Persistent Homology with Stock Prices in Different Sectors

Sierra Cartano

Francis Marion University

Florence, South Carolina

Funded by REAL Grant

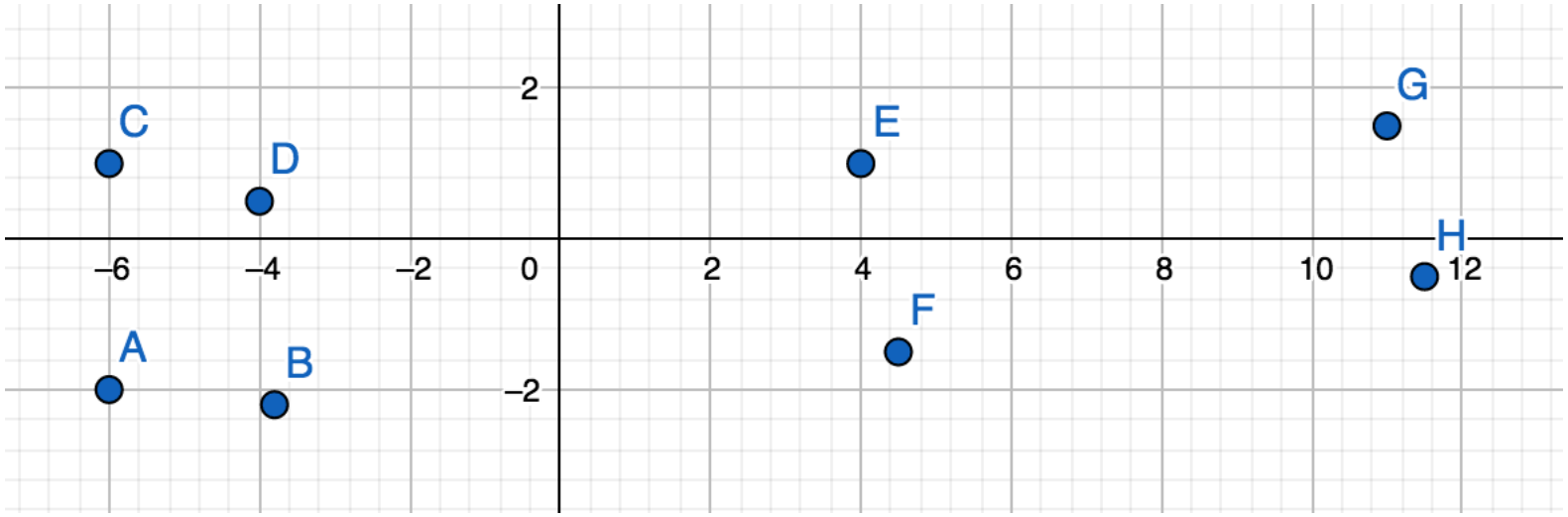
Advised by Dr. Ivan Dungan

Overview

- Determine a way to identify the sector* of any random stock
  - Use topological data analysis, specifically, persistent homology.
    - Homology is the mathematics for identifying holes in a shape
    - Persistent homology identifies holes that persist over time.
  - Determine way to identify a random stock by sector with its unique homology
    - Apply to financial data

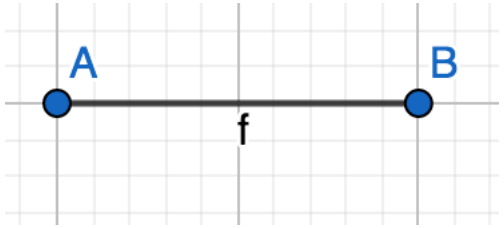
*Financials, Utilities, Energy, Materials, Industrials, Consumer Discretion, Consumer Staples, Health Care, Information Technology, Telecommunication Services, Real Estate

Example of a Point Cloud



Different Dimensional Holes

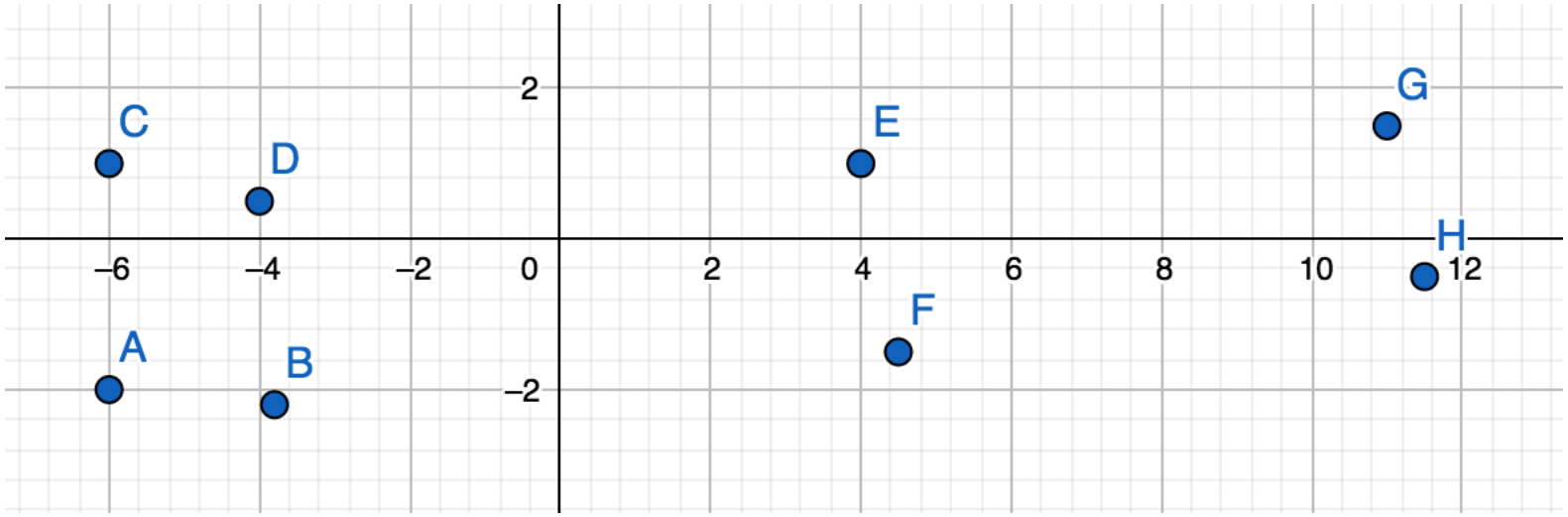
- 0-dimensional holes= connected components



