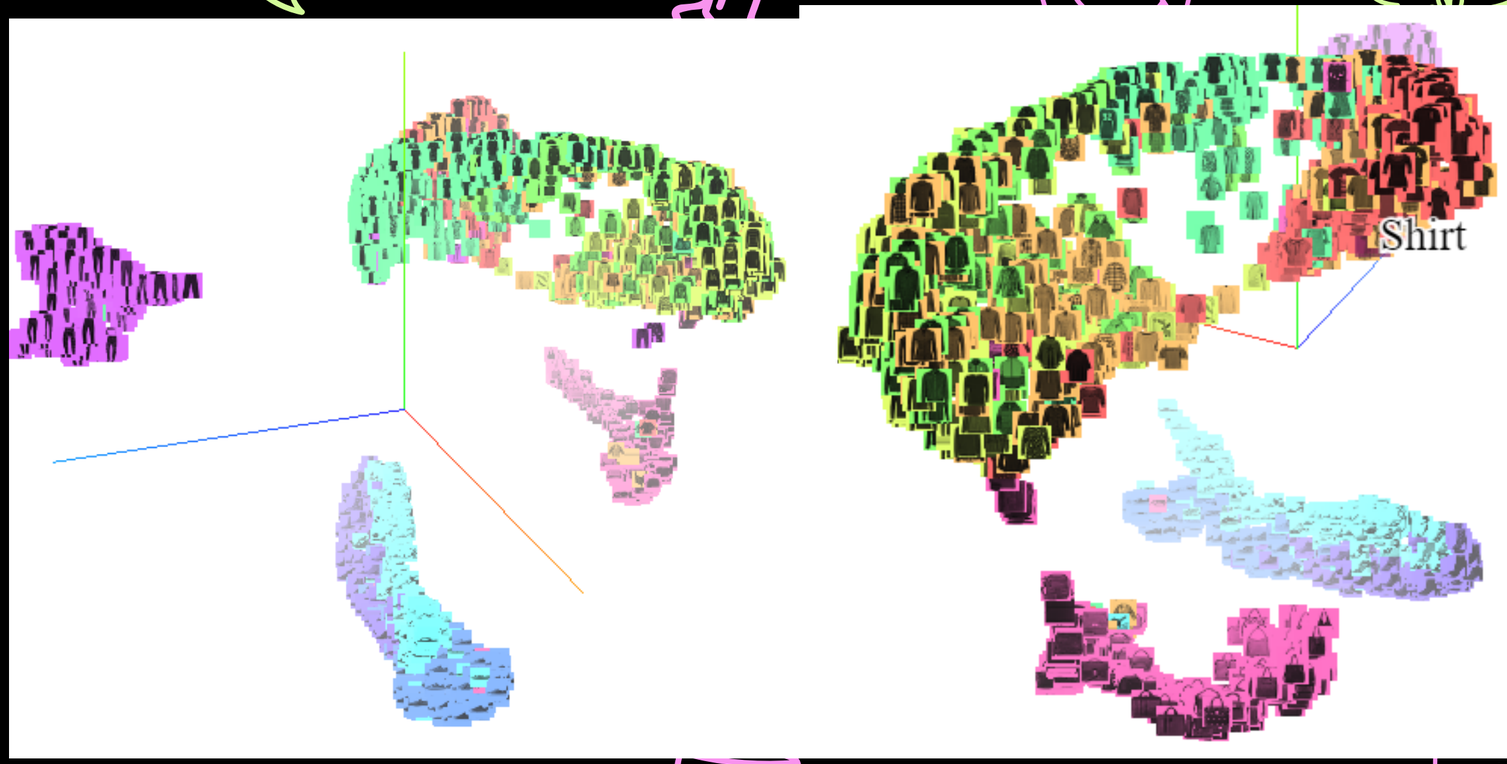
UMAP

Machile learning

Team 6

What is UMAP

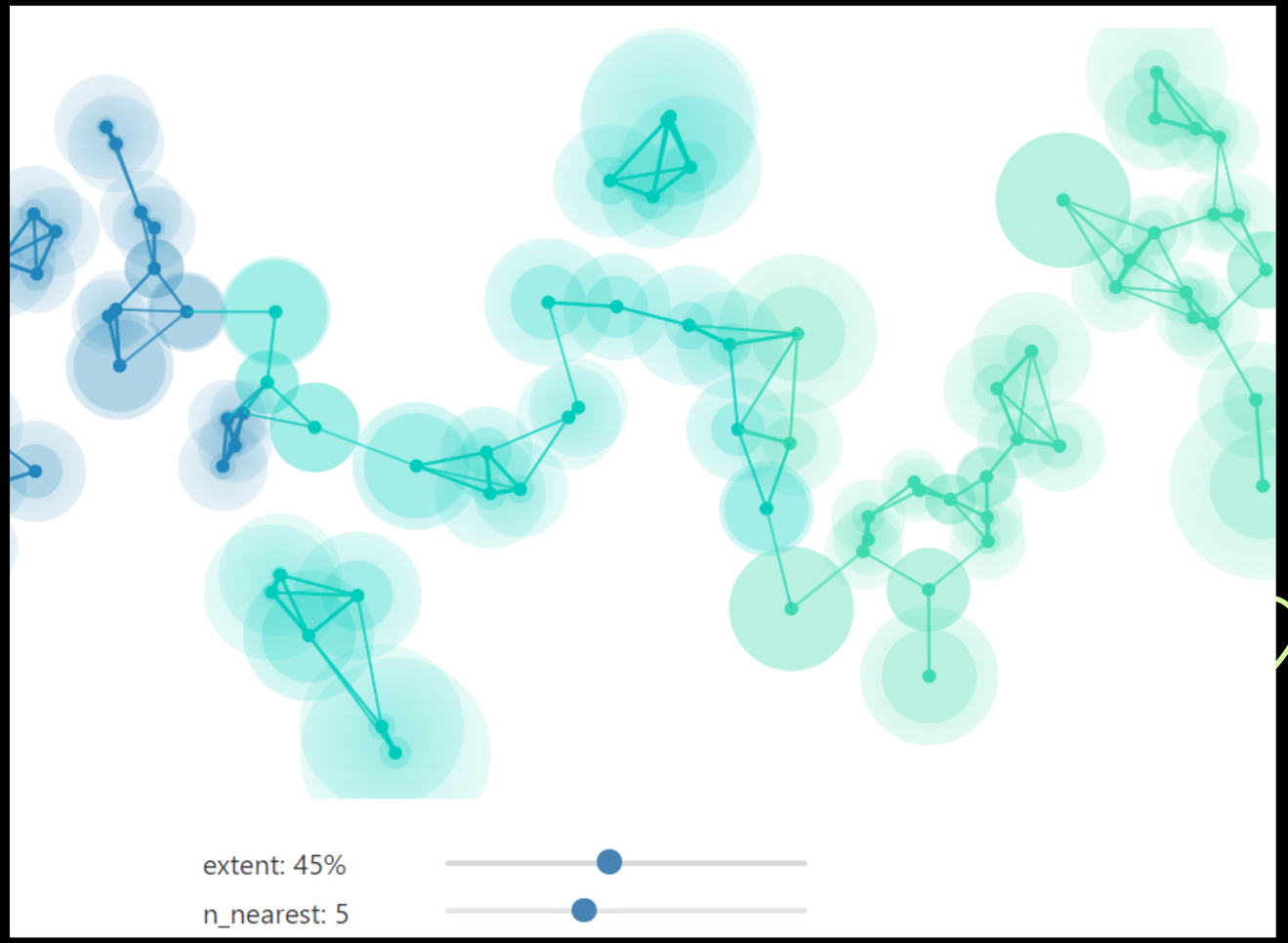
UMAP, or Uniform Manifold Approximation and Projection, is a dimensionality reduction technique used in machine learning. It helps us visualize high-dimensional data by preserving its underlying structure.



T-shirt/top | Shirt | Pullover | Coat | Dress | Sandal | Sneaker | Ankle boot | Trouser | Bag

Basic principles of UMAP

UMAP operates based on fundamental principles such as neighborhoods and neighborhood graphs. It minimizes the graph's energy to find an optimal representation.

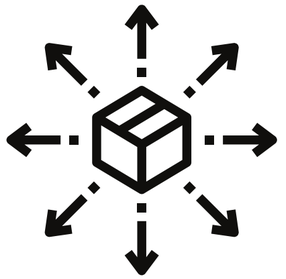


Steps to Implement UMAP

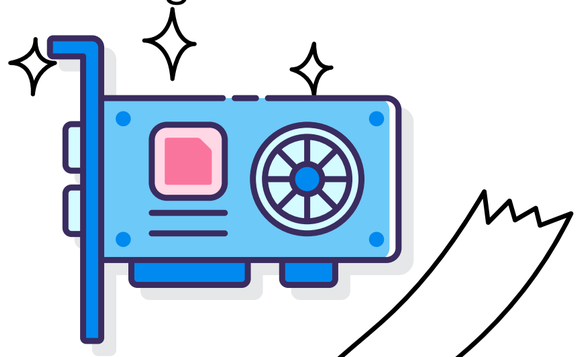
Implementing UMAP involves three main steps

