Laplacian Eigenmap

Group 1

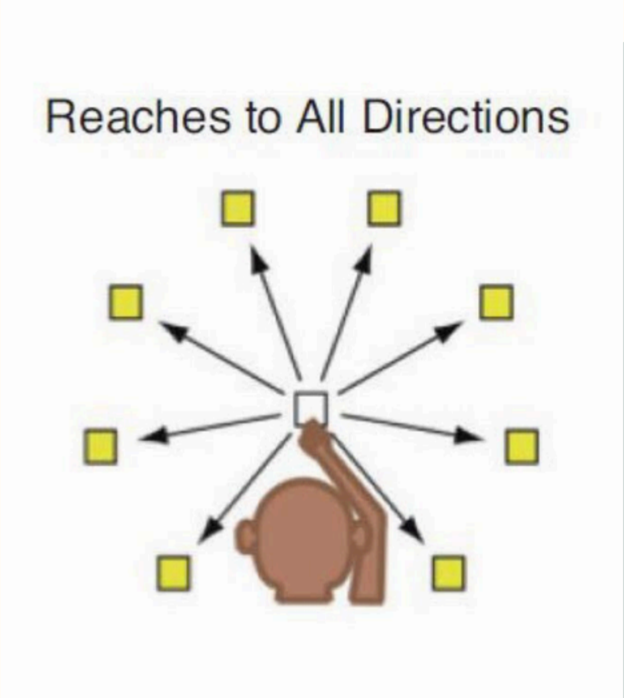
Goal: dimensionality reduction

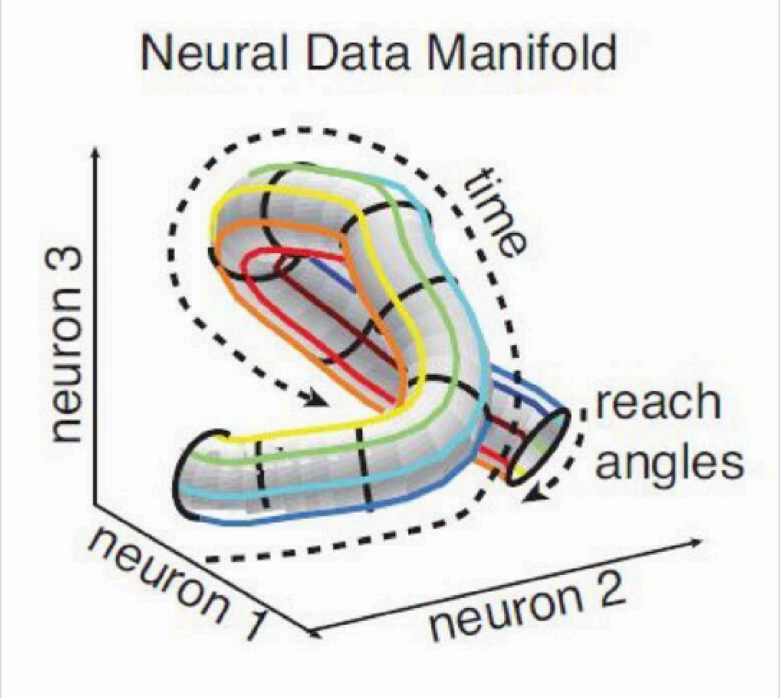
Scenario:
Data in High dimensional space.
But *embedded* in low dimensional manifold.

Why not PCA?
Manifold usually nonlinear.
Objective is not to maximize variances, but to restore the manifold.

Why Laplacian Eigenmap?
Nonlinear, Locality preserving.

image data *embedded* in a manifold whose d.f. is the d.f. of camera





neural activity is also *embedded* in low-dimensional.(Gao et al.2017)

Algorithm

Algorithm 1 Laplacian Eigenmap

---

function LAPLACIAN EIGENMAP(X,k(or ϵ),t,d)
  ▷ X:input data, k: number of neighbors, t: constant of Heat Kernel
  $G \leftarrow GraphConstruction(X, k(or\ \epsilon))$
  $W \leftarrow Heatkernel(G, t)$ ▷ W:Weight matrix
  $D \leftarrow diag(\sum_i W_{1i}, \sum_i W_{2i} \ldots, \sum_i W_{ni})$ ▷ D:degree matrix
  $L \leftarrow D - W$
  eigenvalues, eigenvectors $\leftarrow Solveeigenvalue(D^{-1}L)$ ▷ $Lf = \lambda Df$
  return $f_2, \ldots f_{d+1}$
end function

$$x_i \xrightarrow{Embedding} (f_2(i), f_3(i), \ldots, f_{d+1}(i))$$

$$e^{-\frac{||x_i - x_j||^2}{t}}$$

Heat Kernel

$$D^{-1}L$$

$$D^{-1/2}LD^{-1/2}$$

Eigenvalue problem
Both work

$$\lambda_2, \ldots, \lambda_{d+1}$$
$$f_2, \ldots, f_{d+1}$$

First d smallest nonzero eigen value and eigen functions

Algorithm

How to construct a Graph:

k-NN: select the K closest data.

ϵ-neighborhood: select the data in the ball with radius ϵ.

Determine the edge weight:

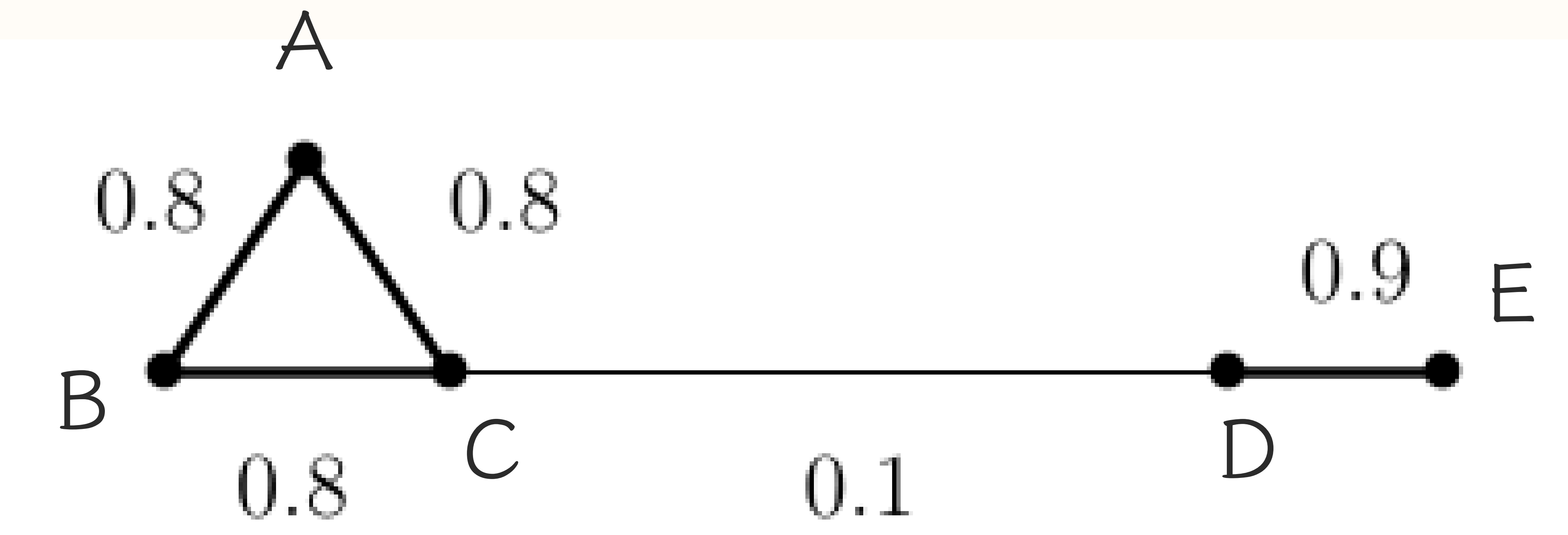
1. connected:

heat kernal: $w_{ij} = \exp -\frac{||x_i - x_j||^2}{t}$

simple version: 1 (t=inf)

2. non-connected : w(i,j)=0

Example



$$D = \begin{pmatrix} 1.6 & & & & \\ & 1.6 & & & \\ & & 1.7 & & \\ & & & 1 & \\ & & & & 0.9 \end{pmatrix}$$

$$\mathbf{W} = \begin{pmatrix} 0 & .8 & .8 & 0 & 0 \\ .8 & 0 & .8 & 0 & 0 \\ .8 & .8 & 0 & .1 & 0 \\ 0 & 0 & .1 & 0 & .9 \\ 0 & 0 & 0 & .9 & 0 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} 1.6 & -0.8 & -0.8 & & \\ -0.8 & 1.6 & -0.8 & & \\ -0.8 & -0.8 & 1.7 & -0.1 & \\ & & -0.1 & 1 & -0.9 \\ & & & -0.9 & 0.9 \end{pmatrix}$$

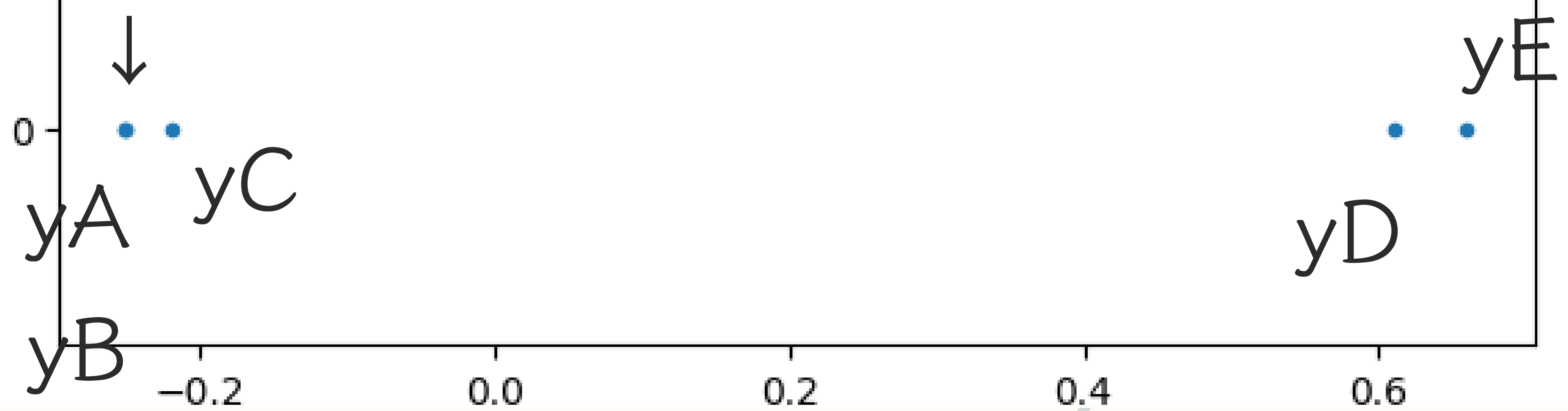
Example

$$\mathbf{D}^{-1}\mathbf{L} = \begin{pmatrix} 1 & -0.5 & -0.5 & & \\ -0.5 & 1 & -0.5 & & \\ -0.4706 & -0.4706 & 1 & -0.0588 & 0 \\ & & -0.1 & 1 & -0.9 \\ & & & -1 & 1 \end{pmatrix}$$

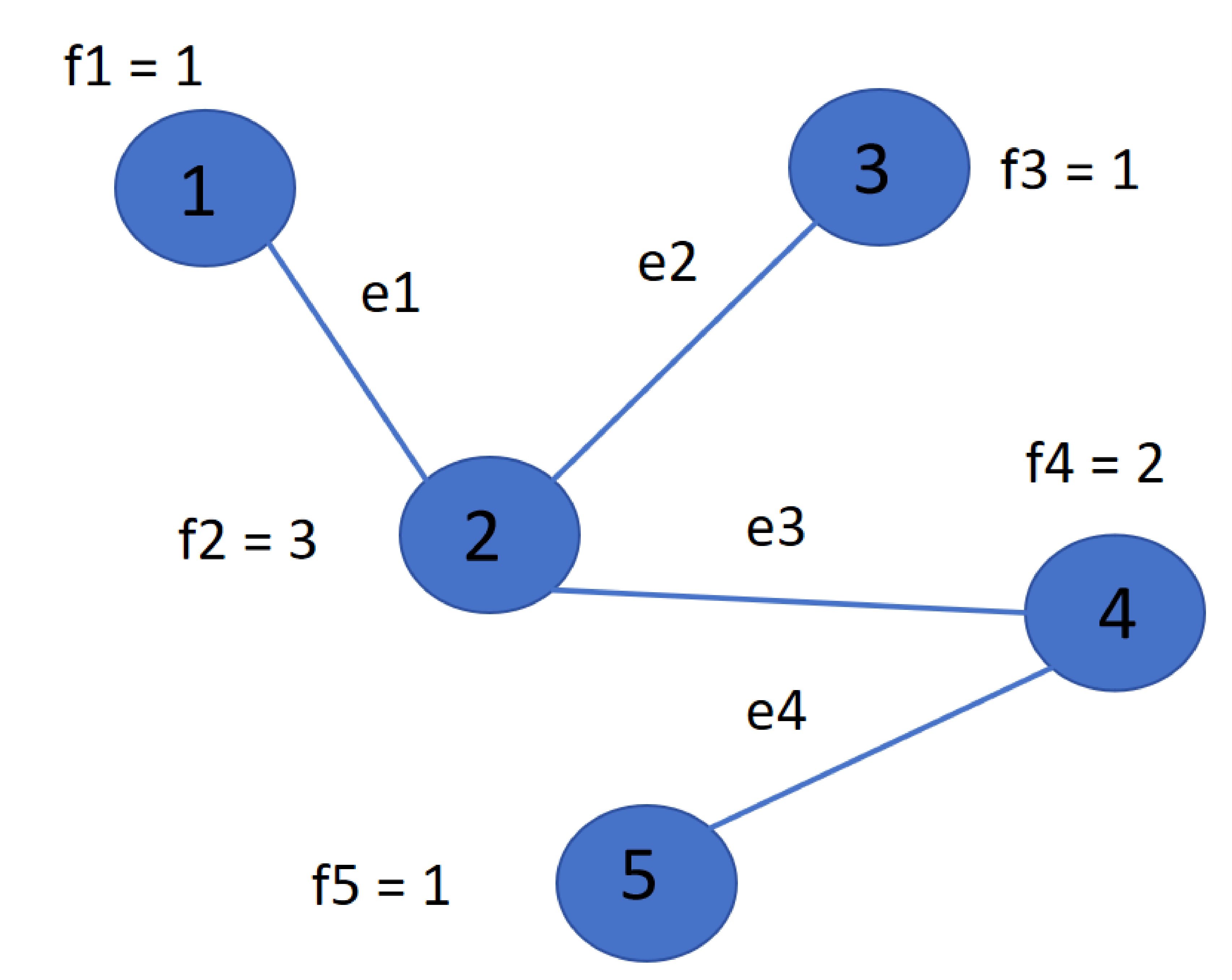
$$\mathbf{v}_2 = (-0.2594, -0.2594, -0.2235, 0.6152, 0.6610).$$

(corresponding to $\lambda_2 = 0.0693$)

