MTH4322 TOPOLOGICAL DATA ANALYSIS: PROJECT

ADAM MCGARRITY, 40331764

1 Paper Summary

In this paper, the author considers a financial system which can be described as a time-varying weighted network. They use techniques from TDA (specifically persistent homology) to analyse the topology of the network as the system approaches a critical transition. The goal is to identify early signs of a critical transition.

1.1 Method Overview

The general overview of the method is as follows:

(1) Given a time-evolving weighted network $G(V, E),\ w_t : E \to [0, \infty)$ as input.
(2) For a certain threshold, at each time t consider the threshold sub-network.
(3) Compute the homology of the Rips complex given by the sub-network.
(4) Construct the persistence diagrams and compute the time series of the distances between each of these and some reference diagram, according to some metric.

Prior to a critical transition, the time series will display significant change.

1.2 Construction of the Network

The network used as an example is the cross-correlation network $C = (c_{i,j})$ of the stock returns for companies in the Dow Jones Industrial Average (DJIA). For every distinct pair of vertices i and j, the edge $e_{i,j}$ exists, and has a weight associated to it.

- The vertices represent individual stocks.
- The weights of the edges are given by $d_{i,j} = \sqrt{2(1 - c_{i,j})}$.

1.3 Distances Between Persistence diagrams

The space of persistence diagrams is a subset of $\mathbb{R}^2$, and can be considered as a metric space with the *degree p Wasserstein distance*:

$$D_p(P_i^1, P_i^2) = \inf_{\varphi} \left[\sum_{q \in P_i^1} ||q - \varphi(q)||_\infty^p \right]^{\frac{1}{p}}, \tag{1.1}$$

where $P_i \subseteq \mathbb{R}^2$ is the i^{th} persistence diagram of the filtration and the superscripts represent two different times.

- When $p = \infty$, D_∞ measures the distance between the most significant features.
- When p is much larger than 1, D_p puts more weight on the significant features.
- When $p > 0$ and is small, D_p puts more weight on the least significant features.

1.4 Persistent Homology of Weighted Networks

To calculate persistent homology of weighted networks, the author first sets $\theta_{\max} = \max(w)$. Then, for every $\theta \in [0, \theta_{\max}]$, they consider sub-level sets of the weight function. I.e. restrict the underlying graph of the weighted network to a subgraph $G(\theta)$ with edges which have weights $w \leq \theta$, and all the necessary vertices. These graphs have the filtration property ($\theta \leq \theta' \Rightarrow G(\theta) \subseteq G(\theta')$). A similar construction is a super-level set, where only edges of weight $w \geq \theta$ are kept in the subgraph.[1]

For every $G(\theta)$, the author then constructs a Rips complex, $K_\theta = X(G(\theta))$. The filtration of threshold subgraphs yields a corresponding filtration of the Rips complexes, $\theta \mapsto K_\theta$.

In this paper, the author only considers persistent homology of dimension 0 and 1.

[1] A super-level set is equivalent to a sub-level set with the weight function $w' = \theta_{\max} - w$.

1.5 Definition of the Weighted Network

The specific data used in this paper comes from the DJIA restricted to the period January 2004 to September 2008. The value associated with each vertex is the adjusted closing price of the stock $S_i(t)$, so

$$x_i(t) = \frac{S_i(t+\Delta T) - S_i(t)}{S_i(t)}$$

where Δt is one day and the i's correspond to individual stocks.

The weight function for the edges is calculated as follows:

- Compute the Pearson correlation coefficient $c_{i,j}(t)$ over a time horizon T:
$$c_{i,j}(t) = \frac{\sum_{\tau=t-T}^{t}(x_i(\tau)-\bar{x}_i)(x_j(\tau)-\bar{x}_j)}{\sqrt{\sum_{\tau=t-T}^{t}(x_i(\tau)-\bar{x}_i)^2}\sqrt{\sum_{\tau'=t-T}^{t}(x_j(\tau')-\bar{x}_j)^2}}.$$
The time horizon used is $T = 15$. Since this is a non-stationary system, a shorter time horizon is preferred as this better describes the current structure of the system (see [3]). The means $\bar{x}_i$ and $\bar{x}_j$ are calculated using the arithmetic mean.
- Compute the distance between i and j:
$$d_{i,j}(t) = \sqrt{2\left(1 - c_{i,j}(t)\right)},$$
and set this value to be the weight: $w(e_{i,j}, t) = d_{i,j}(t)$.

The values for the distance function, $d_{i,j}$ are always in the interval $[0, 2]$.

- If $d_{i,j}(t) = 0$, i and j are perfectly correlated.
- If $d_{i,j}(t) = 2$, i and j are perfectly anti-correlated.

