

# TDA readme

## TDA Computational Project Readme

### Table of Contents

1. [File Structure](#)
2. [Data Description](#)
3. [Usage](#)

### File Structure

- `Main.ipynb` : This Jupyter notebook contains all the code necessary for data analysis, image generation, and report preparation.
- `Preprocess.py` : Python script for preprocessing shape data.

### Data Directory

- `data/`
  - `Nonrigid3d.zip` : Data file available on KLMS.
  - `distance_matrix/` : Contains CSV files of calculated distance matrices.
  - `preprocessing/`
    - `geodesic/` : Preprocessed shapes with geodesic distances.
    - `euclidean/` : Preprocessed shapes with Euclidean distances.

### Source Code Directory

- `src/`
  - `preprocessing/`
    - `Preprocessing.py` : Python module with classes and functions for data preprocessing.
  - `shape/`
    - `Shape.py` : Python class and functions for shape handling and related operations.

### Images Directory

- `images/` : Contains supplementary images not required for submission.

# Data Description

## Shape Data

Defined in `src/shape/Shape.py`, the `Shape` class represents a 3D object as a weighted undirected graph.

```
class Shape:
    def __init__(self, name, dm, coordinates=None):
        """
        Initializes a Shape object.

        Args:
            name (str): Name of the shape.
            dm (numpy.array): Distance matrix of the shape.
            coordinates (numpy.array, optional): Coordinates of the shape.
        Defaults to None.
        """
```

### Attributes:

- `name`: The name of the shape.
- `dm`: Distance matrix (numpy array of shape (200, 200)).
- `coordinates`: 3D coordinates of each point of the shape (numpy array of shape (3, 200)).
- `dgms`: Vietoris–Rips diagram of dimension 0, 1, 2.
- `diameter`: The diameter of the shape.

### Example Usage:

```
shape1 = Shape("cat0", dm, coords)
print(shape1.dm)
```

## Shape Methods

- `Shape.save_to_mat(path)`: Exports the shape to a file at the specified `path`.
- `load_from_mat(path)`: Loads a shape object from a `.mat` file at the given `path`.

## Distance Matrix

Distance matrices are stored as `.csv` files in the `data/distance_matrix/` folder. Each file is a 2D array of shape (200,200), where each element (i, j) represents the pairwise distance between the i-th and j-th shapes. Shapes are sorted alphabetically from a to z.

# Usage

## dependencies

```
pip install scipy
pip install ripser gudhi
pip install scikit-tda
pip install numpy
pip install tqdm
pip install networkx
pip install persim
pip install pot
pip install seaborn
pip install matplotlib
pip install pandas
pip install gudhi
```

0. Install the dependencies. The dependencies are in the first cell of the main.ipynb.
1. Run `preprocess.py` inside the script, you need to properly set the path to nonrigid3d data folder.
2. Run main.ipynb. before running all scripts, set the variables for paths to preprocessed geodesic and euclidean data. (in the second code cell.)