2 points overlap



Demostration of Laplacian



$$f = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

f: mappings

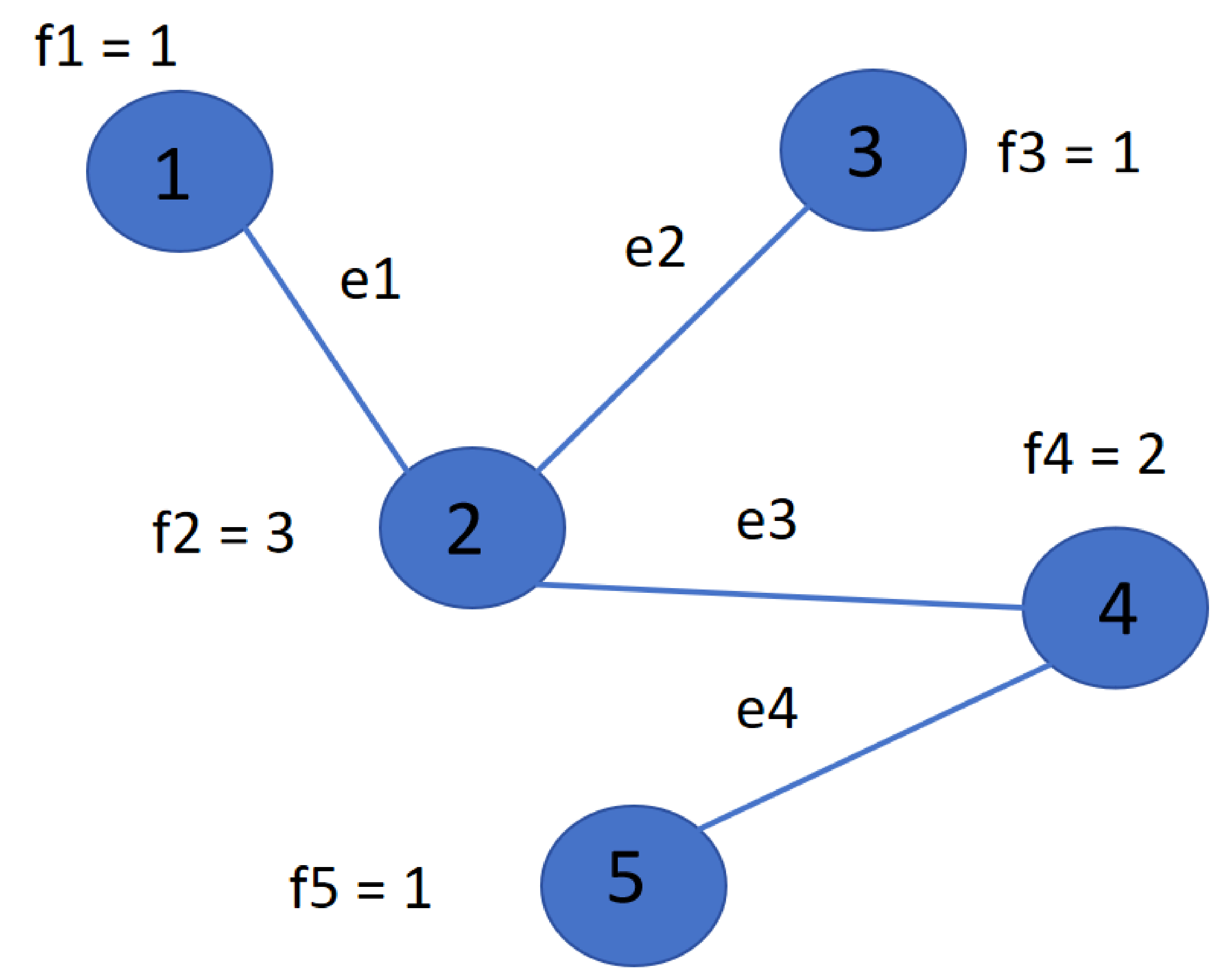
k: incidence matrix, k∈|E|×|V|

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

Demostration of Laplacian

$$\nabla \mathrm{f} = k^T f = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} f(1) - f(2) \\ f(2) - f(3) \\ f(2) - f(4) \\ f(4) - f(5) \end{bmatrix}$$

$$k\nabla \mathrm{f} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$



Demostration of Laplacian

$$Lf = kk^T f$$

Conclude that the Laplacian =

$$L = kk^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} = D - W$$

Observe that we can define W, D directly by

$$W_{i,j} = \begin{cases} 1 & i, j \textit{ are connected} \\ 0 & \textit{otherwise} \end{cases}$$

$$D_{i,j} = \begin{cases} \sum_{k=1}^{n} W_{i,k} & \textit{for } i = j \\ 0 & \textit{otherwise} \end{cases}$$

$$\|k^T f\|^2 = (f^T k)(k^T f) = f^T L f$$

Perspective of preserving locality

The goal is to find a mapping maps the original data $(x_1, x_2, ..., x_n)$ to $(y_1, y_2, ..., y_n)$. We hope such map satisfying $min \sum_{i,j}(y_i - y_j)^2 W_{ij}$, which makes if x_i and x_j are close then W_{ij} is large, there would be a heavy penalty on y_i and y_j, so y_i and y_j need to be close as well, while x_i and x_j are far, the distance of y_i and y_j are not important because W_{ij} is close to 0.

And we have the following result:

$$\min_{\mathbf{f}} \sum_{i,j} (f_i - f_j)^2 W_{ij} \equiv \min_{\mathbf{f}} \mathbf{f}^T L \mathbf{f}$$

Perspective of preserving locality

Proof:

$$
\begin{aligned}
\sum_{i,j}(f_i - f_j)^2 W_{ij} &= \sum_i \left(\sum_j W_{ij}\right) f_i^2 - 2\sum_{i,j} f_i f_j W_{ij} + \sum_j \left(\sum_i W_{ij}\right) f_j^2 \\
&= 2\left(\sum_i d_i f_i^2 - \sum_{i,j} f_i f_j W_{ij}\right) \\
&= 2\left(\mathbf{f}^T D \mathbf{f} - \mathbf{f}^T W \mathbf{f}\right) \\
&= 2\mathbf{f}^T (D - W)\mathbf{f} \\
&= 2\mathbf{f}^T L \mathbf{f}
\end{aligned}
$$

Perspective of preserving locality

we impose the constraint $\|\mathbf{f}\| = 1$

$$\Longrightarrow \min_{\mathbf{f}\neq 0} \frac{\mathbf{f}^T L\mathbf{f}}{\mathbf{f}^T\mathbf{f}}$$

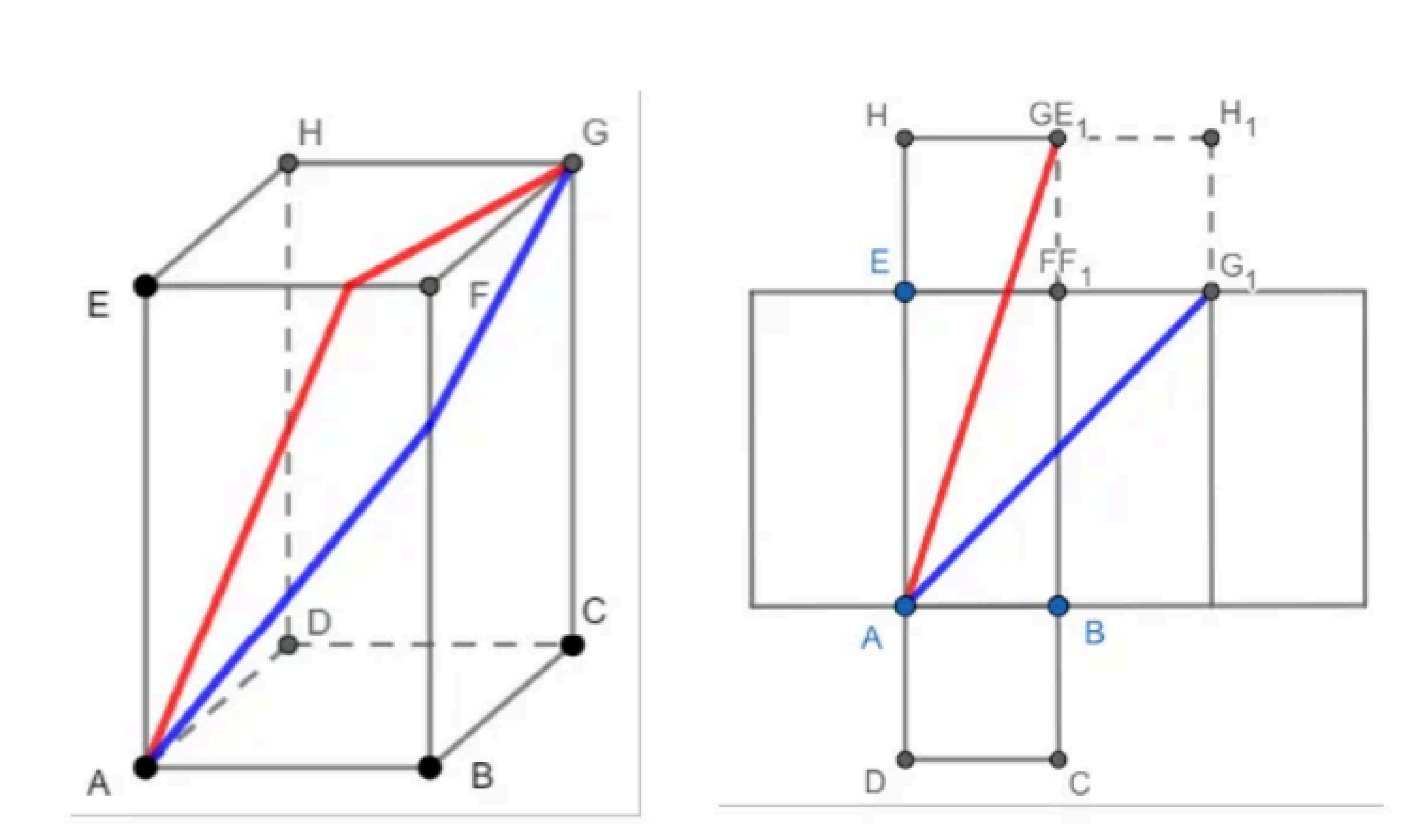
eliminate the trivial solution $\mathbf{f} = \mathbf{1}$, and refine the problem by removing the scaling factor D in $\mathbf{f}$ (Matrix D provides a natural measure on the vertices of the graph. The bigger the value D_{ii} is, the more important is the ith vertex.)

$$\Longrightarrow \min_{\substack{\mathbf{f}\neq 0 \\ \mathbf{f}^T D\mathbf{1}=0}} \frac{\mathbf{f}^T L\mathbf{f}}{\mathbf{f}^T D\mathbf{f}}$$

This is equivalent to solving the generalized eigenvalue problem $L\mathbf{f} = \lambda D\mathbf{f}$ to find the eigenvectors corresponding to the smallest non-zero eigenvalues.

Why it works

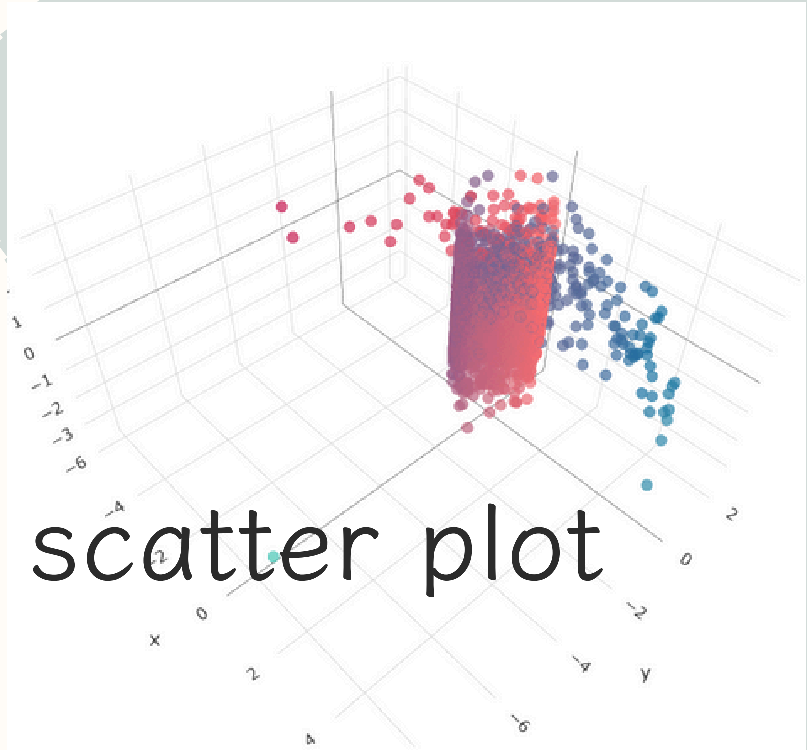
$$\min_{\mathbf{f}} \sum_{i,j} (f_i - f_j)^2 W_{ij}$$



Geodesic

$$|f(x_i) - f(x_j)| \leq \|\nabla f(x)\| \|x_i - x_j\| + o(\|x_i - x_j\|)$$

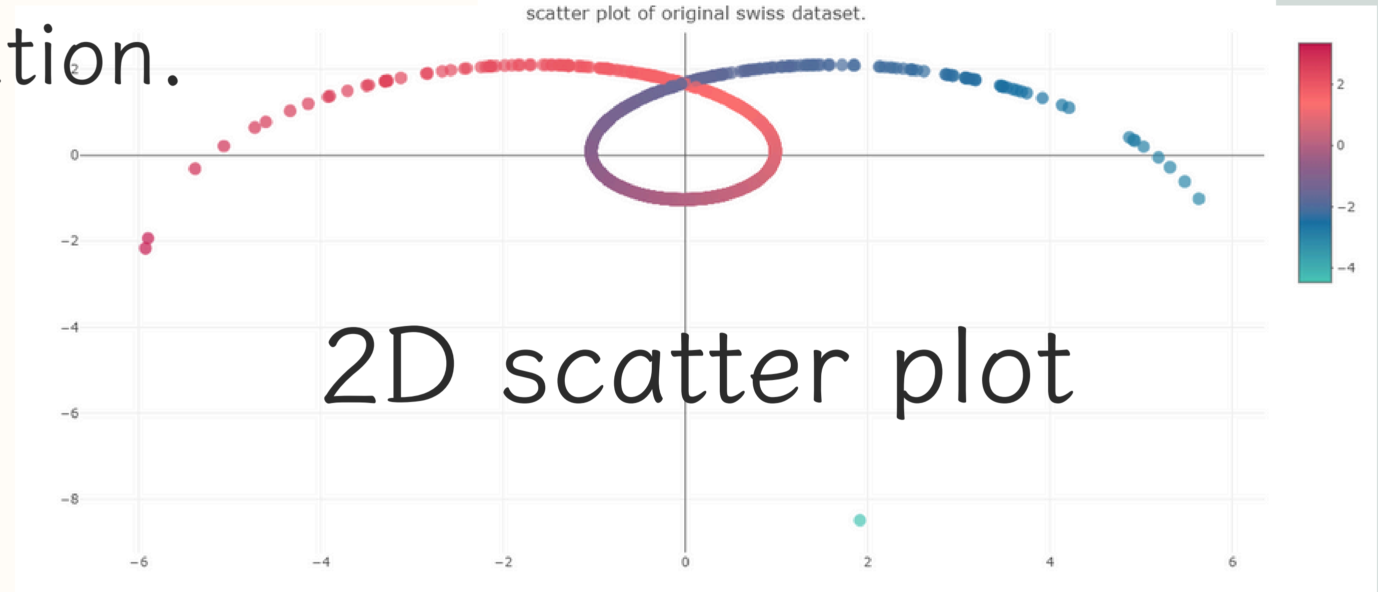
Application: Swiss Roll Data



3D scatter plot

Description

Take the coordinates (x,y) in a 2 dimensional plot, where x, y generated from the standardized normal distribution.



