TOPOLOGY DATA ANALYSIS OF CRITICAL TRANSITIONS IN FINANCIAL NETWORKS

Appu Madathil & Sreya K.K

Index:

Data

Stock price data from 2006 - 2009 of companies listed under DOW index as of the specified period

Introduction

We explore the paper on Topological analysis of Financial networks when approaching a critical transition. The paper proposes a method to detect early signs for critical transitions in financial data. From the time-series of multiple stock prices, time-dependent correlation networks are built , and the topological structures exhibited are analyzed. .We compute the persistent homology associated to these structures in order to track the changes in topology when approaching a critical transition. As a case study, we investigate a portfolio of stocks during a period prior to the US financial crisis of 2007-2008, and show the presence of early signs of the critical transition.

A critical transition refers to an abrupt change in the behavior of a complex system {arising due to small changes in the external conditions {, which makes the system switch from one steady state to some other steady state, after undergoing a rapid transient process

We consider systems that can be described as time-varying weighted networks, and we track the changes in the topology of the network as the system approaches a critical transition.

Theory :

- Persistent homology is a computational method to extract topological features from a given data set(weighted network in our case) and rank them according to some threshold parameter (the distance between data points or the weight of the edges)
- When the threshold parameter is varied, the corresponding simplicial complexes form a filtration(i.e., an ordering of the simplicial complexes that is compatible with the ordering of the threshold values).

- The topological features (e.g., connected components, `holes' of various dimensions) of the simplicial complexes across the ltration, and record for each topological feature the value of the parameter at which that feature appears for the first time (`birth value'), and the value of the parameter at which the feature disappears (`death value').

Persistent Homology of Filtration Networks:

- Weighted network is a pair consisting of a graph and a weight function associated to its edges w :

$$G = G(V, E)$$

$$E \to [0, +\infty);\ \text{let}\ \theta_{\max} = \max(w).$$

- When the threshold parameter is varied, the corresponding simplicial complexes form a filtration(i.e., an ordering of the simplicial complexes that is compatible with the ordering of the threshold values)

- Graphs here are simple and undirected
- Investigate the topology of weighted graphs is via thresholding, by considering only those edges whose weights are below (or above) some suitable threshold, and study the features of the resulting graph.
- Using persistent homology, we can extract the topological features for each threshold graph,and represent all these features, ranked according to their `life span', in a persistent diagram.

Method:

For each

$$\theta \in [0, \theta_{\max}],$$

we consider the sub-level sets of the weight function, that is, we restrict to subgraphs of G, which keeps all edges of weights '**w**' below or equal to the threshold theta.

The graphs obtained by restricting to successive thresholds have the ltration property, i.e.

$$\theta \leqslant \theta' \implies G(\theta) \subseteq G(\theta').$$

In a similar way, we can consider super-level sets, by restricting to subgraphs above or equal to the threshold theta.

Super-level sets can be thought of as sub-level sets of the weight function

$$w' = \theta_{\max} - w.$$

For each graph G, we construct the Rips complex (clique complex) :

$$K = X(G(\theta))$$

This is defined as the simplicial complex of all complete subgraphs(cliques) of G, as its faces. That is the 0-skeleton of K consists of just the vertices of G as its faces

The 1-skeleton of all vertices and edges which is the graph G itself the 2 skeleton of all vertices edges and filled triangles etc

High dimensional cliques correspond to highly interconnected clusters of nodes with similar characteristics(as encoded by the weight function)

The filtration of the threshold subgraphs yields a corresponding filtration of the Rips complexes :

$$\theta \mapsto K_\theta := X(G(\theta)); \text{ thus, } \theta \leqslant \theta' \implies K_\theta \subseteq K_{\theta'}$$

The homology groups associated with the filtration satisfy the filtration property

$$\text{i.e., } \theta \leqslant \theta' \implies H_i(K_\theta) \subseteq H_i(K_{\theta'})$$

Correlation Networks:

Compute the Pearson correlation coefficient between the nodes i and j at time t, over a time horizon T, by

$$c_{i,j}(t) = \frac{\sum_{\tau=t-T}^{t}(x_i(\tau)-\bar{x}_i)(x_j(\tau)-\bar{x}_j)}{\sqrt{\sum_{\tau=t-T}^{t}(x_i(\tau)-\bar{x}_i)^2}\sqrt{\sum_{\tau'=t-T}^{t}(x_j(\tau')-\bar{x}_j)^2}}$$

where $\bar{x}_i$, $\bar{x}_j$ denote the averages of $x_i(t)$, $x_j(t)$ respectively, over the time interval $[t-T, t]$;

Compute the distance between the nodes i and j as :

$$d_{i,j}(t) = \sqrt{2(1-c_{i,j}(t))}$$

Assign the weight **'w'** to the edge between i and j

$$w(e,t) = d_{i,j}(t)$$

The range of values of **d(i,j) is [0,2]**

If the nodes are perfectly correlated then **d(i,j) = 0**

If the nodes are perfectly anti-correlated, then **d(i,j) = 2**

Edges between correlated edges have higher weights and uncorrelated edges have lower weights

Sublevel & Superlevel sets:

Each **sub-level set** of weight function **w,** at a threshold level yields a subgraph G containing only those edges for which :

$$0 \leqslant d_{i,j} \leqslant \theta,$$

$$G(\theta) = \{e = e(i,j) \mid 1 - \frac{1}{2}\theta^2 \leqslant c_{i,j} \leqslant 1.\}$$

When the parameter theta is small **G** contains only edges between edges between highly-correlated nodes. As theta is increased upto the critical value **1.414,** edges between low correlated nodes are progressively added to the network. As theta is increased further, edges between anti correlated nodes appear in the network.

Each **super-level** set of the weight function **w = d,** can be conceived as a sub-level set of the dual function :

:

$$w' = 2 - d.$$

The **sub-level** set **G** for this weight function contains only those edges for which :

$$d_{i,j} \geqslant 2 - \theta, \text{ hence } G_{w'}(\theta) = \{e = e(i,j)\,|$$

Hence :

$$G_{w'}(\theta) = \{e = e(i,j)\,| \quad 1 \leqslant c_{i,j} \leqslant 1 - \frac{1}{2}(2 - \theta)^2.\}$$

When the parameter **theta** is small **G** , contains only edges between anti-correlated nodes. When **theta** crosses the critical value : **0.5857,** edges between low correlated nodes are progressively added to the network. As theta is increased further and approaches **2**, highly correlated nodes are added to the network.

Results:

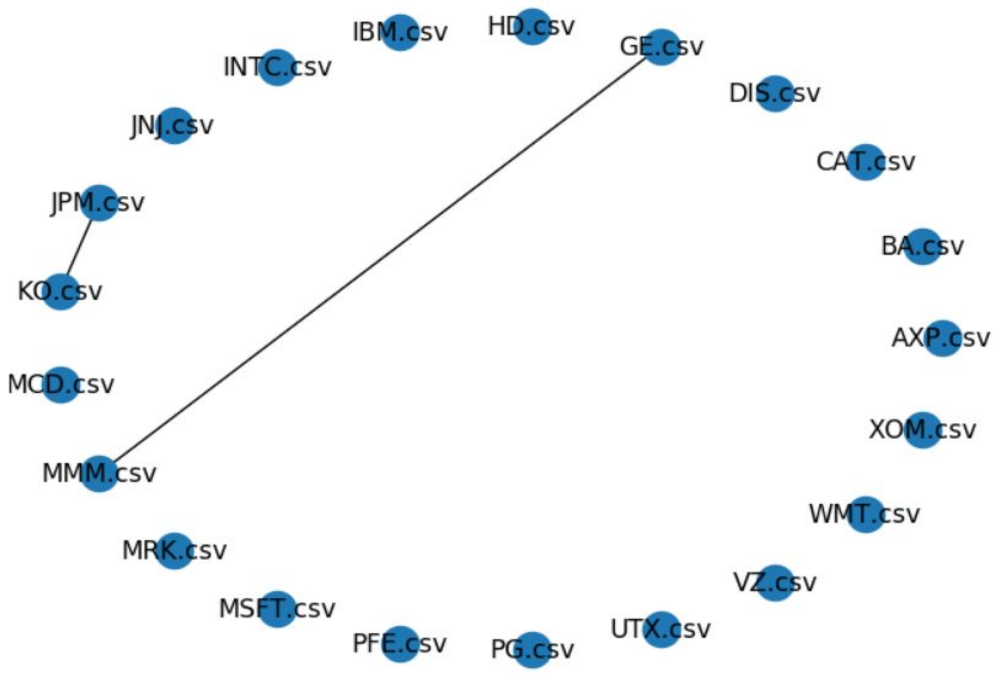
We compute the correlation matrix and then compute the weight matrix. This is then used to plot the persistence diagrams corresponding to the sublevel and superlevel sets of the Weight function. The time period for analysing correlation is taken as **t = 10 days** since the paper states that the empirical studies indicate non stationarity in time series with higher time periods and we would want to eliminate seasonal dependence in the computation of the covariance matrix.

The Correlation network associated with sample periods are also provided below. Since we anticipate lack of highly correlated share price movements in the time periods away from crisis, we expect lesser number of edges between nodes and higher number of edges in the period of late 2007 - 2008 leading upto the crisis.

Correlation networks :

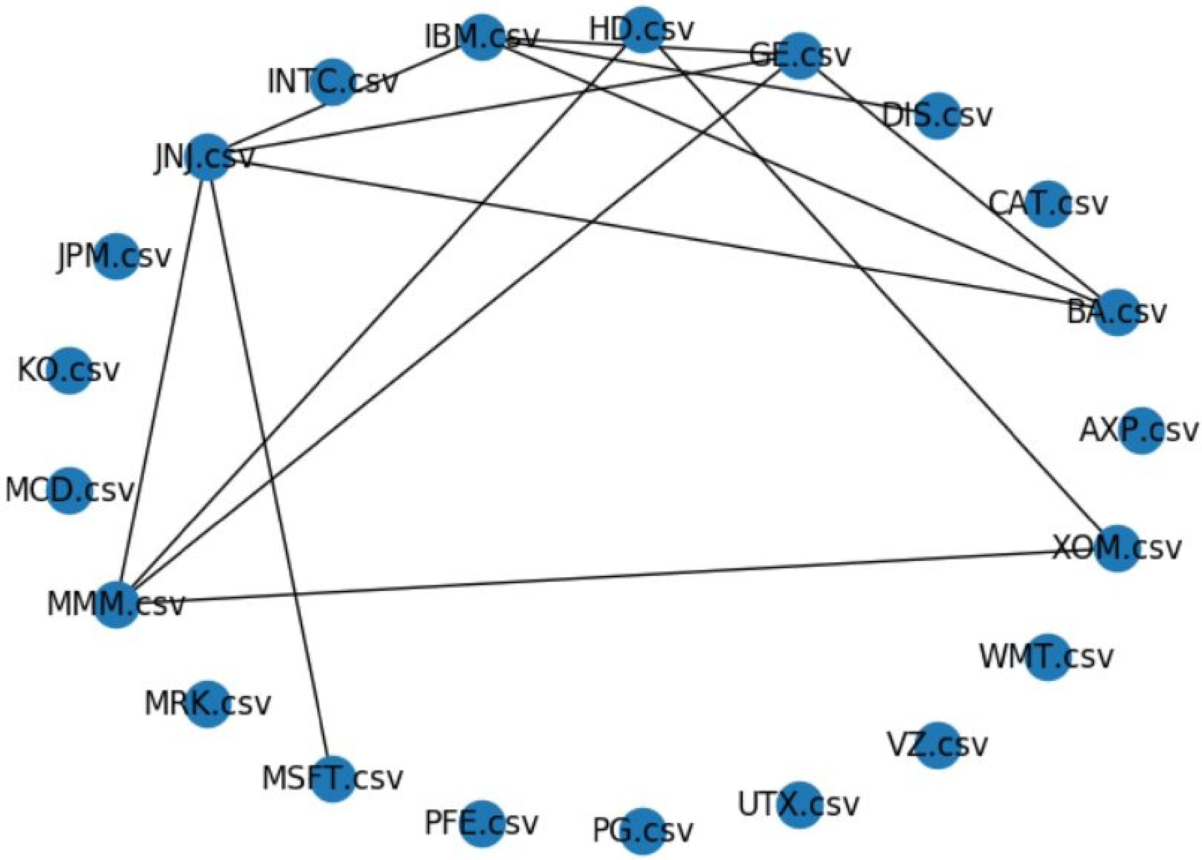
2006 Jan:

```
Number of edges :
2
```



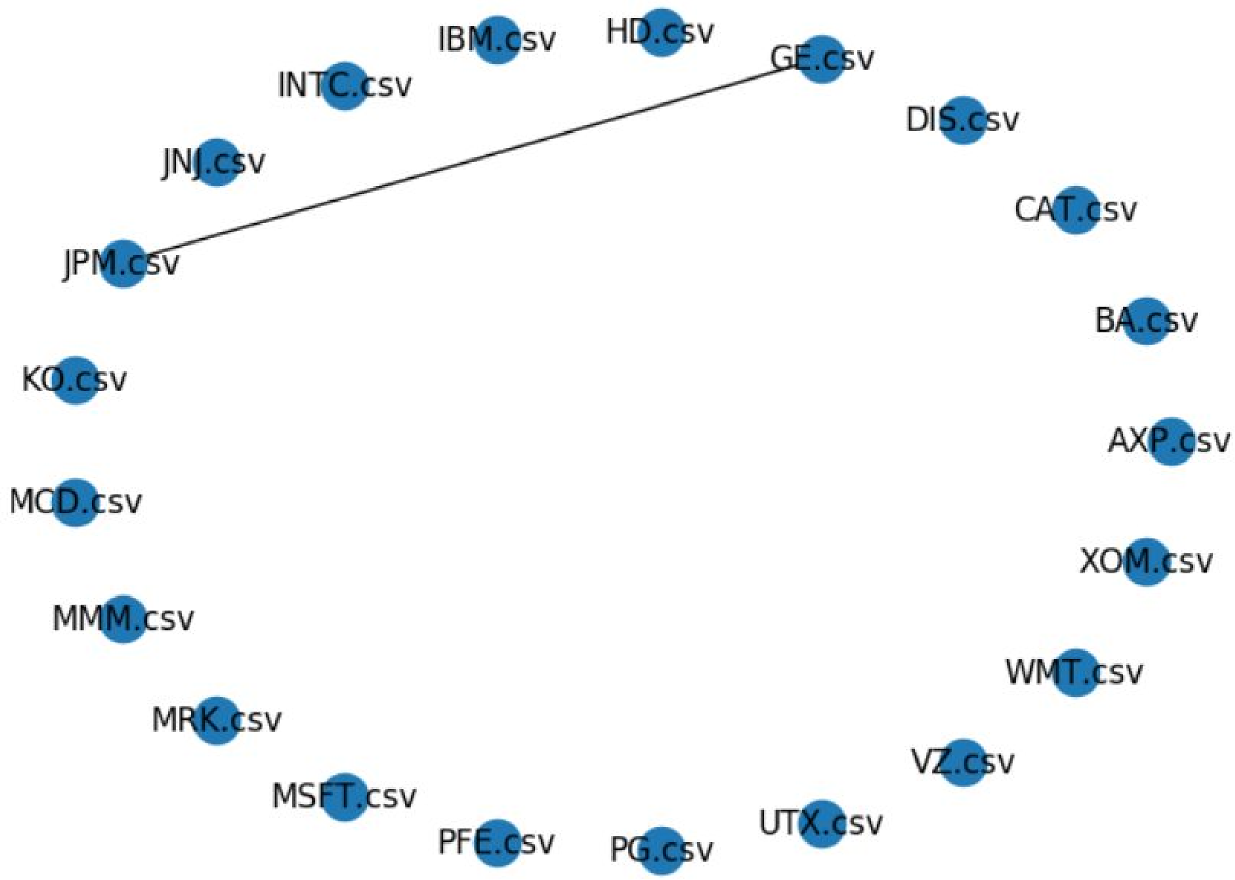
2006 March :

```
Number of edges :
13
```



