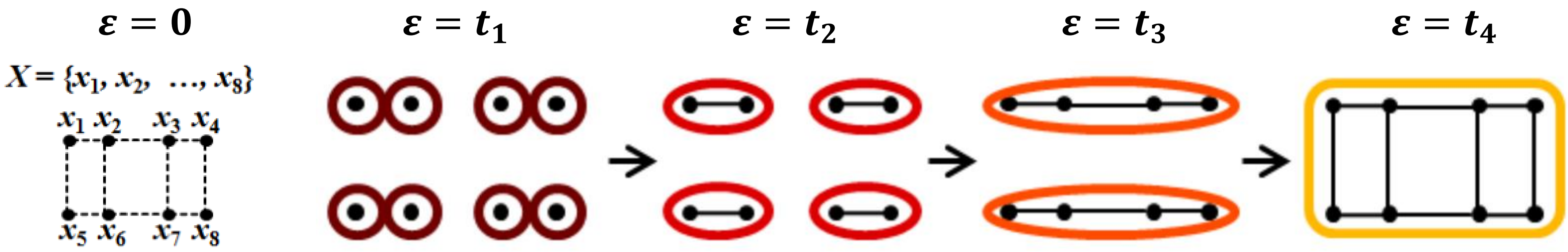
Method: dendrogram

Dendrogram :
The diagram that shows the hierarchical relationship between objects such as stem of a tree.