1 Creating a neighborhood graph

2 Optimizing point distribution



3 Transforming the data

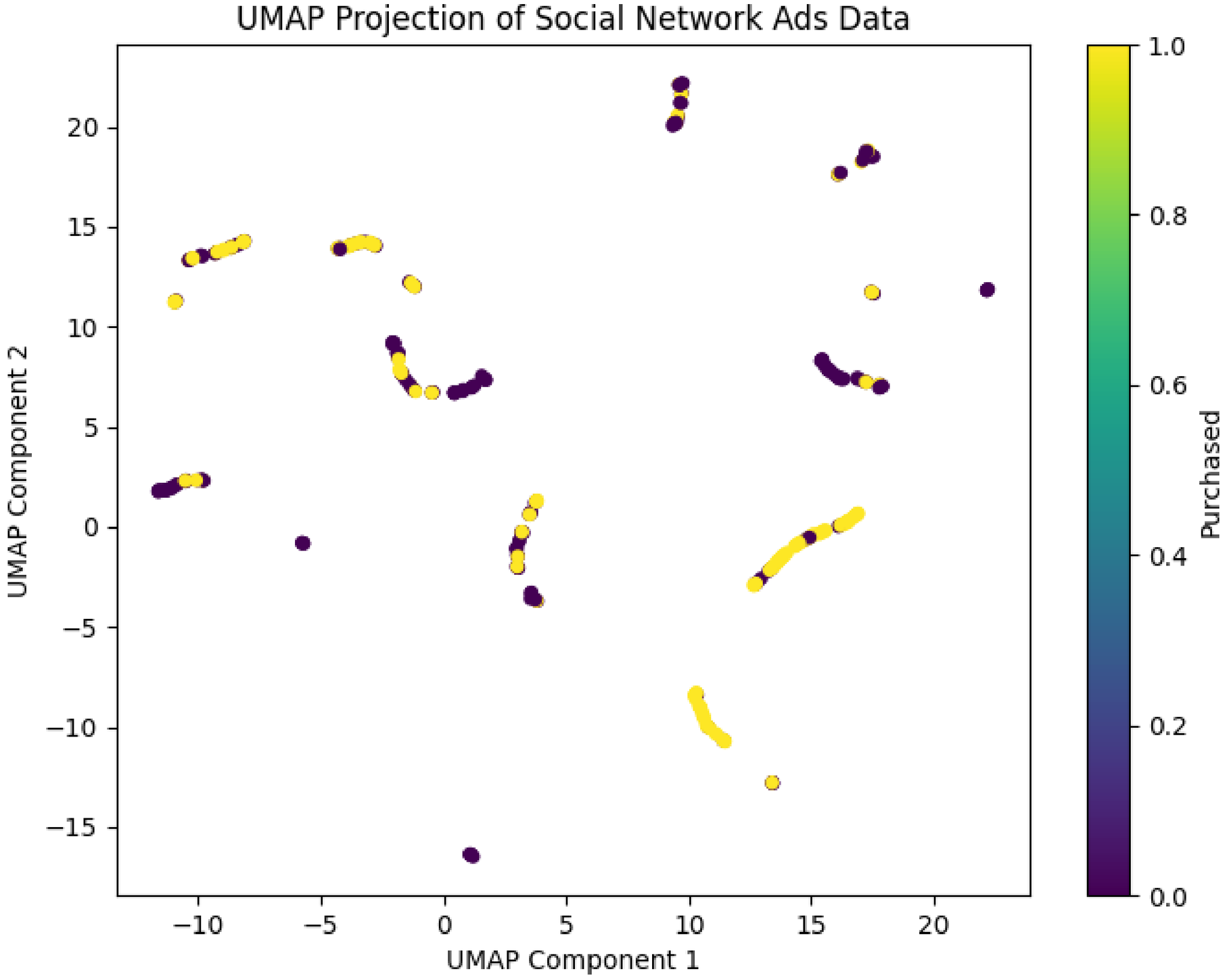


Advantages of UMAP

preservation of global structure

visualizing high-dimensional data preprocessing

efficient dimensionality reduction.