2D scatter plot

Map it to a 3 dimensional table:

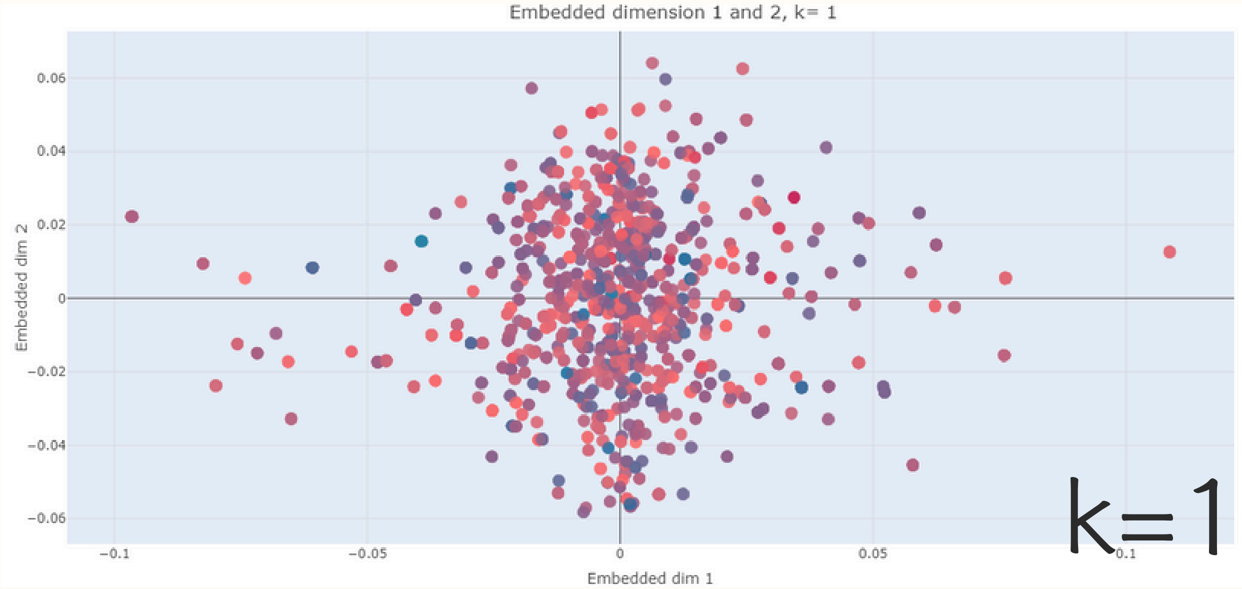
(x,y) -> (xcos x, xsin x, y),

A standard "hello world" 3 dimensional data set for testing dimensionality reduction techniques and algorithms.

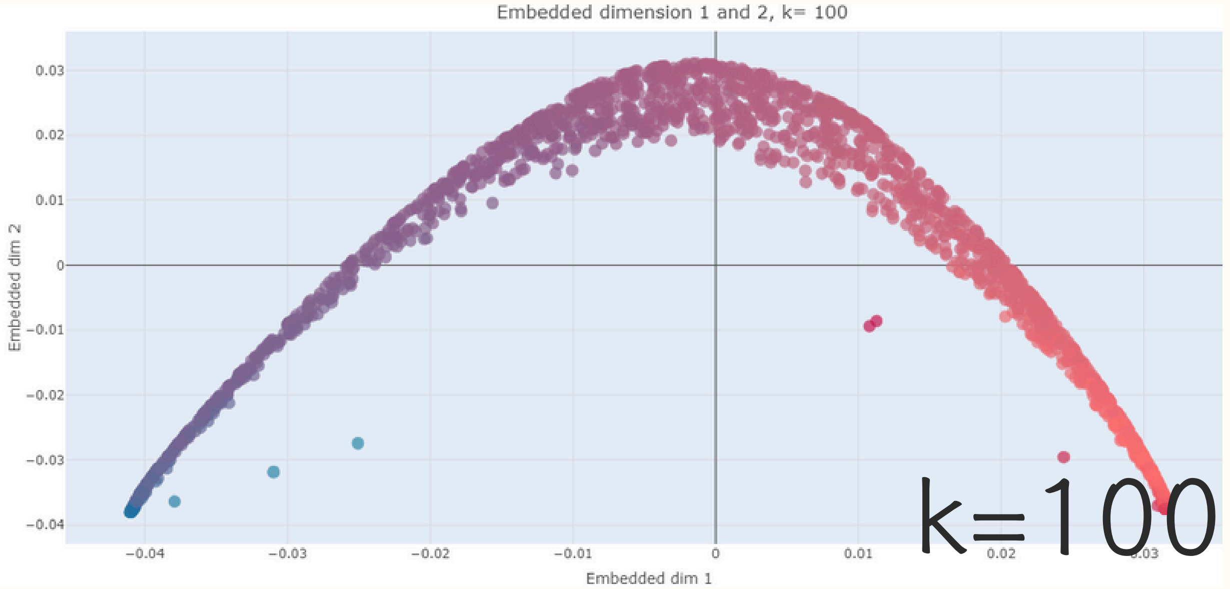
3D scatter plot

Some Experiments

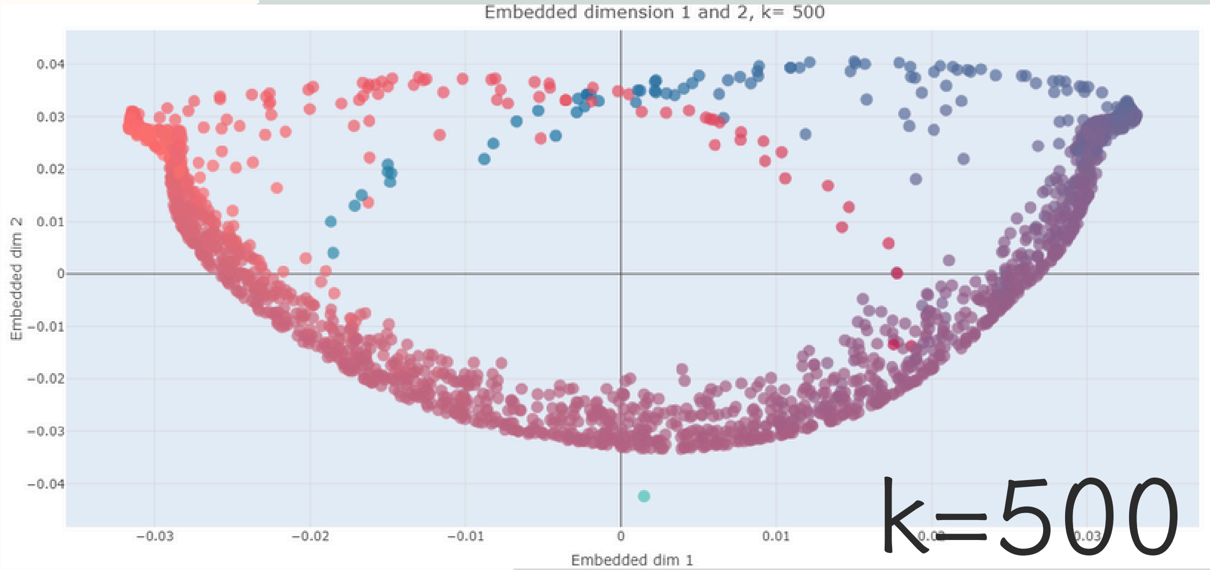
data number:2000



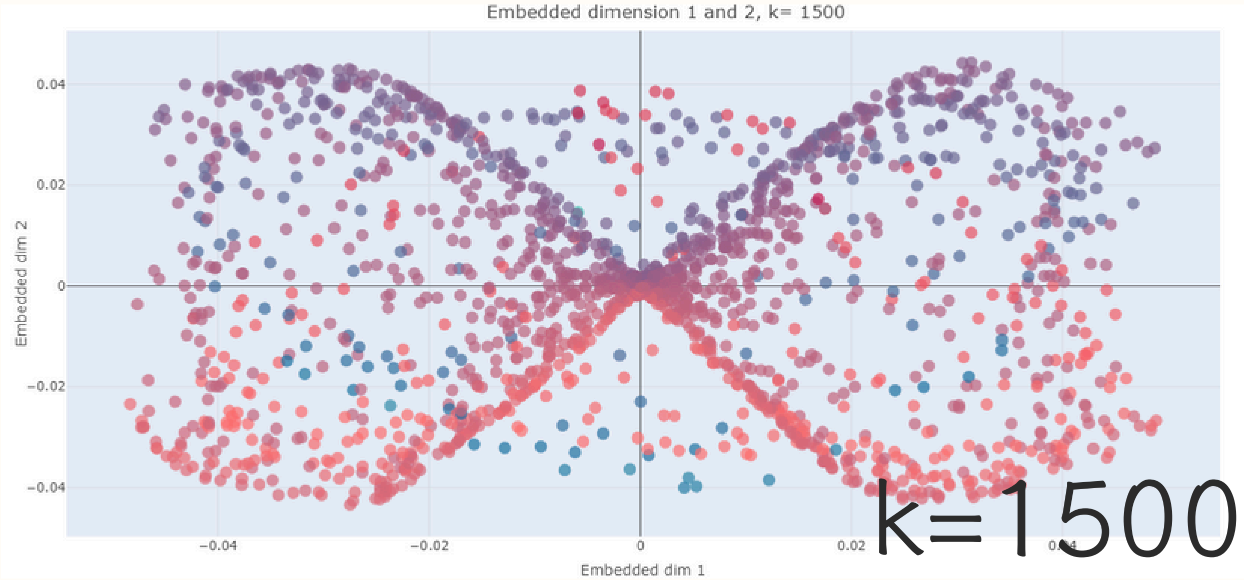
tend to overlap



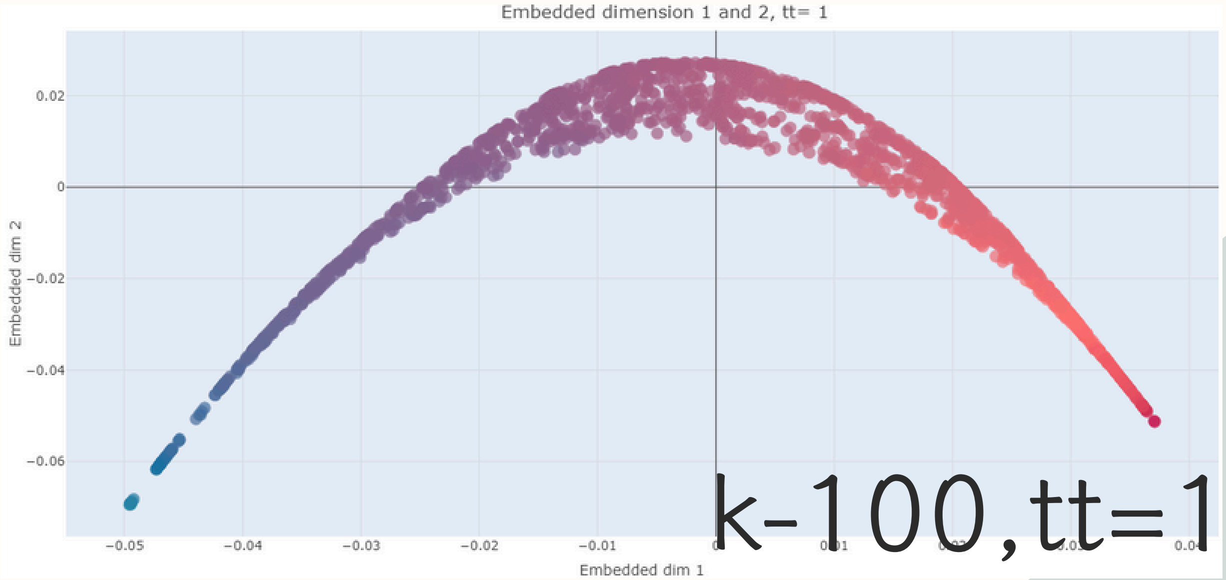
best!!!!!



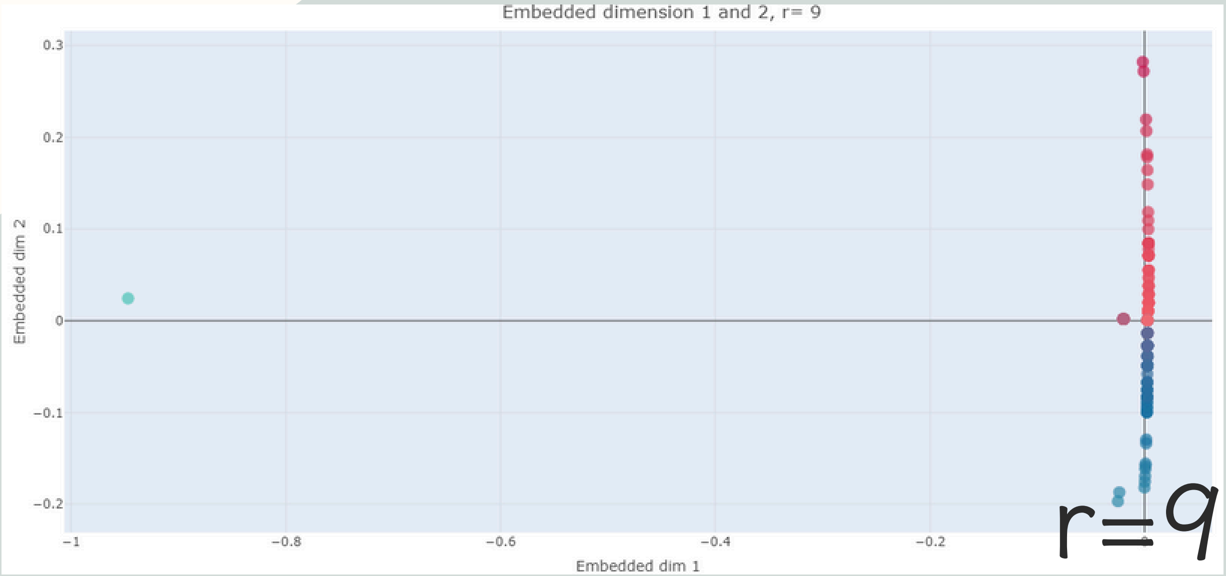
capture overlapping and curved neighboring variations