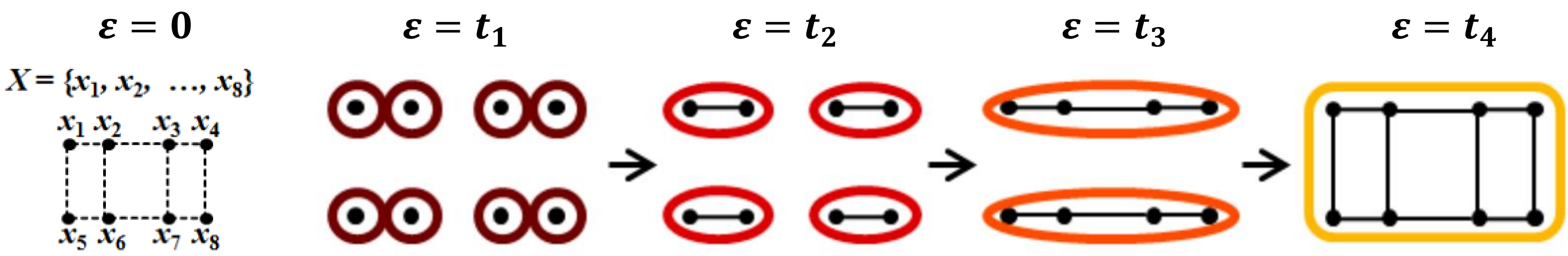


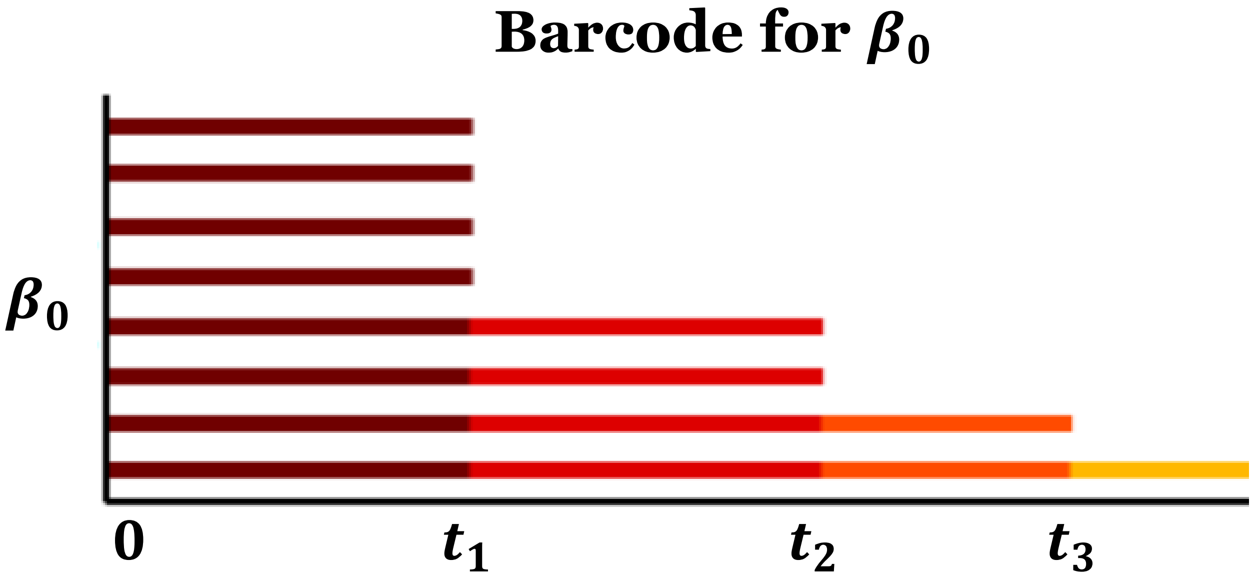
Method: barcode and dendrogram

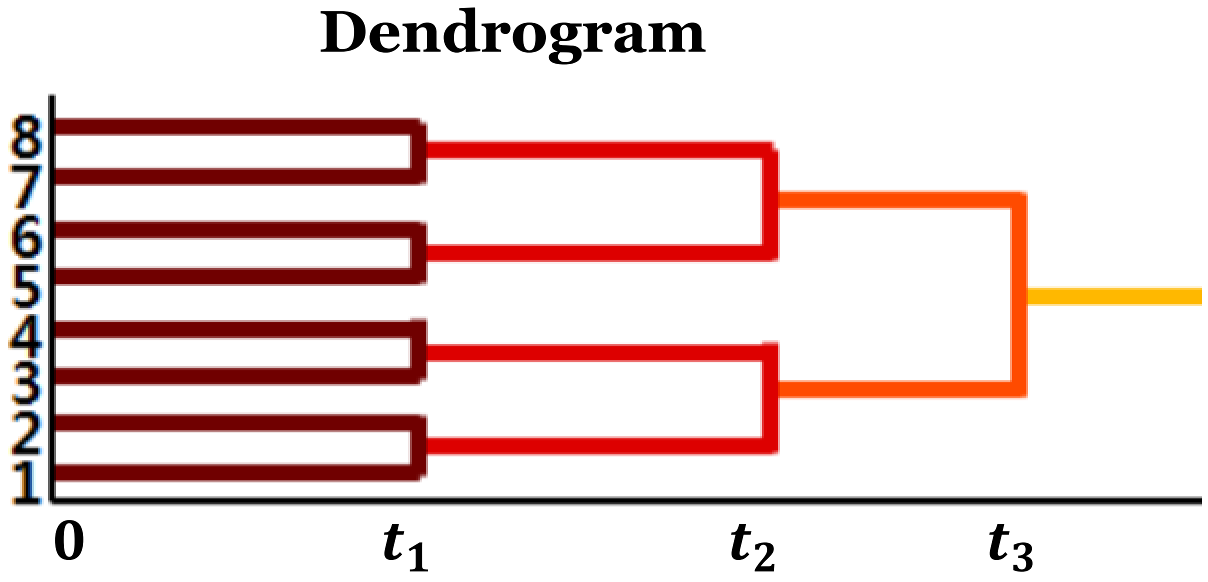


Reference & Figure : Lee, H., Kang, H., Chung, M. K., Kim, B. N., & Lee, D. S. (2012). Persistent brain network homology from the perspective of dendrogram. *IEEE transactions on medical imaging*, *31*(12), 2267-2277.