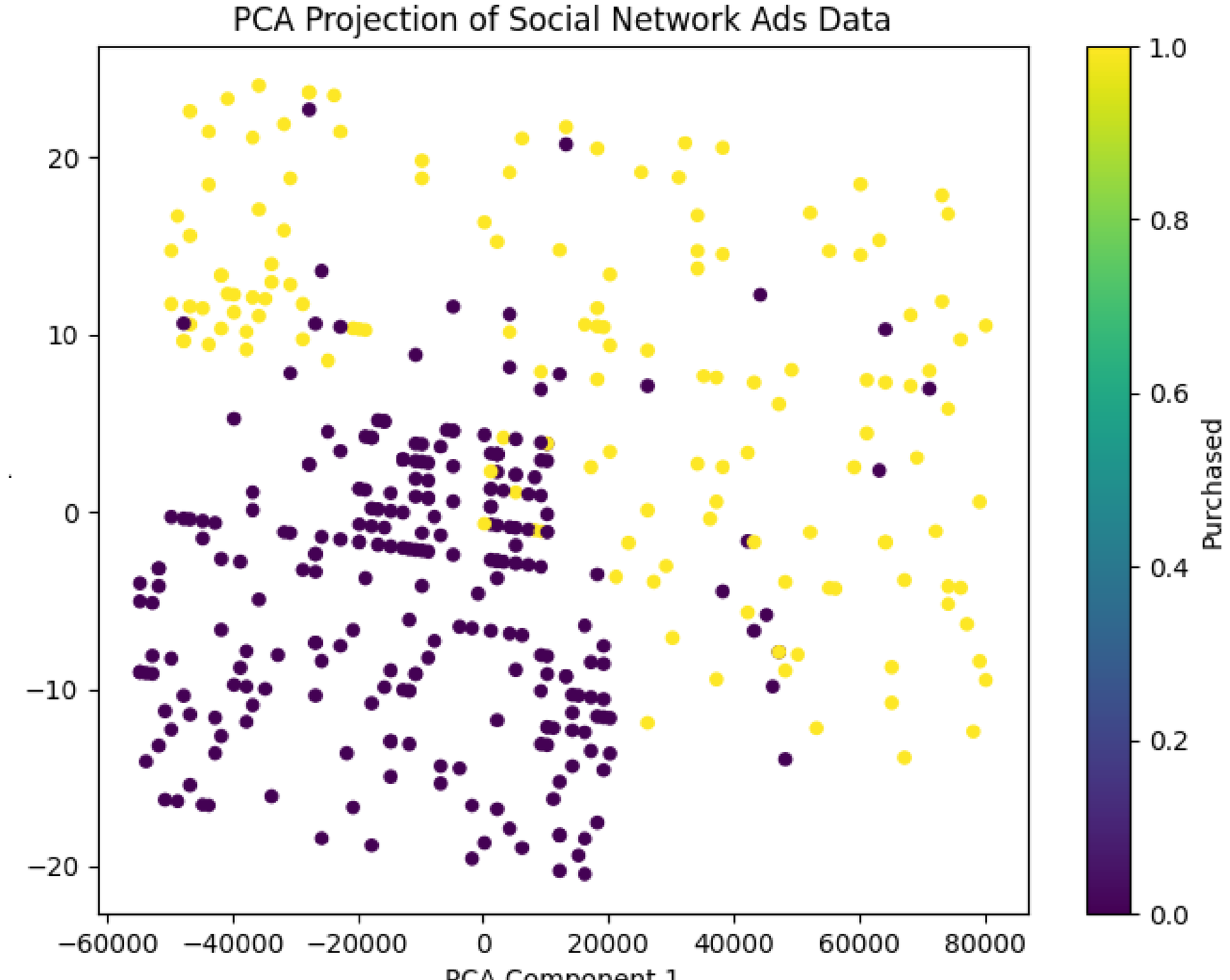


Output

Dimension 1

Dimension 2

"UMAP “1" is a dimension in the new space that captures a portion of the variability or structure in your data.

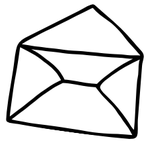

PCA Projection of Social Network Ads Data
20
10
0
−10
−20
−60000
−40000
−20000
0
20000
40000
60000
80000
PCA Component 1
1.0
0.8
0.6
0.4
0.2
0.0
Purchased