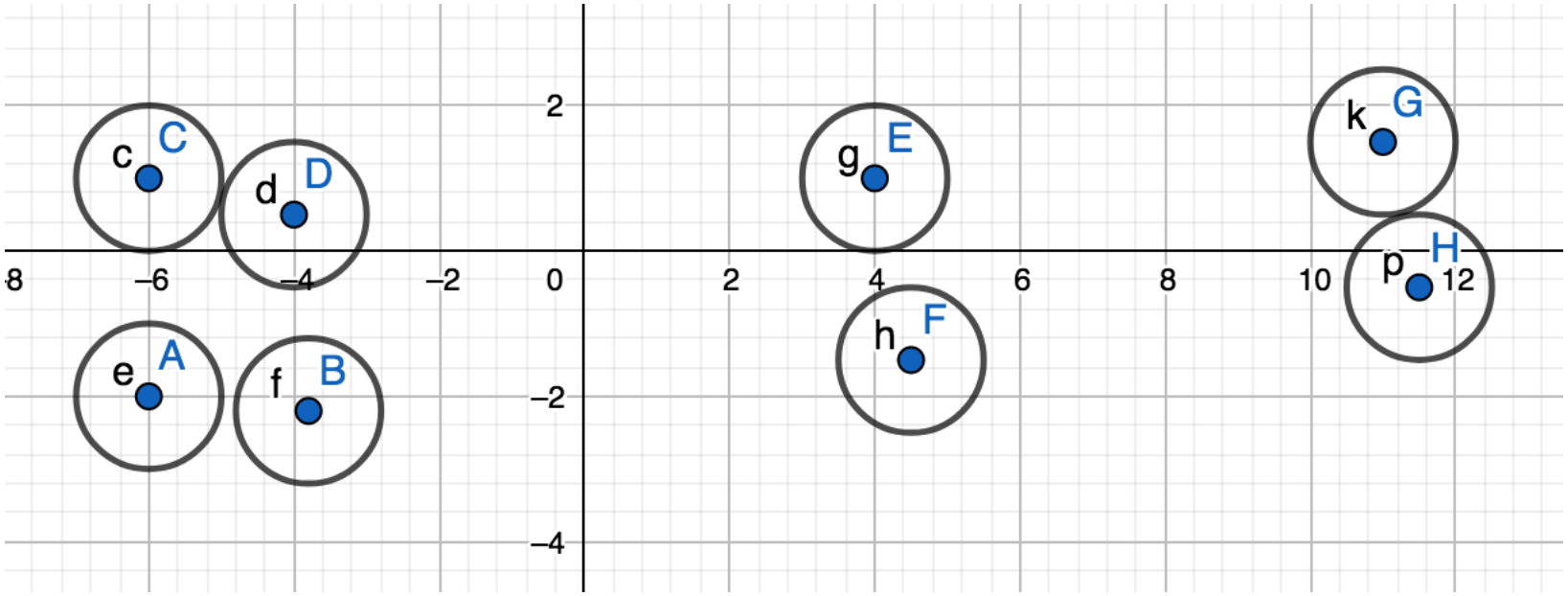
- 1-dimensional holes= circles

- 2-dimensional holes= spheres

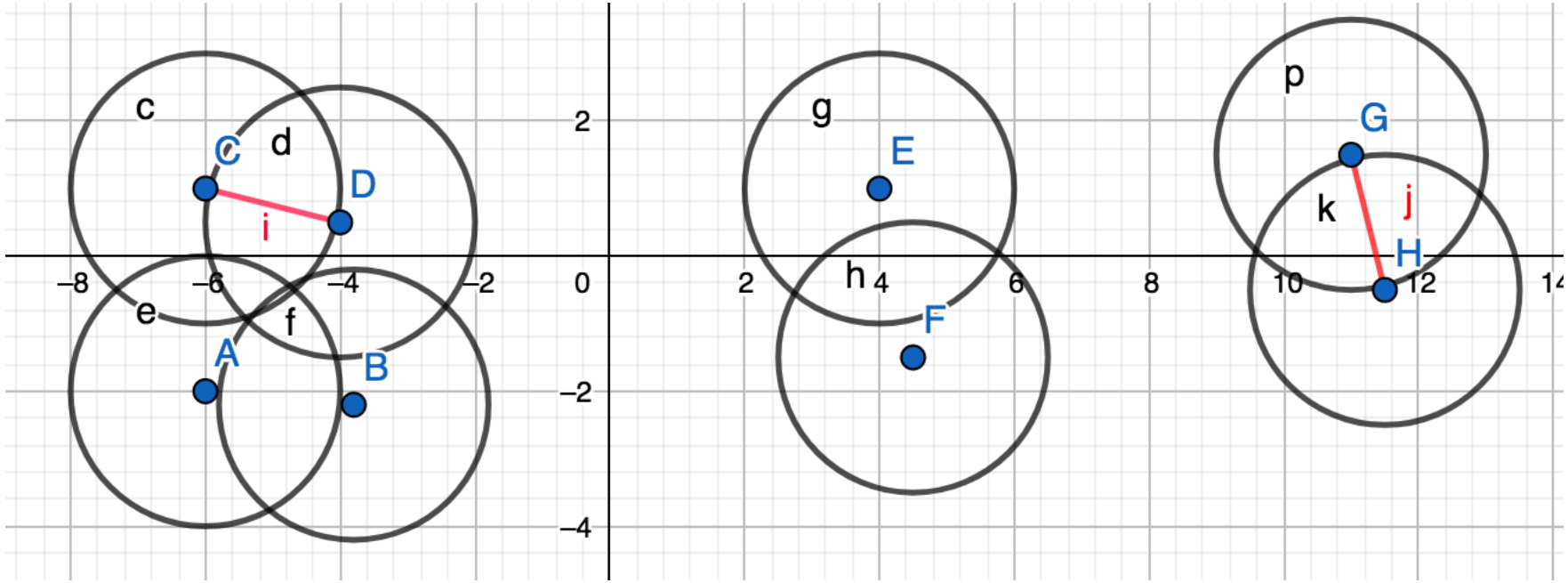
Example of a Point Cloud



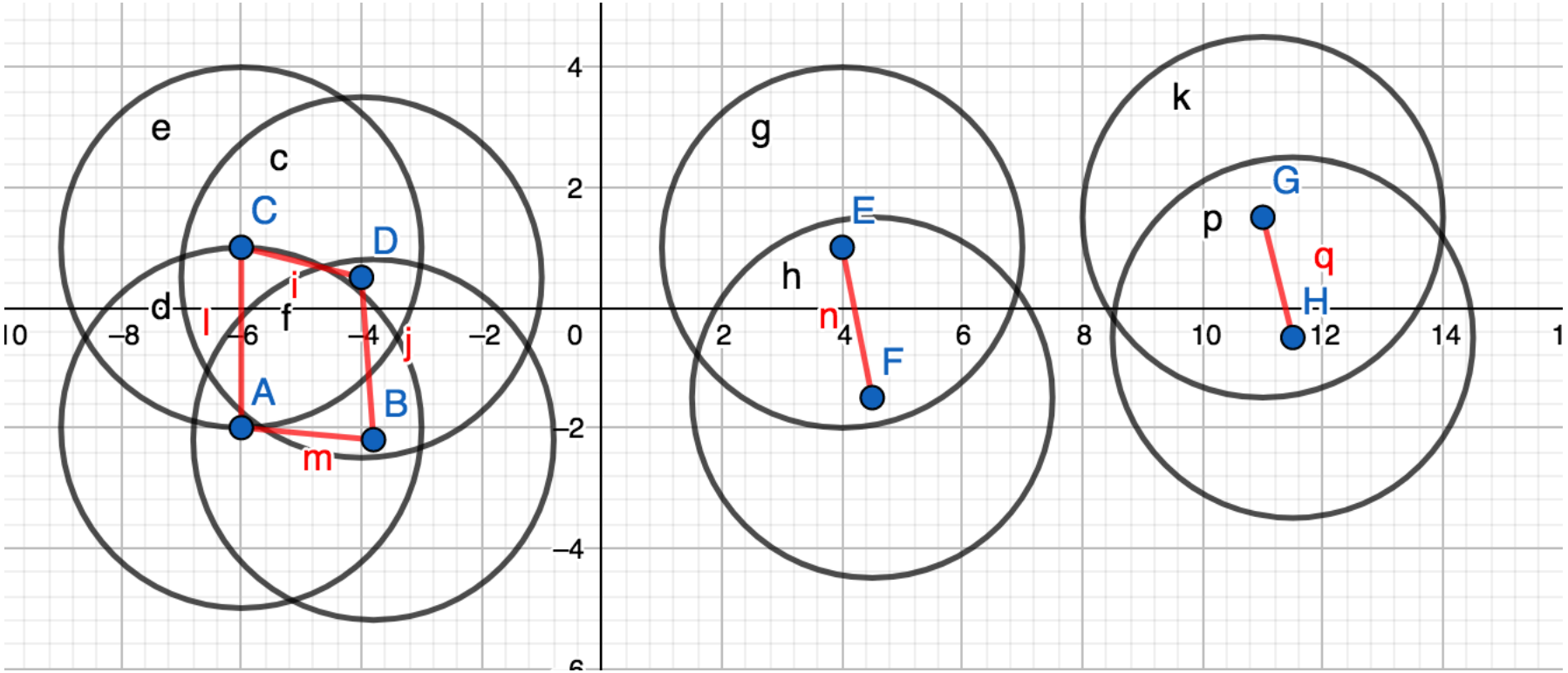
Radius=1



Radius=2



Radius=3



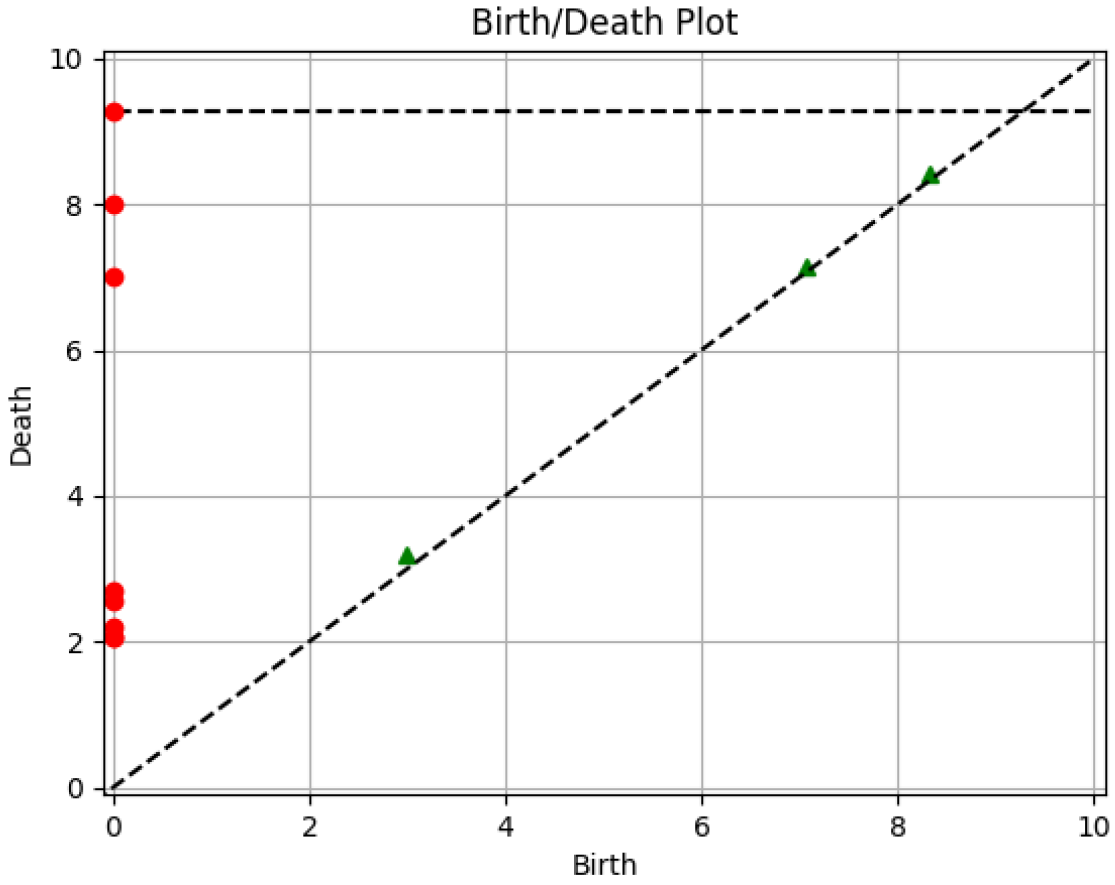
Method of Persistent Homology

- A(-6, -2)
- B(-3.8, -2.2)