Method: barcode and dendrogram



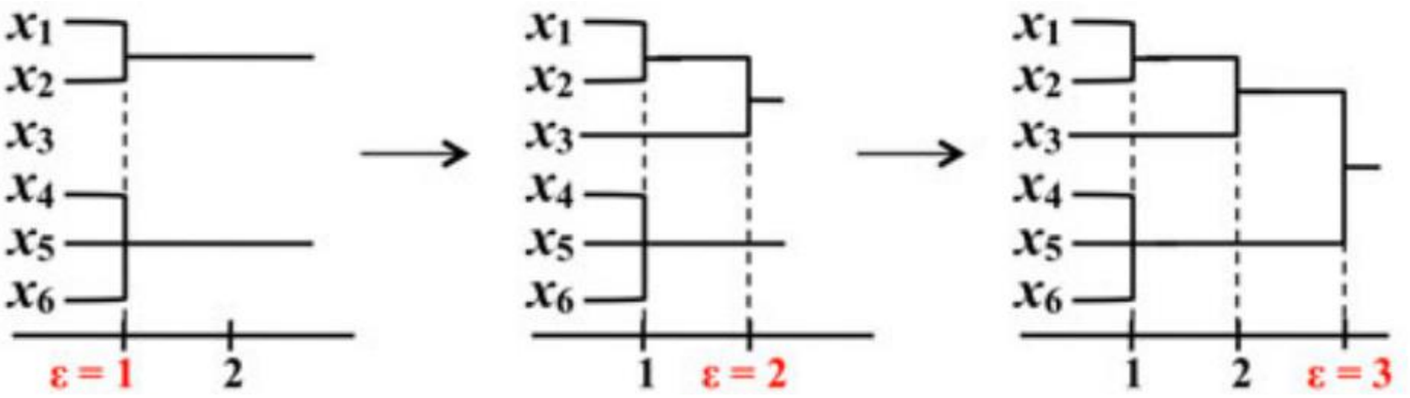




Reference & Figure : Lee, H., Kang, H., Chung, M. K., Kim, B. N., & Lee, D. S. (2012). Persistent brain network homology from the perspective of dendrogram. *IEEE transactions on medical imaging*, *31*(12), 2267-2277.

Method: comparison between subjects

For each dataset of subject $\boldsymbol{X_i}$, $\boldsymbol{i = 1, \cdots n,}$ we obtain $\boldsymbol{d_{X_i}}$ as follows.



Save the matrix form

$\boldsymbol{d_{X_i}}$

$$
\begin{array}{c|cccccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
\hline
x_1 & 0 & 1 & 2 & 3 & 3 & 3 \\
x_2 & 1 & 0 & 2 & 3 & 3 & 3 \\
x_3 & 2 & 2 & 0 & 3 & 3 & 3 \\
x_4 & 3 & 3 & 3 & 0 & 1 & 1 \\
x_5 & 3 & 3 & 3 & 1 & 0 & 1 \\
x_6 & 3 & 3 & 3 & 1 & 1 & 0
\end{array}
$$

Reference & Figure : Lee, H., Chung, M. K., Kang, H., Kim, B. N., & Lee, D. S. (2011, September). Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 302-309). Springer, Berlin, Heidelberg.

Method: comparison between subjects

We want to compute the distance between two subjects $\boldsymbol{X_i}, \boldsymbol{X_j} \in \mathbb{R}^{p \times n}$.

Gromov-Hausdorff distance (GH distance)

$$\boldsymbol{d_{GH}}(\boldsymbol{X_i}, \boldsymbol{X_j}) = \frac{1}{2} \max_{\forall i,j} \left| \boldsymbol{d_{X_i}} - \boldsymbol{d_{X_j}} \right|$$

$d_{X_i}, d_{X_j} \in \mathbb{R}^{n \times n}$ **:** the matrices which are converted subjects for $\boldsymbol{X_i}, \boldsymbol{X_j}$.

Reference: Lee, H., Chung, M. K., Kang, H., Kim, B. N., & Lee, D. S. (2011, September). Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 302-309). Springer, Berlin, Heidelberg.

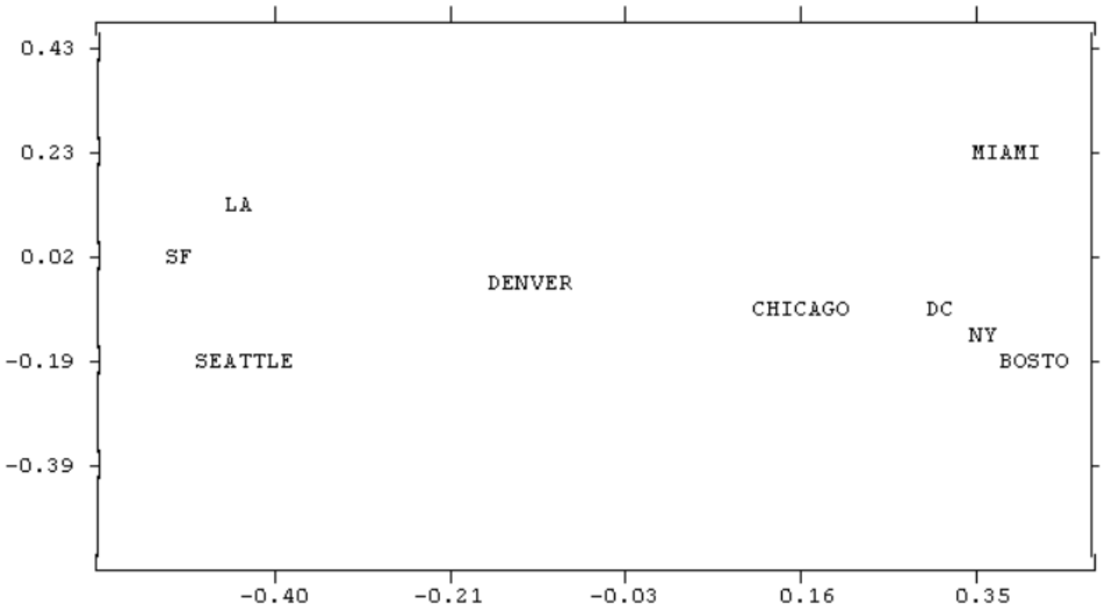
Method: multidimensional scaling (MDS)

- A technique that represents proximities among objects as distances among points in a low-dimensional space (with given dimensionality).