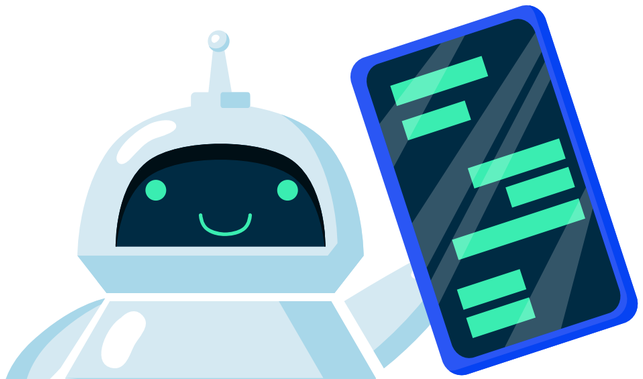
Considerations and Limitations

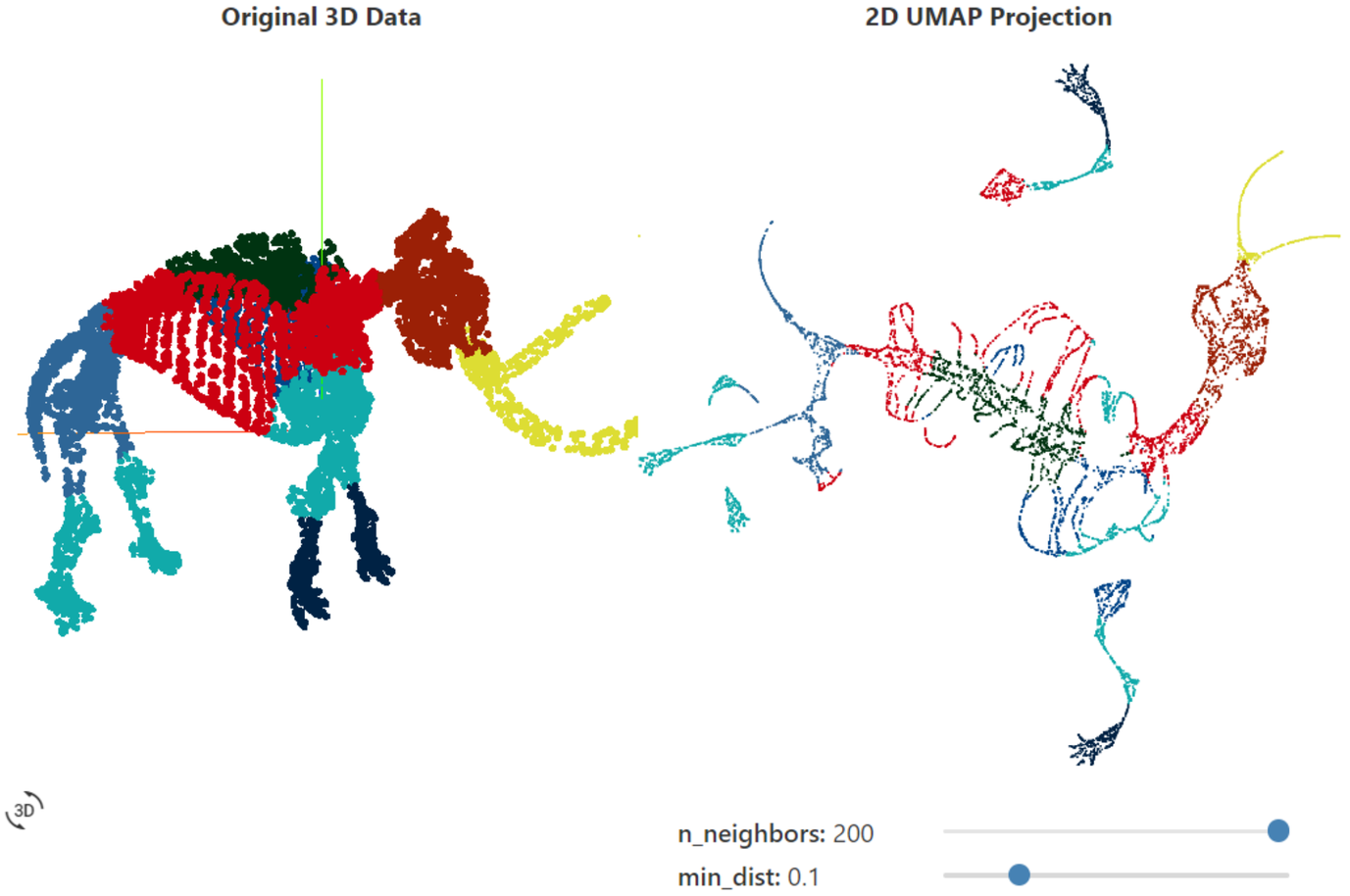
1. parameter sensitivity
2. The challenge of interpreting the reduced representation
3. Parameter Tuning Required

1

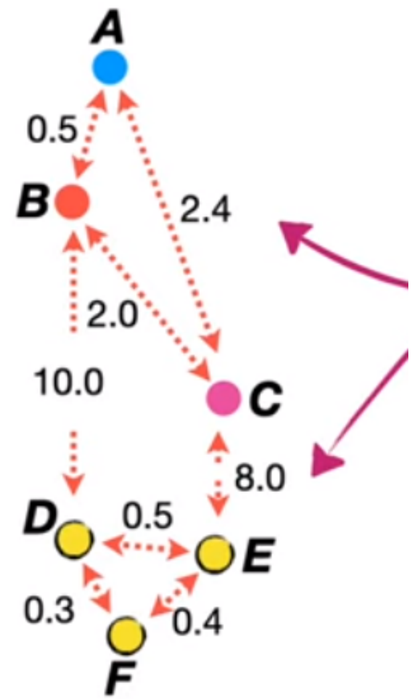
The visualization is an exploration of the impact of UMAP parameters on a 2D projection of 3D data.