Edges between correlated nodes have smaller weights, and since correlation of stocks is of interest the paper focuses on edges with low values of d.

1.6 Sub-Level Sets and Super-Level Sets

By analysing the sub-level sets from various subgraphs, $G(\theta) = \left\{e(i,j) \middle| 1 - \frac{1}{2}\theta^2 \leq c_{i,j} \leq 1\right\}$, the author makes the following observations:

- For small θ, $G(\theta)$ contains only edges between highly-correlated nodes.
- As θ approaches $\sqrt{2}$ edges between low-correlated nodes are added.
- As θ increases further edges between anti-correlated nodes appear in the network.

In a similar way, for super-level sets from subgraphs $G_{w'}(\theta) = \left\{e(i,j) \middle| -1 \leq c_{i,j} \leq 1 - \frac{1}{2}(2-\theta)^2\right\}$:

- For small θ, $G_{w'}(\theta)$ contains only edges between anti-correlated nodes.
- As θ crosses $2 - \sqrt{2}$, edges between low-correlated nodes are added.
- As θ increases towards 2, edges between highly-correlated nodes are added.

Sub-level and super-level sets give complementary information, so it is important to consider both.

1.7 Analysis of Data

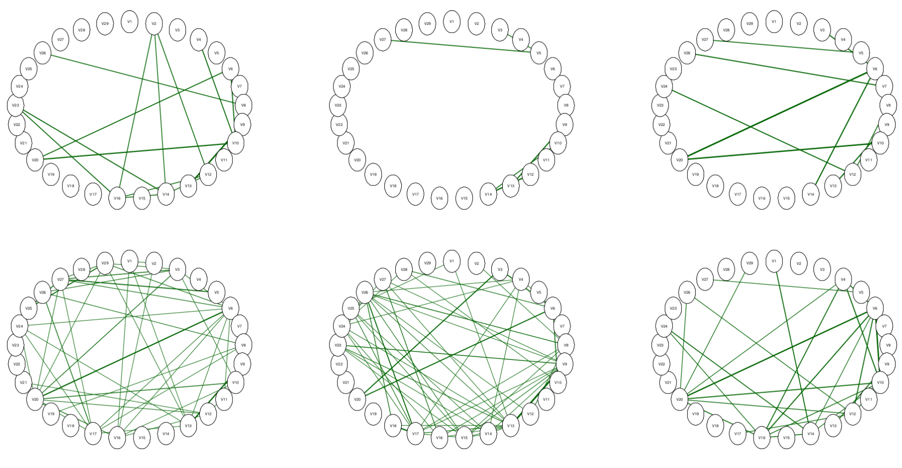

FIGURE 1.1. Threshold correlation networks. The top networks represent times far from the beginning of the 2007-08 financial crisis, while the bottom networks represent times preceding the crisis. (Taken from Fig. 2 in the original paper)

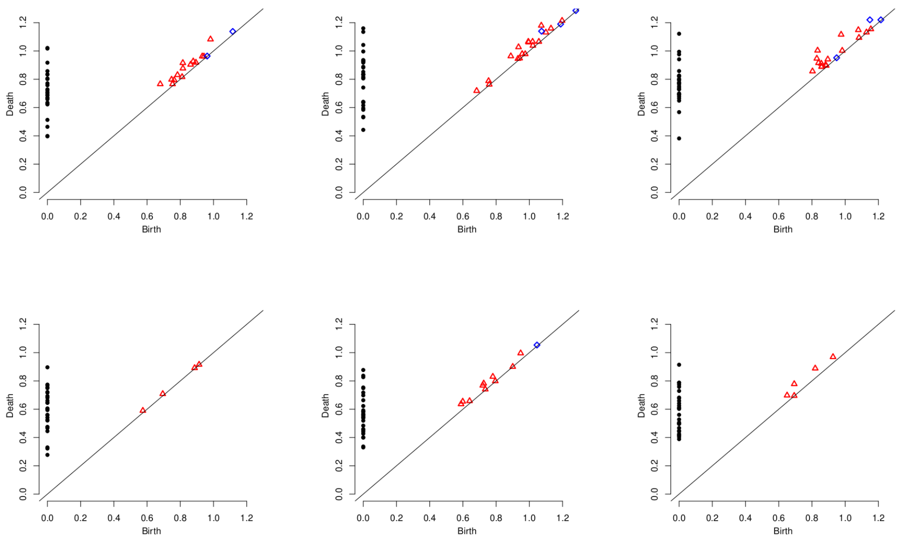

FIGURE 1.2. Persistence diagrams of sub-level sets of the weight function. The top diagrams correspond to times far from the beginning of the crisis, while the bottom correspond to times preceding the crisis. (Taken from Fig. 3 in the original paper)

The graphs in Fig. 1.1 and the persistence diagrams in Fig. 1.2 tell us that there is less correlation in the network in the period far from the beginning of the crisis compared with that directly preceding the crisis.