- C(-6, 1)
- D(-4, .5)
- E(4, 1)
- F(4.5, -1.5)
- G(11, 1.5)
- H(11.5, -.5)

- Download data into Excel spreadsheet
- Save as text file
- Run through Ripser

-6	-2
-3.8	-2.2
-6	1
-4	0.5
4	1
4.5	-1.5
11	1.5
11.5	-0.5

Persistence Diagram



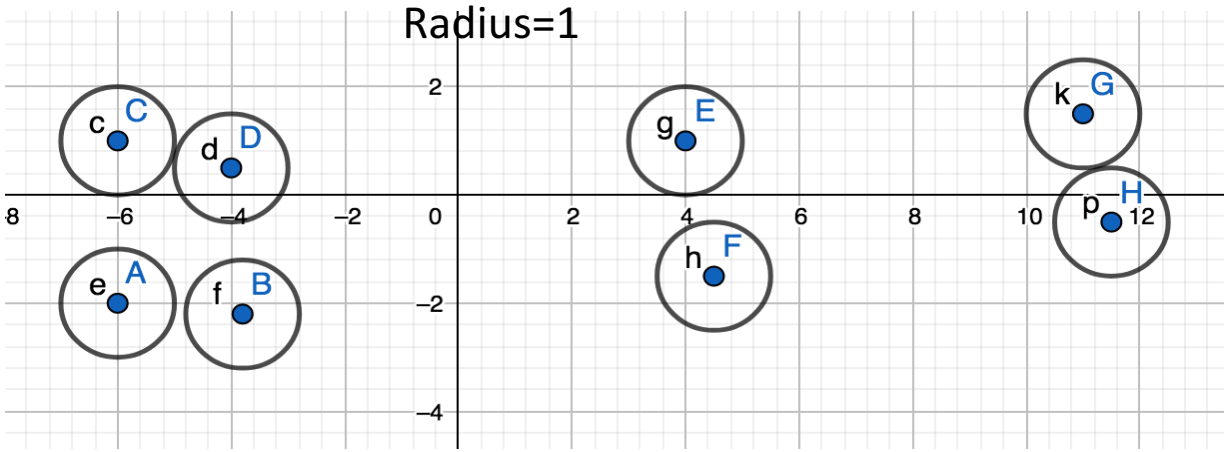
Persistence refers to topological features persist over a period of time.