middle data disappear



outliers: affected by kernal weight

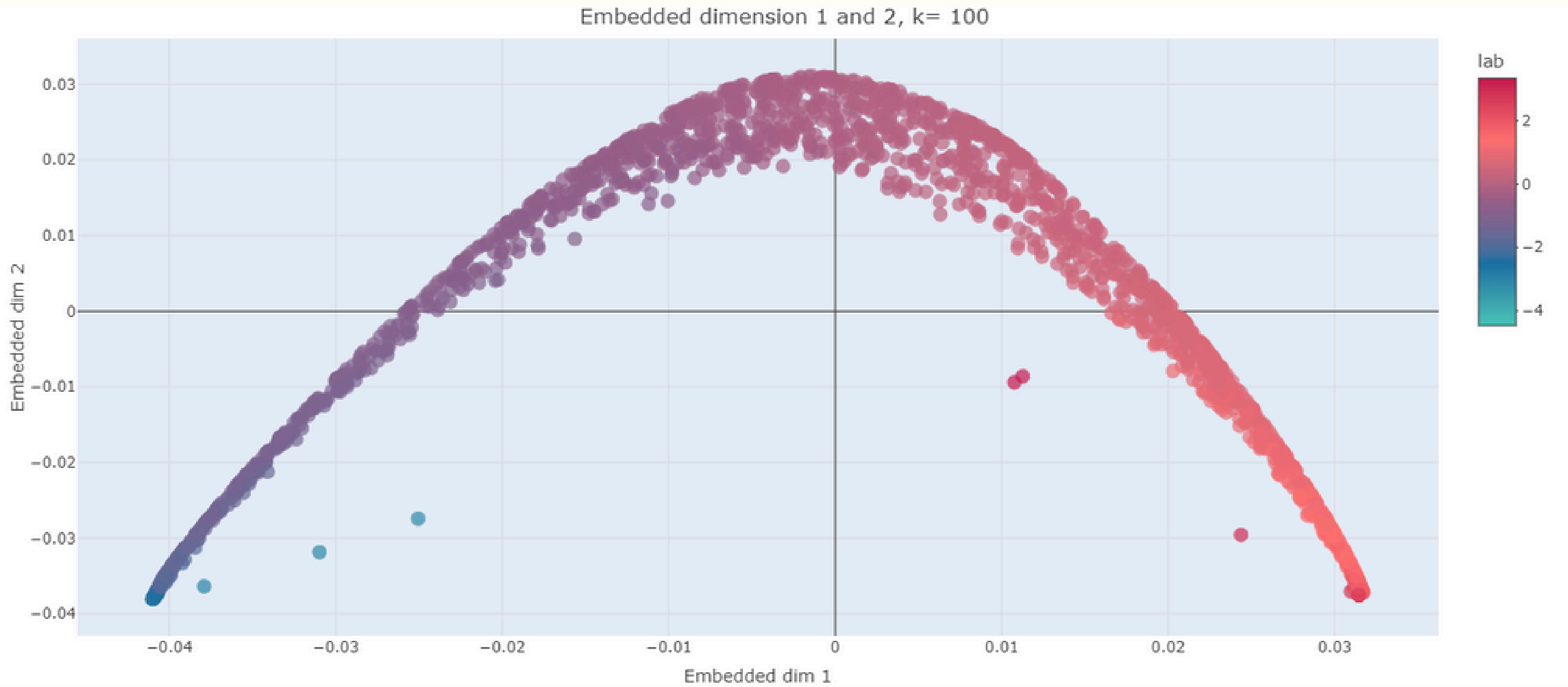


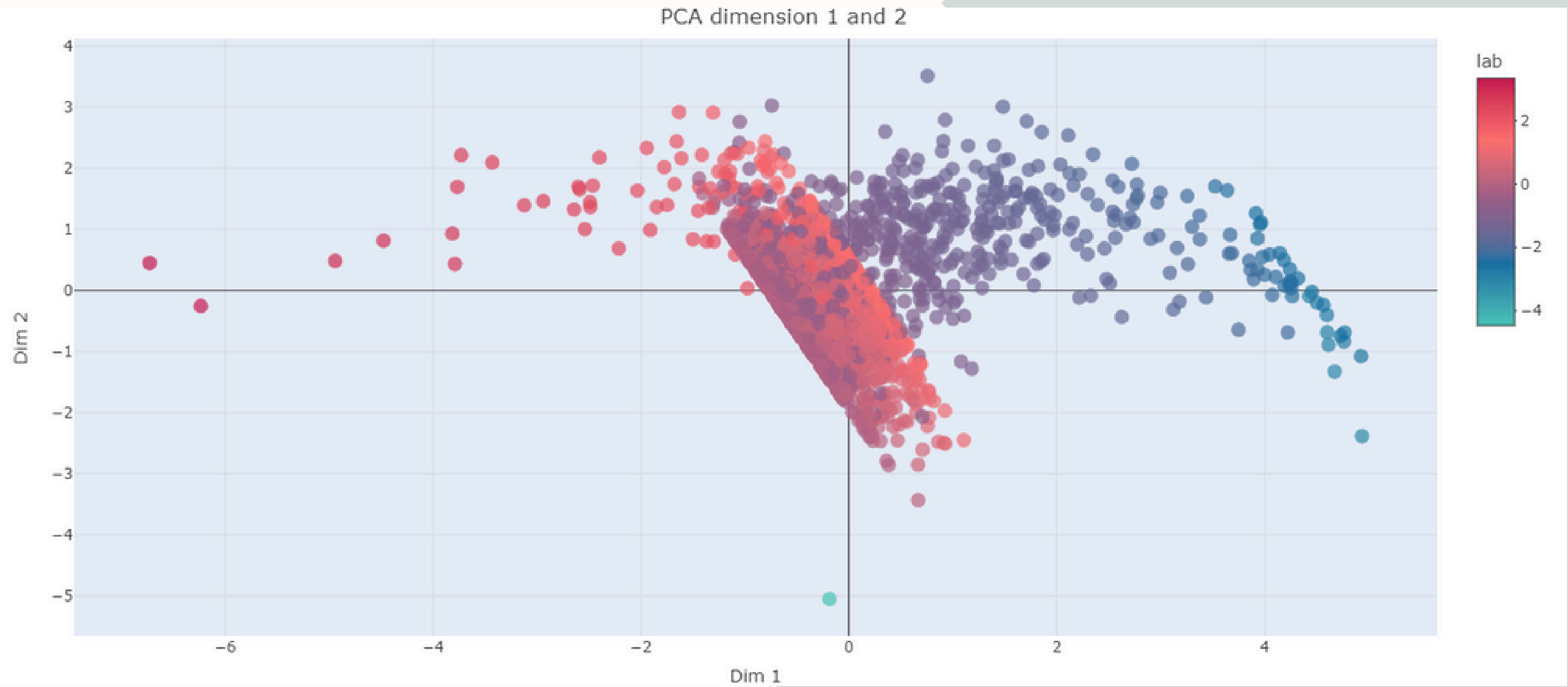
outlier makes the 1st dimension fail

Some conclusions

1. The value k in knn method sholdn't be too large if we want to observe the overall neighboring position pattern. But when we want to observe the position pattern of the subsets in the data, a larger k would be better.
2. The kernal weight wouldn't affect the transformation of the connected part, but would affect the unconnected peripheral points.
3. The enn method would be largely affected by the unconnected peripheral points and easily collapsed in the first dimension. If the set is not totally connected, the knn method may be better.

Compared with PCA and MDS





LE Result (KNN, k=100)

The neighboring trend: smooth curve

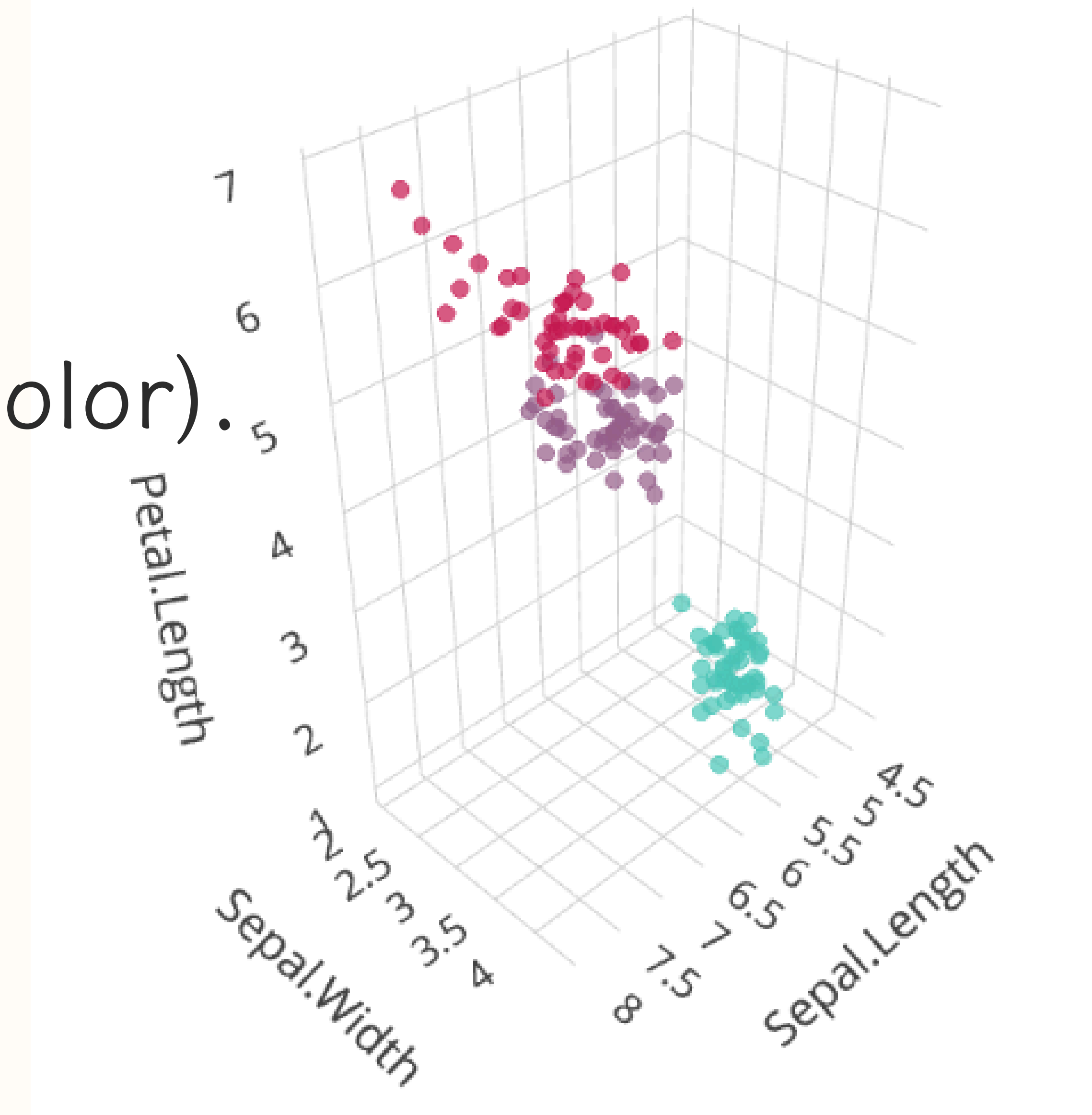
PCA & MDS Result

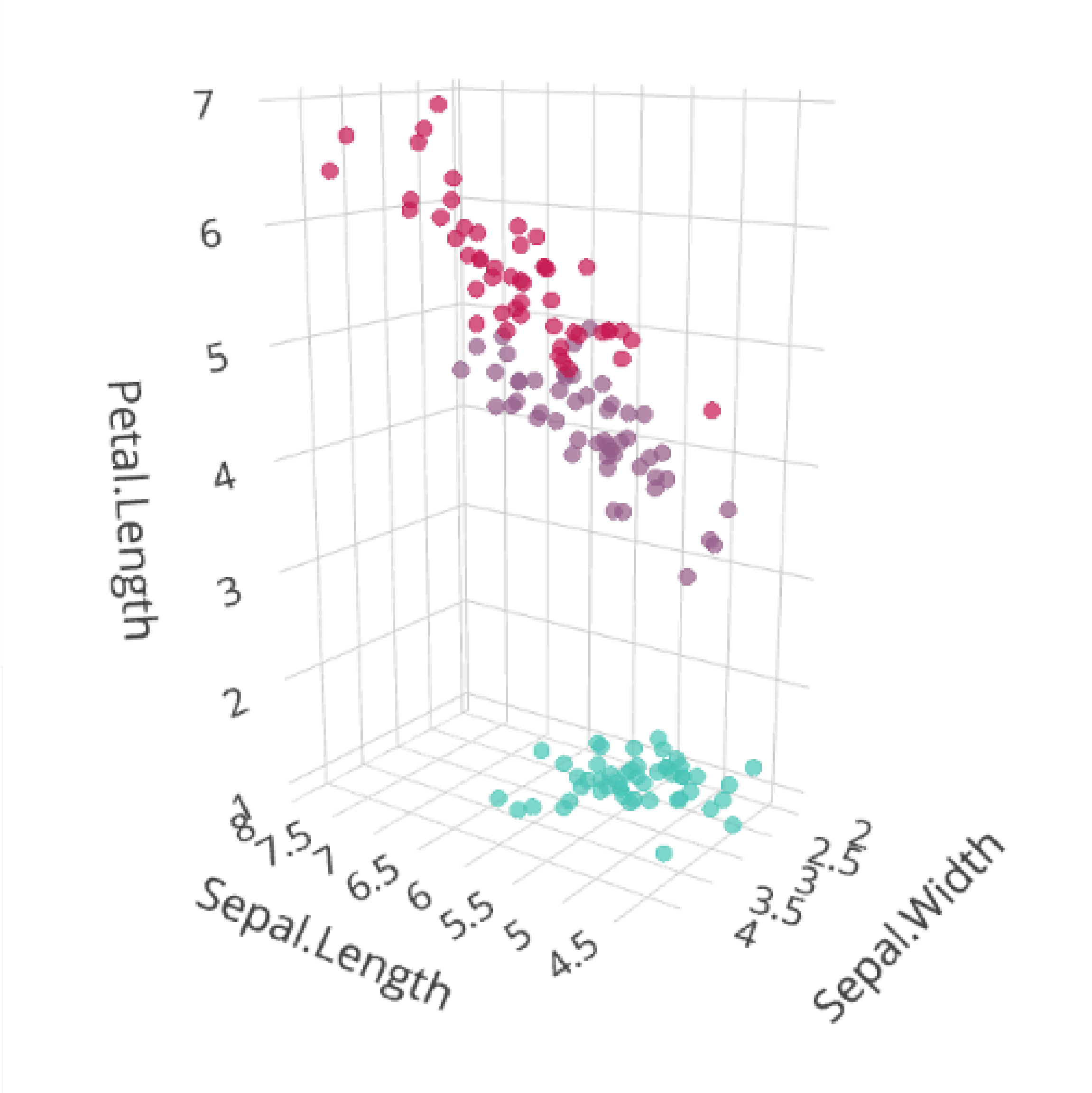
The overall trend: interception

Application: Iris Data

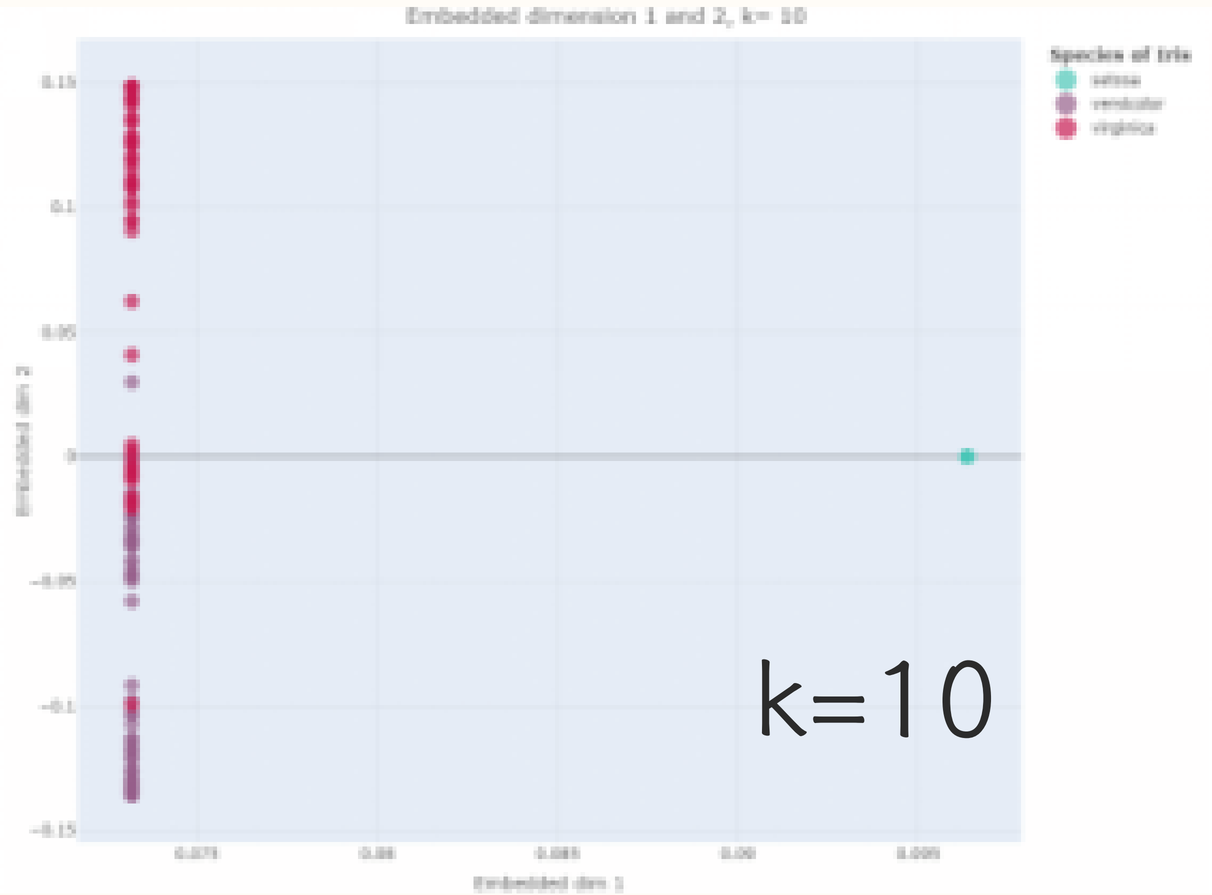
<u>Description</u>:
150 samples.
3 species of Iris.
(setosa, virginica and versicolor).