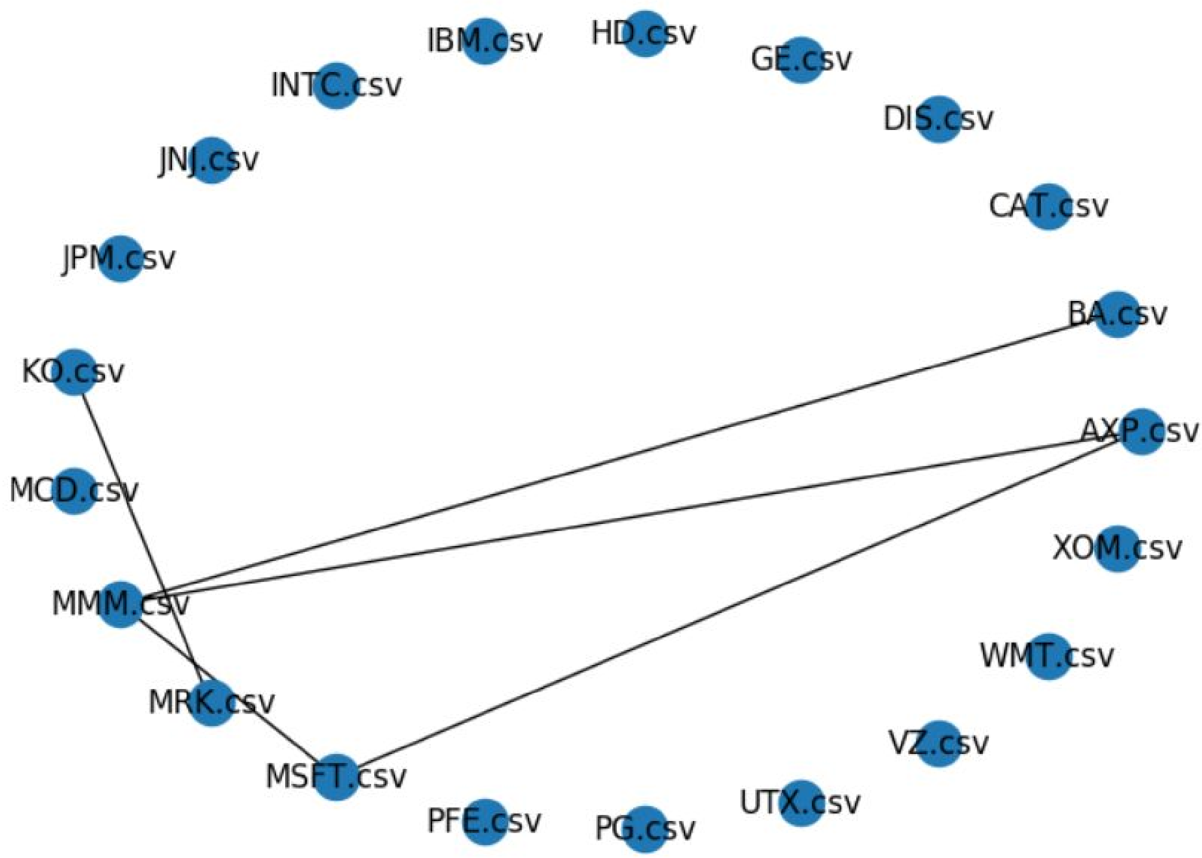
2006 April :

```
Number of edges :
1
```



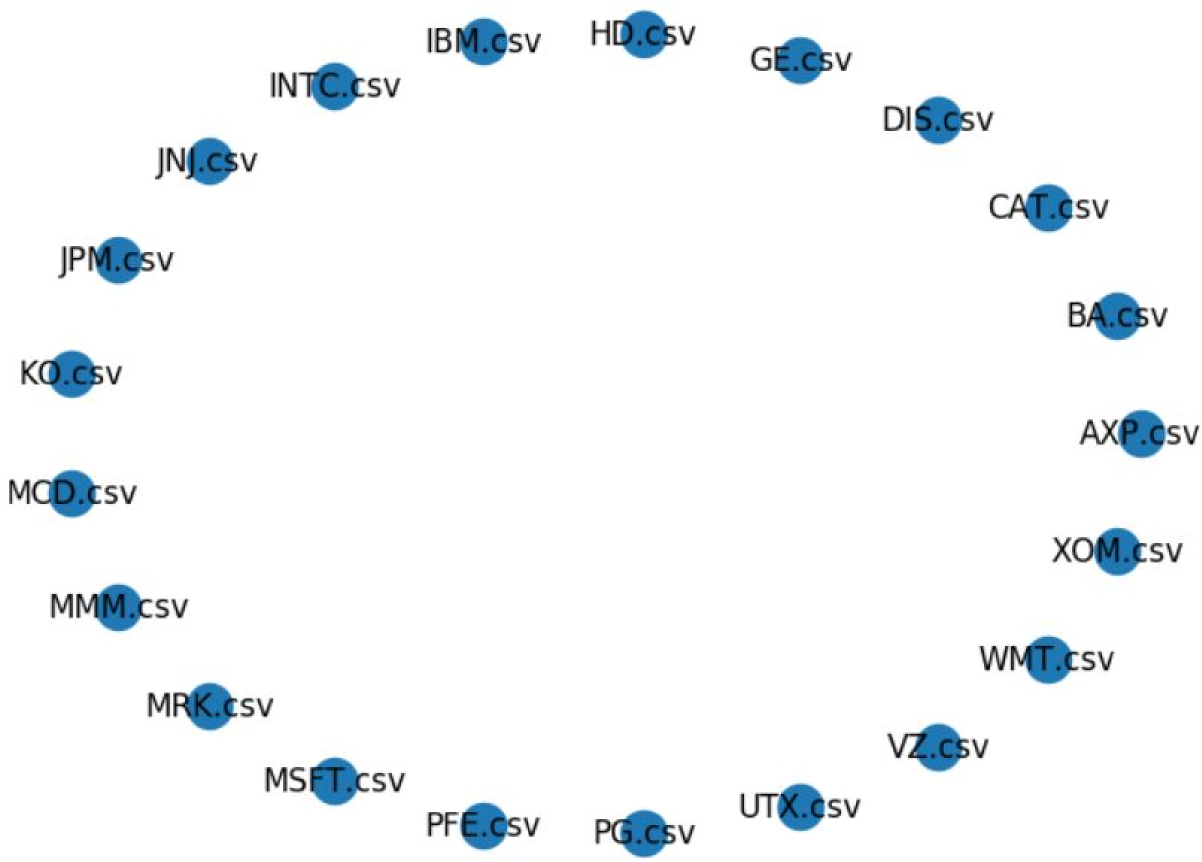
2009 April:

```
Number of edges :
5
```



2009 September:

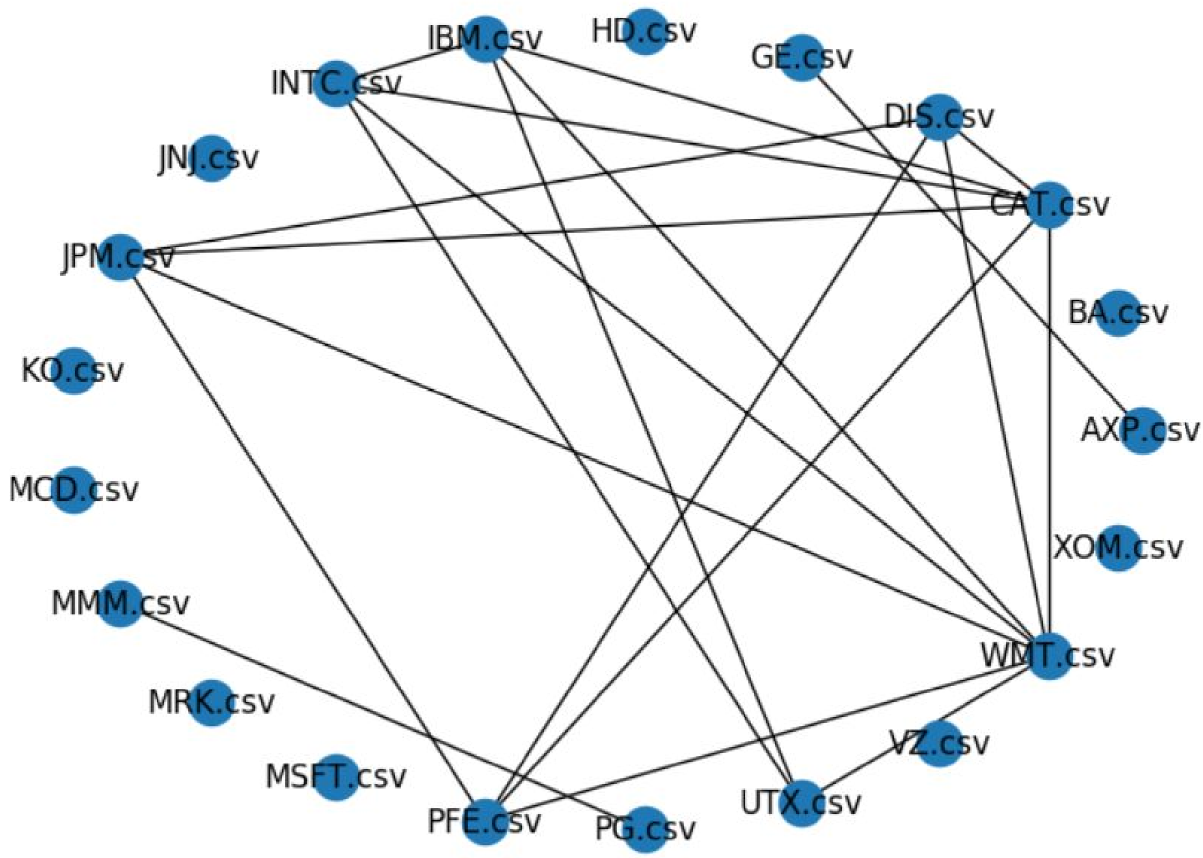
```
Number of edges :
0
```



Approaching Crisis & During the Crisis (2007-2008)

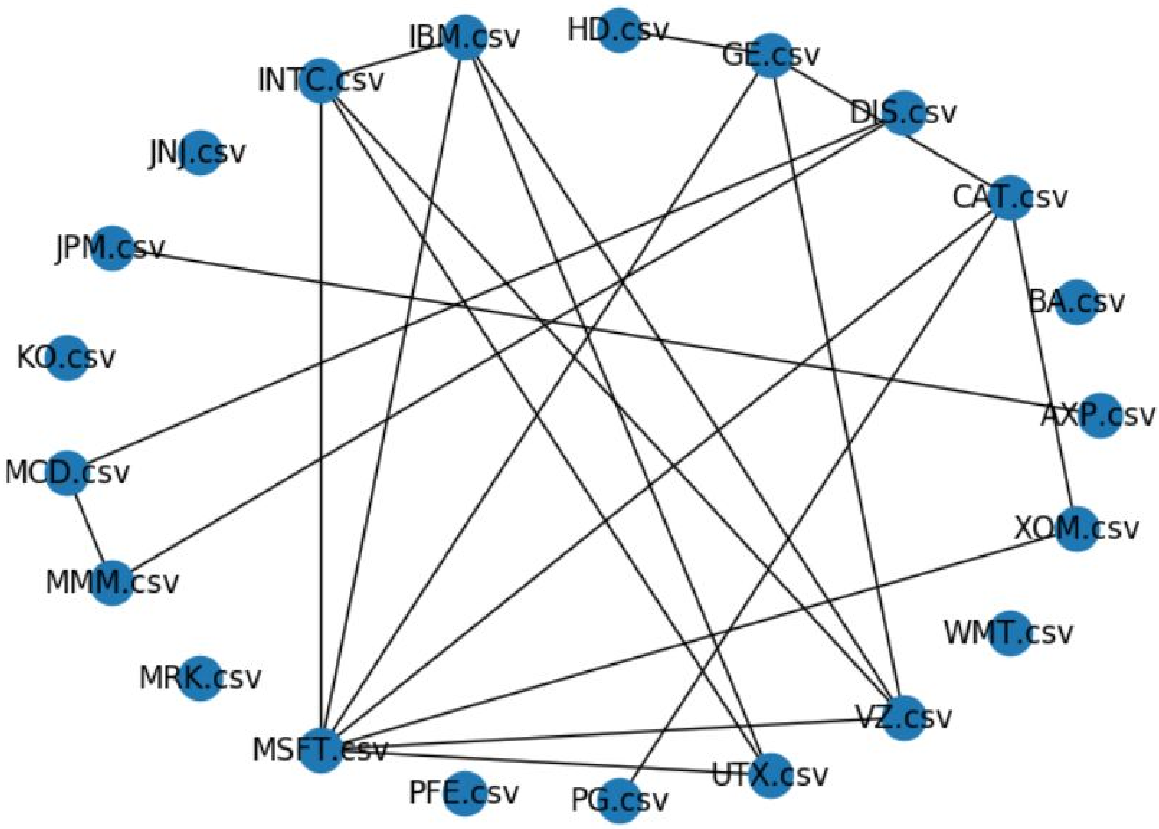
2007 November :

```
Number of edges :
20
```



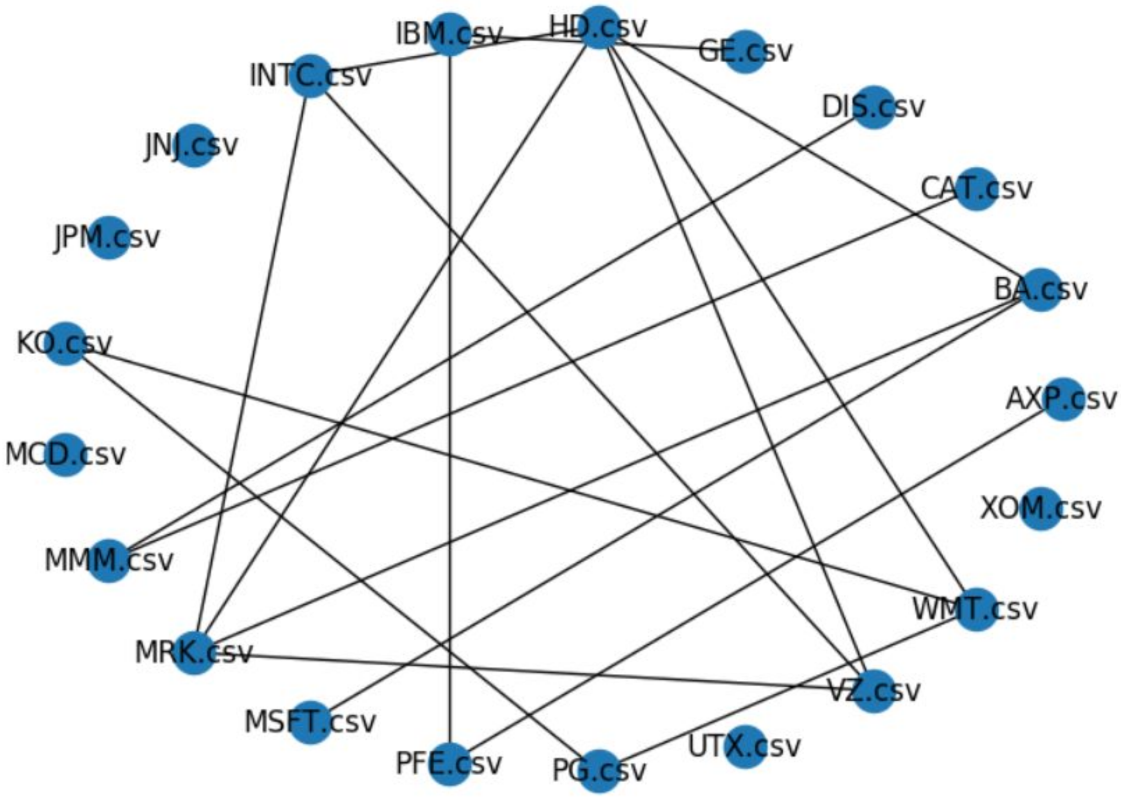
2007 December :

```
Number of edges :
21
```