These observations can be quantified through the time-series representing the distances between the diagram at time t and some initial time t_0. The sampling interval is $\Delta t = 10$, and the Wasserstein distance (Eq. (1.1)) of degree $p = 2$ is used.

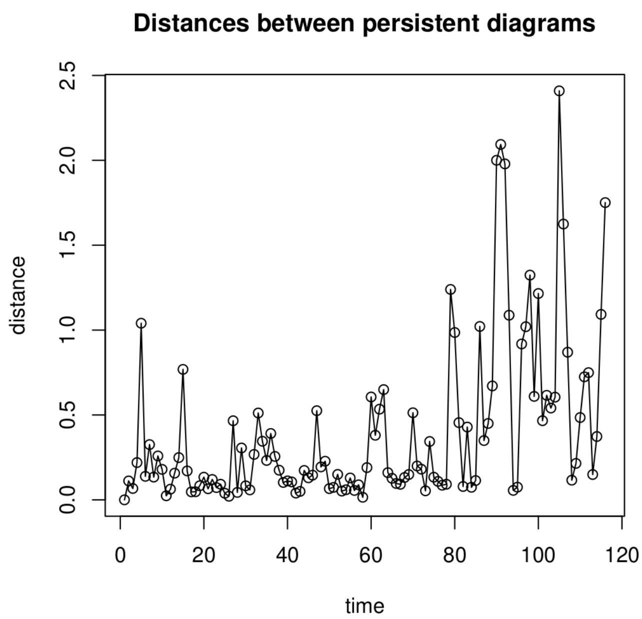

FIGURE 1.3. Distances between 0-dimensional persistence diagrams, using sub-level sets. (Taken from Fig. 4 in the original paper)

From Fig. 1.3, the author makes the following findings regarding the 0-dimensional persistent homology:

- The oscillations in the second half of the time series almost double compared with those in the first half.
- This represents a change in the topology in the network (it becomes more connected) when approaching the critical transition.

Similarly, they make the following findings about the 1-dimensional persistent homology from Fig. 1.4

- The oscillations in the second half of the time series are smaller.
- This is a change in the topology in the network - the number of loops of correlated nodes stabilises.

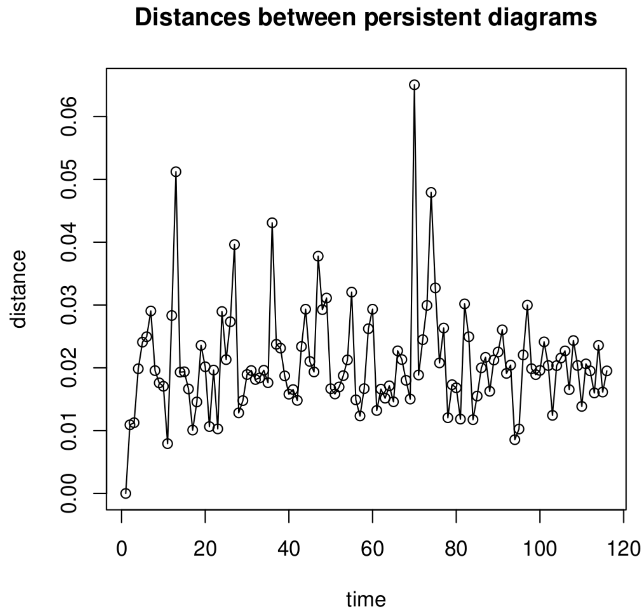

FIGURE 1.4. Distances between 1-dimensional persistence diagrams, using sub-level sets. (Taken from Fig. 4 in the original paper)

1.8 Conclusions

Based on the above analysis, the author concludes that early signs of the financial crisis became apparent around February 2007. The market changes come about due to an increase in the cross correlation between various stocks, along with the creation of sub-networks of cross-correlated stocks.

2 Analysis of Financial Data Around Brexit

Using similar techniques to those described in Sec. 1, we will analyse the cross-correlation networks around the time of the Brexit referendum in June 2016. The data we are using is from the DJIA, between the 1st of January 2015, and the 29th of December 2017. The data came from [2], and has been pre-processed to contain just the useful data needed. A copy of the pre-processed data has been uploaded to Canvas with this document. The main difference between the analysis here and that carried out in the paper is that the correlation of nodes in our network is calculated using the standard closing price, rather than the adjusted closing price, as this is what was available.