Input as the distance matrix:

		1 BOST	2 NY	3 DC	4 MIAM	5 CHIC	6 SEAT	7 SF	8 LA	9 DENV
1	BOSTON	0	206	429	1504	963	2976	3095	2979	1949
2	NY	206	0	233	1308	802	2815	2934	2786	1771
3	DC	429	233	0	1075	671	2684	2799	2631	1616
4	MIAMI	1504	1308	1075	0	1329	3273	3053	2687	2037
5	CHICAGO	963	802	671	1329	0	2013	2142	2054	996
6	SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
7	SF	3095	2934	2799	3053	2142	808	0	379	1235
8	LA	2979	2786	2631	2687	2054	1131	379	0	1059
9	DENVER	1949	1771	1616	2037	996	1307	1235	1059	0

Output as proximities:



Reference: Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate behavioral research*, *51*(6), 772-789..
Figure: http://www.analytictech.com/borgatti/mds.htm

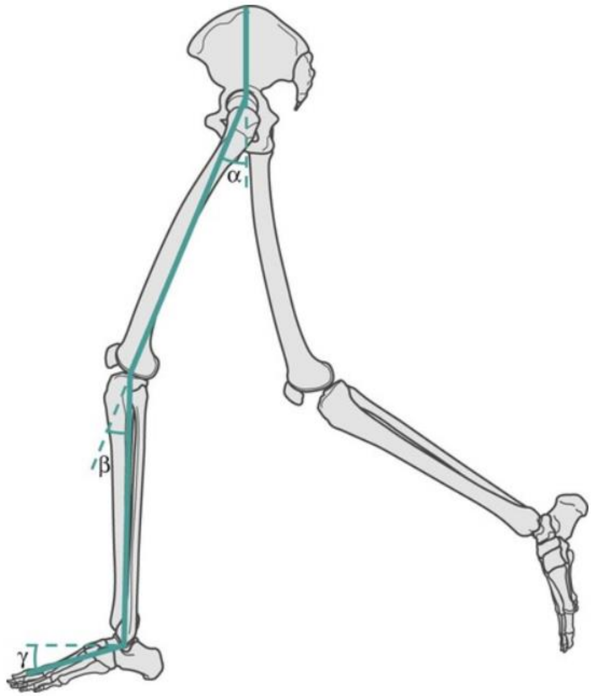
Application

- Dataset: biomechanical data (kinematics)