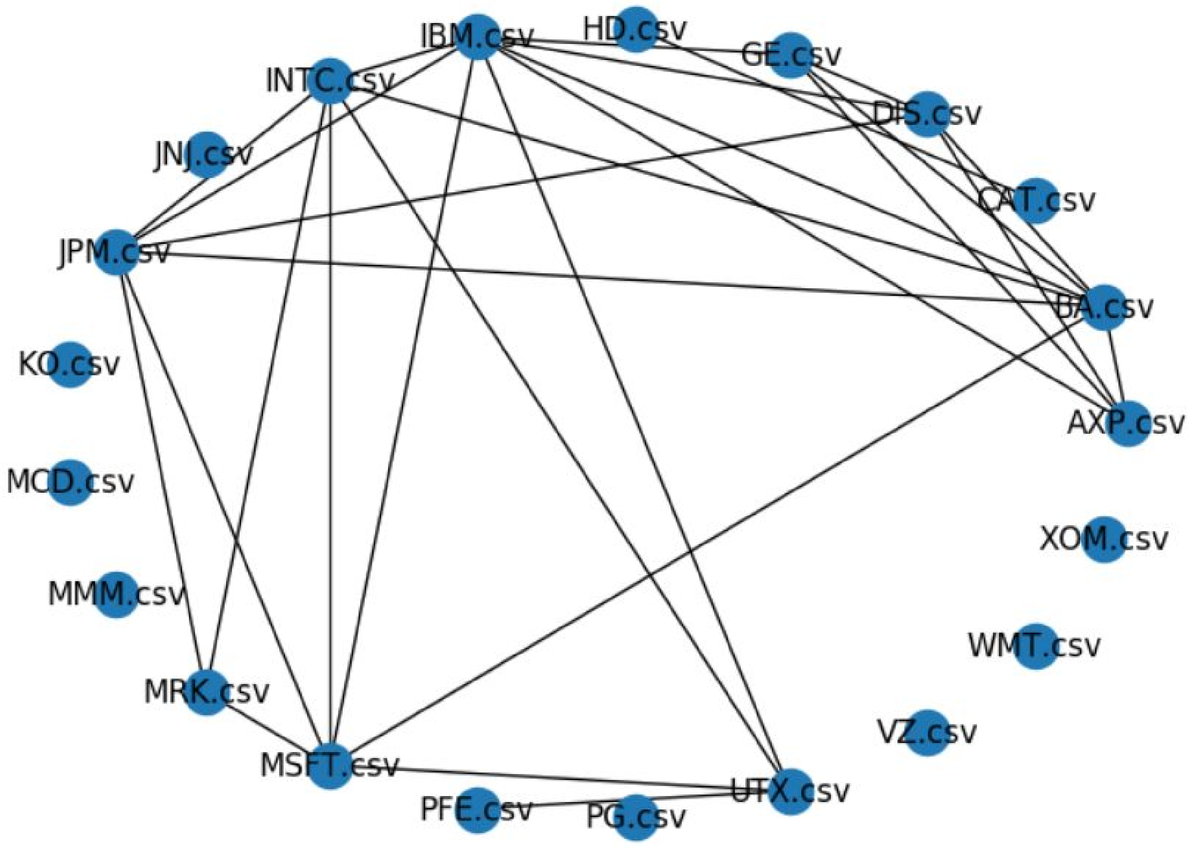
2008 September

```
Number of edges :
18
```



2008 November :

```
Number of edges :
28
```



Sublevel sets : Persistence diagrams

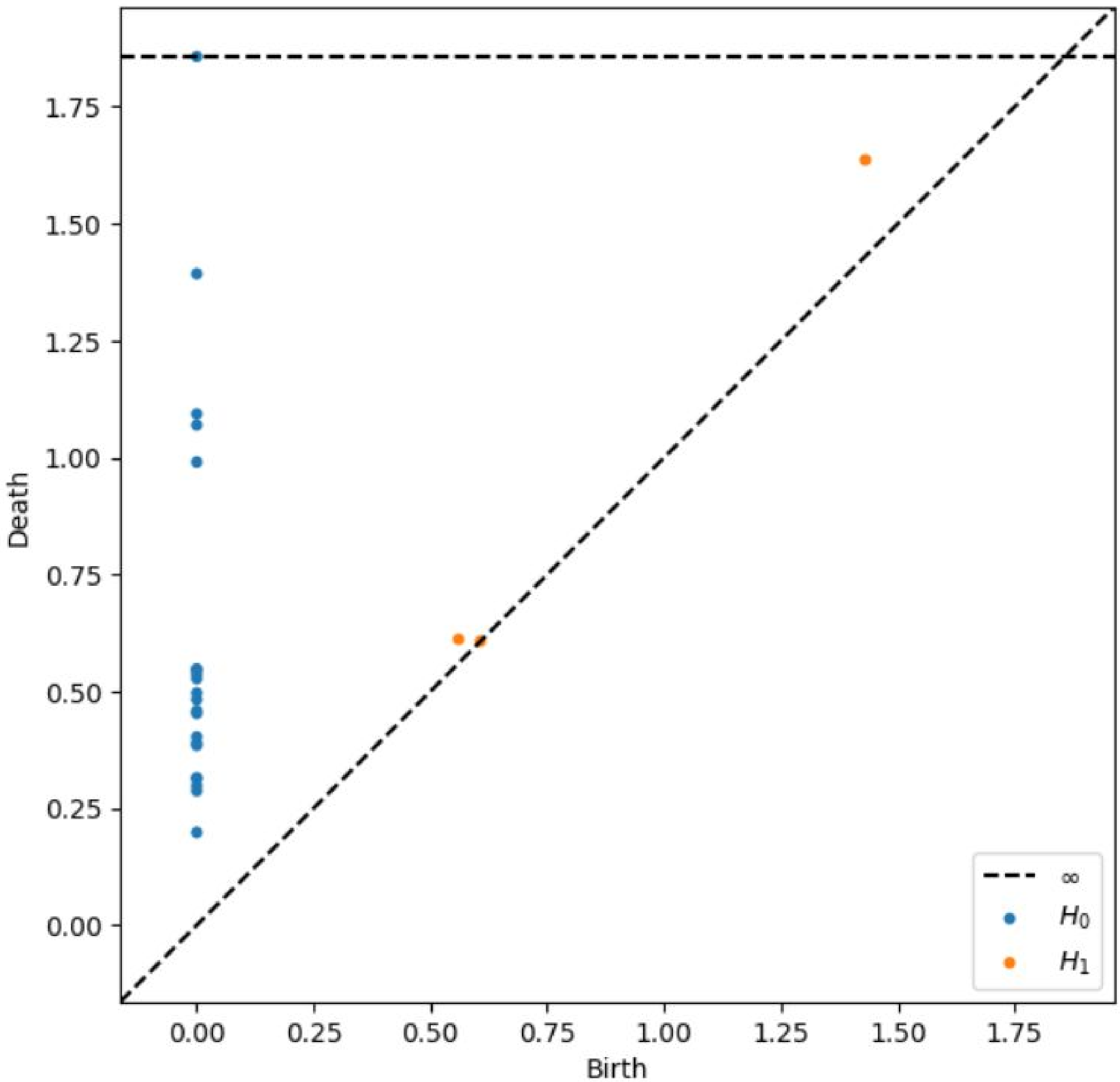
The 0-dimensional persistent homology provides information on how the network connectivity changes as the value of theta in increased from 0 to 2. Each black dot on the persistent diagram corresponds to one (or several) connected component of the graph. The horizontal coordinate of each dot is 0, since all components are born at threshold value **theta** = 0. The vertical coordinate of a dot gives the threshold value, theta, at which a connected component dies, by joining together with another connected component. The dot with highest vertical coordinate (other than 2) gives the threshold value, **theta,** for which the graph becomes really connected. A dot at 2 (the maximum value) indicates that once the graph is fully connected, it remains fully connected (hence the component never dies) as , **theta,** is further increased. Dots with lower vertical coordinates indicate threshold values for which smaller connected components consisting of highly correlated nodes die, i.e., coalesce together into larger components. Dots with higher vertical coordinates correspond to death of connected components due to the appearance of edges between uncorrelated or anti-correlated nodes. We recall that the critical value of that marks the passage from correlation to anti-correlation is 1.41

Now we interpret the 1-dimensional persistent homology, illustrated in by red marks. The horizontal coordinate of a mark gives the birth value of a loop in the network, and the vertical coordinate gives the death value of that loop. The death of a loop occurs when edges between the nodes of the loop appear and form complete 2-simplices (filled triangles) that filledup the loop. Dots with low coordinates indicate the presence of cliques that are highly correlated. Marks with higher vertical coordinates indicate the death of loops due to edges between low-correlated or anti-correlated nodes.

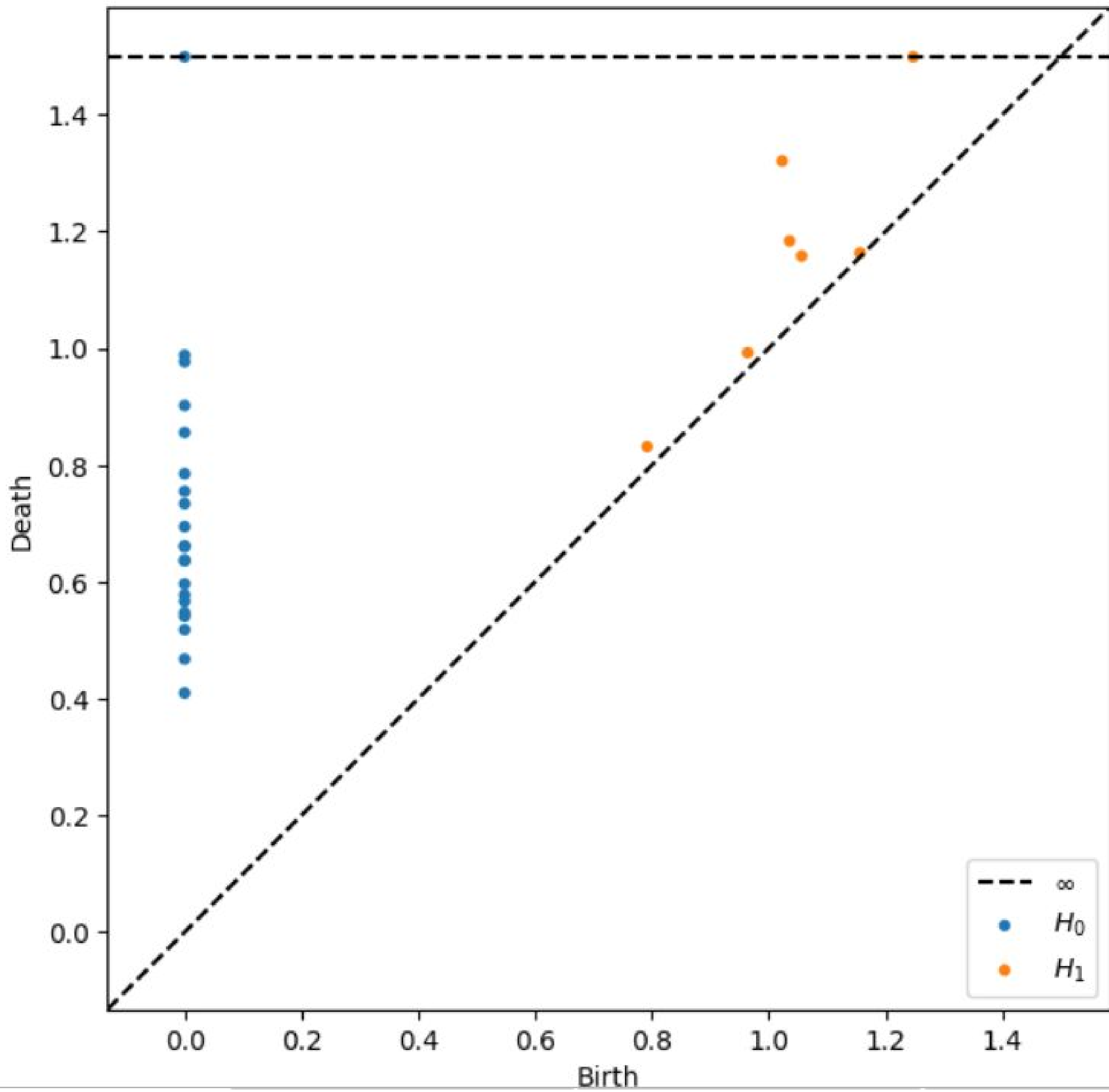
There is a concentration of ‘dots’ in the higher coordinates for periods away from crisis and distance of dots away from the diagonal indicating lack of correlation in the network prior to crisis in comparison with the periods 2007-2008 ,

Time periods away from crisis :