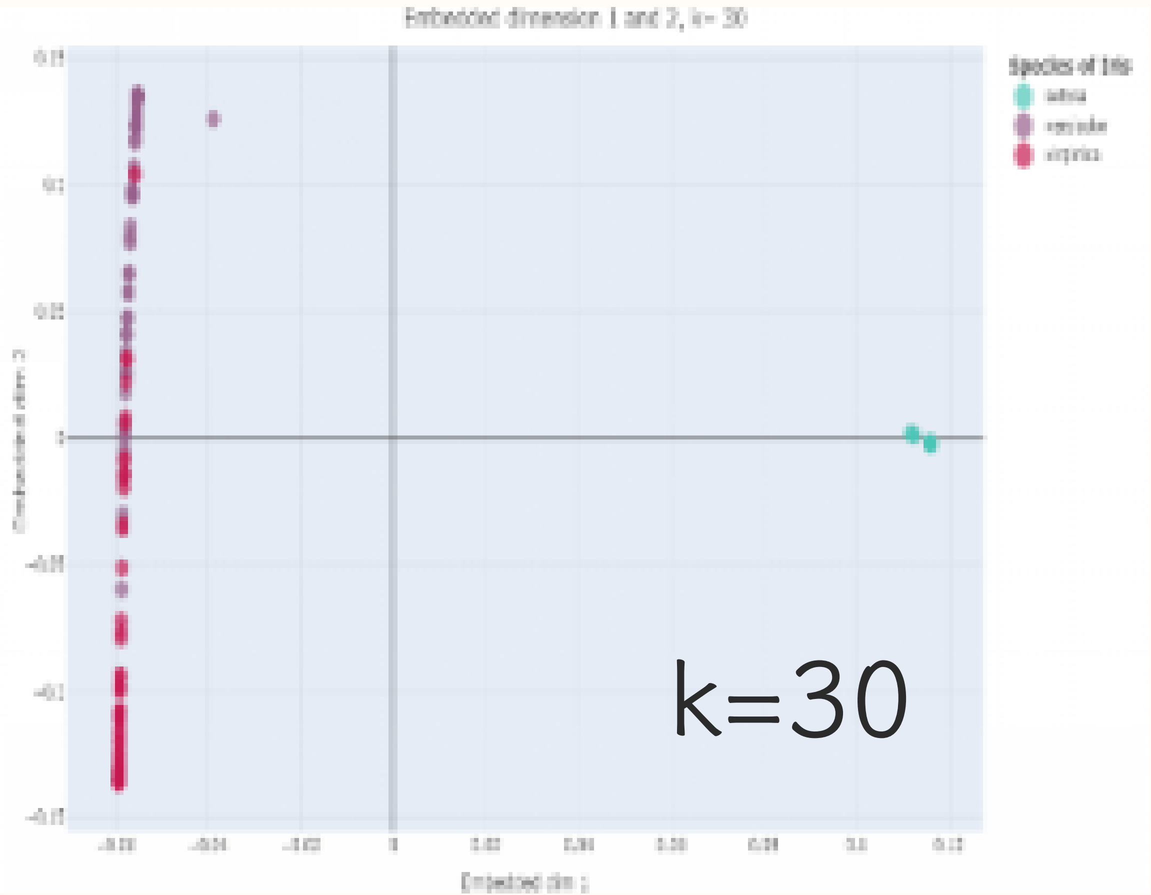
4 features(in centimeters):
Sepal length; Sepal width;
Petal length; Petal width;

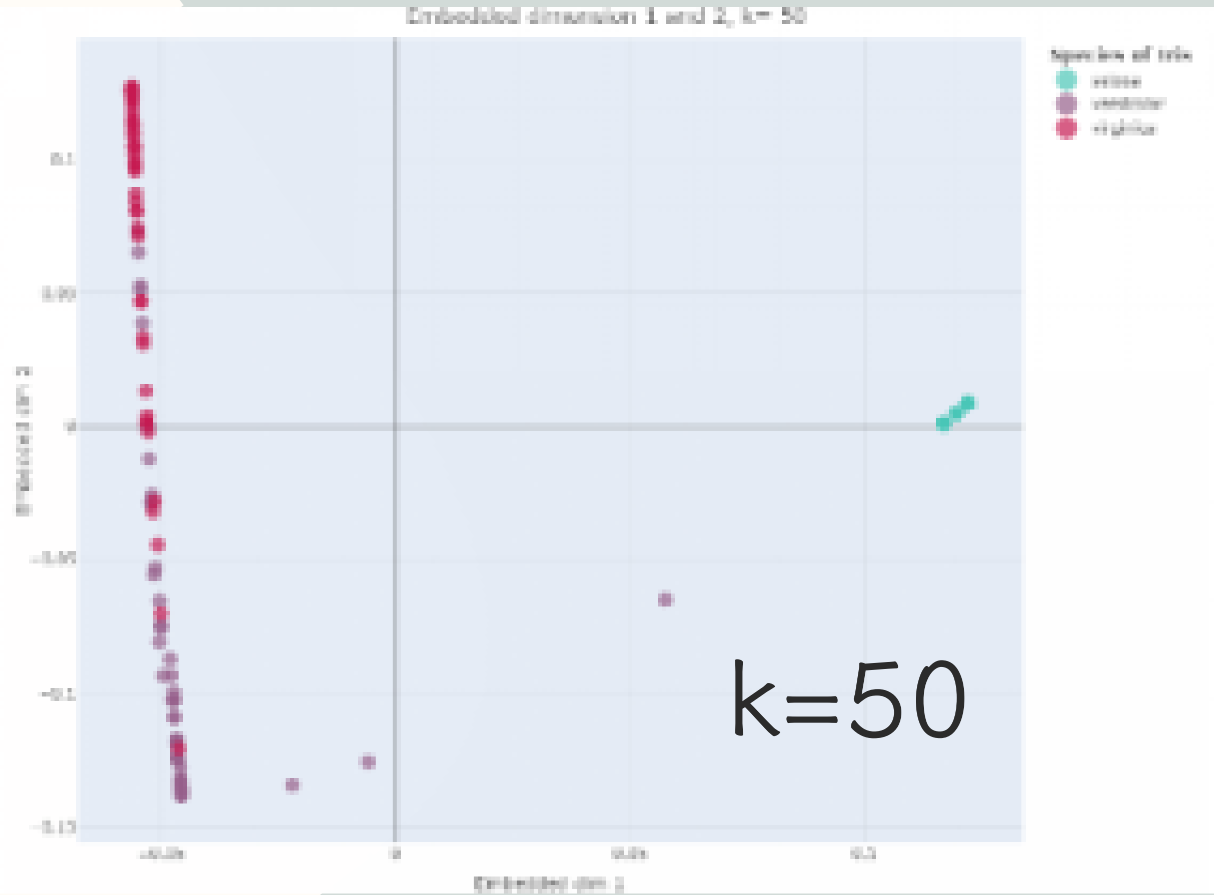


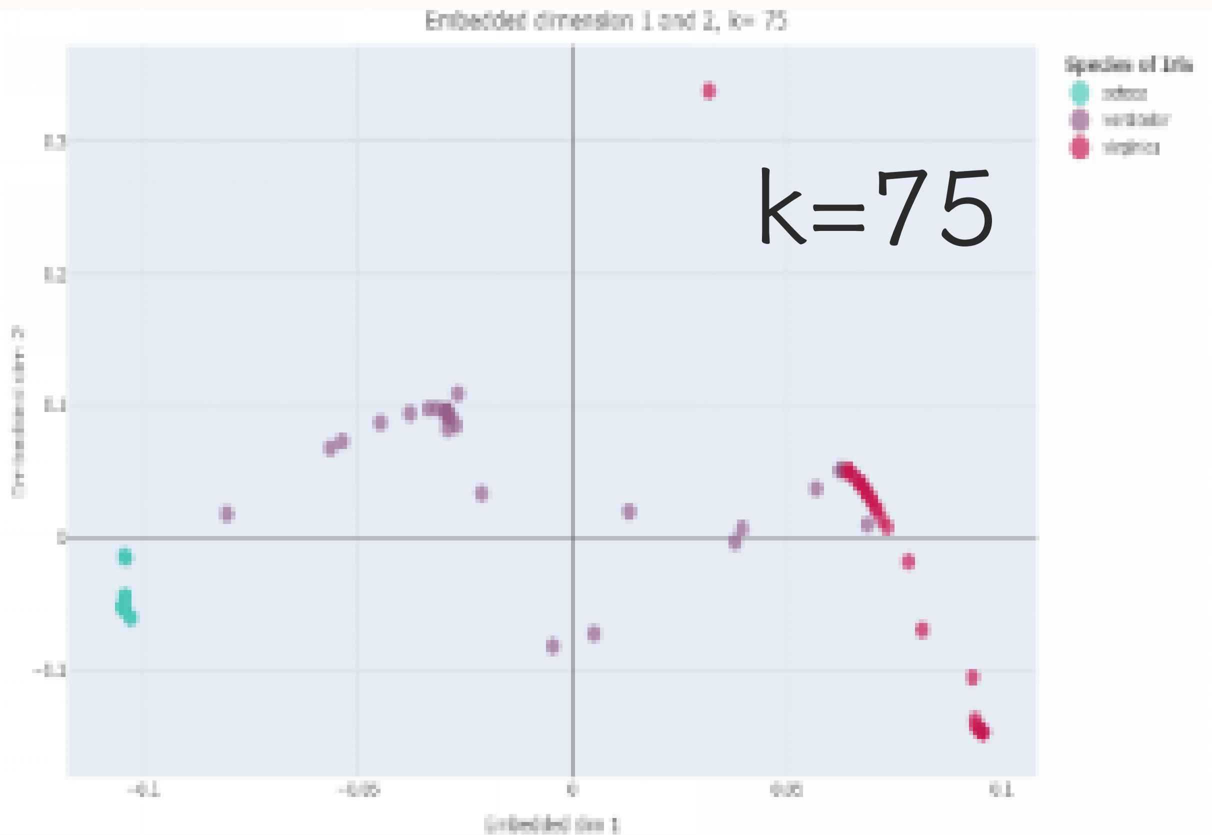


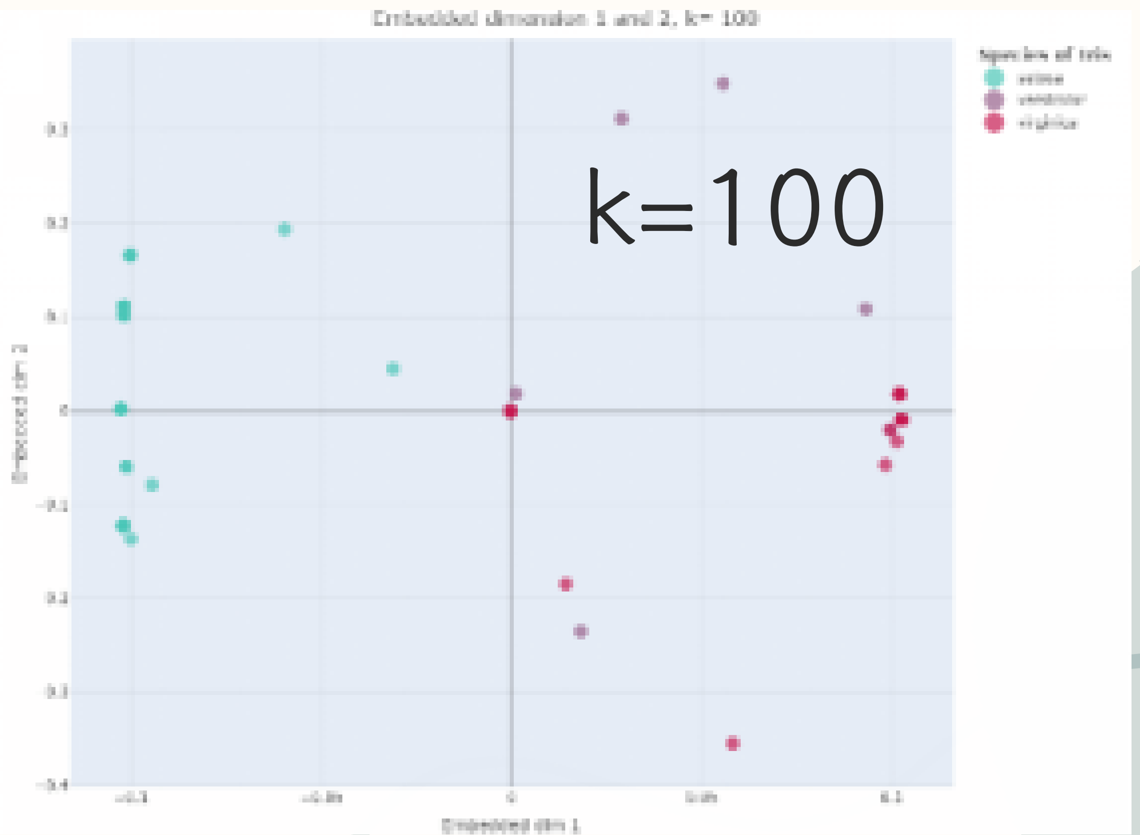
Application: Iris Data (trying different k)

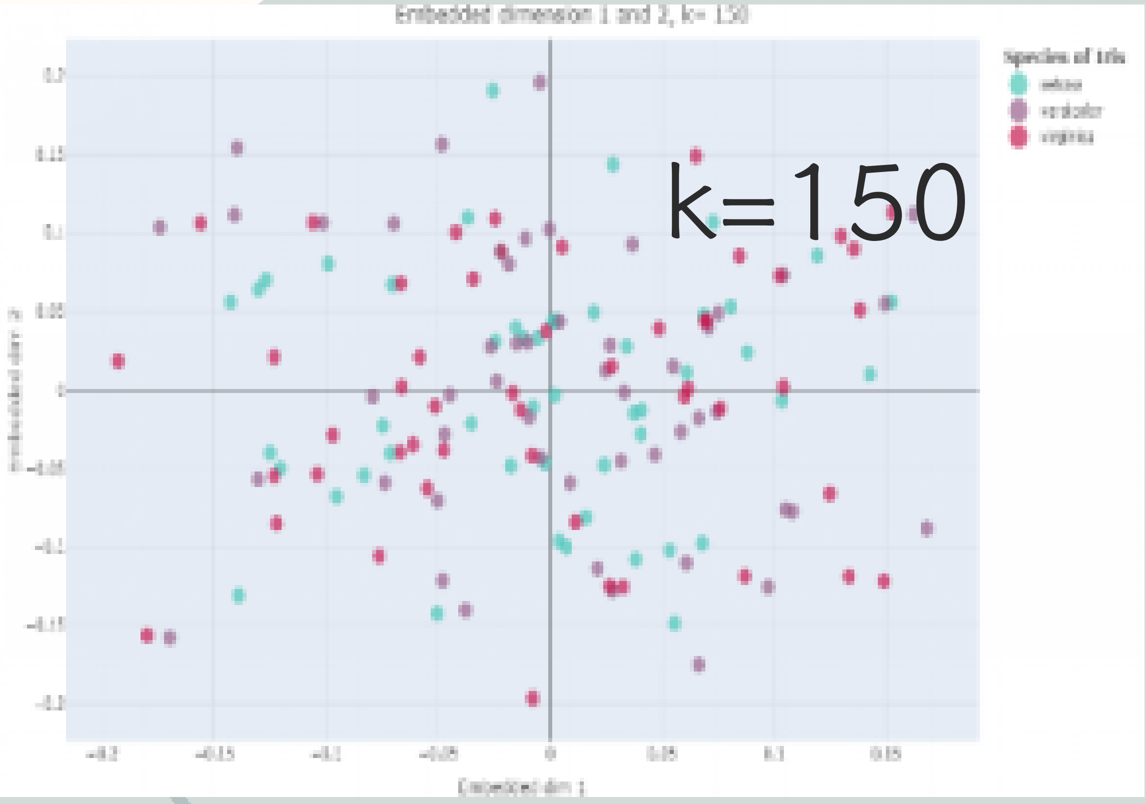












Application: Iris Data (trying different k)

What happened?