  - Joint angle measurement dataset

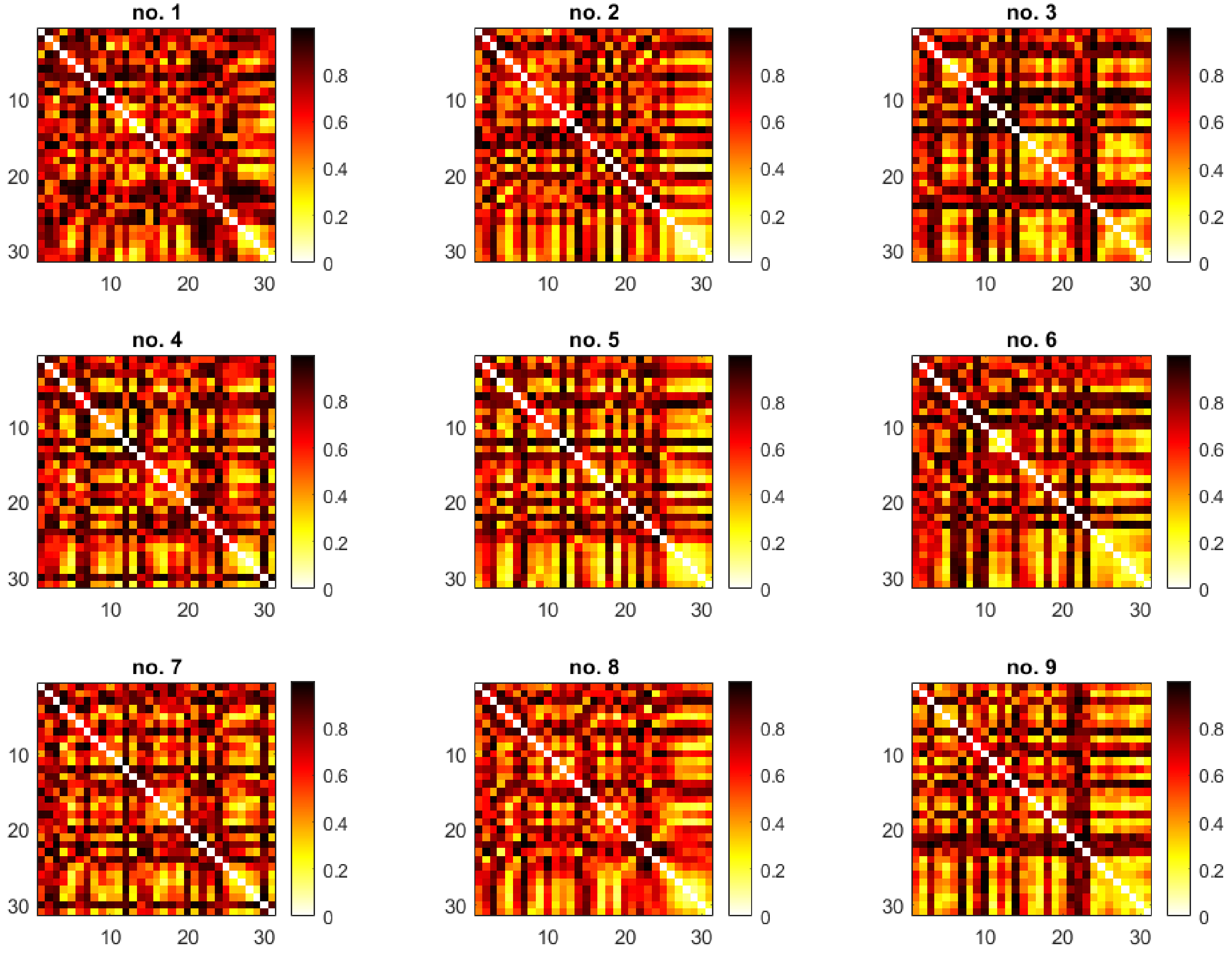
  - Data size : (100 times) × (31 variables) × (9 subjects)

10 joint angles (x-axis) + 10 joint angles (y-axis) + 10 joint angles (z-axis) + X factor