Connected Components

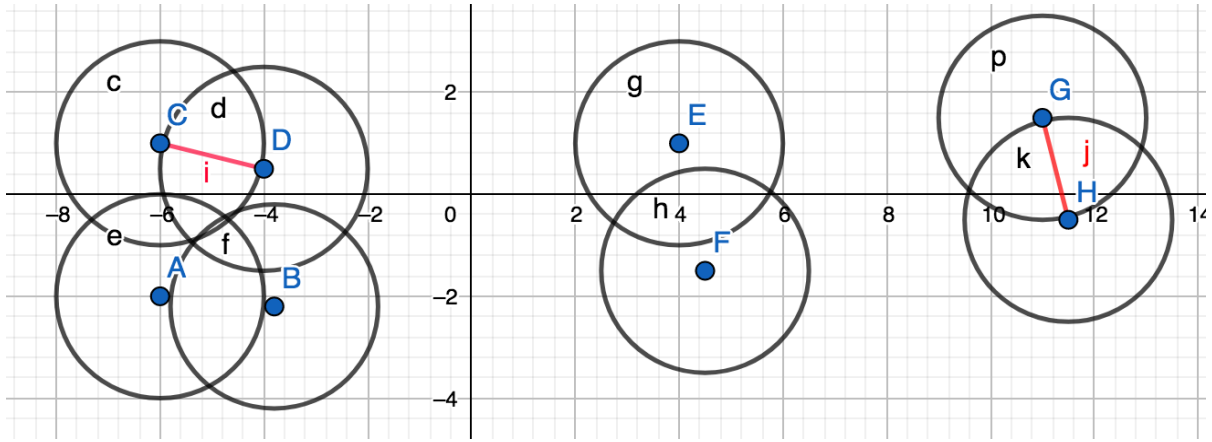
Circles

Radius=1, 2, 3

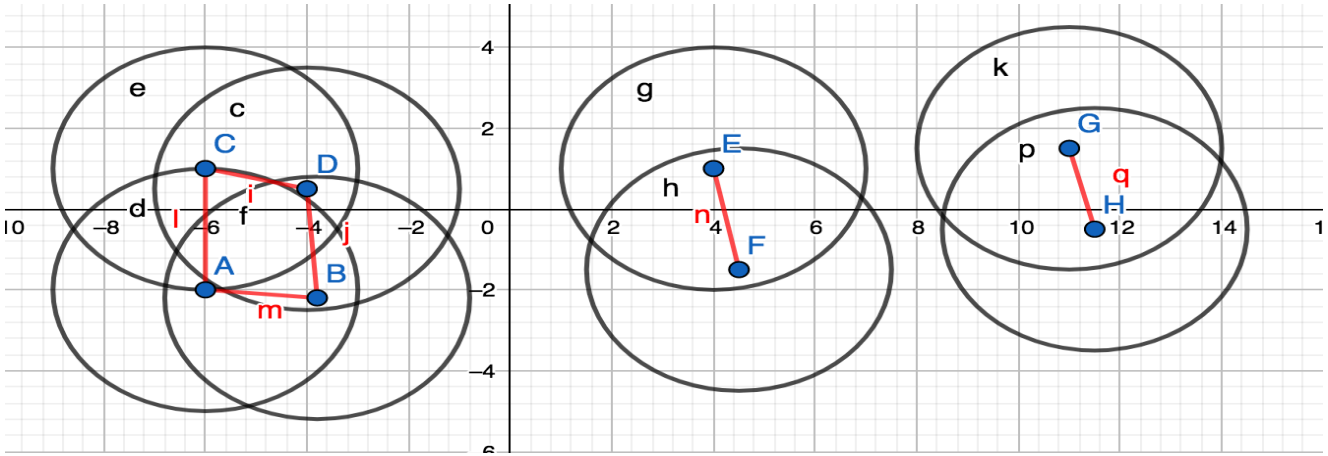
Radius=1



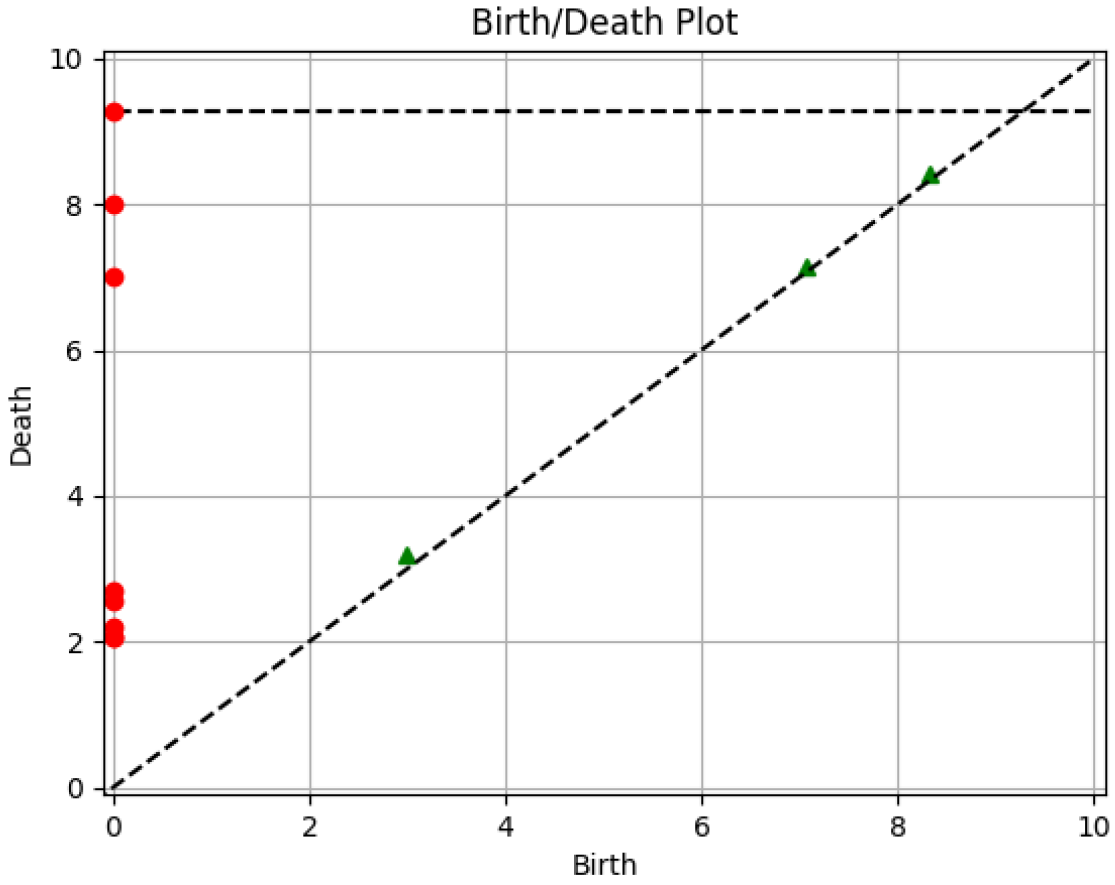
Radius=2



Radius=3



Persistence Diagram

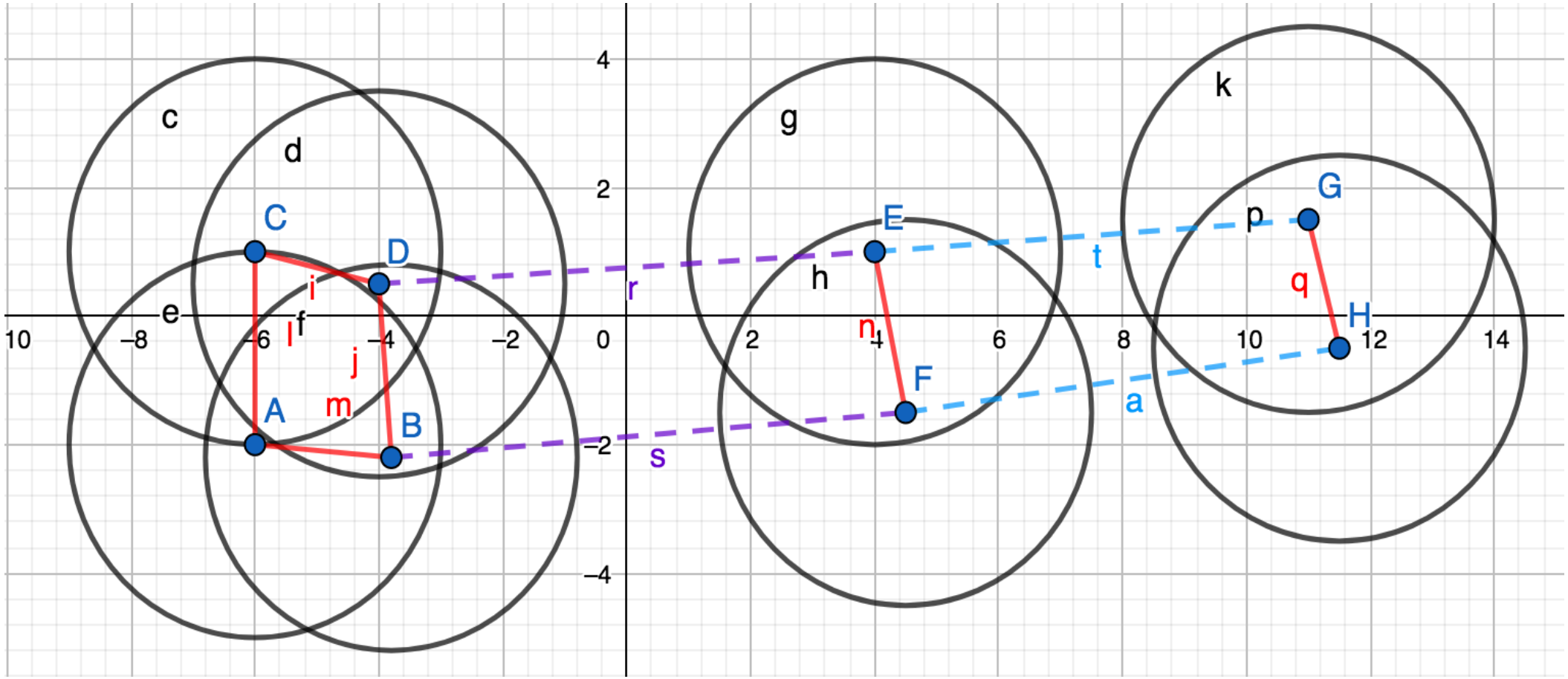


Persistence refers to topological features persist over a period of time.

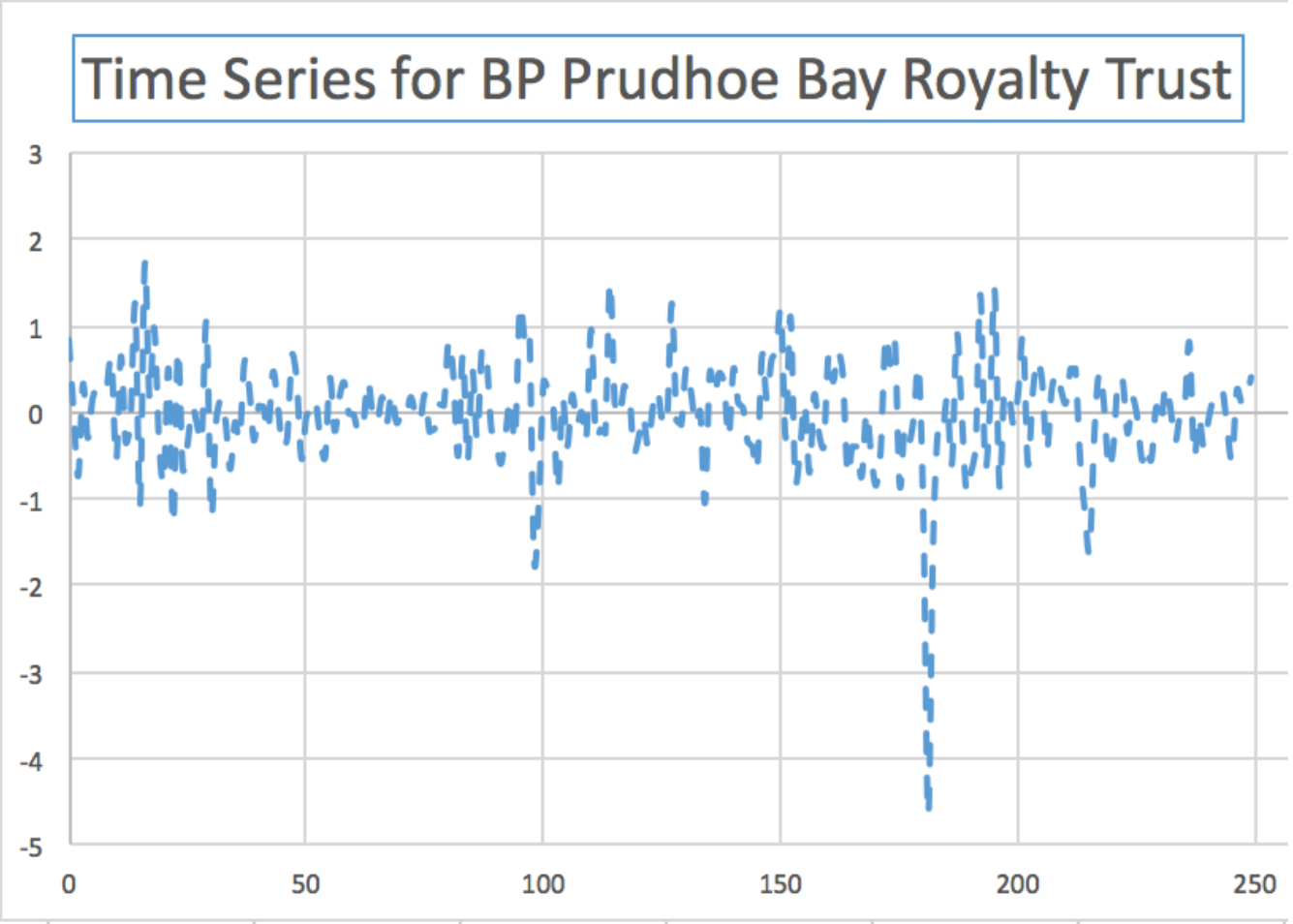
Connected Components

Circles