Financial data from this time period is interesting to analyse since it was an event which had a pre-determined date, but the market response was dependent on the outcome of the referendum. For example, a poll conducted on 14th of June 2016 suggested the UK was more likely to leave, causing the FTSE 100 to fall by 2% (losing £98 billion) with other markets having similar downturns [5, 1]. They recovered after further polling suggested otherwise. Furthermore, on the day of the referendum, the markets were quite volatile, with the strength of the pound falling substantially as a leave victory looked more likely [5].

2.1 Method Summary

The overall method is the same as that described in the paper. In this section we will discuss some details of the implementation in `R`.

All dates are stored as number days since 1970-01-01 (i.e. a UNIX time-stamp without the full precision), which allows us to easily perform arithmetic with them. For each day in a given range of dates, we calculate the full weighted network (over a time horizon of 15 days) as an adjacency matrix. We then use this matrix to generate a Rips complex.

Using the earliest complex as our reference, we calculate the distances as previously described using Eq. (1.1), and plot these distances against the corresponding time period.

2.2 Results

The networks in Fig. 2.1 show us that the correlation between stocks in the network increases closer to a critical transition. We can also see from the barcodes in Fig. 2.2 that both 0 and 1 dimensional features of the network generally persist for longer as we approach a critical transition.

Figures 2.3-2.10 show the distances over various time periods and for sub-/super-level sets. In each of these plots, the black line shows the distances of persistence diagrams. Some notable dates which may have had

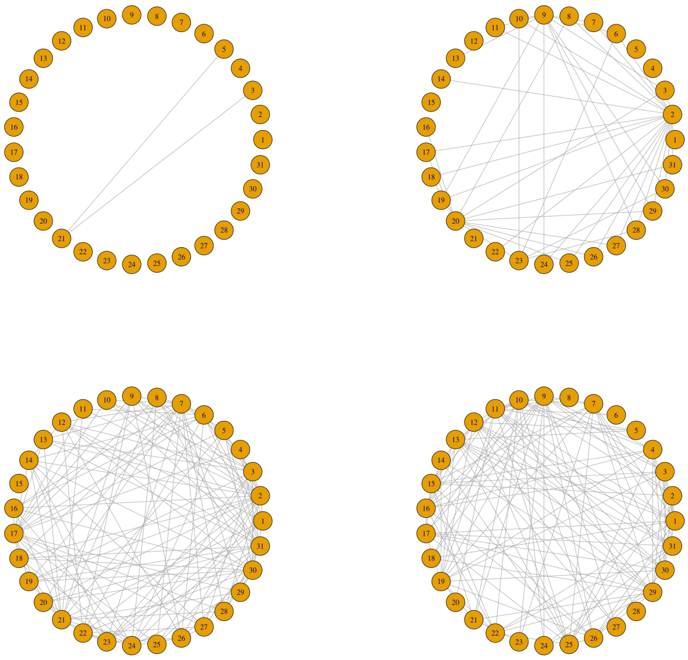

FIGURE 2.1. Some example threshold correlation networks (sub-level sets). The top networks correspond to 2015-01-02 and 2015-12-31 respectively, while the bottom networks correspond to 2016-05-23 and 2016-06-22 respectively. The threshold used for each network is $\theta = 1.5$.

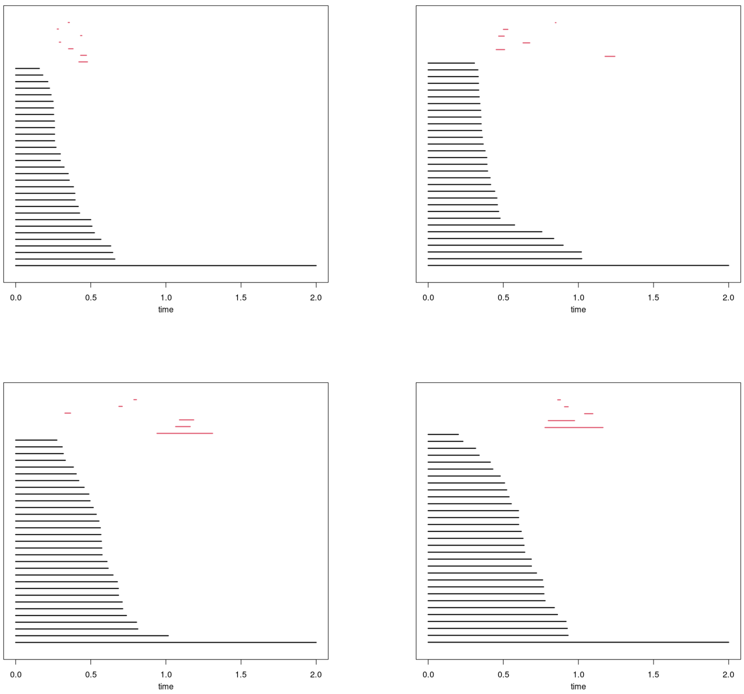

FIGURE 2.2. Persistence barcodes corresponding to the example networks shown in Fig. 2.1.