The first k eigen vector "collapse".

Lemma1: the eigen space corresponding to eigen value 0 in a fully connected graph's Laplacian is of 1 dimension, with 1n is its eigen vector. ↓

$$1_{m_i}^T L_i 1_{m_i} = 0$$

Property: If a graph has k disjoint connected components, its Laplacian matrix has k zero eigenvalues.

Proof:

Consider a graph with k connected components. We can get a block diagonal Laplacian. →

$$L = \begin{bmatrix} L_1 & 0 & \dots & 0 \\ 0 & L_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & L_k \end{bmatrix}$$

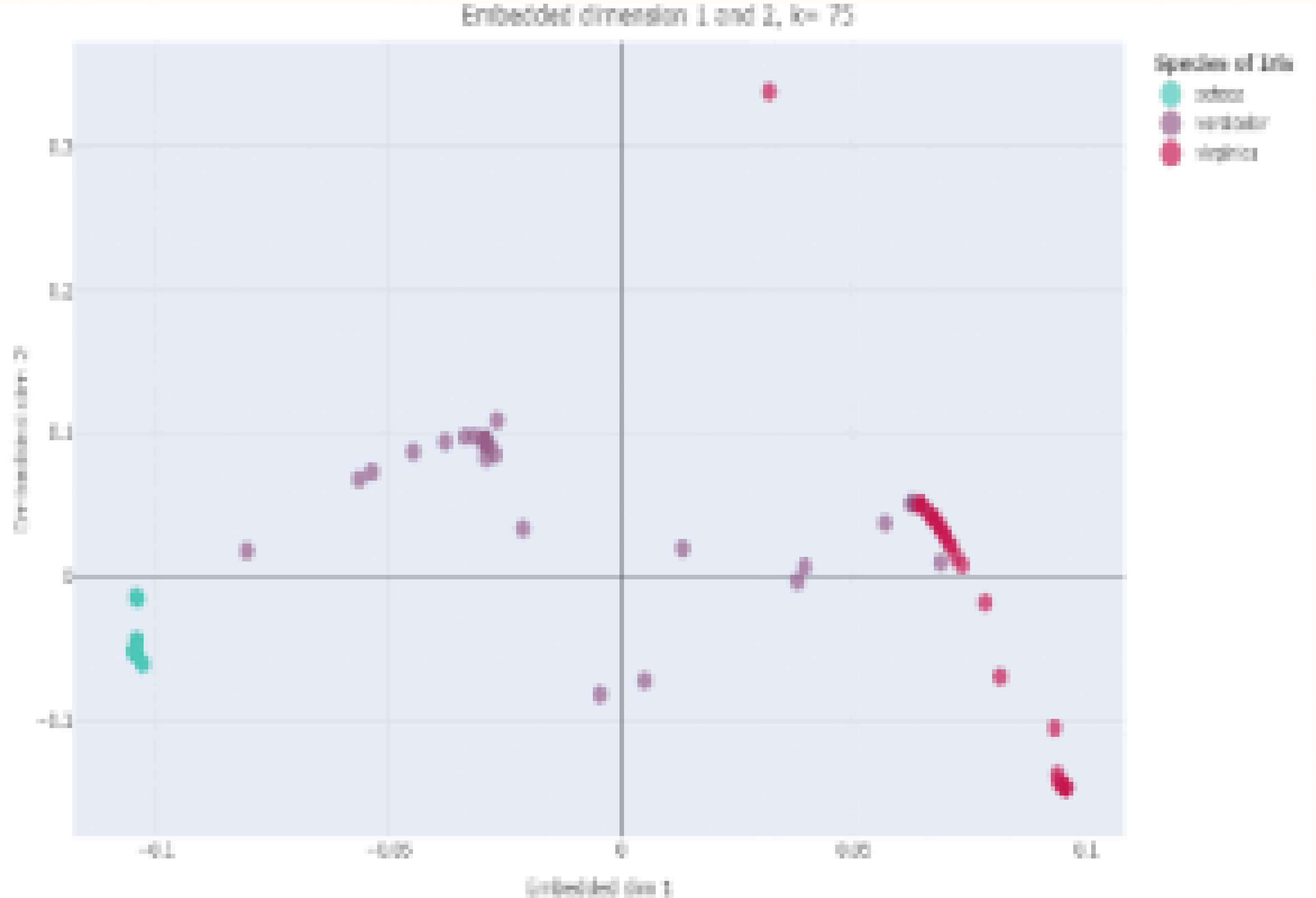
$$e^{(i)} = \frac{1}{\sqrt{m_i}} \begin{bmatrix} 0 \\ \dots \\ 0 \\ 1 \\ \dots \\ 1 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

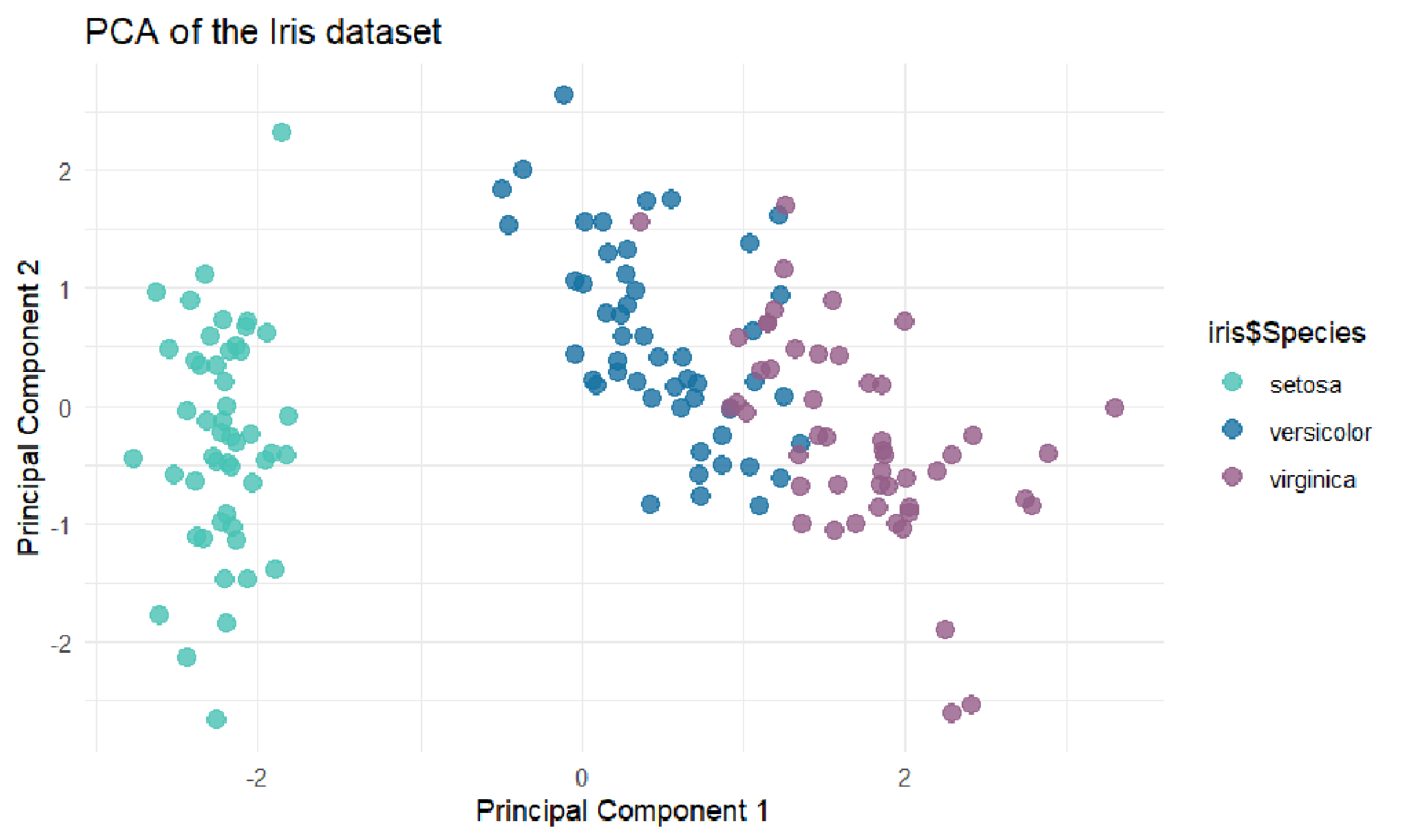
e(i)'s are L's eigenvector of eigenvalue 0.

$$e^{(i)} \perp e^{(j)}, i \neq j, i, j = 1, 2, \dots k$$

$$e^{(i)T} L e^{(i)} = 0$$

Compared with PCA





LE Result (KNN, k=75)

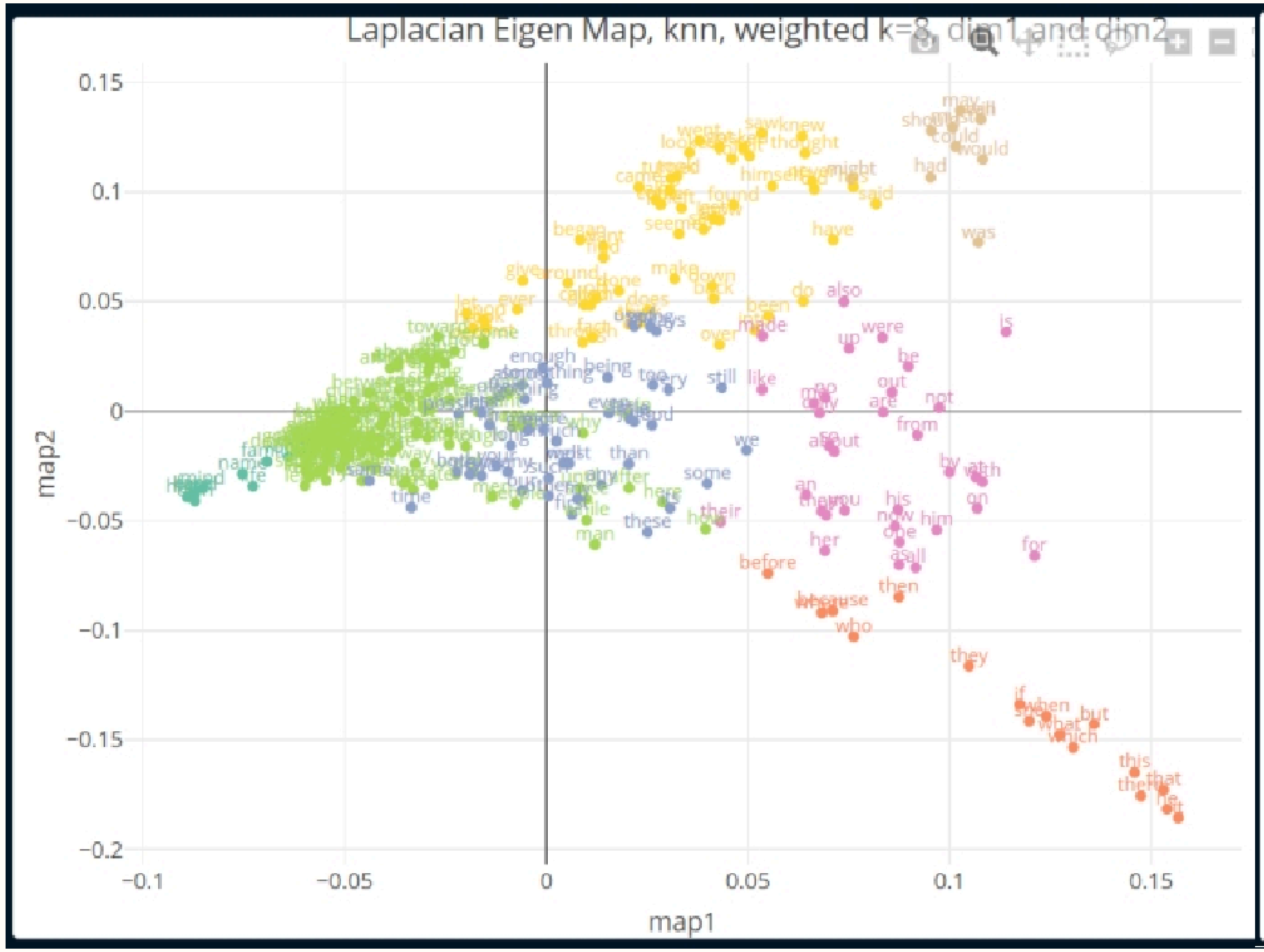
PCA Result

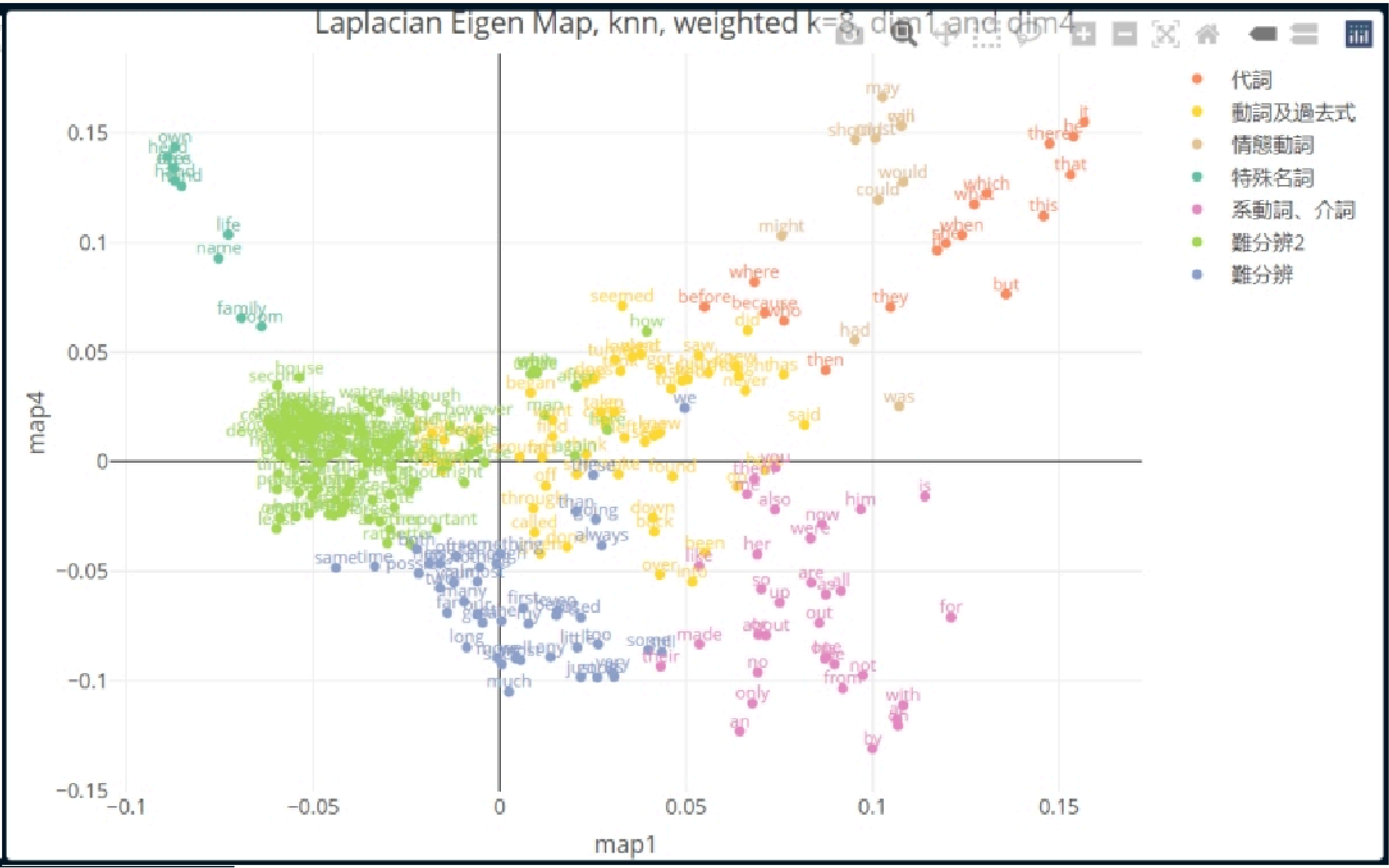
Application: Brown Corpus Data

<u>Description:</u>
Pick top 300 most frequent words in the Brown corpus.
Count their co-occurence times as attributes.
Get a 300×600 data frame.