Figures in this slide: https://musculoskeletalkey.com/biomechanics-2/

Point cloud

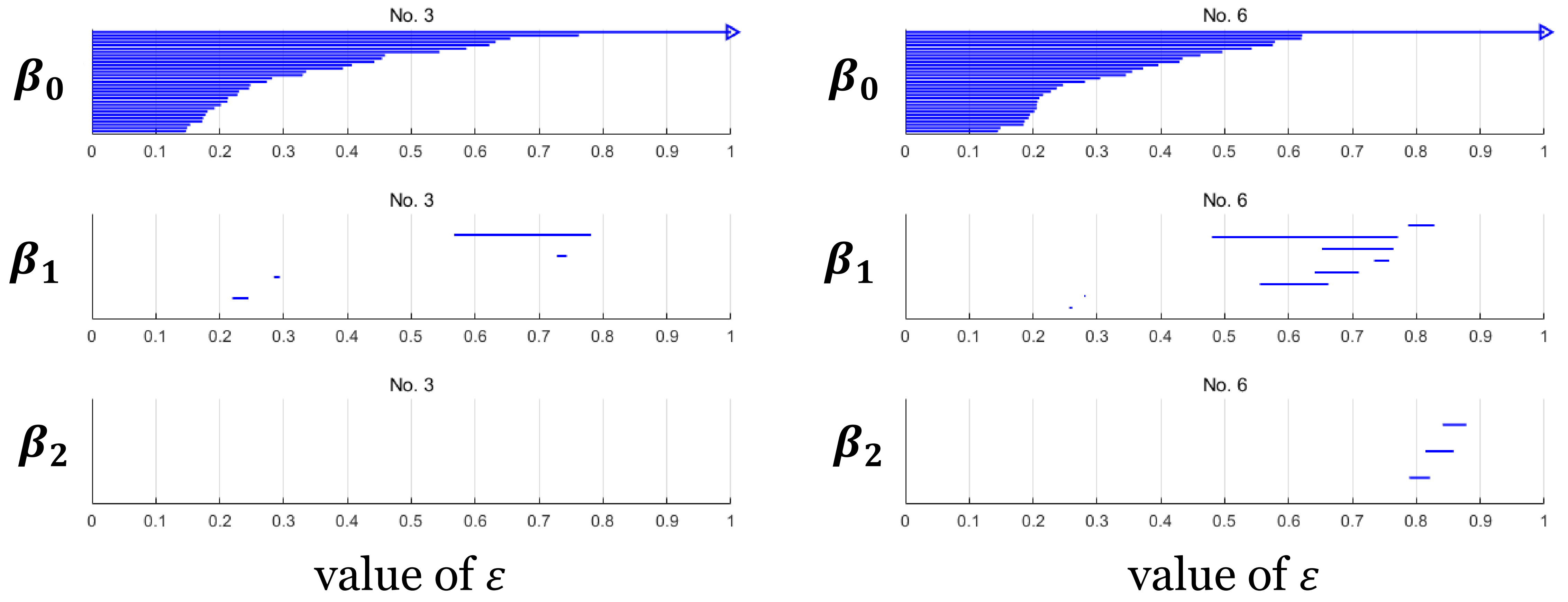


Define a point cloud.

For the subject X_i , $i = 1, \cdots 9$,

$$c_{X_i} = \sqrt{1 - |corr(X_i)|} \in \mathbb{R}^{31\times 31}$$

Barcode



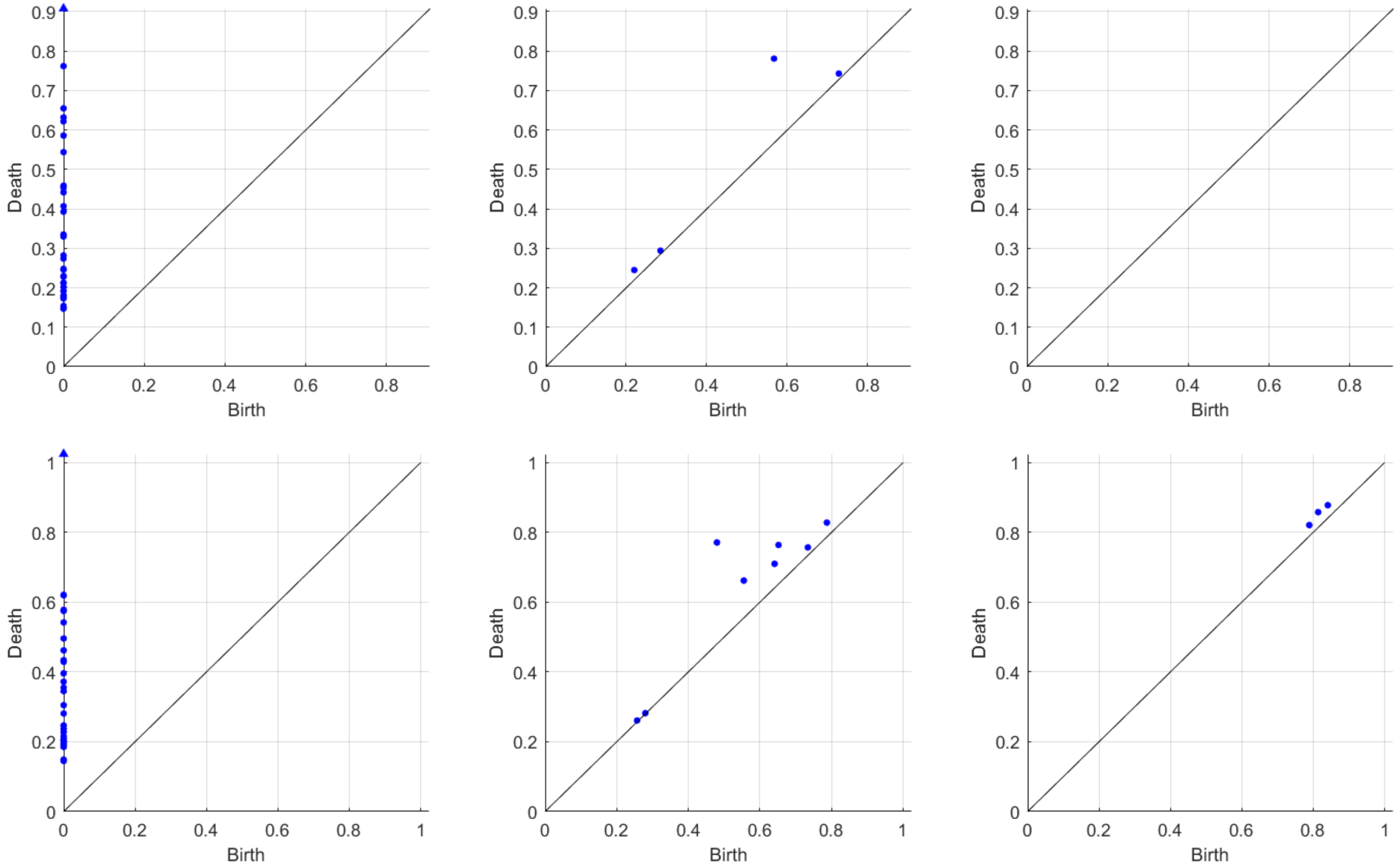
No. 3
No. 6
β_0
β_1
β_2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
value of ε
value of ε