an influence on the data are also shown: the green dashed line shows the day the referendum was announced (2016-02-20); the blue dashed line shows the day the referendum campaign began (2016-04-15); the orange dashed line shows the day the FTSE 100 fell by 2% (2016-06-14, previously mentioned); and the red dashed line shows the day of the referendum (2016-06-23).

Some notable features of these plots include

- In Fig. 2.3 (Left) we can see a large spike just after the poll on the 14th of June.
- From Fig. 2.4 (Left), we can see that in the weeks leading up to the referendum there is an increase in the distances between 1-dimensional persistence diagrams. Also visible in this plot is a change around July 2015, caused by the 2015-2016 Stock Market Selloff [4].
- Fig. 2.5 (Right) tells that there were lots large of changes in the days after the referendum, which gradually got smaller throughout the second half of 2016 and into 2017.
- Figures 2.3, 2.5, and 2.6 all show a large change just after the referendum was announced.
- In Fig. 2.8 we can see that there is an increase in the distances just before the referendum, which is maintained after the referendum. There is a drop at the end of 2017, which may suggest the beginning of a return to normal stock market conditions.
- From Fig. 2.9 we see that, aside from a large peak just after the referendum, the amplitudes of the distances tend to decrease relative to a "ceiling" at ≈ 3.
- Finally, in Fig. 2.10 we can see that the oscillations tend to be larger after the referendum than before it.

Each of these observations suggests a change in the topology of the network either just before or after the referendum. Also present (in Figs. 2.4 (Right), 2.5 (Left)) are large spikes, many times larger than anything else on the plot. Whether these represent significant changes, or are just quirks in the data is unknown, and a topic of further investigation. Also in Fig. 2.4 (Left), we can see a difference before and after July 2015 in the distances, again this may be significant and is a topic of further investigation.

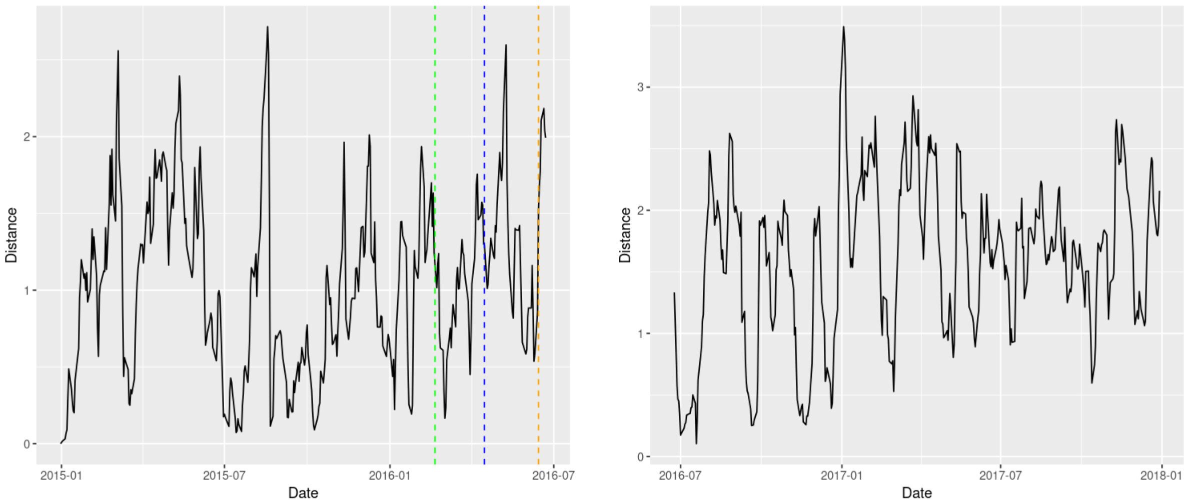

Figure 2.3. Left: The distances between 0-dimensional persistence diagrams (sub-level sets), before 23rd June 2016. Right: The same as the left but after 23rd June 2016.