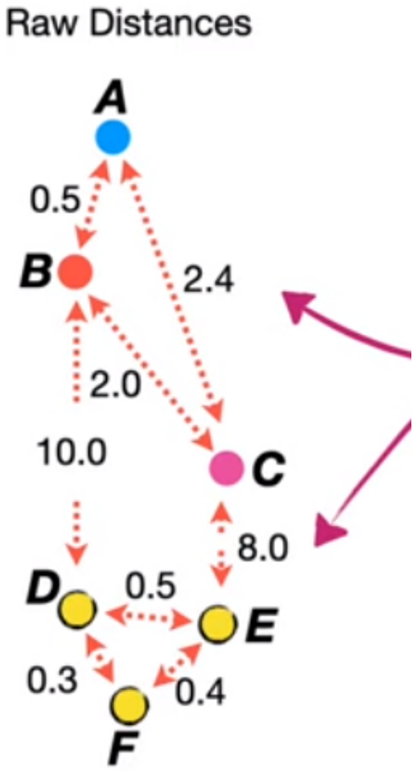


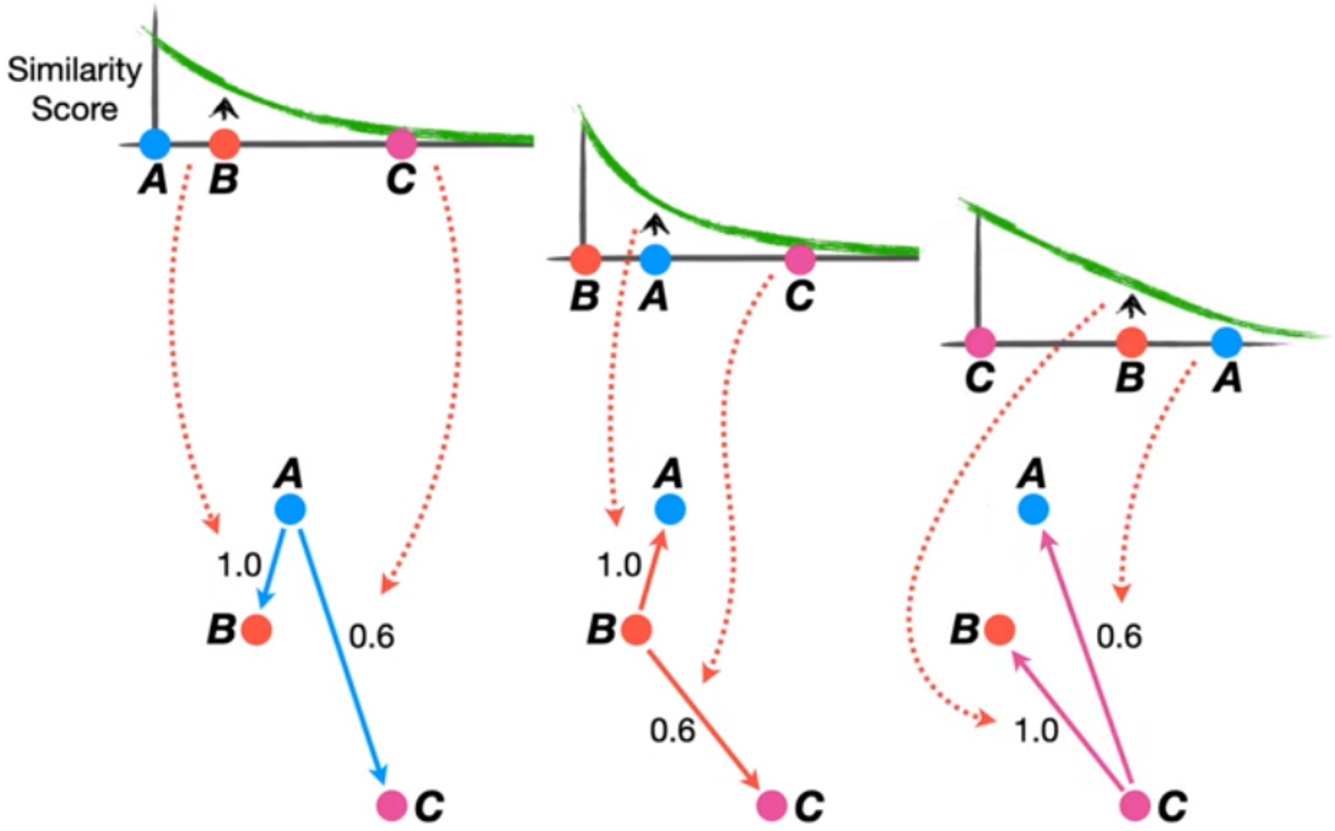


Raw Distances

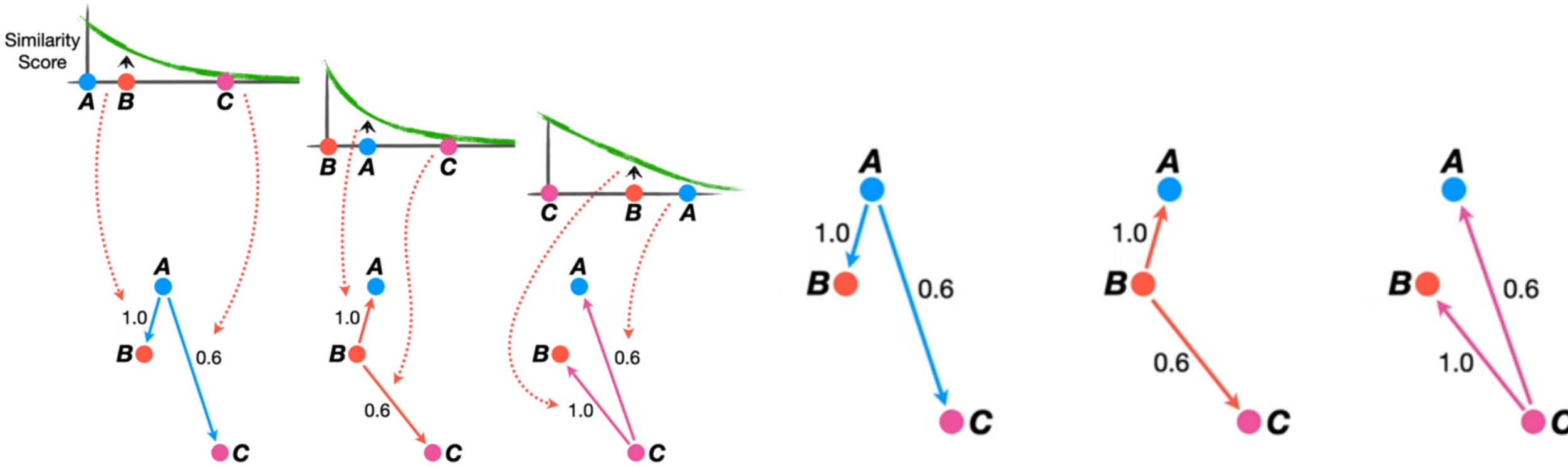


Similarity Score = $e^{-(\text{raw dist.} - \text{dist. to nearest neighbor})/\sigma}$