Circles



BP Prudhoe Bay Royalty Trust

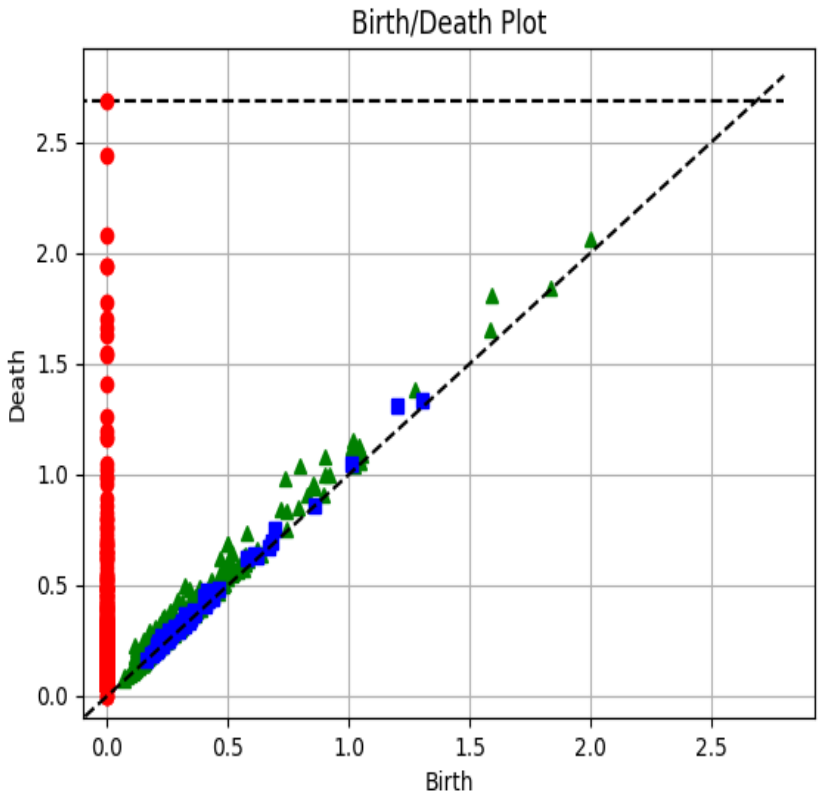


BP Prudhoe Bay Royalty Trust (BPT)

0.83
0.03
-0.739999
0.32
-0.3
0.15
0.3
0.24
0.29
0.6
-0.559999
0.640001
-0.35
-0.13
1.239999
-1.08
1.699999
0.17
0.97
-0.319999
-0.809999
0.5
-1.300002
0.710001
-0.65
-0.380001
0.050002
-0.140002
-0.33
1.030001

$$SW_{M,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+M\tau) \end{bmatrix}$$