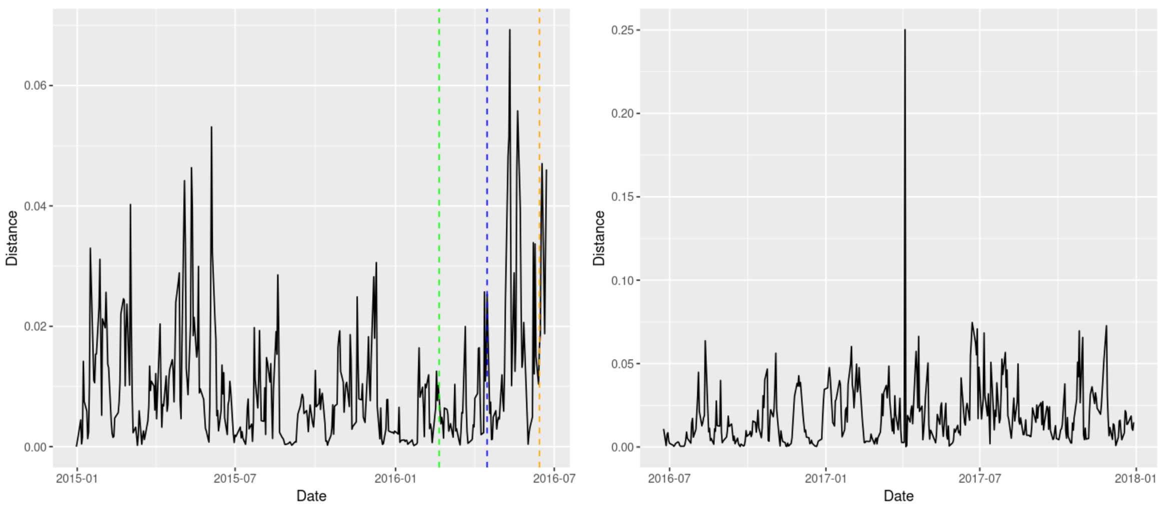

Figure 2.4. Left: The distances between 1-dimensional persistence diagrams (sub-level sets), before 23rd June 2016. Right: The same as the left but after 23rd June 2016.

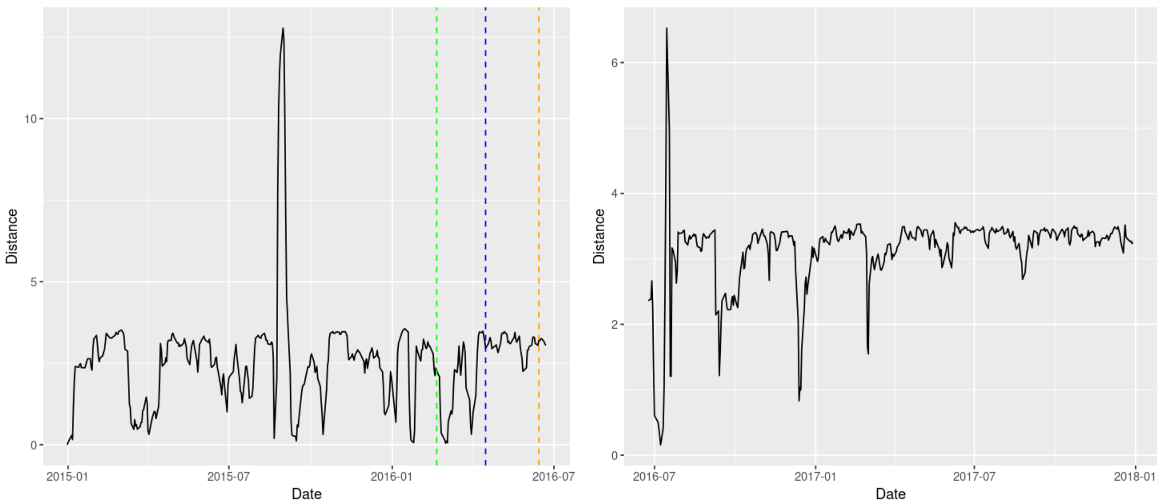

Figure 2.5. Left: The distances between 0-dimensional persistence diagrams (super-level sets), before 23rd June 2016. Right: The same as the left but after 23rd June 2016.