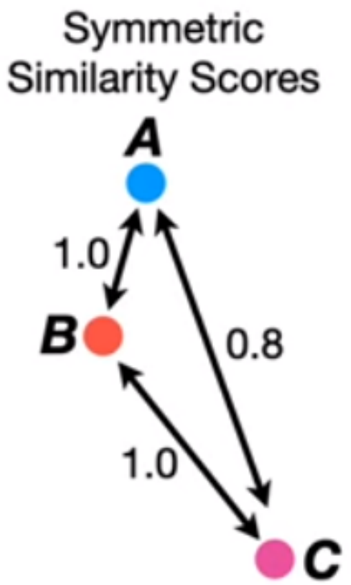


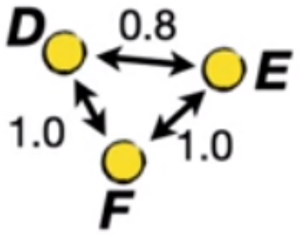


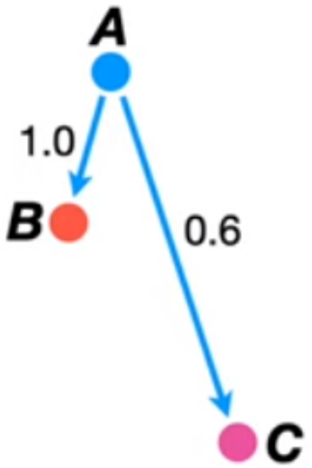
Similarity Score = $e^{-(\text{raw dist. - dist. to nearest neighbor})/\sigma}$

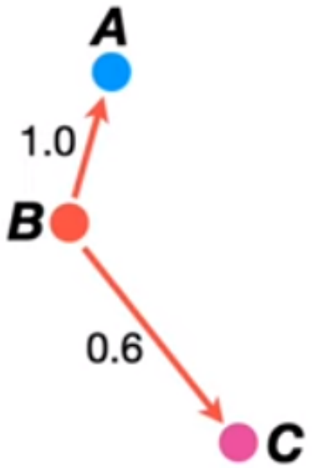


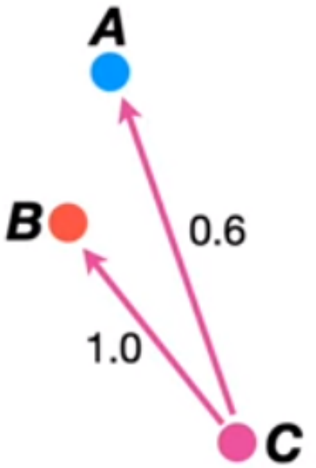
Symmetrical Score = $(S_1 + S_2) - S_1S_2$