0.83	0.03	-0.739999
0.03	-0.739999	0.32
-0.739999	0.32	-0.3
0.32	-0.3	0.15
-0.3	0.15	0.3
0.15	0.3	0.24
0.3	0.24	0.29
0.24	0.29	0.6
0.29	0.6	-0.559999
0.6	-0.559999	0.640001
-0.559999	0.640001	-0.35
0.640001	-0.35	-0.13
-0.35	-0.13	1.239999
-0.13	1.239999	-1.08
1.239999	-1.08	1.699999
-1.08	1.699999	0.17
1.699999	0.17	0.97
0.17	0.97	-0.319999
0.97	-0.319999	-0.809999
-0.319999	-0.809999	0.5
-0.809999	0.5	-1.300002
0.5	-1.300002	0.710001
-1.300002	0.710001	-0.65
0.710001	-0.65	-0.380001
-0.65	-0.380001	0.050002
-0.380001	0.050002	-0.140002
0.050002	-0.140002	-0.33
-0.140002	-0.33	1.030001
-0.33	1.030001	-1.190001
1.030001	-1.190001	-0.129999



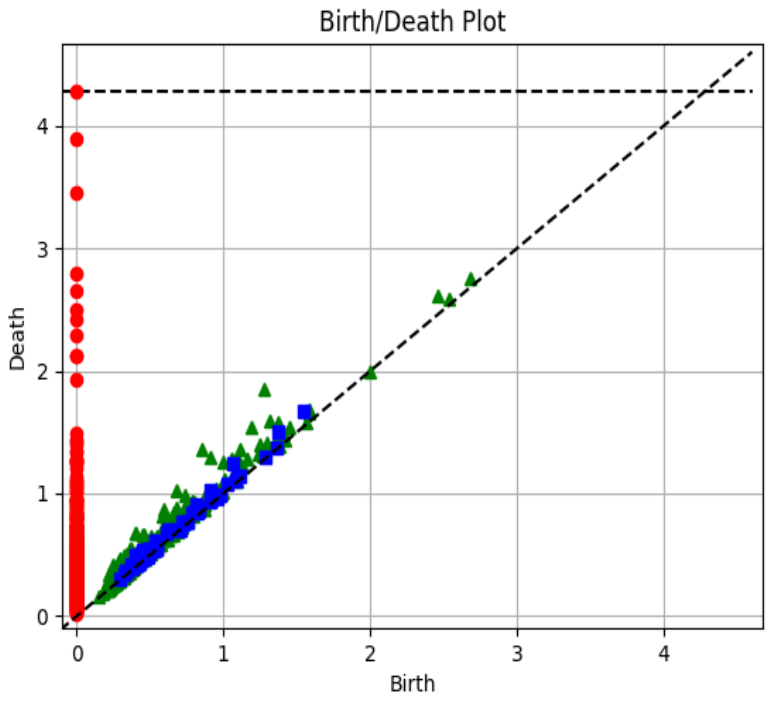
*Blue squares are spheres

Phillips 66 (PSX)

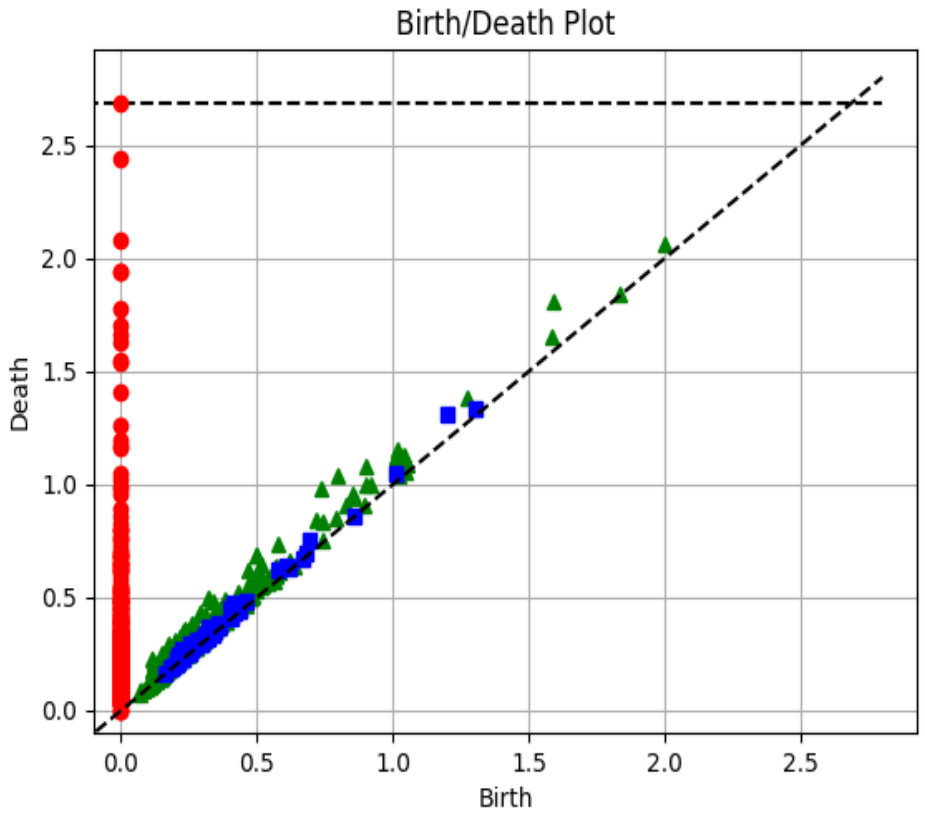
0.150001
0.449997
0.790001
-0.520005
-0.18
-0.93
0.270005
0.510002
0.959999
-0.839996
-0.470001
0.049995
0.169998
0.850006
0.330002
-1.29
0.75
0.489998
-0.060006
0.709999
-1.040001
-0.889999
-2.319999
0.309997
0.360001
0.410004
0.169998
0.25
-1.25
-0.119995

$$SW_{M,\tau}f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+M\tau) \end{bmatrix}$$