<u>Goal:</u>
Do <u>clustering</u> among these words.
(First reduce dimensionality, then do kmeans)

Overview,knn,k=8,no weighted

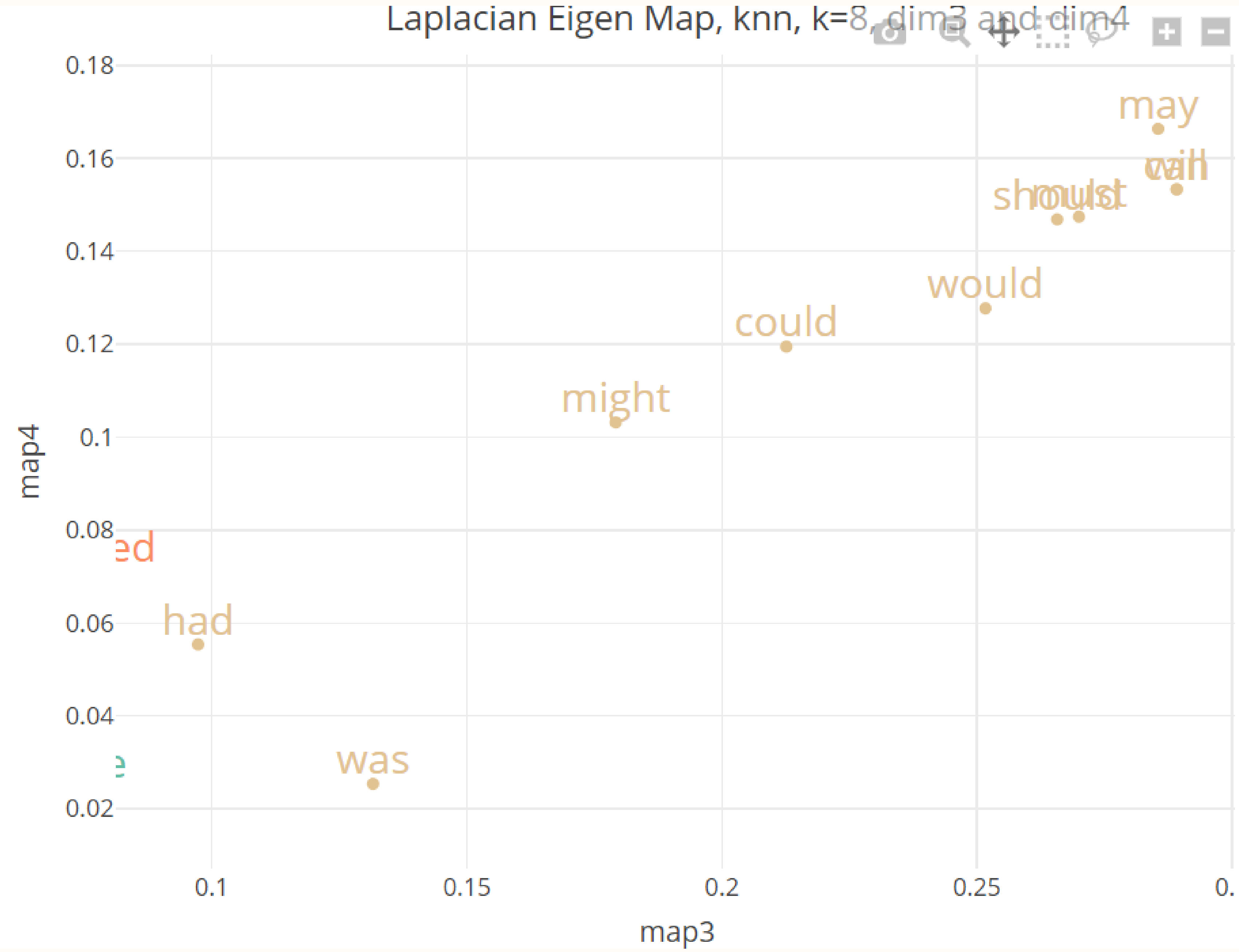


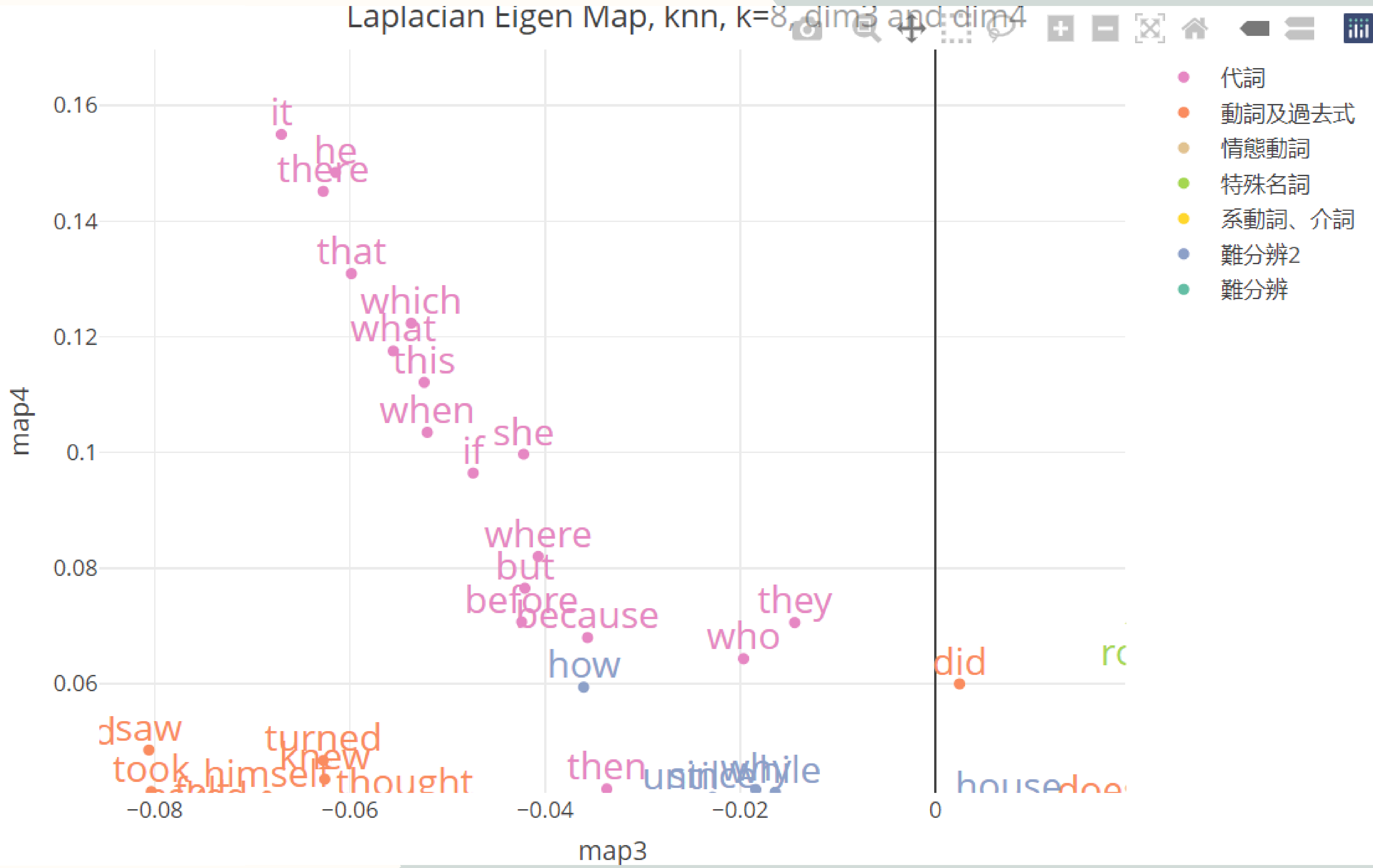


dim 1 and dim 2

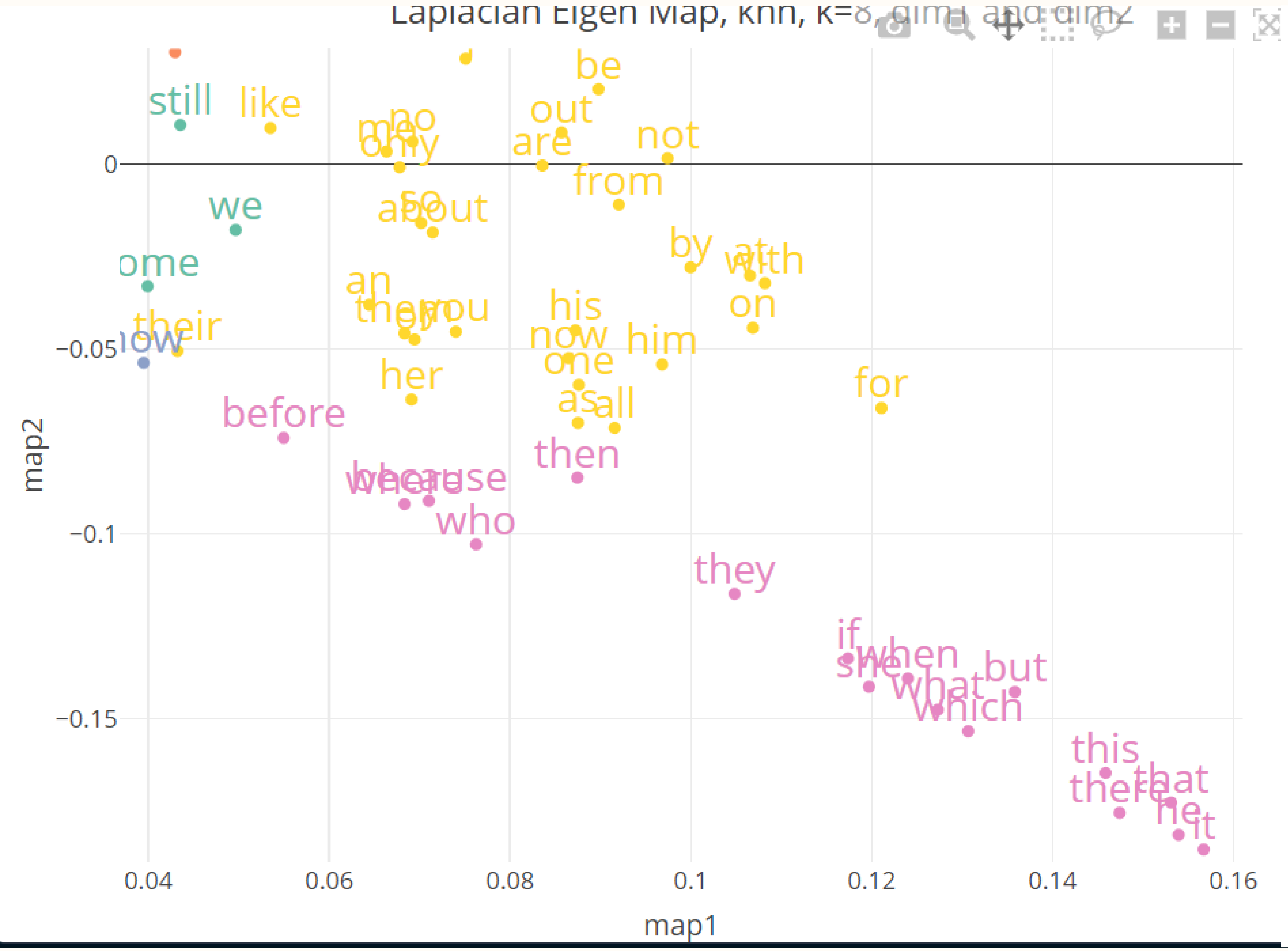
dim 1 and dim 4

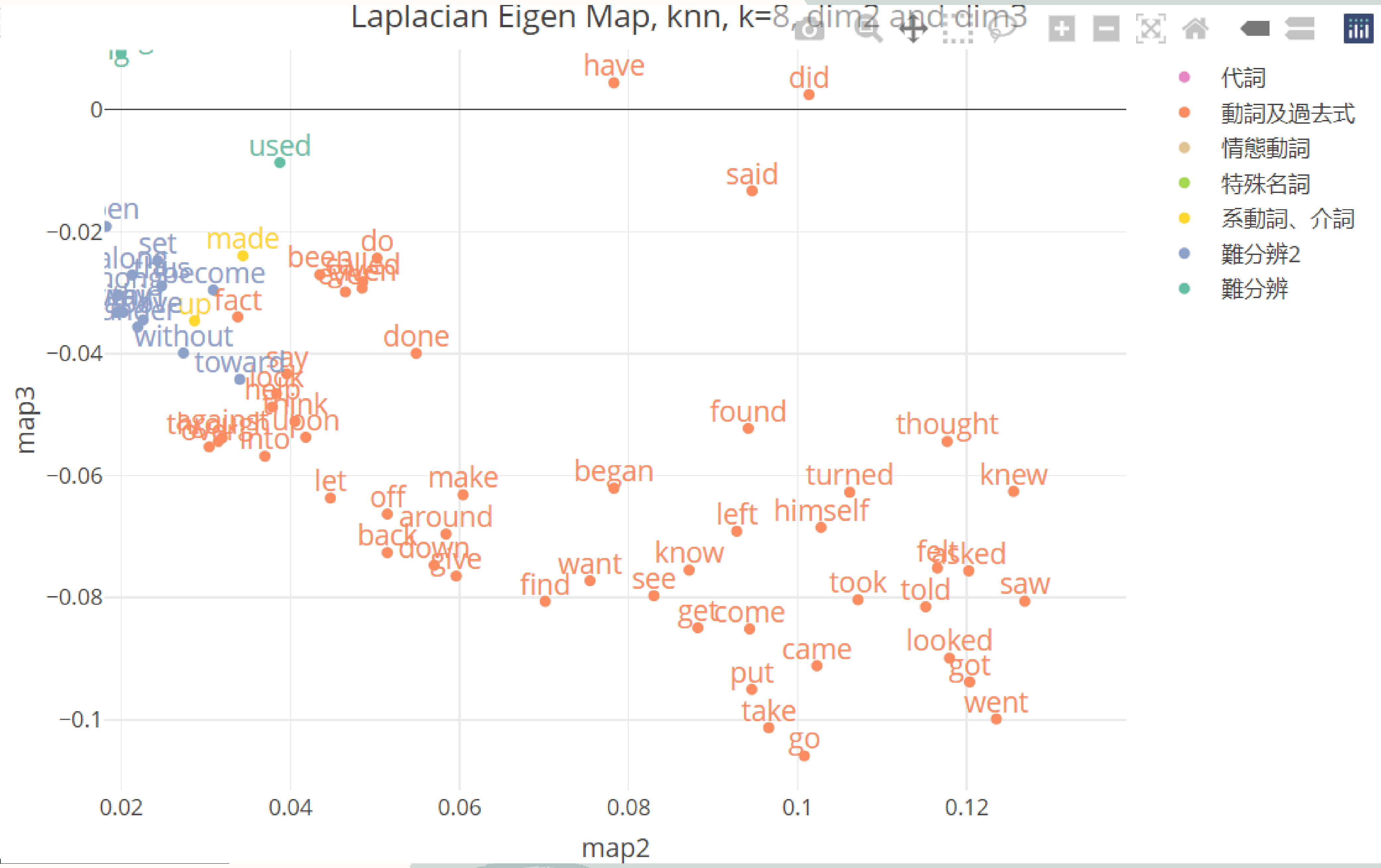
Modal verbs and Pronouns





Prepositions and verbs





Discussion: Graph Embedding

General Graph Embedding

$$\min_{Y} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}||y_i - y_j||^2$$
$$s.t.\ \mathbf{Y}^{\mathbf{T}}\mathbf{B}\mathbf{Y} = \mathbf{I}$$

$$\min_{\mathbf{Y}} tr(\mathbf{Y}^{\mathbf{T}}\mathbf{L}\mathbf{Y})$$
$$s.t.\ \mathbf{Y}^{\mathbf{T}}\mathbf{B}\mathbf{Y} = \mathbf{I}$$

$$\mathbb{R}^{n\times n} \ni \frac{\partial \mathcal{L}}{\partial \mathbf{Y}} = 2\mathbf{L}\mathbf{Y} - 2\mathbf{B}\mathbf{Y}\mathbf{\Lambda} \stackrel{set}{=} \mathbf{0}$$
$$\Rightarrow \mathbf{L}\mathbf{Y} = \mathbf{B}\mathbf{Y}\mathbf{\Lambda}$$