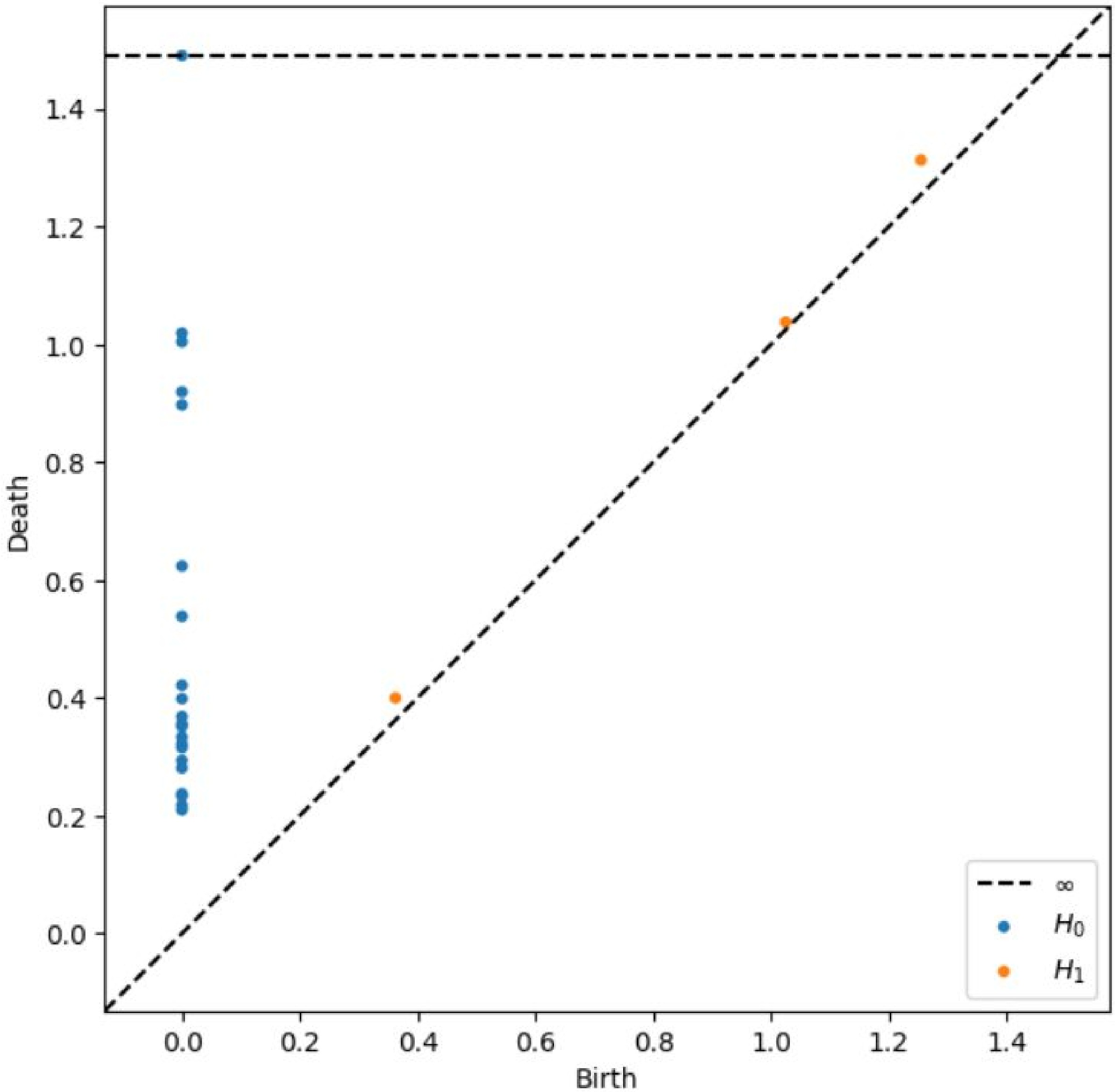
2006 March



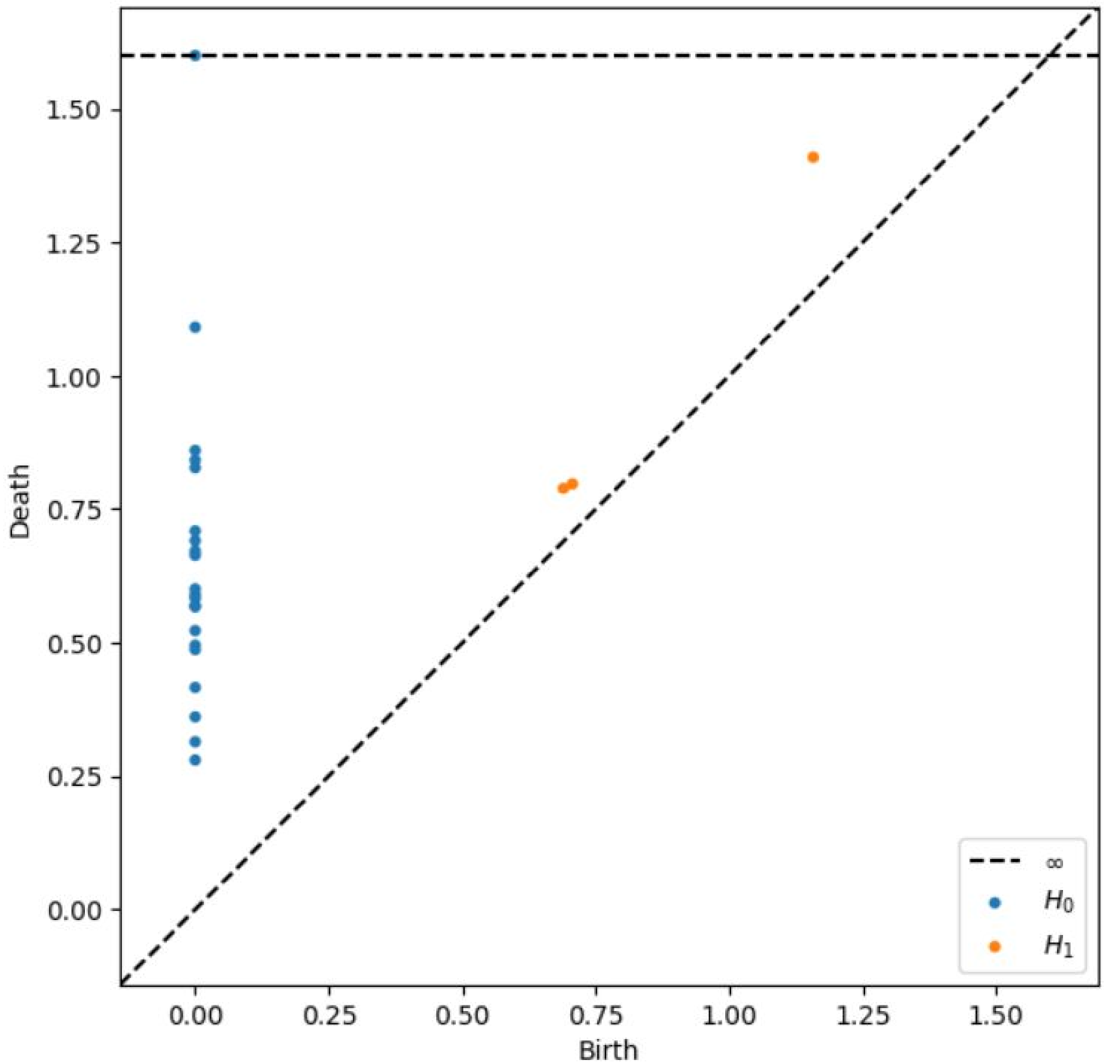
2006 April



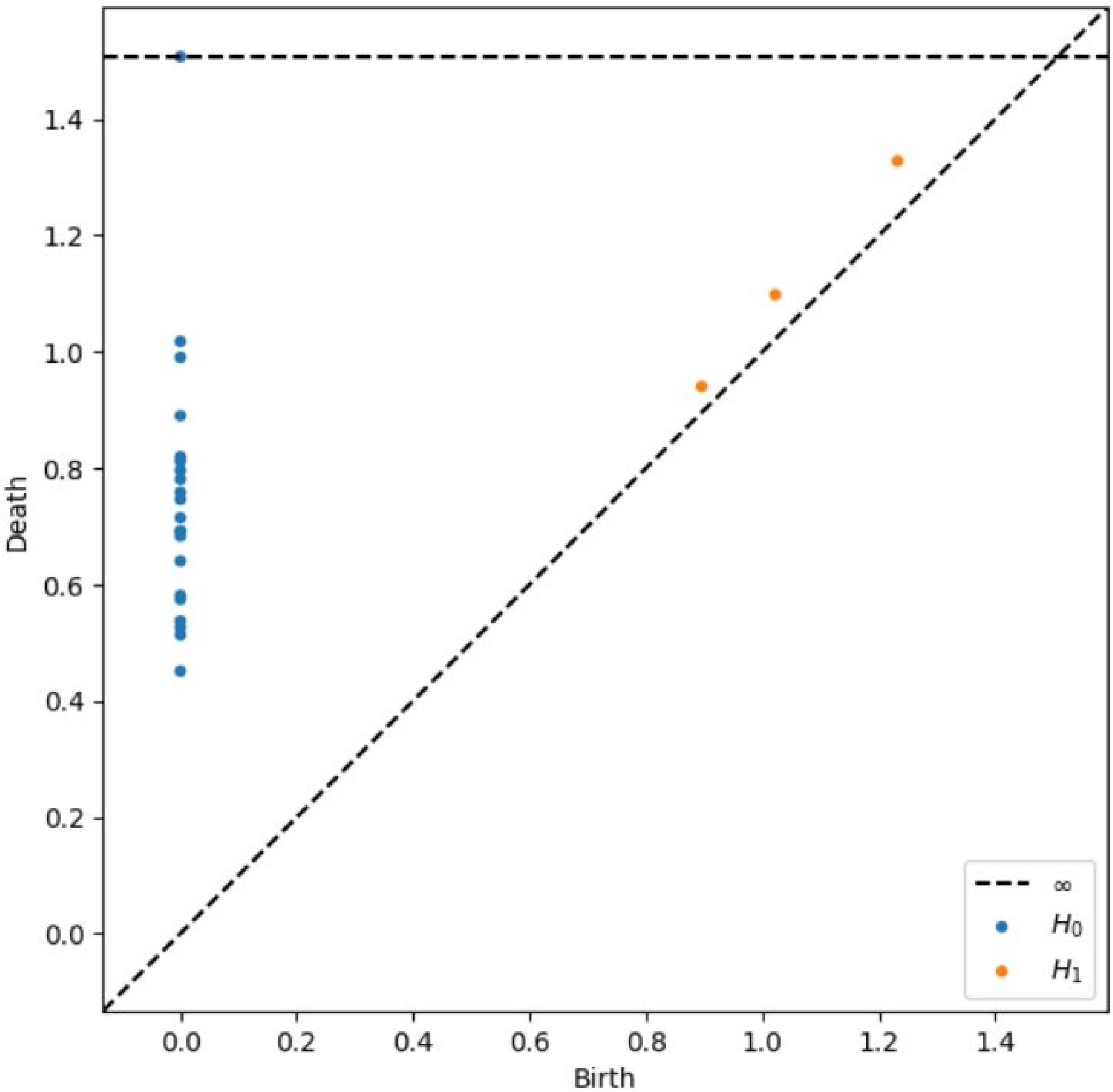
2006 May



2009 April:

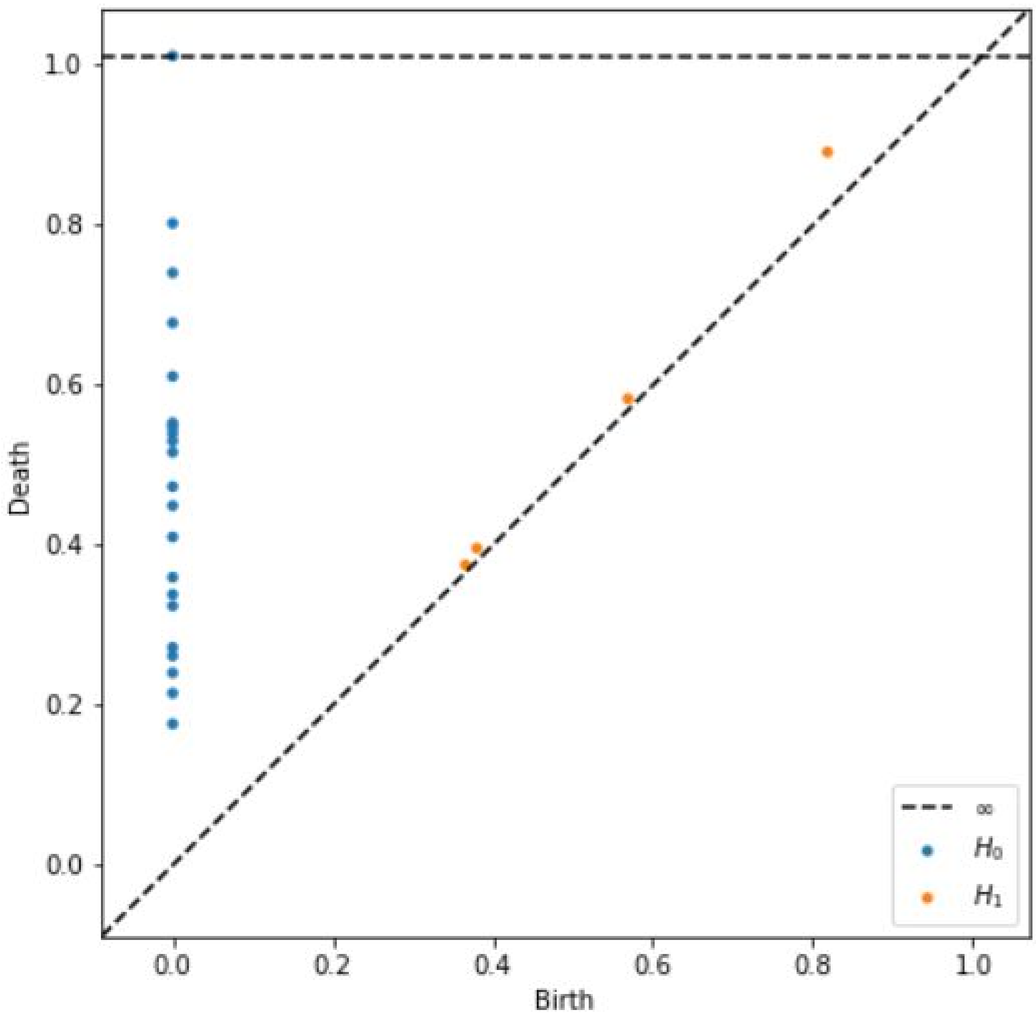


2009 SEPTEMBER :

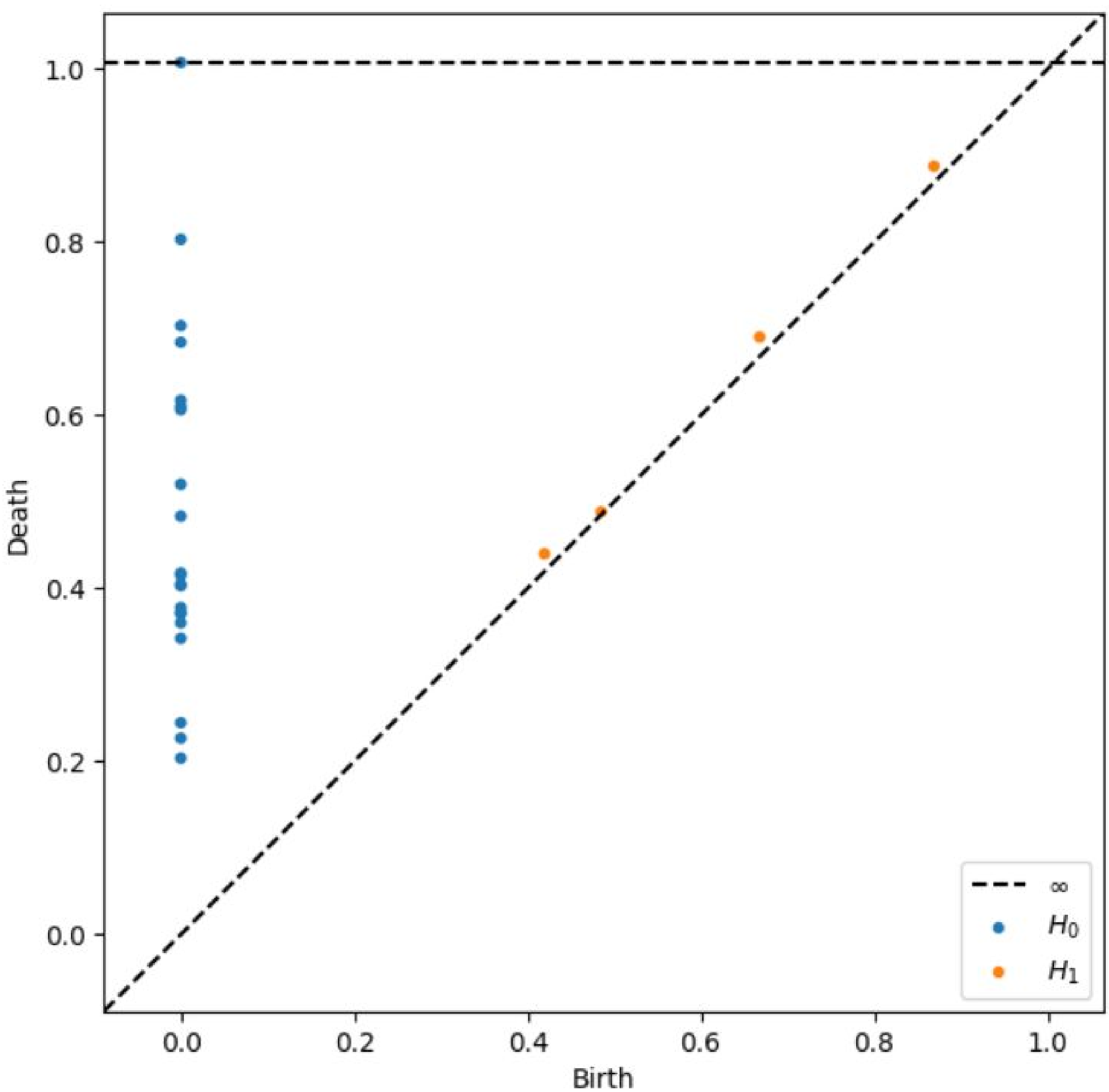


Sublevel sets(Entering into crisis in 2007 - 2008)