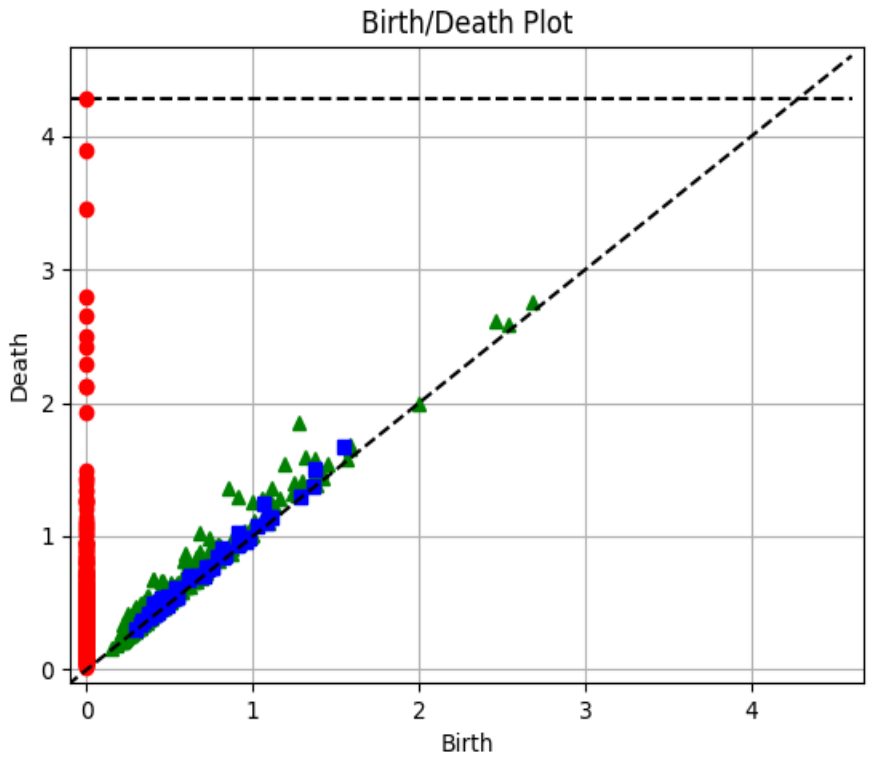
0.150001	0.449997	0.790001
0.449997	0.790001	-0.520005
0.790001	-0.520005	-0.18
-0.520005	-0.18	-0.93
-0.18	-0.93	0.270005
-0.93	0.270005	0.510002
0.270005	0.510002	0.959999
0.510002	0.959999	-0.839996
0.959999	-0.839996	-0.470001
-0.839996	-0.470001	0.049995
-0.470001	0.049995	0.169998
0.049995	0.169998	0.850006
0.169998	0.850006	0.330002
0.850006	0.330002	-1.29
0.330002	-1.29	0.75
-1.29	0.75	0.489998
0.75	0.489998	-0.060006
0.489998	-0.060006	0.709999
-0.060006	0.709999	-1.040001
0.709999	-1.040001	-0.889999
-1.040001	-0.889999	-2.319999
-0.889999	-2.319999	0.309997
-2.319999	0.309997	0.360001
0.309997	0.360001	0.410004
0.360001	0.410004	0.169998
0.410004	0.169998	0.25
0.169998	0.25	-1.25
0.25	-1.25	-0.119995
-1.25	-0.119995	0.009995
-0.119995	0.009995	-0.400002



BPT vs. PSX Birth/Death Plots



BPT



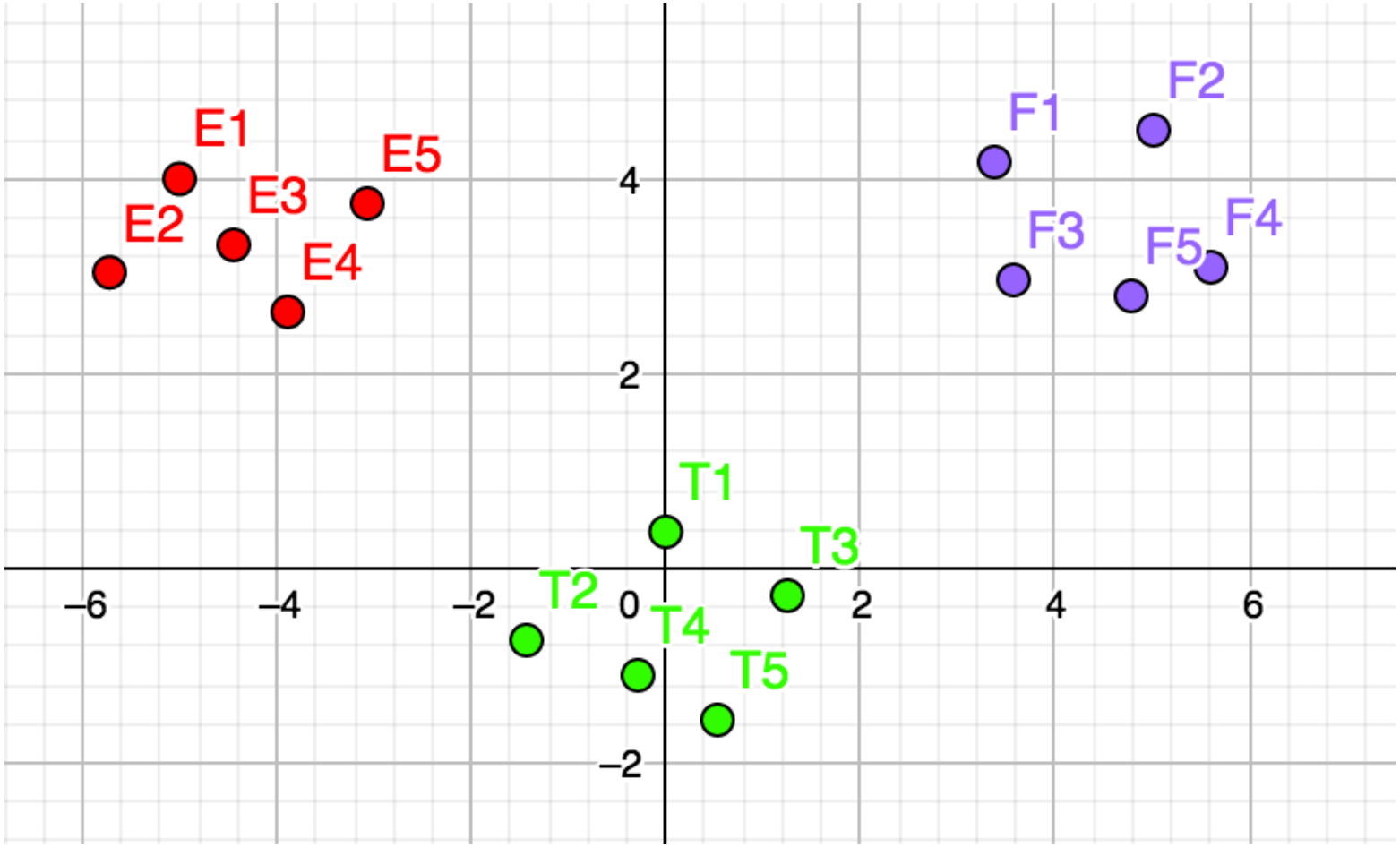
PSX

Persistence diagrams can be compared using the bottleneck distance.