↑ Generalized eigen value problem of (L, B)

Linearized Graph Embedding

$$\min_{\mathbf{U}} tr(\mathbf{U}^{\mathbf{T}}\mathbf{X}\mathbf{L}\mathbf{X}^{\mathbf{T}}\mathbf{U})$$
$$s.t.\ \mathbf{U}^{\mathbf{T}}\mathbf{X}\mathbf{B}\mathbf{X}^{\mathbf{T}}\mathbf{U} = \mathbf{I}$$

↑ Generalized eigen value problem of (XLX^T, XBX^T)

PCA

$$\max_{\mathbf{U}} tr(\mathbf{U}^{\mathbf{T}}\mathbf{X}\mathbf{X}^{\mathbf{T}}\mathbf{U})$$
$$s.t.\ \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I}$$

↑ Generalized igen value problem of (XX^T, I)

Classical MDS

Can you prove the result?

→

$$\max_{\mathbf{U}} tr(\mathbf{Y}^{\mathbf{T}}\mathbf{K}\mathbf{Y})$$
$$s.t.\ \mathbf{Y}^{\mathbf{T}}\mathbf{Y} = \mathbf{I}$$
$$\mathbf{K} = -\frac{1}{2}\mathbf{H}\mathbf{D}\mathbf{H}$$
$$\mathbf{H} := \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathbf{T}}, \mathbf{D}_{ij} = (d(x_i, x_j))^2$$

Connections to Clustering

Remark: Clustering the original data is equivalent to finding a partition of the associated graph:$V = A_1 \cup A_2 \cup \cdots \cup A_c$where $A_i \cap A_j = \emptyset$ for $i \neq j$

Definition:

let A,B be subsets of V

(1) $Vol(A) = \sum_{i \in A} d_i$

(2) $W(A,B) = \sum_{i \in A, j \in B} w_{ij}$

(3) if $B = \bar{A}$, $W(A,B)$ is call a cut $Cut(A,B) = W(A,B) = \sum_{i \in A, j \in B} w_{ij}$

(4) $NCut(A,B) = Cut(A,B)(\frac{1}{Vol(A)} + \frac{1}{Vol(B)})$

Connections to Clustering

Define $\mathbf{f} = \frac{1}{a}\mathbf{1}_A - \frac{1}{b}\mathbf{1}_B$ with:

$$f_i = \begin{cases} \frac{1}{a} & , i \in A \\ -\frac{1}{b} & , i \in B \end{cases}$$

a = Vol(A)
b = Vol(B)

For this f, we have:

$$\begin{aligned} f^T L f &= \sum_{i,j} (f_i - f_j)^2 W_{ij} \\ &= \sum_{i \in A, j \in B} W_{ij} \left(\frac{1}{a} + \frac{1}{b}\right)^2 \\ &= Cut(A,B)\left(\frac{1}{a} + \frac{1}{b}\right)^2 \end{aligned}$$

$$\begin{aligned} f^T D f &= \sum_i f_i^2 d_{ii} \\ &= \sum_{i \in A} \frac{1}{a^2} d_{ii} + \sum_{i \in B} \frac{1}{b^2} d_{ii} \\ &= Vol(A)\frac{1}{a^2} + Vol(B)\frac{1}{b^2} \\ &= \frac{1}{a} + \frac{1}{b} \end{aligned}$$

$$\implies \frac{f^T L f}{f^T D f} = Cut(A,B)\left(\frac{1}{a} + \frac{1}{b}\right) = NCut(A,B)$$

We also have

$$f^T D \mathbf{1} = \sum_i f_i d_{ii} = \frac{1}{a} Vol(A) - \frac{1}{b} Vol(B) = 0$$

Connections to Clustering

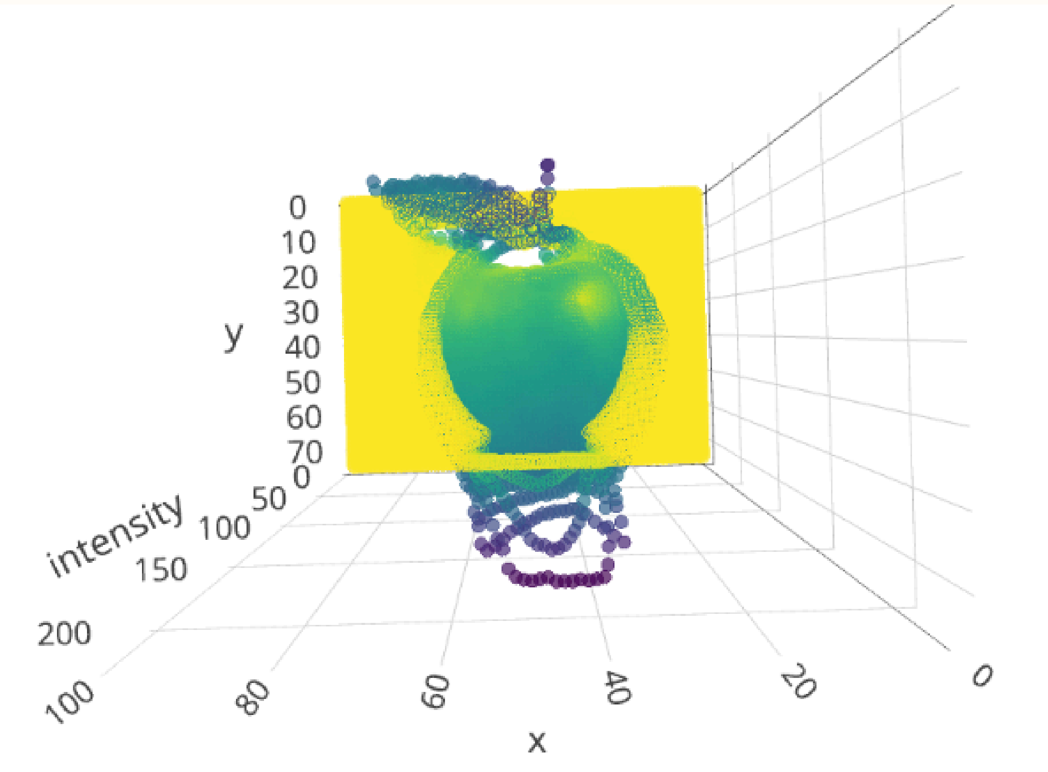
Therefore, we can obtain the following equivalent problem

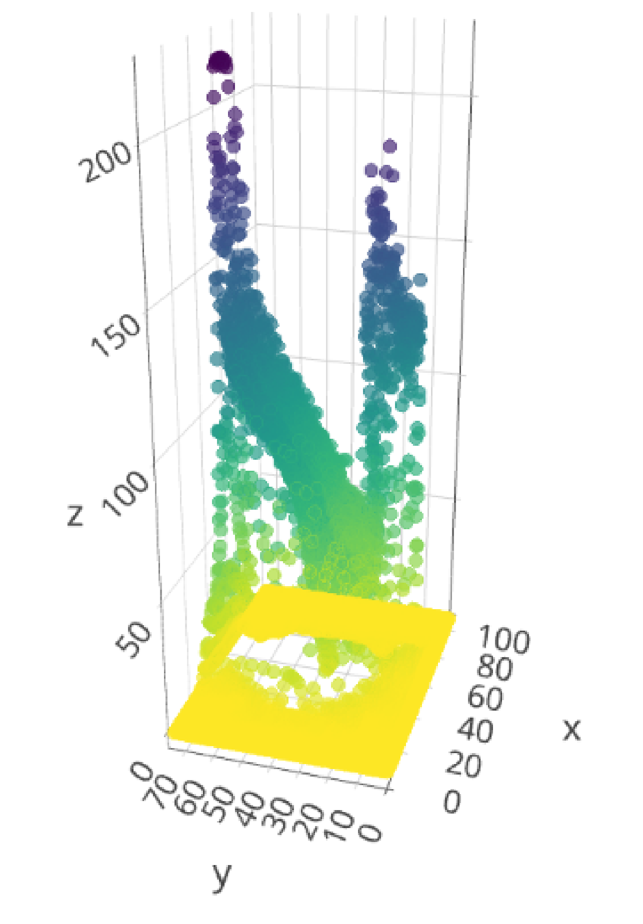
$$\min_{\substack{A \cup B = V \\ A \cap B = \emptyset}} NCut(A, B) \Longleftrightarrow \min_{\substack{\mathbf{f} \neq 0 \\ \mathbf{f}^T D\mathbf{1} = 0}} \frac{\mathbf{f}^T L\mathbf{f}}{\mathbf{f}^T D\mathbf{f}}$$

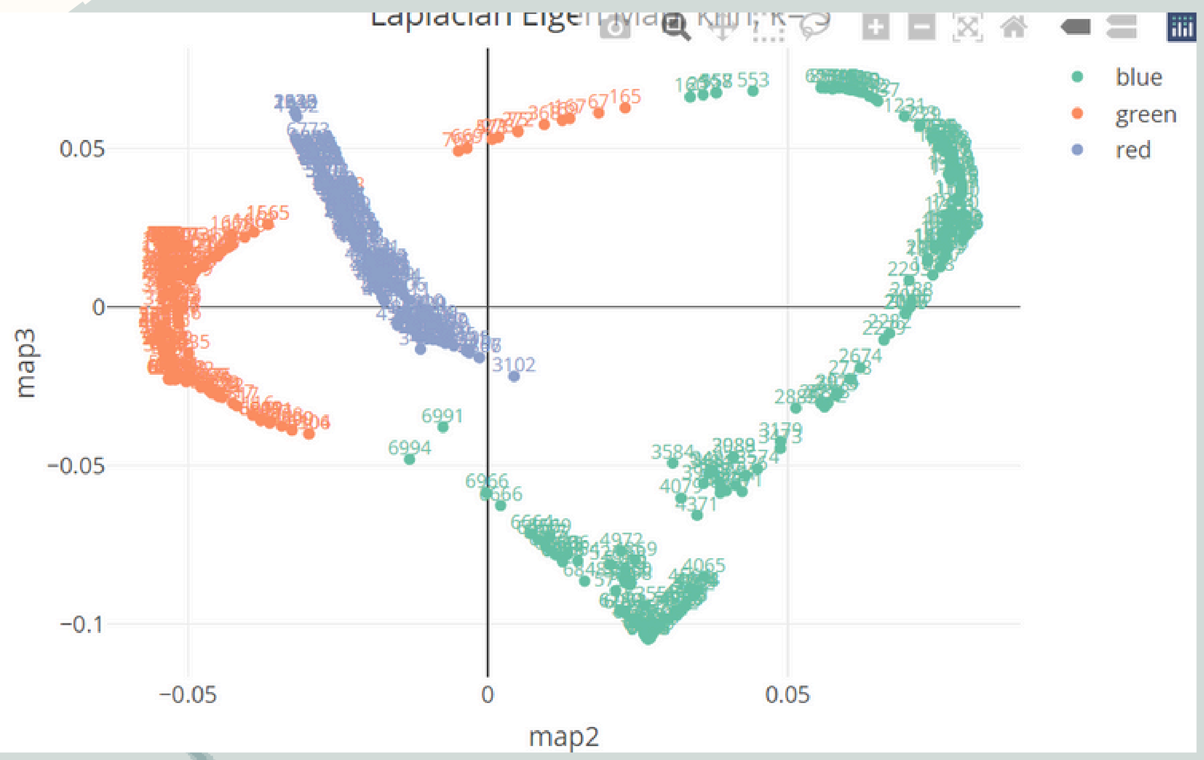
The minimizer f^* represents an approximate solution to the Ncut problem, providing information about the labels of the data.

Questions

1. Why sometimes dimension 1 collapse?
2. For the given data “data_apple_pixel.csv”, apply LE method to reduce its dimensionality to 2. You can also try to do image segmentation.







Thanks!

Reference

Belkin, Mikhail, and Partha Niyogi. "Laplacian eigenmaps and spectral techniques for embedding and clustering." Advances in neural information processing systems 14 (2001).

Belkin, Mikhail, and Partha Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation." Neural computation 15, no. 6 (2003): 1373-1396.

Ghojogh, Benyamin, Ali Ghodsi, Fakhri Karray, and Mark Crowley. "Laplacian-based dimensionality reduction including spectral clustering, Laplacian eigenmap, locality preserving projection, graph embedding, and diffusion map: Tutorial and survey." arXiv preprint arXiv:2106.02154 (2021).

Tai, Mariko, Mineichi Kudo, Akira Tanaka, Hideyuki Imai, and Keigo Kimura. "Kernelized supervised laplacian eigenmap for visualization and classification of multi-label data." Pattern Recognition 123 (2022): 108399.

Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, Surya Ganguli. “A theory of multineuronal dimensionality, dynamics and measurement”. bioRxiv 214262; doi: https://doi.org/10.1101/214262

Reference

Joram Keijser,Low-dimensional manifolds in neuroscience and evolution,https://joramkeijser.github.io/2022/02/04/manifolds.html
Muni Sreenivas Pydi. "What's the intuition behind a Laplacian matrix? " https://www.quora.com/Whats-the-intuition-behind-a-Laplacian-matrix-Im-not-so-much-interested-in-mathematical-details-or-technical-applications-Im-trying-to-grasp-what-a-laplacian-matrix-actually-represents-and-what-aspects-of-a-graph-it-makes-accessible/answer/Muni-Sreenivas-Pydi(2021).
Guangliang Chen. “Laplacian Eigenmaps.” https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec12laplacian.pdf.
Juan Orduz. “On Laplacian Eigenmaps for Dimensionality Reduction.” https://www.youtube.com/watch?v=U31TIICsHiA
“Laplacian Eigenmaps.”https://www.cs.cmu.edu/~aarti/Class/10701/slides/Lecture21_1.pdf