Topological Data Analysis Financial Network during Critical Transition

NOV 26 2019

Overview

A critical transition refers to an abrupt change in the behavior of a complex system arising due to small changes in the external conditions

Examples include ubiquitous, including market crashes, abrupt shifts in ocean circulation and climate,regime changes in ecosystems

We use topological study to detect early signs, that is, to identify significant changes in the structure of the time-series data emitted by the system prior to a sharp transition.

The system is described as time-varying weighted networks and we track changes in the topology of the network as the system approaches a critical transition

Persistent Homology

- 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 rst 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 $G = G(V, E)$ and a weight function associated to its edges w : $E \rightarrow [0, +\infty)$; let $\theta_{\max} = \max(w)$.
- ❖ 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.

- ❖ 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)

Method

- For each , $\theta \in [0, \theta_{\max}]$, we consider the sub-level sets of the weight function, that is, we restrict to subgraphs $G(\theta)$ which keep all edges of weights ‘w’ below or equal to threshold θ.
- 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 θ. Super-level sets can be thought of as sub-level sets of the weight function $w' = \theta_{\max} - w$.
- For each threshold graph $G(\theta)$ we construct the Rips complex (clique complex) $K = X(G(\theta))$

Method

- This is defined as the simplicial complex of with all complege subgraphs (cliques) of $G(\theta)$,as its faces
- That is, the 0-skeleton of K consists of just the vertices of $G(\theta)$ as its faces
- The 1-skeleton of all vertices and edges which is the graph $G(\theta)$ 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 $\theta \mapsto K_\theta := X(G(\theta))$; thus, $\theta \leqslant \theta' \implies K_\theta \subseteq K_{\theta'}$
- The homology groups associated with the filtration satisfy the filtration property 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 $[t-T, t]$;

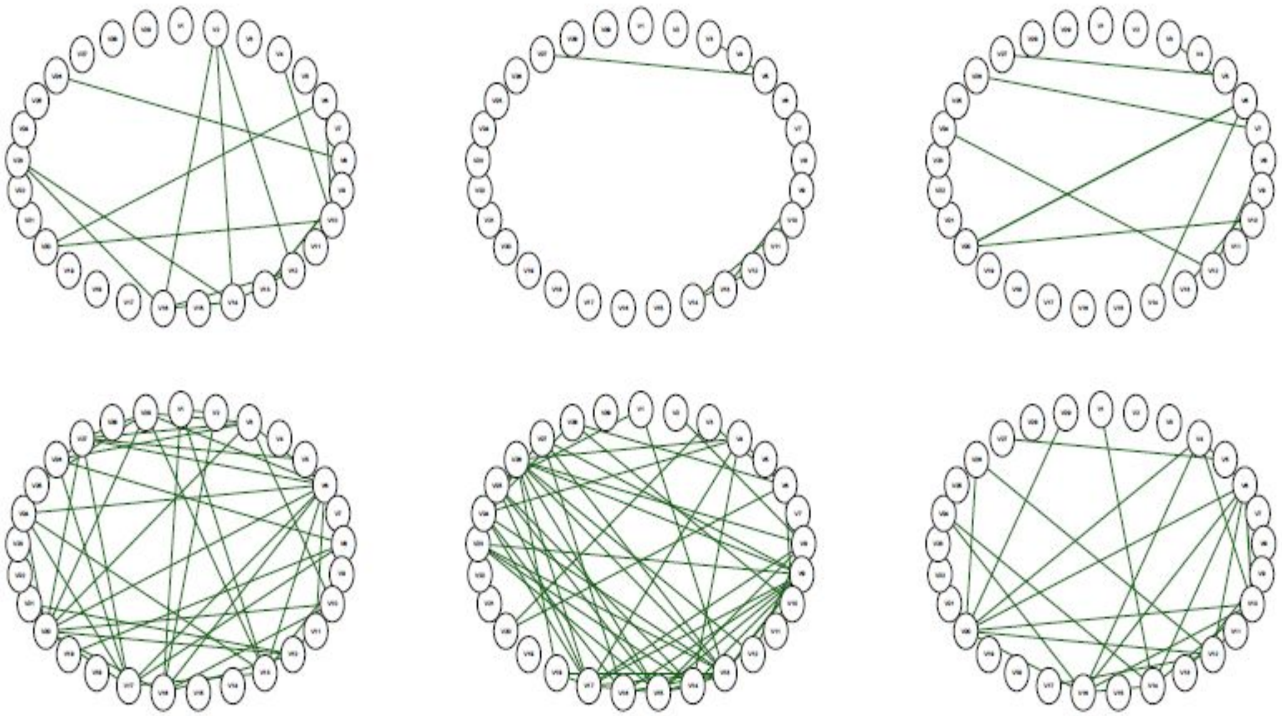
- Compute the distance between the nodes i and j $d_{i,j}(t) = \sqrt{2(1 - c_{i,j}(t))}$

