TOPOLOGICAL DATA ANALYSIS

ASSIGNMENT 1
DUE DATE: FEBRUARY 16TH

Definition: Recall that **clustering** is a task of dividing data points into a number of groups such that data points in the same group are more similar compared to those in other groups.

The task: For this assignment you are expected to write a clustering algorithm based on persistent homology. Recall that the 0-dimensional homology detects the number of connected components of the underlying Vietoris-Rips complex. So, the number of "reasonably long" bars in the 0-dimensional barcodes should be a rough estimate. However, this number is far from accurate and something else should be done. Hence, your clustering algorithm should be an improvement over this (you can check your results against DBSCAN).

Note: Your submission should consists of a report that, at the beginning, clearly describes your approach and a pseudo-code. Followed by analysis of each individual data sets. This analysis should consists of the number of clusters, a plot of the data that indicates clusters with different colors (if the dimension allows it!) and comparison with the DBSCAN output. Make sure to upload your code as a Jupyter Notebook (an HTML download should suffice).

The data set: Uploaded on Moodle as clusterdata.zip file.

Date: January 24, 2022.