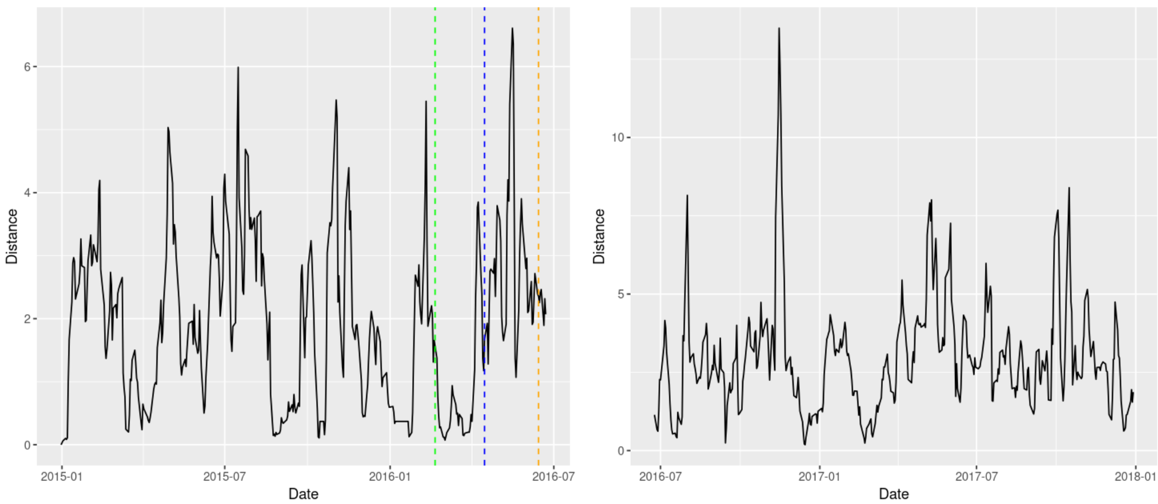

Figure 2.6. Left: The distances between 1-dimensional persistence diagrams (super-level sets), before 23rd June 2016. Right: The same as the left but after 23rd June 2016.

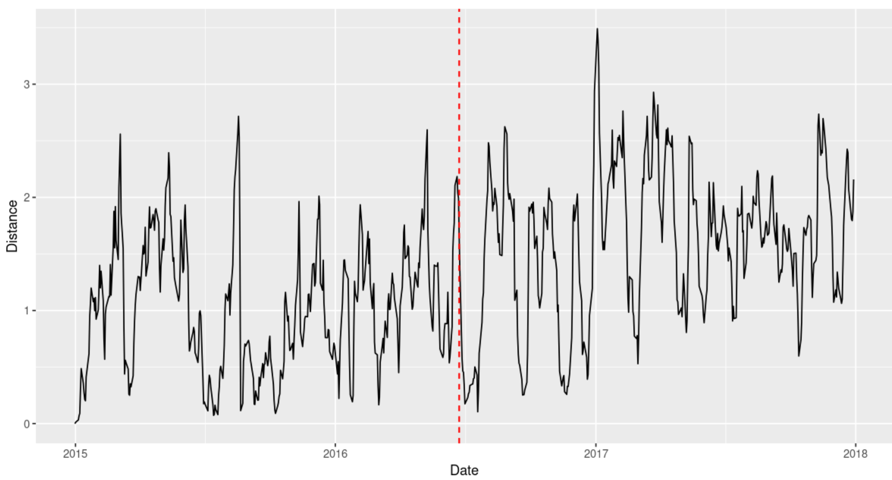

Figure 2.7. The distances between 0-dimensional persistence diagrams (sub-level sets), for the full time period.

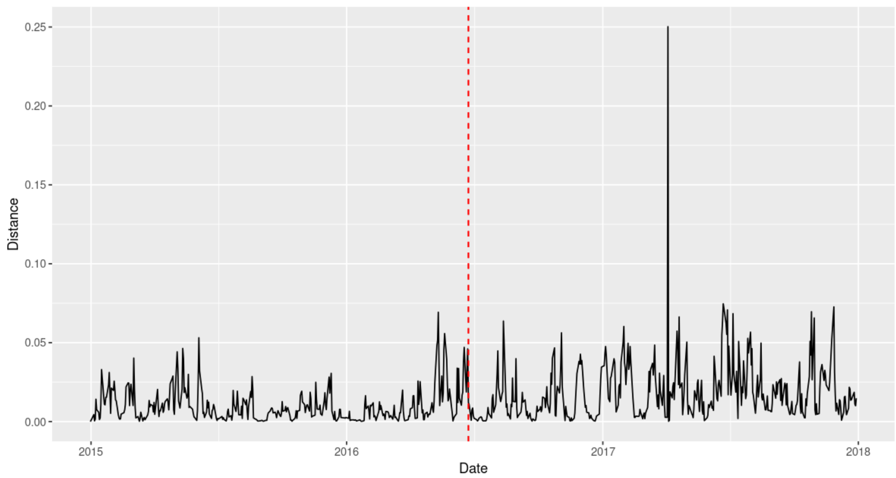

FIGURE 2.8. The distances between 1-dimensional persistence diagrams (sub-level sets), for the full time period (black line).

