CASE WESTERN RESERVE
UNIVERSITY EST. 1826

Neuro-Integrative Connectivity (NIC) Workflow Component III b: Topological Data Analysis

Version 1.0

Section 0 – Licensing

Section 1 – Uses

This software is intended to be used to create and explore Vietoris-Rips Complexes from the correlation matrices obtained using the NIC Correlator. All data cleaning steps are handled internally; thus, input files are expected to be in the format of those output by the NIC Correlator. All persistent homology computation is handled using Inria's GUDHI.

This program reads correlation matrices output by the NIC Correlator to generate a list of simplicial complexes, their dimension, their birth filtration, their death filtration, and their persistence or lifespan value. A single csv is output for each correlation matrix processed.

Section 2 – Requirements

A requirements.txt file is provided for your convenience so that you may install all dependencies in a virtual environment from the terminal using the command pip install -r requirements.txt. It is recommended, but not necessary, that this setup is done within a virtual environment for optimal package version control. All development was conducted using python version 3.9.4

The required python dependencies are as follows:

- python ≥ 3.8.12
- pandas ≥ 1.5.2
- gudhi ≥ 3.7.1

Text files output by the correlator are generally expected to all be in one directory, with one patient per directory. They do not need to be in the same directory as EDF files or CSF files. All files within a given directory will be analyzed unless a specific file name is specified. For simplicity, all output files are in csv format and share the same name as the txt file which was used to generate them.

Section 3 – Processing

From the command line, navigate to where the python project is stored and enter:

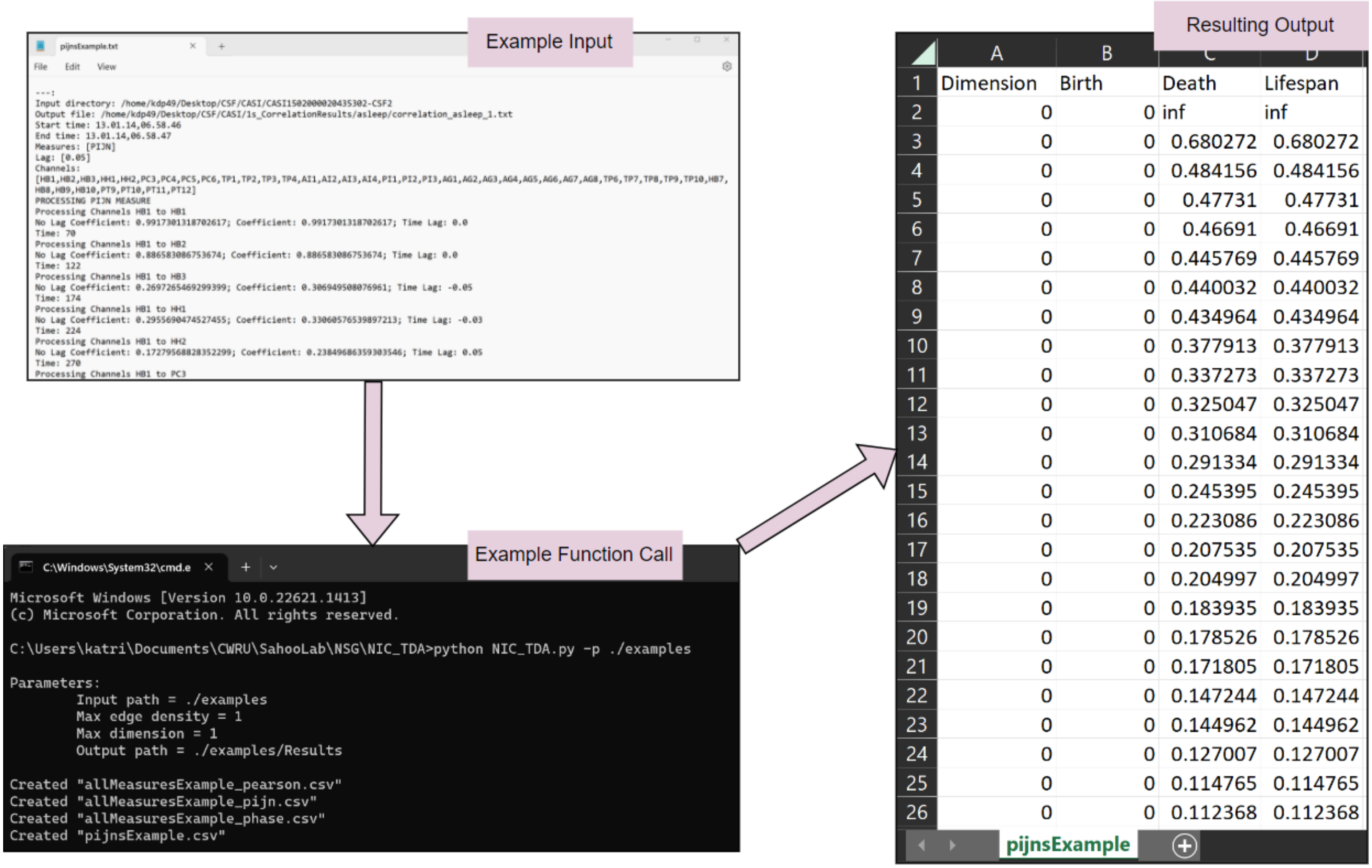
```
python3 NICTDA.py -p <input path> -f <file> -o <output path> -e <max edge density {1} > -d <max dimension {1}>
```

with the following arguments

- **Input Path *(mandatory):*** The path to the folder containing any correlation matrices (output by the NIC Correlator) that you would like to process.