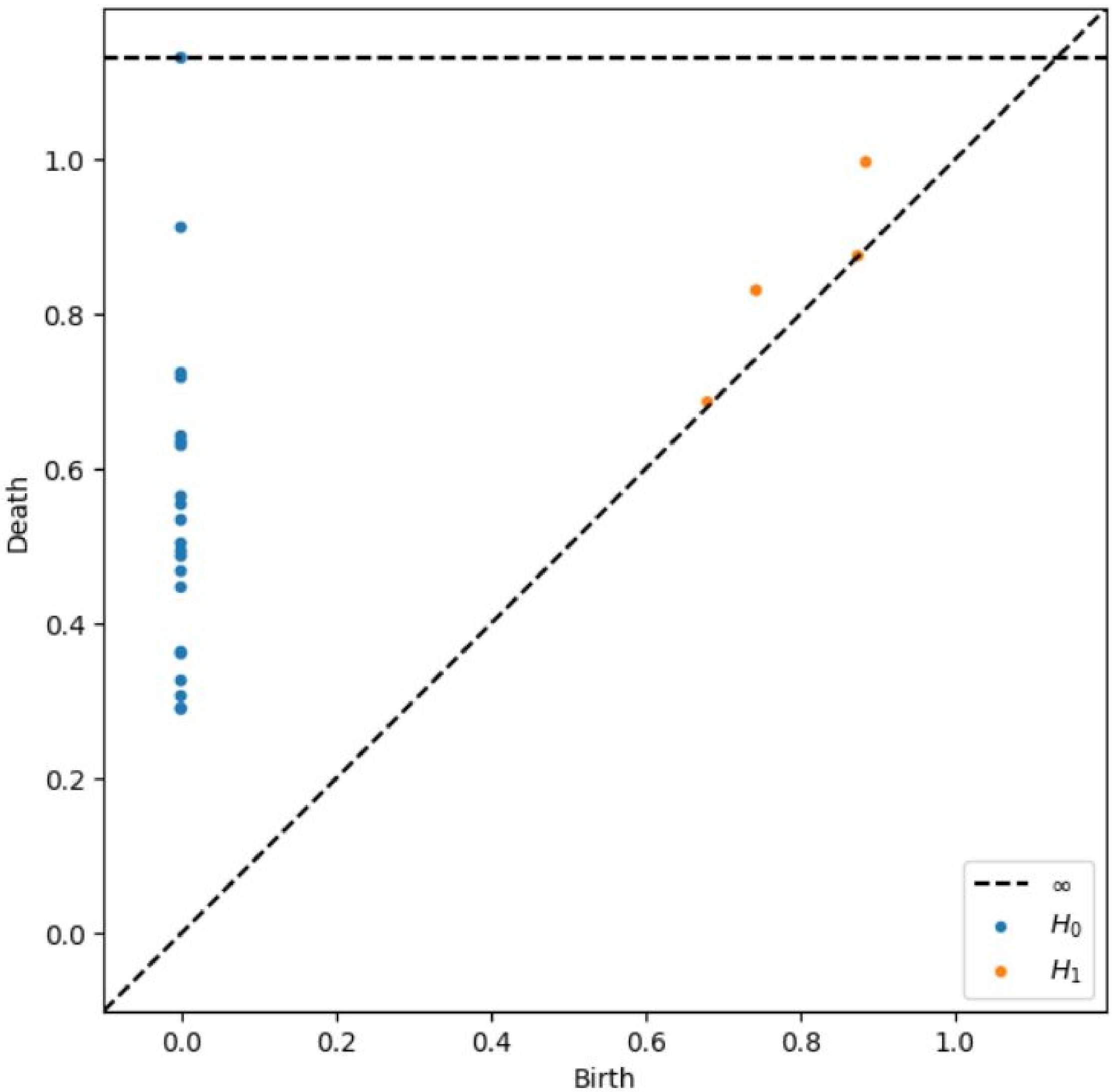
2007 November :



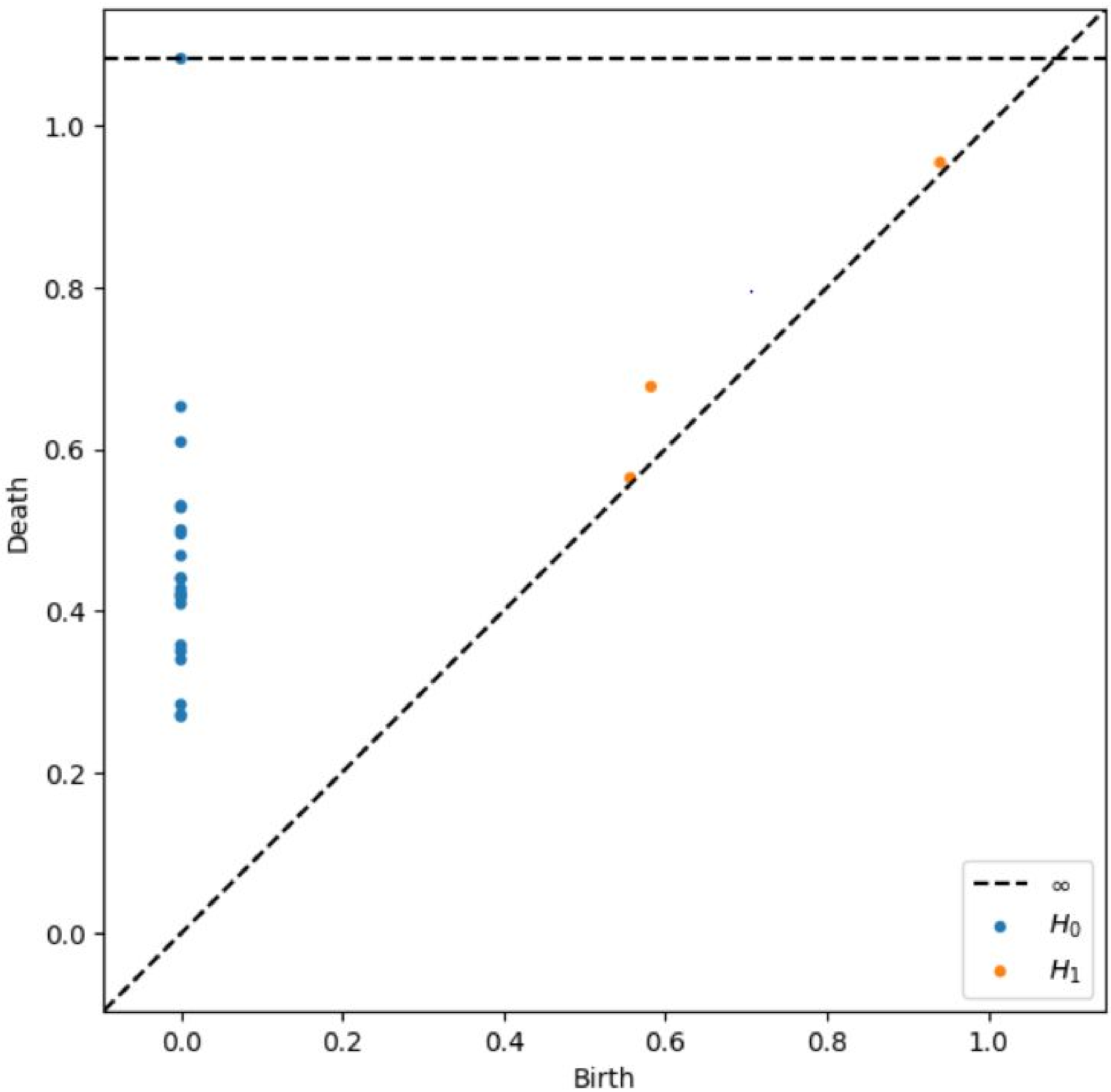
2007 December :



2008 Feb :

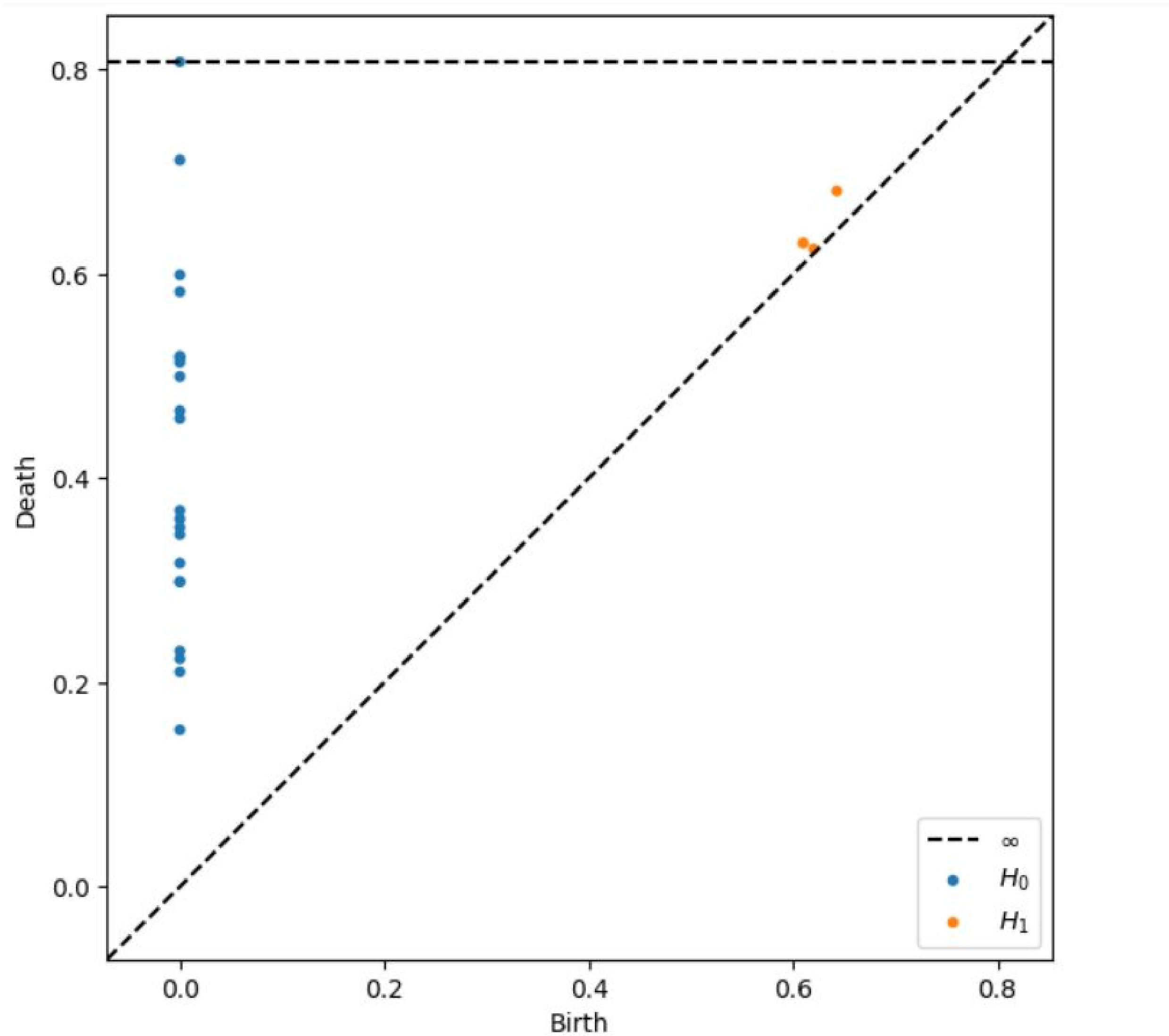


2008 September :



1.0
0.8
0.6
0.4
0.2
0.0
Death
0.0
0.2
0.4
0.6
0.8
1.0
Birth
∞
H_0
H_1

2008 November :



Superlevel sets : Persistence Bar diagrams

We now compute the super-levels sets of w, which are sub-level set of w1. The resulting persistent diagrams have a different interpretation. The critical value of the threshold **'theta'** for the switch from anti-correlation to correlations is 0:5857864. Points in the persistent diagram with low vertical coordinates correspond to anti-correlation/non-correlation, and points with higher value of the vertical coordinate (other than 2) indicates the appearance of edges between correlated nodes. A point on the persistent diagram with higher vertical coordinate represents the death of a connected component (or a loop), possibly formed by anti-correlated or low correlated nodes, when an edge between correlated nodes is added to the networks.Thus, the homology generators identified by the persistent diagrams represent cliques of stocks associated to `normal market conditions (which are associated with lack of correlation). The death of these generators is caused by the addition of correlated edges to the threshold network (in dimension 0, by joining together different connected components, and in dimension 1 by closing the loops).

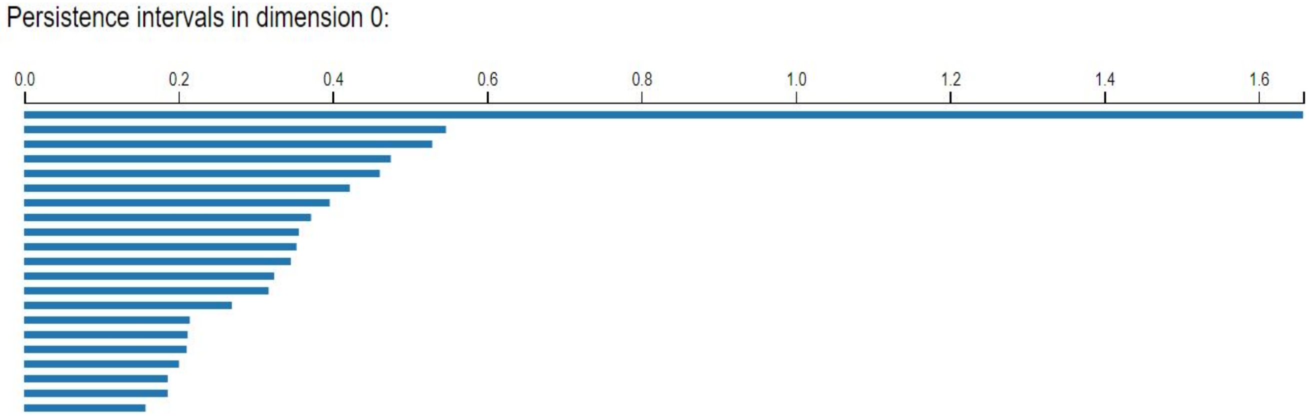
Note that here we have plotted persistent bar diagrams which have the similar interpretation but plotted as horizontal bars instead of dots in the persistence diagrams. The interpretation remains the same : The lifespan of a component is now represented as bars for the 0th and 1st homology groups instead of dots.