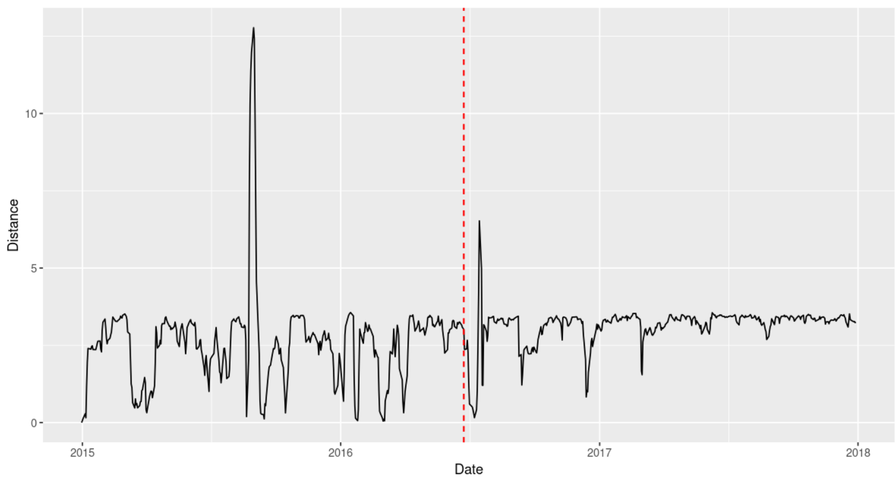

FIGURE 2.9. The distances between 0-dimensional persistence diagrams (super-level sets), for the full time period (black line).

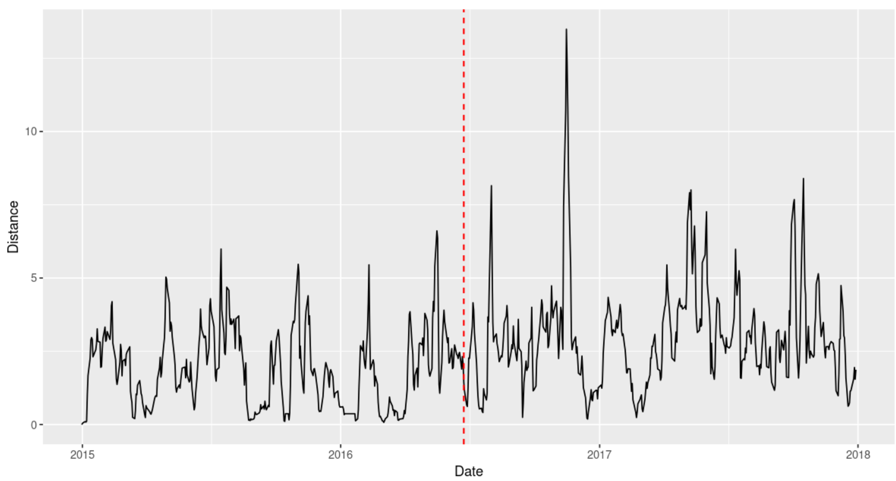

FIGURE 2.10. The distances between 1-dimensional persistence diagrams (super-level sets), for the full time period (black line).

3 Conclusion

We have seen, through persistent homology, that there were significant changes in the topological structure of the network both before and after the UK's decision to leave the EU. This analysis was carried out using a US stock market index, as it was the most readily available. It would also be interesting to look at data from a London stock market index, or another European stock market index and carry out the same analysis, as these would likely be more volatile during this time period, and may give more insightful results.

References

[1] Chris Johnston. *FTSE 100 falls 2% as investors fret over Brexit.* June 14, 2016. URL: https://www.bbc.co.uk/news/business-36526008.

[2] Kaggle.com. *DJIA 30 Stock Time Series.* 2018. URL: https://www.kaggle.com/datasets/szrlee/stock-time-series-20050101-to-20171231.

[3] M C Münnix et al. "Identifying States of a Financial Market". In: *Scientific Reports* (2012).

[4] Wikipedia contributors. *2015–2016 stock market selloff — Wikipedia, The Free Encyclopedia.* [Online; accessed 15-December-2023]. 2023. URL: https://en.wikipedia.org/w/index.php?title=2015%E2%80%932016_stock_market_selloff&oldid=1172118233.

[5] Wikipedia contributors. *2016 United Kingdom European Union membership referendum — Wikipedia, The Free Encyclopedia.* [Online; accessed 11-December-2023]. 2023. URL: https://en.wikipedia.org/w/index.php?title=2016_United_Kingdom_European_Union_membership_referendum&oldid=1188901076.