- **File *(optional):*** The name of a single file within the provided input folder if you would like to limit your processing to single-file processing. This should not include the full file path as the path has already been specified.
- **Output Path *(optional: default = {input path}/Results):*** The path to a folder to store your results. If no folder is provided, a folder called "Results" will be created inside of the specified input path. If a "Results" subdirectory already exists, you will be prompted for further action.
- **Max Edge Density *(optional: default = 1.0):*** The floating-point value that represents the maximum edge density to be used by GUDHI. Since all inputs in a correlation matrix are less than 1, the default is set to 1.0. For more information, see https://gudhi.inria.fr/python/latest/rips_complex_user.html.
- **Dimension *(optional: default = 1):*** The integer value representing the maximum dimension (non-inclusive) of the simplices to be returned by GUDHI. For more information, see https://gudhi.inria.fr/python/latest/rips_complex_user.html.

The program will begin by locating the destination folder for the text files holding correlation matrices. For observational purposes, the input parameters will be output to the screen. Each file within the provided folder will be processed using GUDHI's Vietoris-Rips Complex functions to generate a list of the dimension, birth filtration, death filtration, and lifespan value of all simplicial complexes in the provided data up to the given dimension.



```
---:
Input directory: /home/kdp49/Desktop/CSF/CASI/CASI1502000020435302-CSF2
Output file: /home/kdp49/Desktop/CSF/CASI/1s_CorrelationResults/asleep/correlation_asleep_1.txt
Start time: 13.01.14,06.58.46
End time: 13.01.14,06.58.47
Measures: [PIJN]
Lag: [0.05]
Channels:
[HB1,HB2,HB3,HH1,HH2,PC3,PC4,PC5,PC6,TP1,TP2,TP3,TP4,AI1,AI2,AI3,AI4,PI1,PI2,PI3,AG1,AG2,AG3,AG4,AG5,AG6,AG7,AG8,TP6,TP7,TP8,TP9,TP10,HB7,
HB8,HB9,HB10,PT9,PT10,PT11,PT12]
PROCESSING PIJN MEASURE
Processing Channels HB1 to HB1
No Lag Coefficient: 0.9917301318702617; Coefficient: 0.9917301318702617; Time Lag: 0.0
Time: 70
Processing Channels HB1 to HB2
No Lag Coefficient: 0.886583086753674; Coefficient: 0.886583086753674; Time Lag: 0.0
Time: 122
Processing Channels HB1 to HB3
No Lag Coefficient: 0.2697265469299399; Coefficient: 0.306949508076961; Time Lag: -0.05
Time: 174
Processing Channels HB1 to HH1
No Lag Coefficient: 0.2955690474527455; Coefficient: 0.33060576539897213; Time Lag: -0.03
Time: 224
Processing Channels HB1 to HH2
No Lag Coefficient: 0.17279568828352299; Coefficient: 0.23849686359303546; Time Lag: 0.05
Time: 270
Processing Channels HB1 to PC3
```

```
Microsoft Windows [Version 10.0.22621.1413]
(c) Microsoft Corporation. All rights reserved.

C:\Users\katri\Documents\CWRU\SahooLab\NSG\NIC_TDA>python NIC_TDA.py -p ./examples

Parameters:
        Input path = ./examples
        Max edge density = 1
        Max dimension = 1
        Output path = ./examples/Results

Created "allMeasuresExample_pearson.csv"
Created "allMeasuresExample_pijn.csv"
Created "allMeasuresExample_phase.csv"
Created "pijnsExample.csv"
```

	A	B	C	D
1	Dimension	Birth	Death	Lifespan
2	0	0	inf	inf
3	0	0	0.680272	0.680272
4	0	0	0.484156	0.484156
5	0	0	0.47731	0.47731
6	0	0	0.46691	0.46691
7	0	0	0.445769	0.445769
8	0	0	0.440032	0.440032
9	0	0	0.434964	0.434964
10	0	0	0.377913	0.377913
11	0	0	0.337273	0.337273
12	0	0	0.325047	0.325047
13	0	0	0.310684	0.310684
14	0	0	0.291334	0.291334
15	0	0	0.245395	0.245395
16	0	0	0.223086	0.223086
17	0	0	0.207535	0.207535
18	0	0	0.204997	0.204997
19	0	0	0.183935	0.183935
20	0	0	0.178526	0.178526
21	0	0	0.171805	0.171805
22	0	0	0.147244	0.147244
23	0	0	0.144962	0.144962
24	0	0	0.127007	0.127007
25	0	0	0.114765	0.114765
26	0	0	0.112368	0.112368

Section 4 – Output

The output of the program will be a csv file of the same name as each input correlation matrix. Each csv file will contain a list of the dimension, birth filtration, death filtration, and lifespan value of all simplicial complexes in the provided data up to the given dimension. By default, all csv files are stored in the same directory as the input files within a subdirectory called “Results” unless a separate output location is specified by the user. In the case that the “ALL” parameter was selected when running the correlator, three files will be created (one for each correlation measure). Each of the three output files will share the name of the original correlator output but will append “PJIN”, “PHASE”, and “PEARSON” depending on which matrix within the original file was processed.

Credits

This software has been developed at Case Western Reserve University as part of a research project and includes contributions by Katrina Prantzalos and Dr. Satya Sahoo (project PI).