Persistence diagram

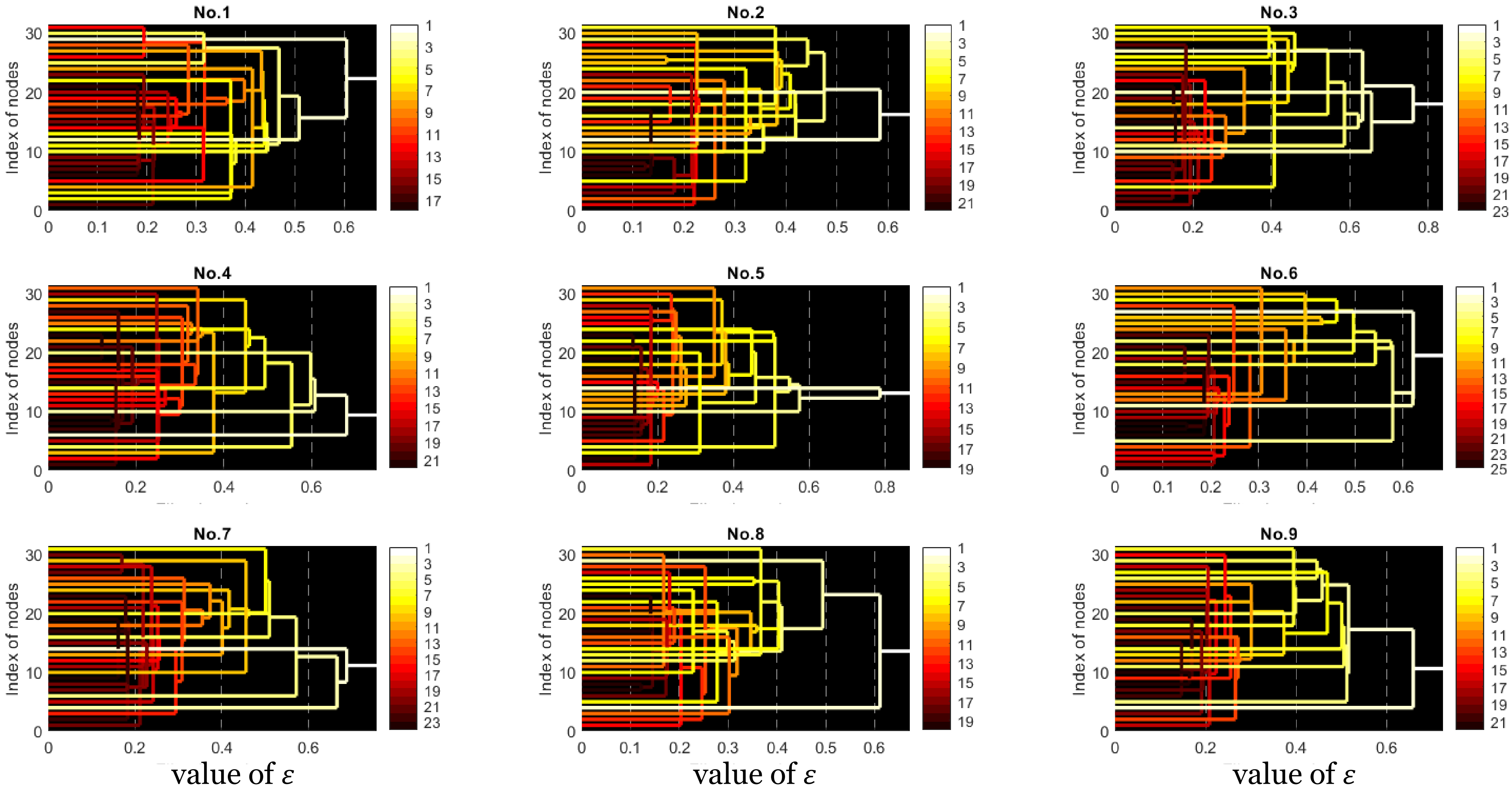
$\boldsymbol{\beta_0}$ $\boldsymbol{\beta_1}$ $\boldsymbol{\beta_2}$