Method of Sector Classification

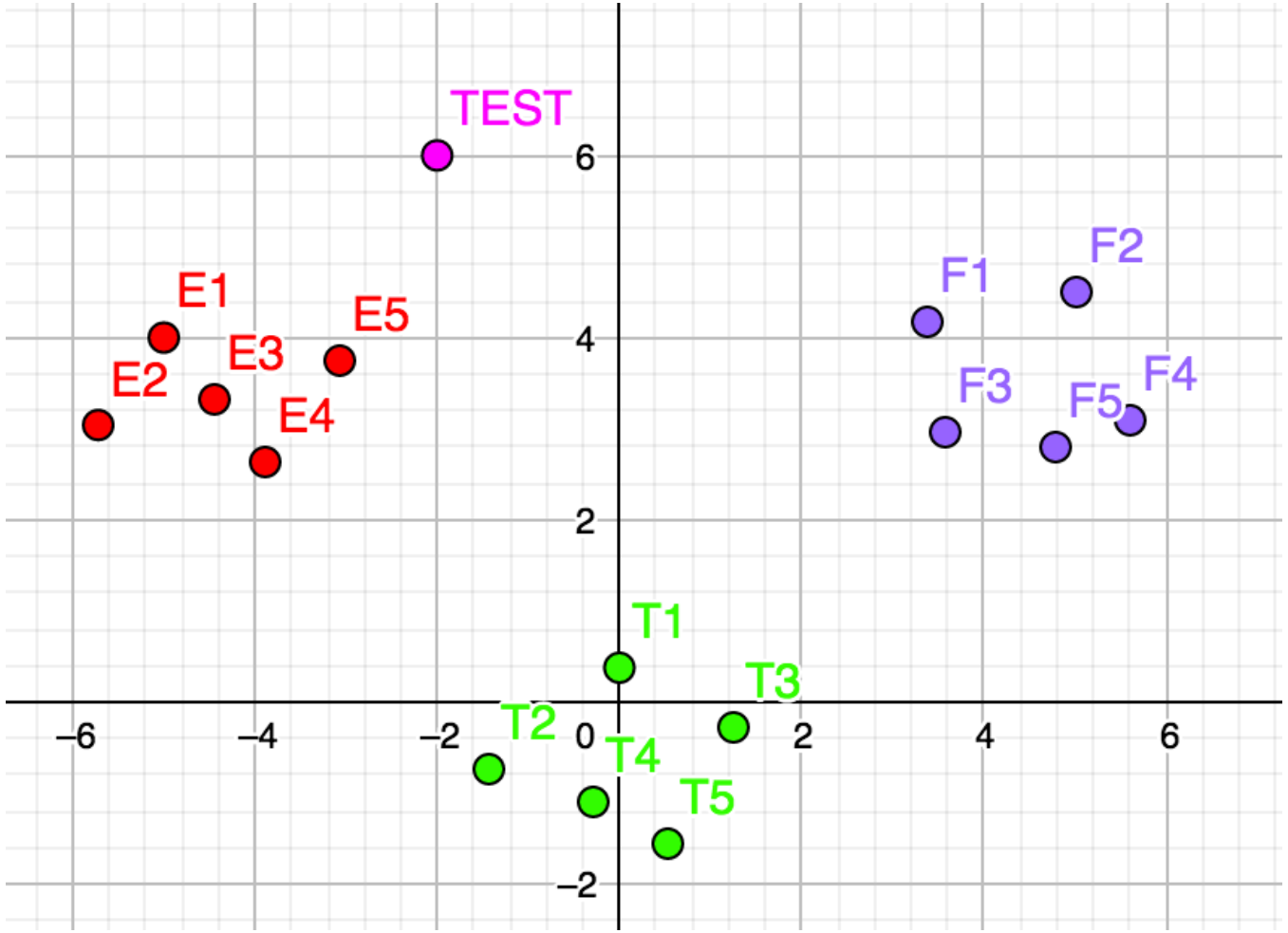
- Companies from energy, financial and technology sectors
- Data gathered from NASDAQ and Yahoo
- 60 companies from each sector
  - 30 **training data**
  - 30 **test data**

Ideal Situation



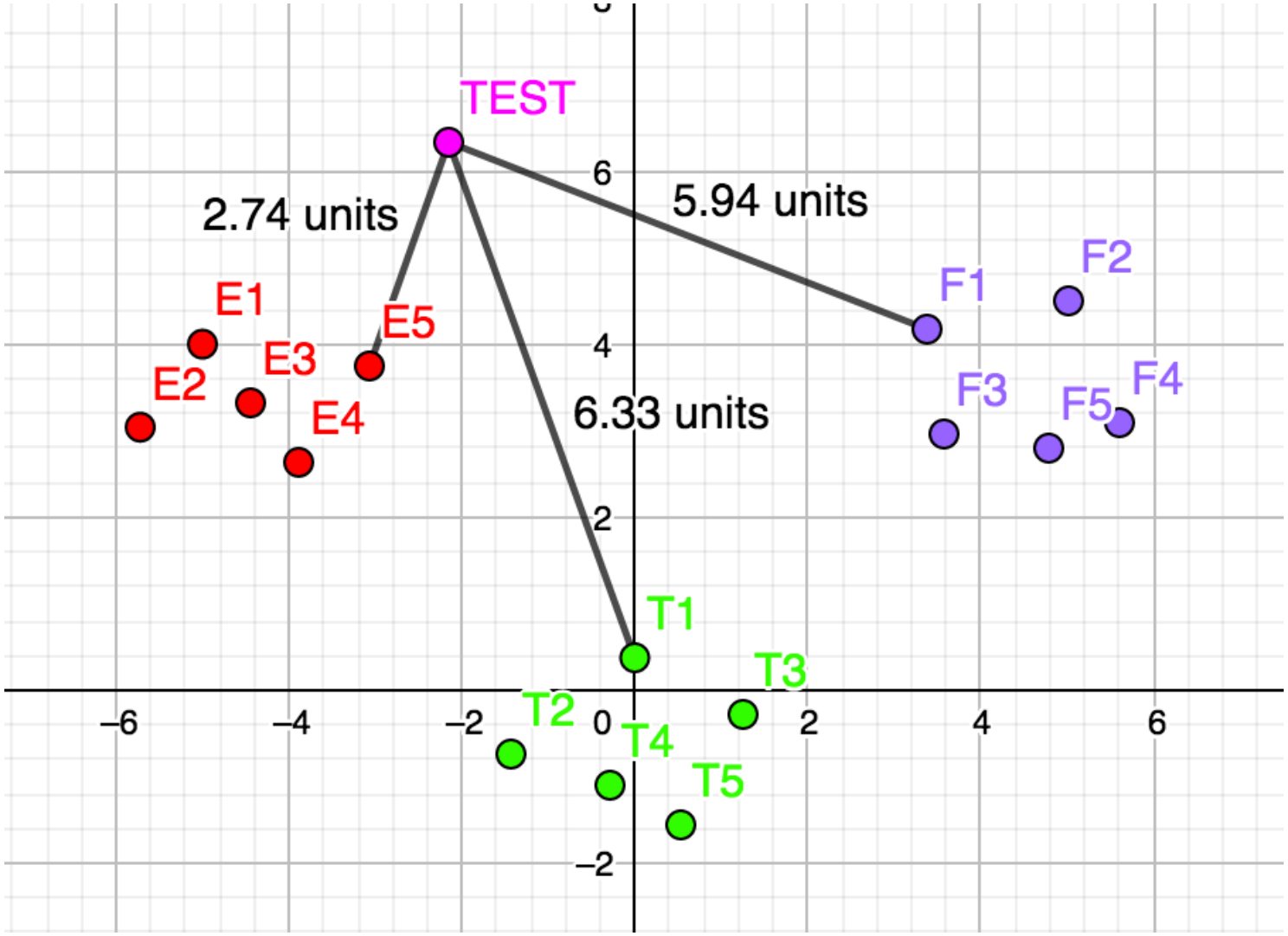
Each point is a persistent diagram from each sector.

Ideal Situation



Distance between persistence diagrams calculated with bottleneck distance.

Ideal Situation



Output of Results

```
E P(MIN AVG IQR SSD): [26.67, 3.33, 53.33, 60.00]
F P(MIN AVG IQR SSD): [26.67, 100.00, 23.33, 50.00]
T P(MIN AVG IQR SSD): [36.67, 0.00, 16.67, 3.33]
```

The program to calculate the bottleneck distance was run 100 times.

Output of Results

```
E Mean(Min Avg IQRavg SDavg): [26.03, 9.37, 50.60, 33.27]
E SD(Min Avg IQRavg SDavg): [7.28, 24.38, 28.18, 34.80]

F Mean(Min Avg IQRavg SDavg): [34.03, 89.00, 29.17, 61.40]
F SD(Min Avg IQRavg SDavg): [9.02, 28.50, 25.17, 34.70]

T Mean(Min Avg IQRavg SDavg): [28.50, 3.47, 11.57, 5.50]
T SD(Min Avg IQRavg SDavg): [7.81, 10.16, 10.25, 9.64]
```

Future Work

- What is the correct statistic to describe the results?
- What is the best way to adjust the sliding window function?
  - Is there a way to match up the parameters to produce a valid result?
- Is there a way to include H0, H1 and H2 when running the data?

References

Joshua Dean, Spring 2014, *Homology using linear algebra.*

Gunnar Carlsson, February 2005, *Computing Persistent Homology*

Robert Ghrist, January 2008, Barcodes: *The Persistent Topology of Data*

Jose A. Perea and John Harer, November 2013, *Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis*

Saba Emrani, Thanos Gentimis, and Hamid Krim, September 2014, *Persistent Homology of Delay Embeddings and its Application to Wheeze Detection*