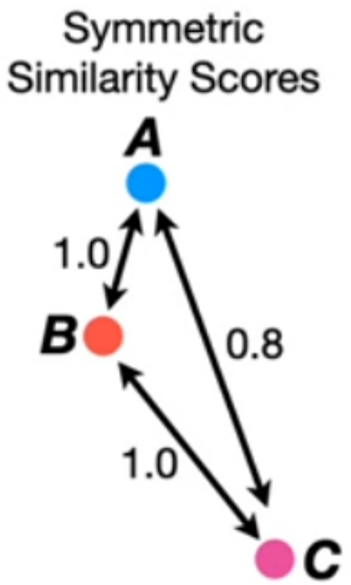


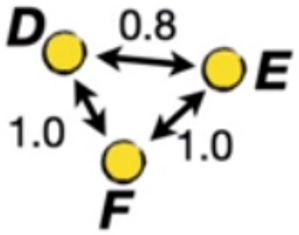


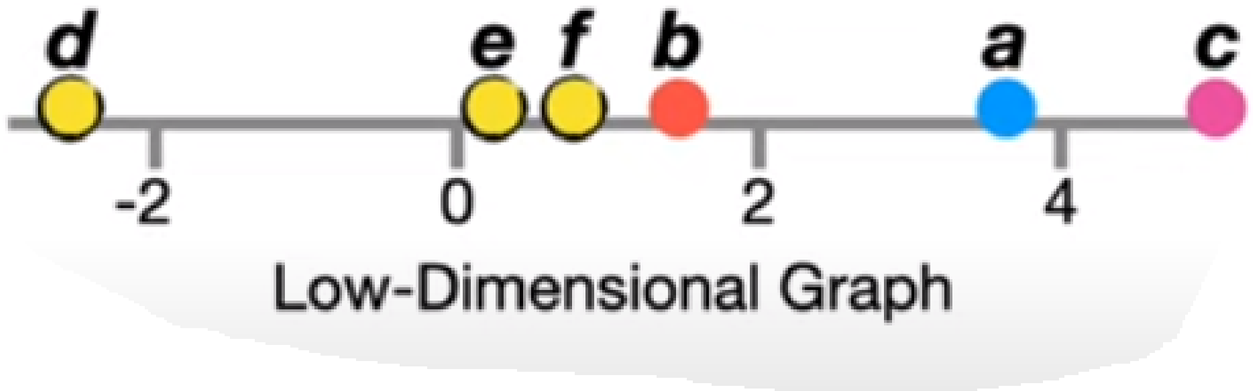




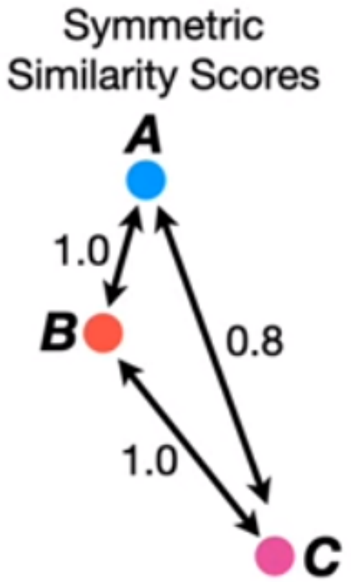
$$\text{Symmetrical Score} = (S_1 + S_2) - S_1S_2$$

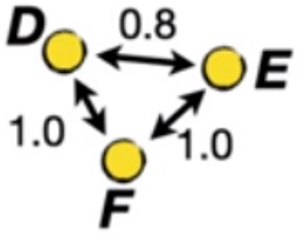


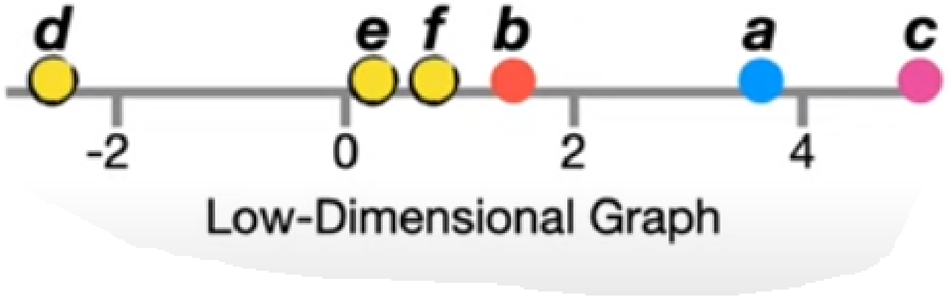


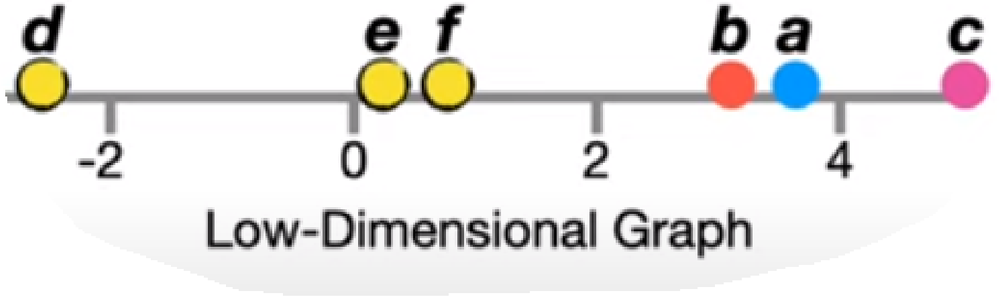


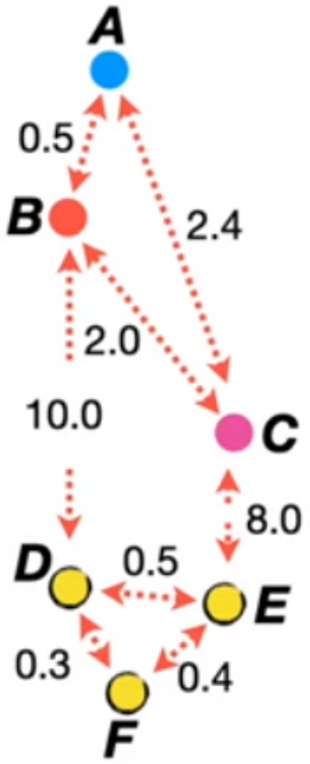
Symmetrical Score = $(S_1 + S_2) - S_1S_2$

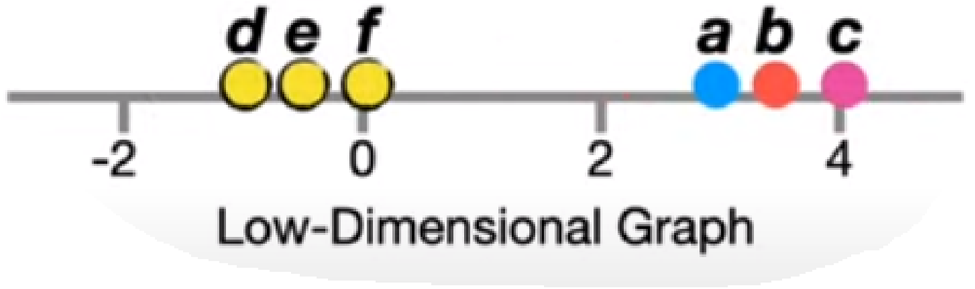









A
0.5
B
2.4
2.0
10.0
C
8.0
D
0.5
E
0.3
0.4
F


d e f
a b c
-2
0
2
4
Low-Dimensional Graph