Subject 3

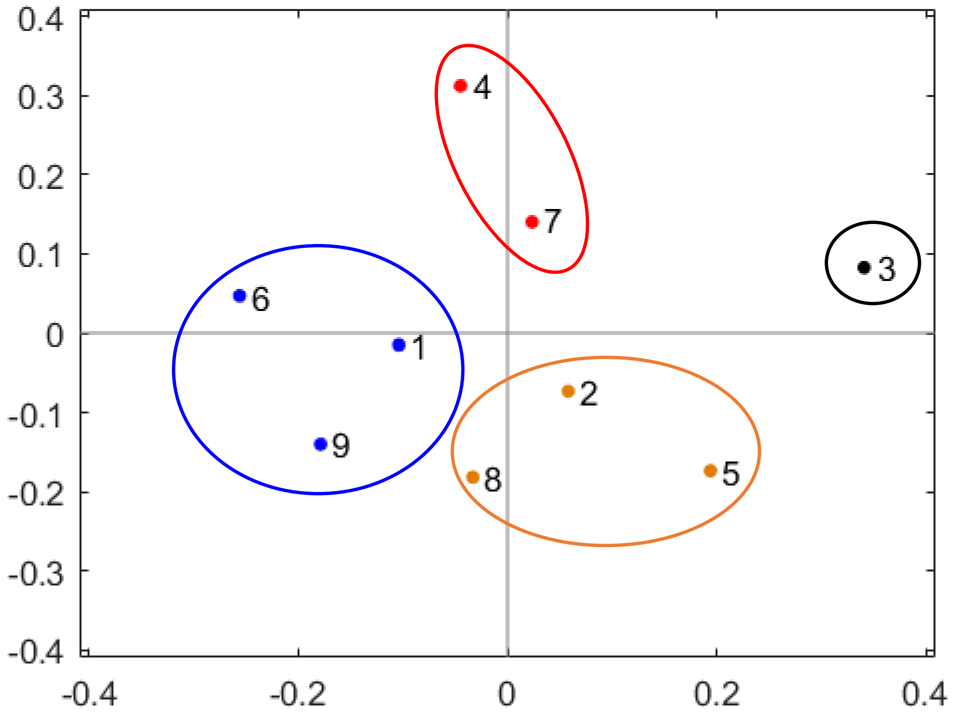
Subject 6



Dendrogram


No.1
No.2
No.3
No.4
No.5
No.6
No.7
No.8
No.9
Index of nodes
value of ε

Multidimensional scaling



Multidimensional scaling result

The indices 1 ~ 9 = subject number 1 ~ 9

Input distance matrix:

Gromov_Hausdorff_distance =

0	0.3018	0.5194	0.4992	0.4728	0.2639	0.3911	0.3000	0.2788
0.3018	0	0.4062	0.5444	0.4371	0.3945	0.4484	0.3936	0.4408
0.5194	0.4062	0	0.5255	0.4502	0.5555	0.4630	0.5336	0.5980
0.4992	0.5444	0.5255	0	0.5229	0.4942	0.2629	0.5074	0.5333
0.4728	0.4371	0.4502	0.5229	0	0.5708	0.4148	0.3840	0.4108
0.2639	0.3945	0.5555	0.4942	0.5708	0	0.4717	0.4525	0.3769
0.3911	0.4484	0.4630	0.2629	0.4148	0.4717	0	0.3993	0.4252
0.3000	0.3936	0.5336	0.5074	0.3840	0.4525	0.3993	0	0.3374
0.2788	0.4408	0.5980	0.5333	0.4108	0.3769	0.4252	0.3374	0

Output:

MDS_2D_coordinates =

-0.1038	-0.0147
0.0580	-0.0729
0.3408	0.0826
-0.0450	0.3121
0.1941	-0.1732
-0.2561	0.0471
0.0235	0.1402
-0.0331	-0.1813
-0.1785	-0.1399

Summary

- Overall, there is a distinct difference in β_0 between subject.
- Especially in β_0, the connectivity of subject 3,6, and 9 is slower than others.
- As a result of the MDS, all the subjects were separated by 3 partial groups.
- The subject 3 had the largest dissimilarity than others.

- It is necessary to find the key variables representing difference among subjects

Thank you !

Contact me : sangmanjung@khu.ac.kr