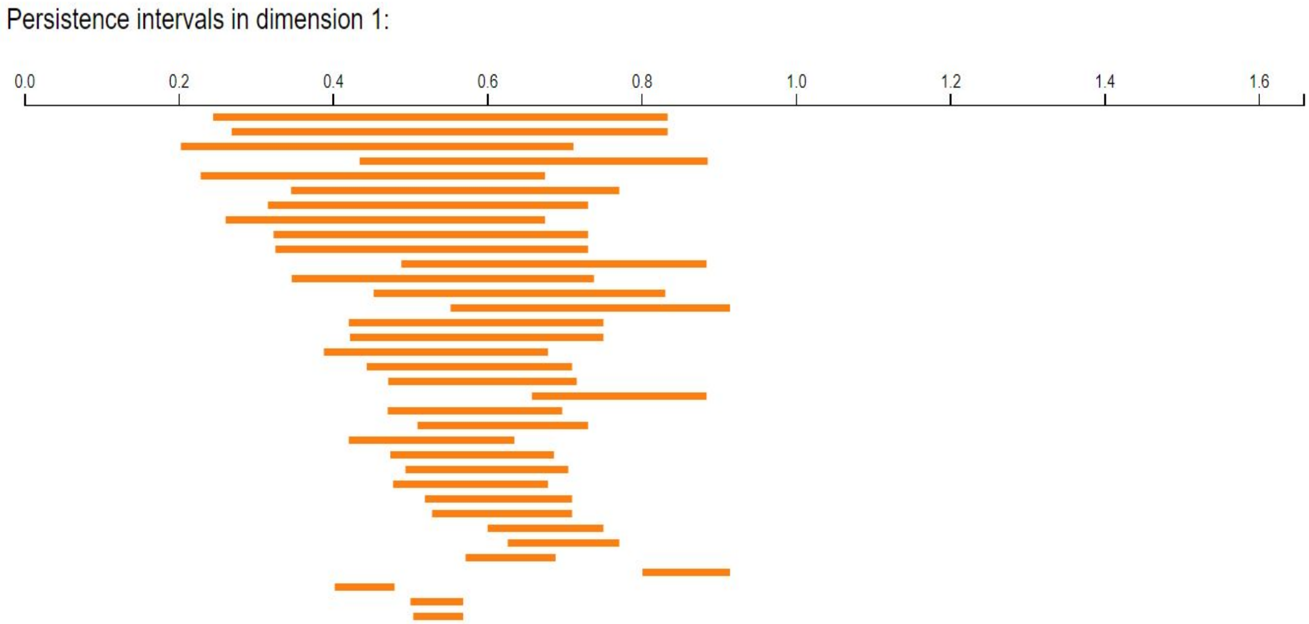
In contrast with the sublevel sets of the weight function now we see that periods away from the crisis have smaller coordinate value for the **0-homology group** indicating once again that there is lesser correlation in comparison with the periods near crisis. This is because all the edges gets added (based on the new weight function **2-w**) as there is lesser extent of anti correlation in periods away from crisis and we approach the critical value of non correlation within a smaller distance in comparison to the critical transition periods

We cannot make a general observation regarding the persistence of the **1-homology** groups indicating the birth and death of loops, when comparing time periods closer to the transition and away from transition based on the samples analysed.

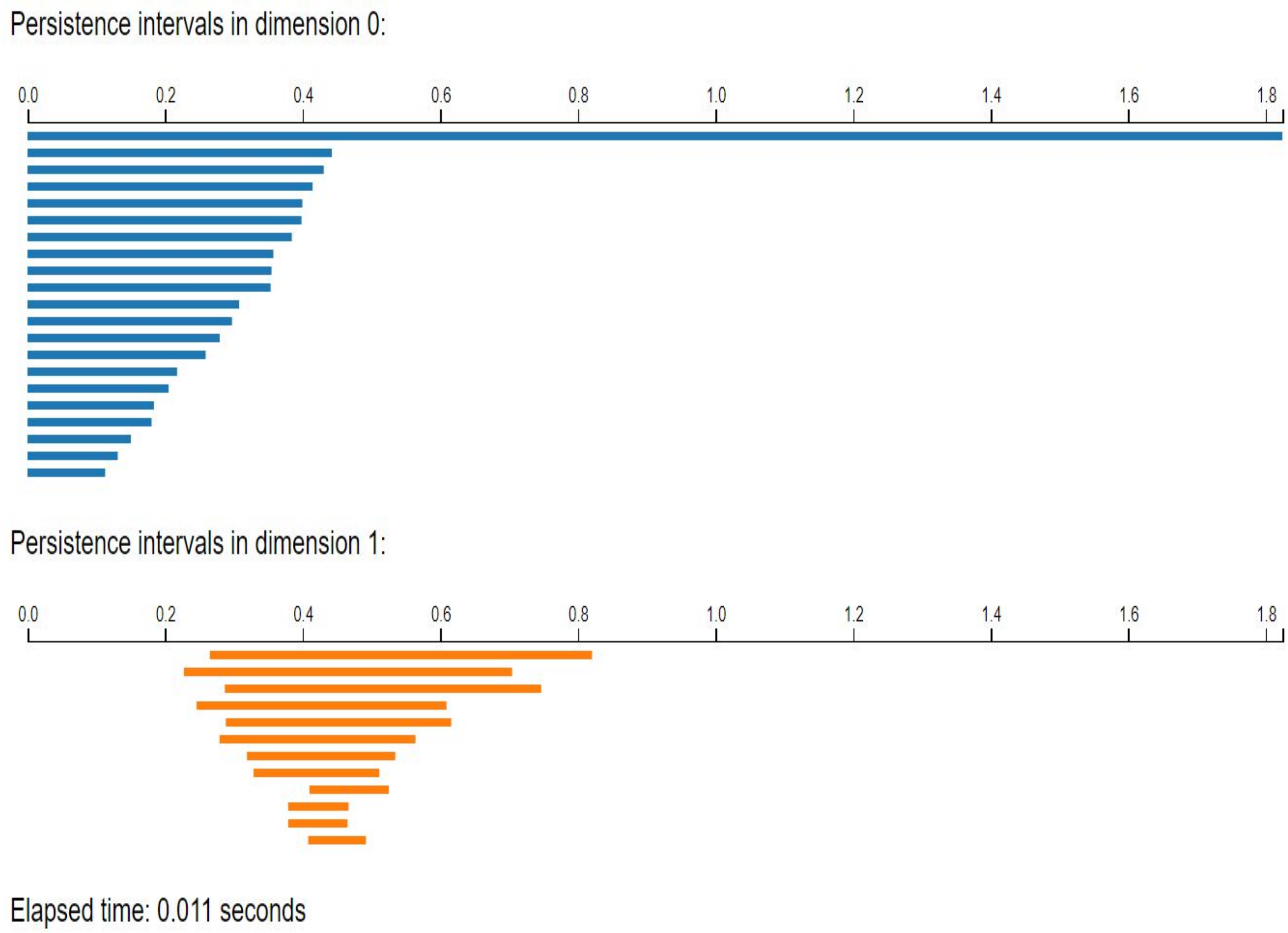
Time periods away from crisis :

2006 -Jan:


Persistence intervals in dimension 0:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6

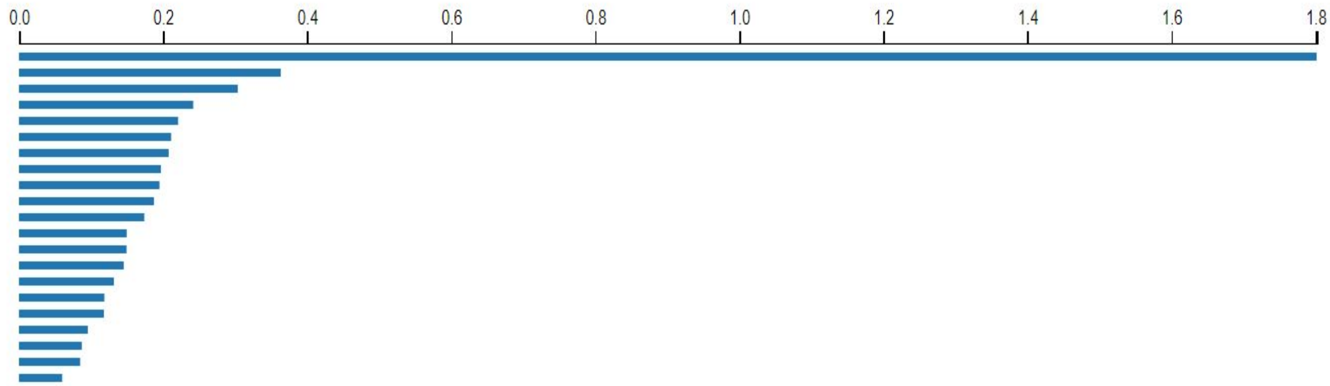

Persistence intervals in dimension 1:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6

2006 -Feb

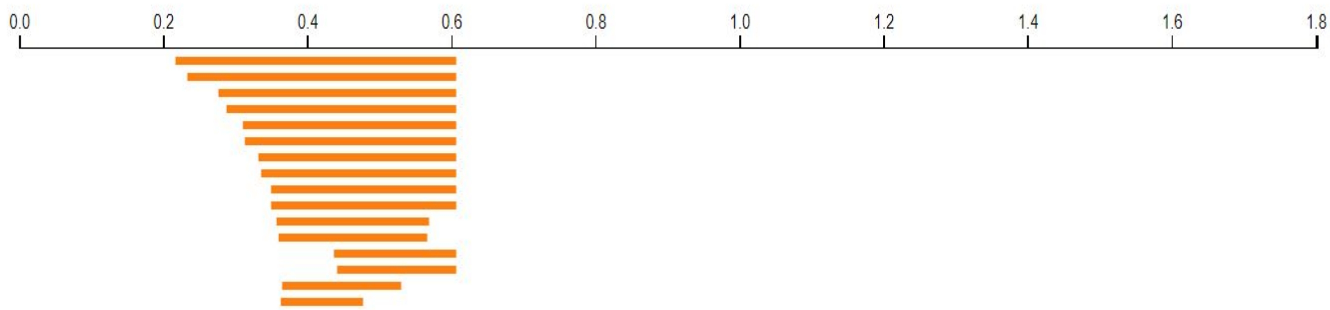

Persistence intervals in dimension 0:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Persistence intervals in dimension 1:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Elapsed time: 0.011 seconds

2006 - March

Persistence intervals in dimension 0:



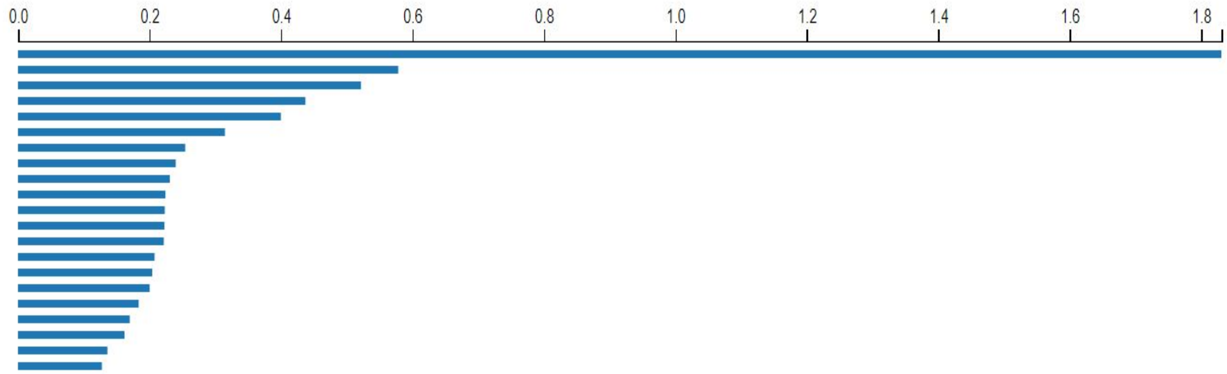
Persistence intervals in dimension 1:



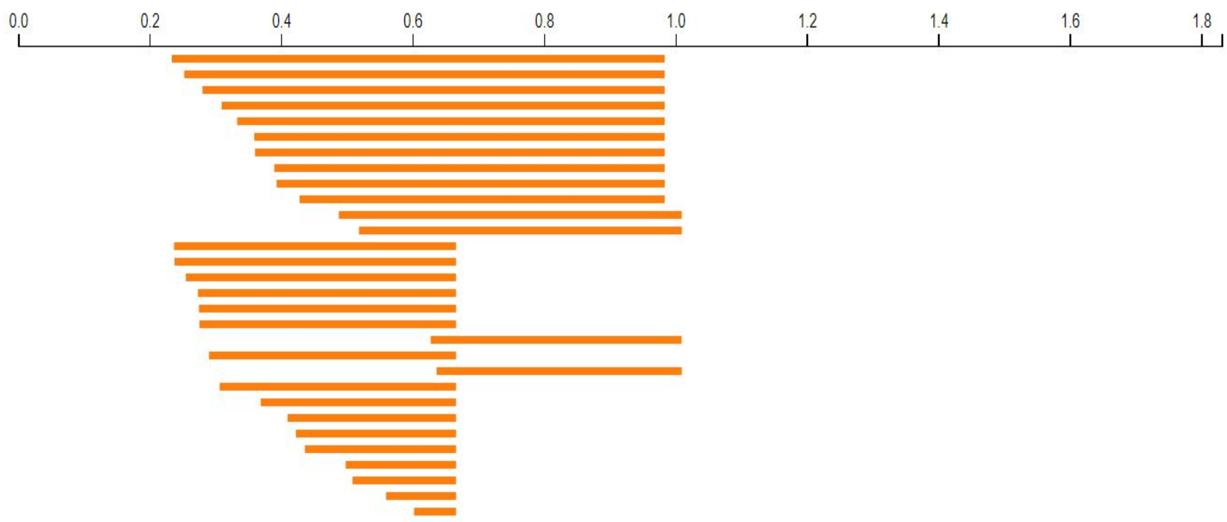
Elapsed time: 0.011 seconds

2009 JUNE :

Persistence intervals in dimension 0:

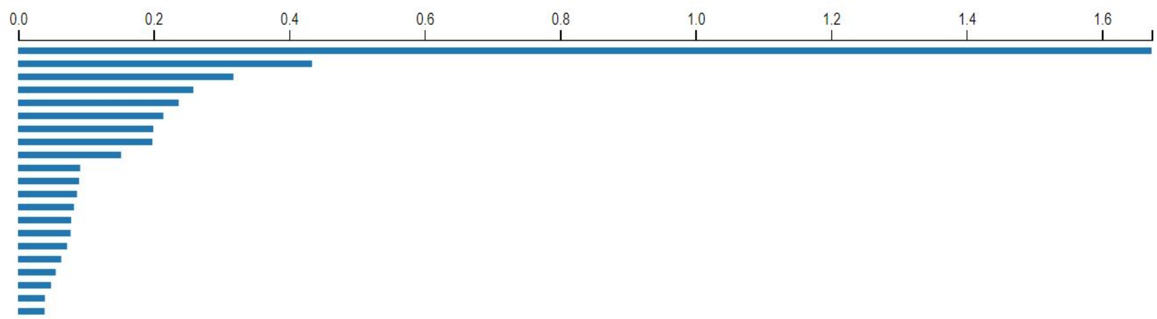


Persistence intervals in dimension 1:

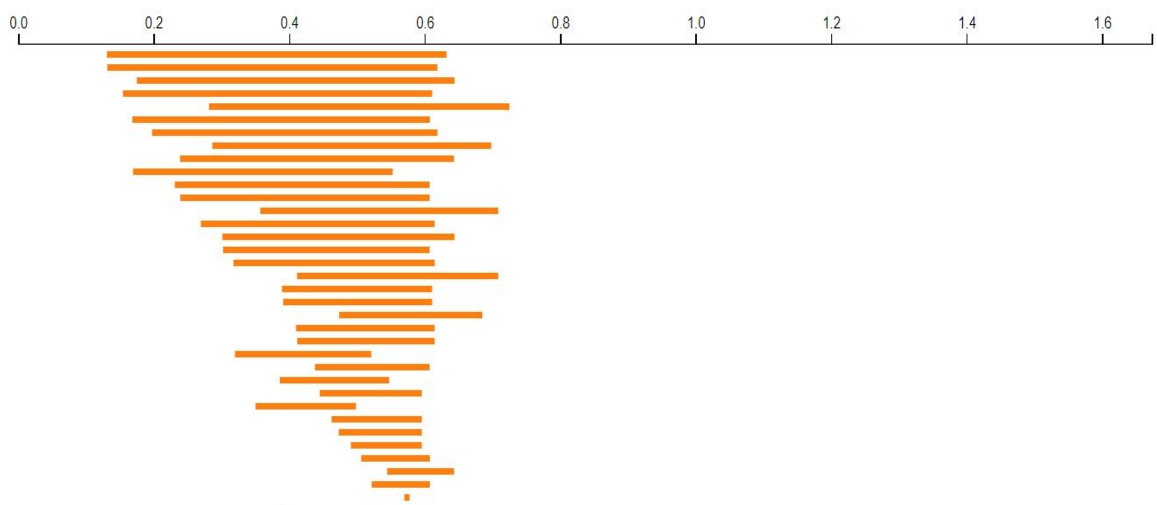


2009 SEPTEMBER :

Persistence intervals in dimension 0:



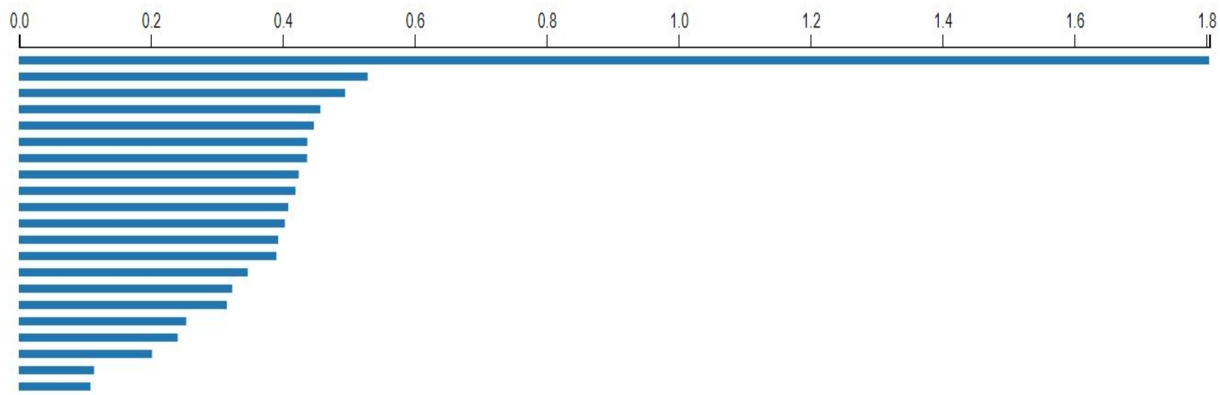
Persistence intervals in dimension 1:



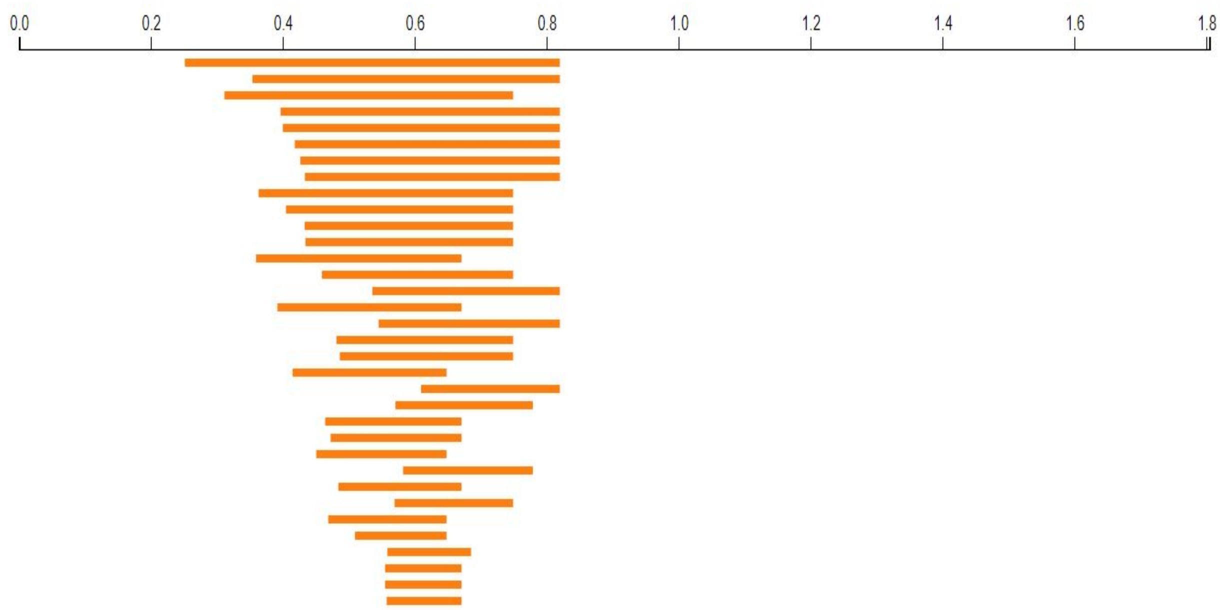
Super level sets(During the crisis & leading upto 2008) :

2007 November

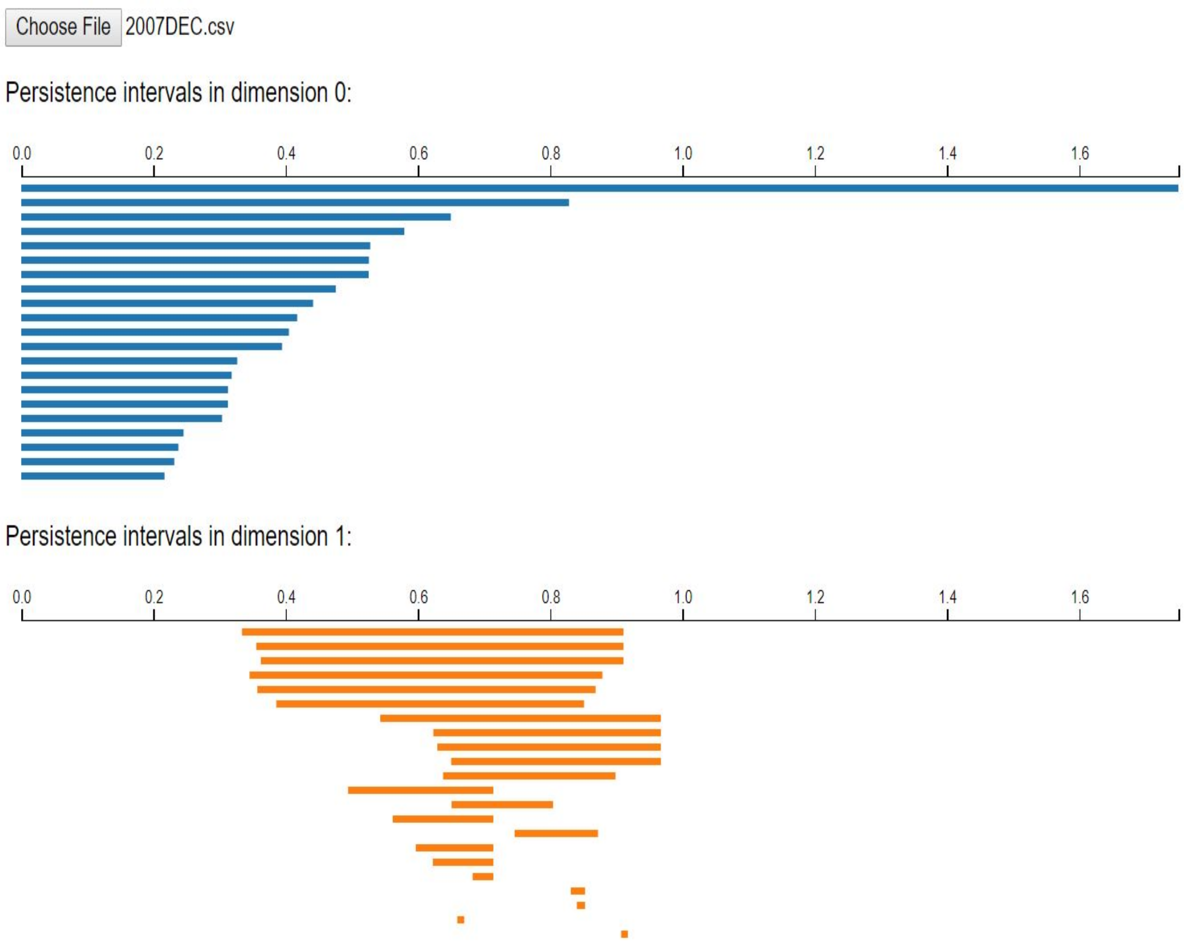
Persistence intervals in dimension 0:



Persistence intervals in dimension 1:

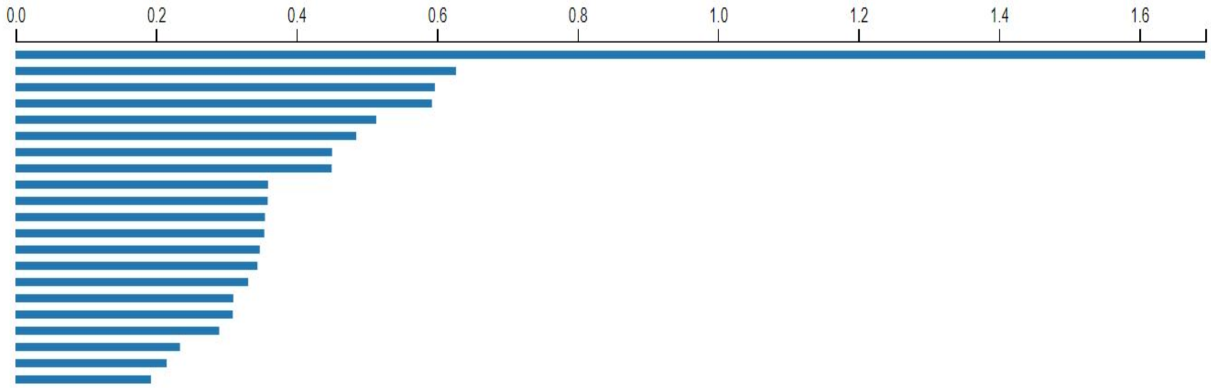


2007 December

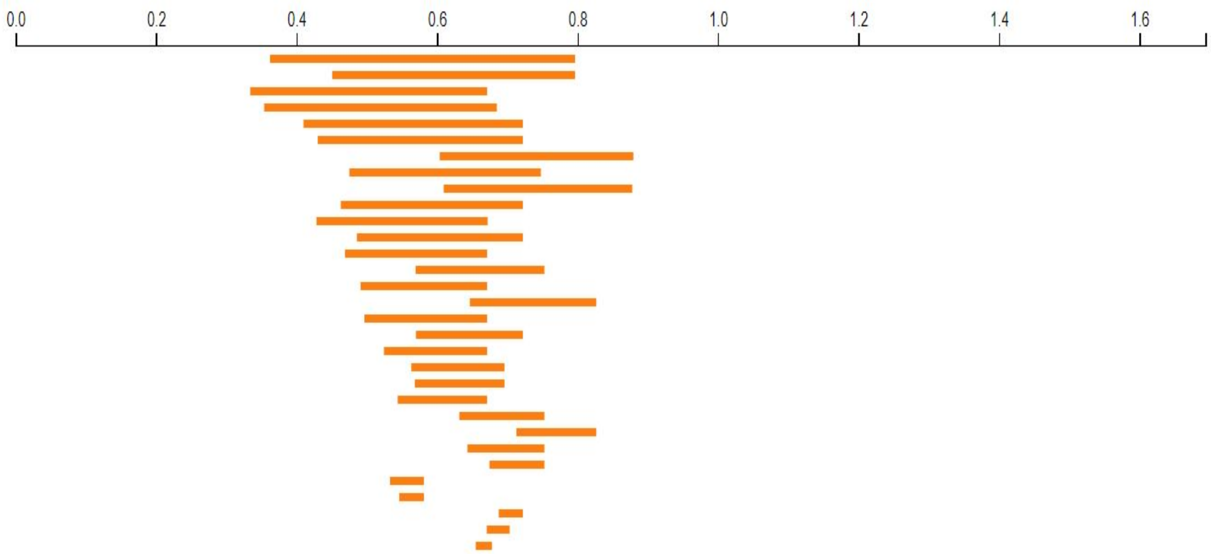


2008 Feb:

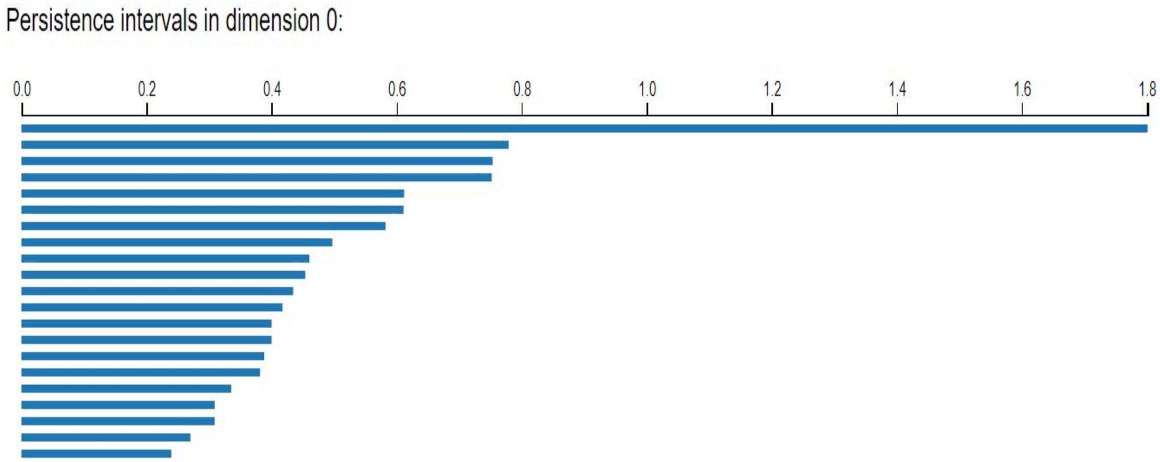
Persistence intervals in dimension 0:

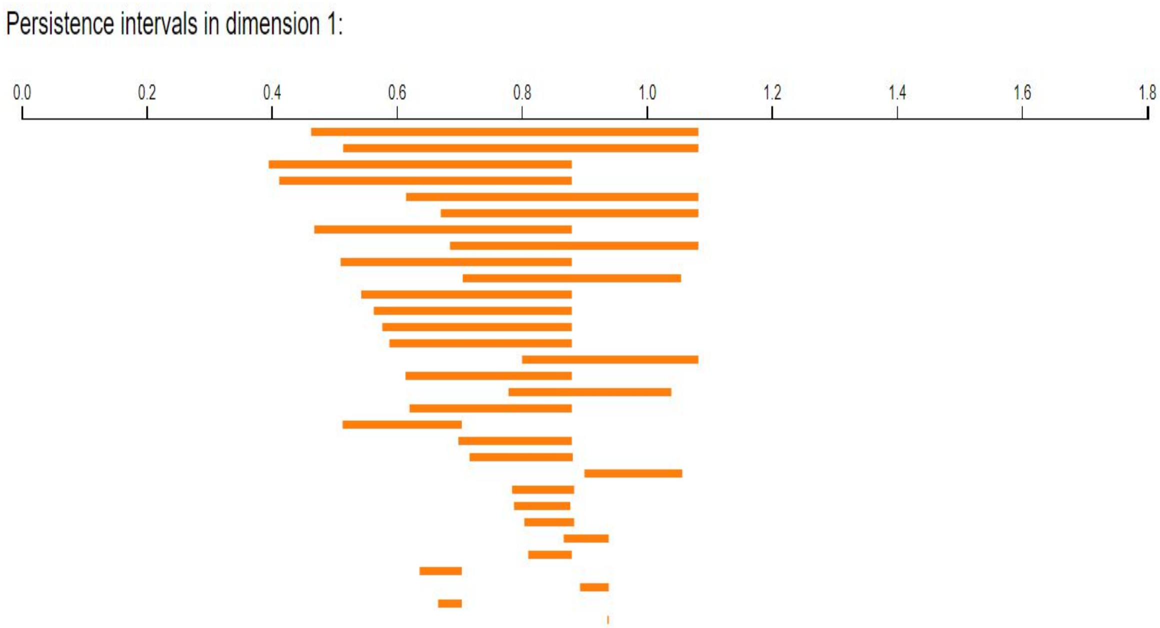


Persistence intervals in dimension 1:

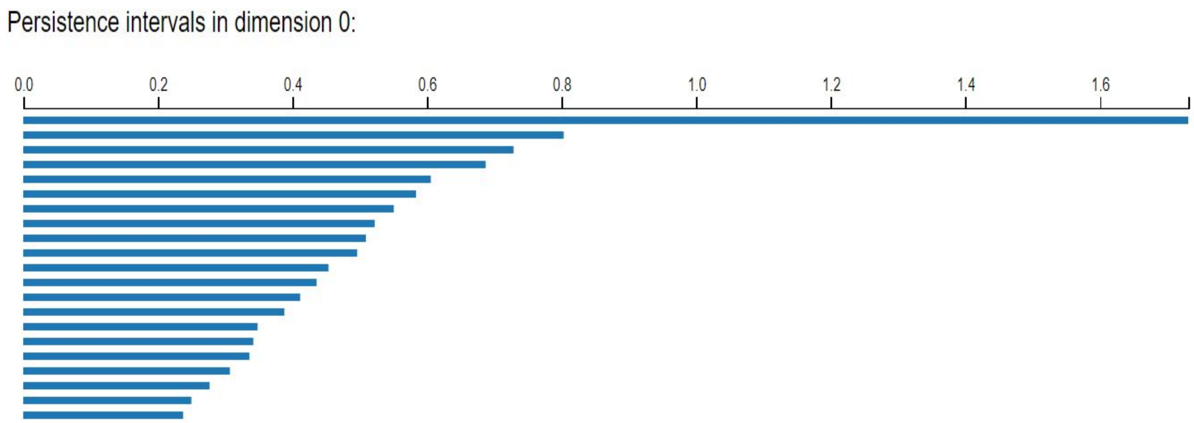


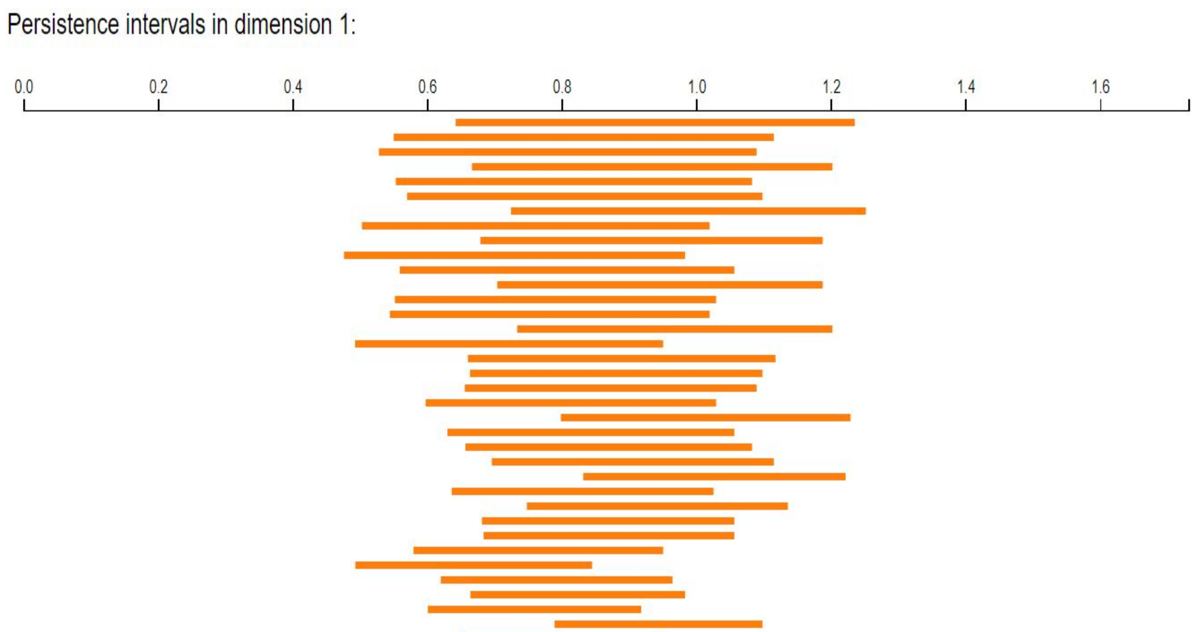
2008 March:


Persistence intervals in dimension 0:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8

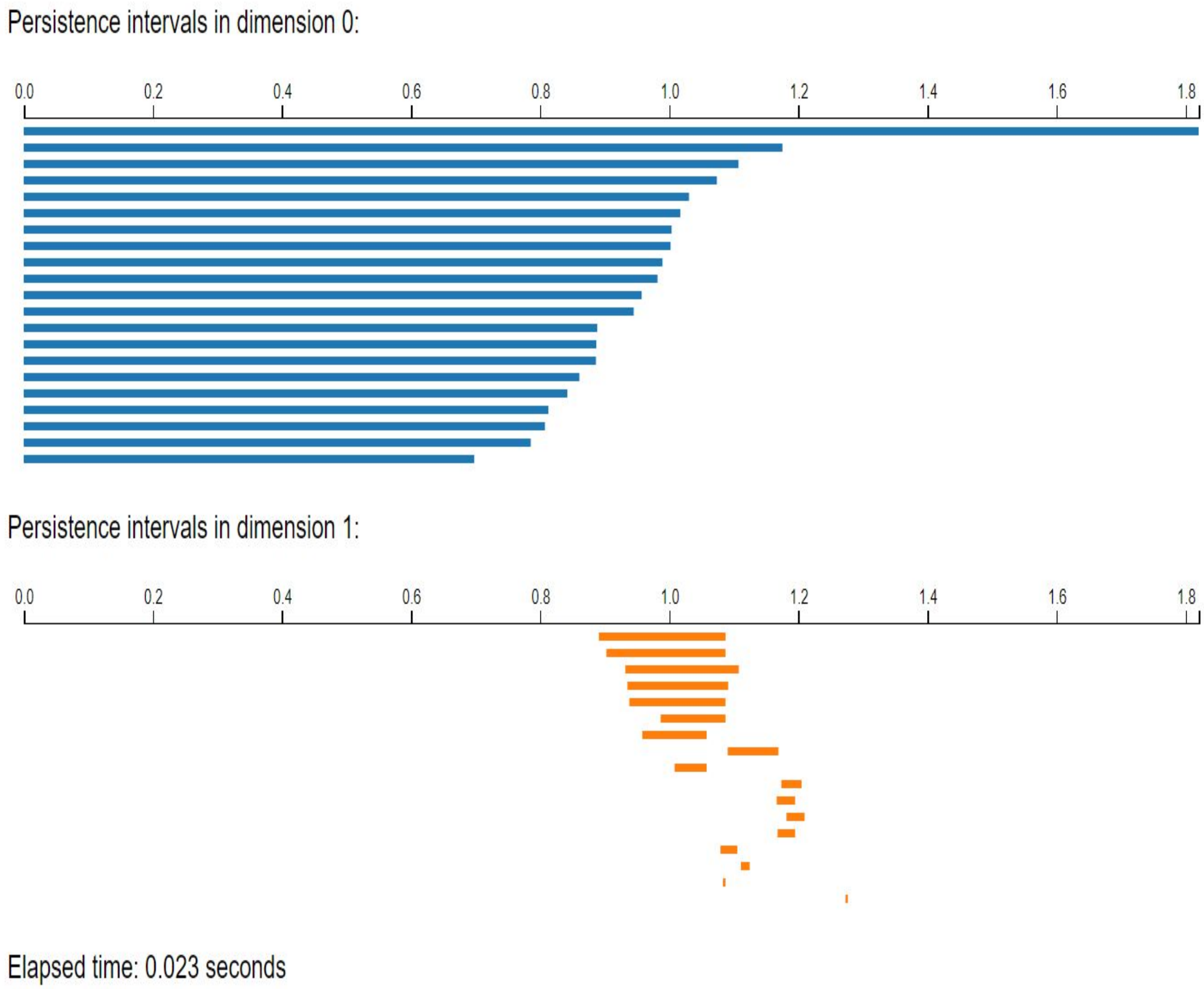

Persistence intervals in dimension 1:
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8

2008 September:

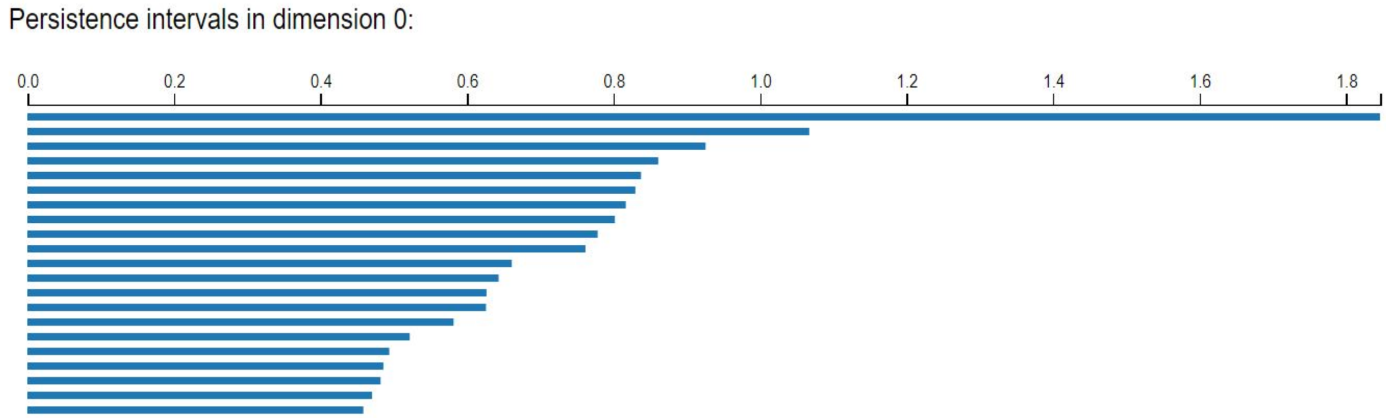


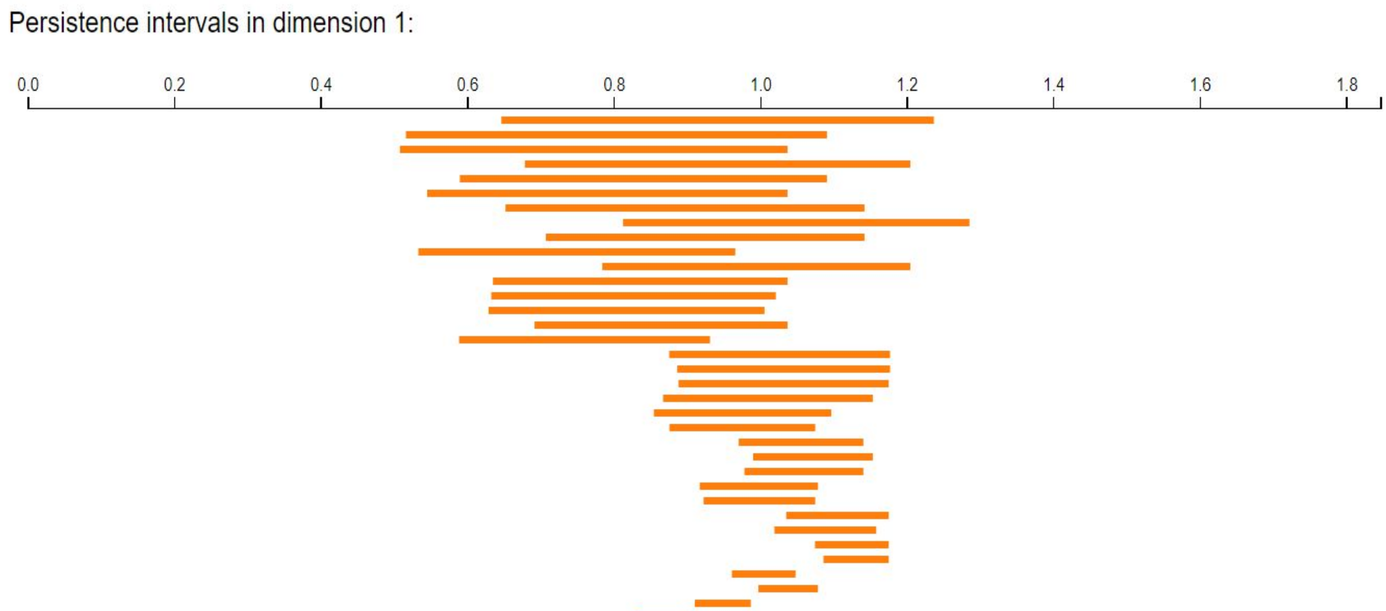


2008 October:



2008 November:





Conclusion :

The analysis persistence diagrams indicates significant changes in the structure of the correlation graph network when evaluated at a period of critical transition.

However, we have observed instances where the persistence diagrams indicate characteristics similar to that of a crisis in some samples at time periods away from the financial crash date. These could be due to some other minor events that could have occurred during the specific market window.

Analysis of consecutive time periods where the persistent diagrams indicate a change in topology could be worthy of a closer look for patterns that might be missed in usual analysis as it could potentially indicate a pattern in the markets. This method can be studied in conjunction with other statistical analysis methods to provide an accurate prediction of anomalous behaviour.