- Assign the weight $w(e,t) = d_{i,j}(t)$ to the edge e between i and

Correlation Networks

- 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

Illustration : Correlation Networks



Correlation Networks at time points away from critical transition

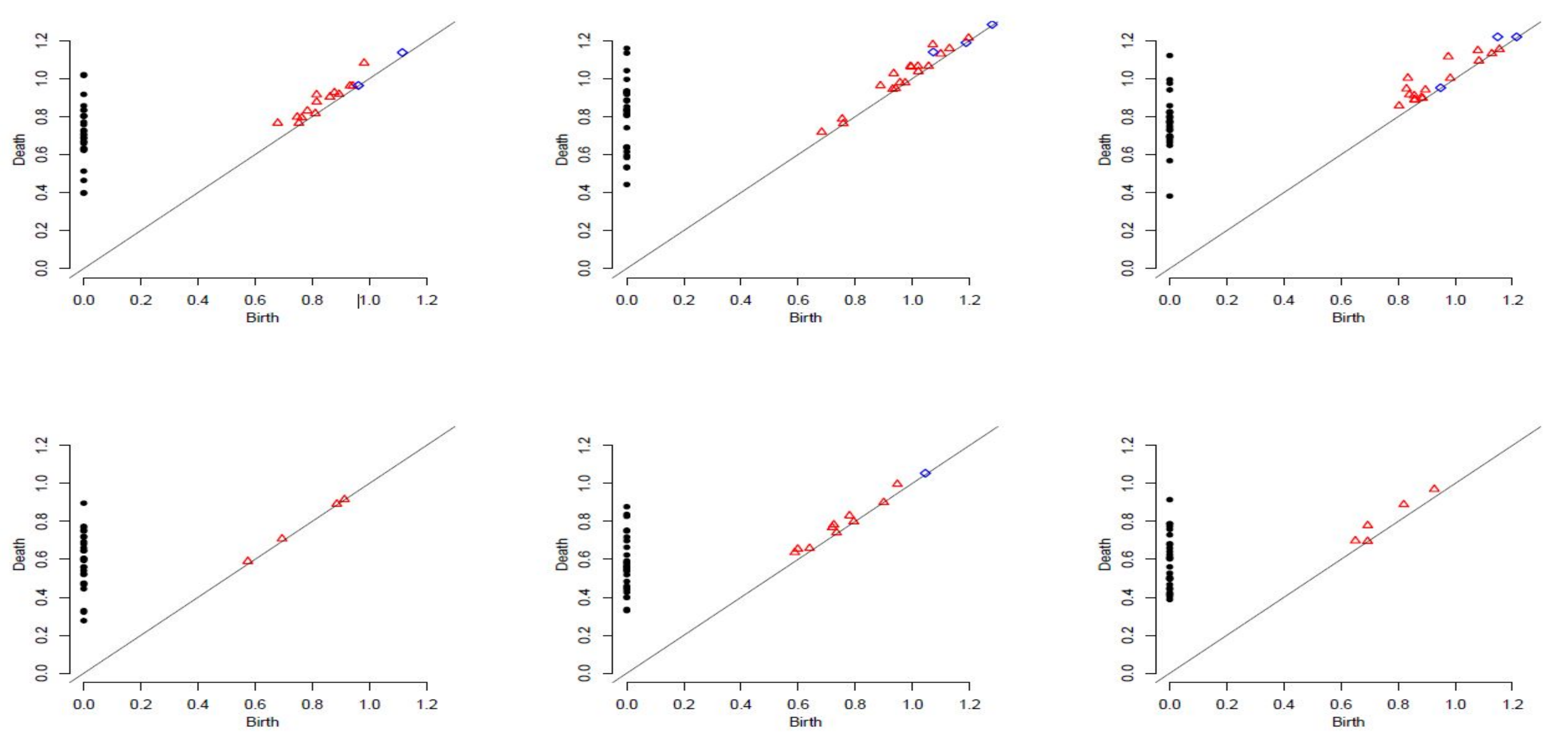
Correlation Networks near critical transition

Sublevel & Superlevel sets

Each sub-level set of the weight function w, at a threshold level $\theta \in [0, 2]$, yields a subgraph $G(\theta)$ containing only those edges for which $0 \leqslant d_{i,j} \leqslant \theta$, that is, $G(\theta) = \{e = e(i,j) \,|\, 1 - \frac{1}{2}\theta^2 \leqslant c_{i,j} \leqslant 1.\}$ When θ is small, $G(\theta)$ contains only edges between highly-correlated nodes. As θ is increased up to the critical value $\sqrt{2} = 1.414214$ edges between low-correlated nodes are progressively added to the network. As θ 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 sets for the dual weight function $w' = 2 - d$. The sub-level set $G(\theta)$ for this weight-function contains only those edges for which $d_{i,j} \geqslant 2 - \theta$, hence $G_{w'}(\theta) = \{e = e(i,j) \,|\, -1 \leqslant c_{i,j} \leqslant 1 - \frac{1}{2}(2 - \theta)^2.\}$ When θ is small, $G_{w'}(\theta)$ contains only edges between anti-correlated nodes. When θ crosses the critical value $2 - \sqrt{2} = 0.5857864$, edges between low-correlated nodes are progressively added to the network. As θ is increased further towards the highest possible value of 2, highly-correlated nodes are added to the network.

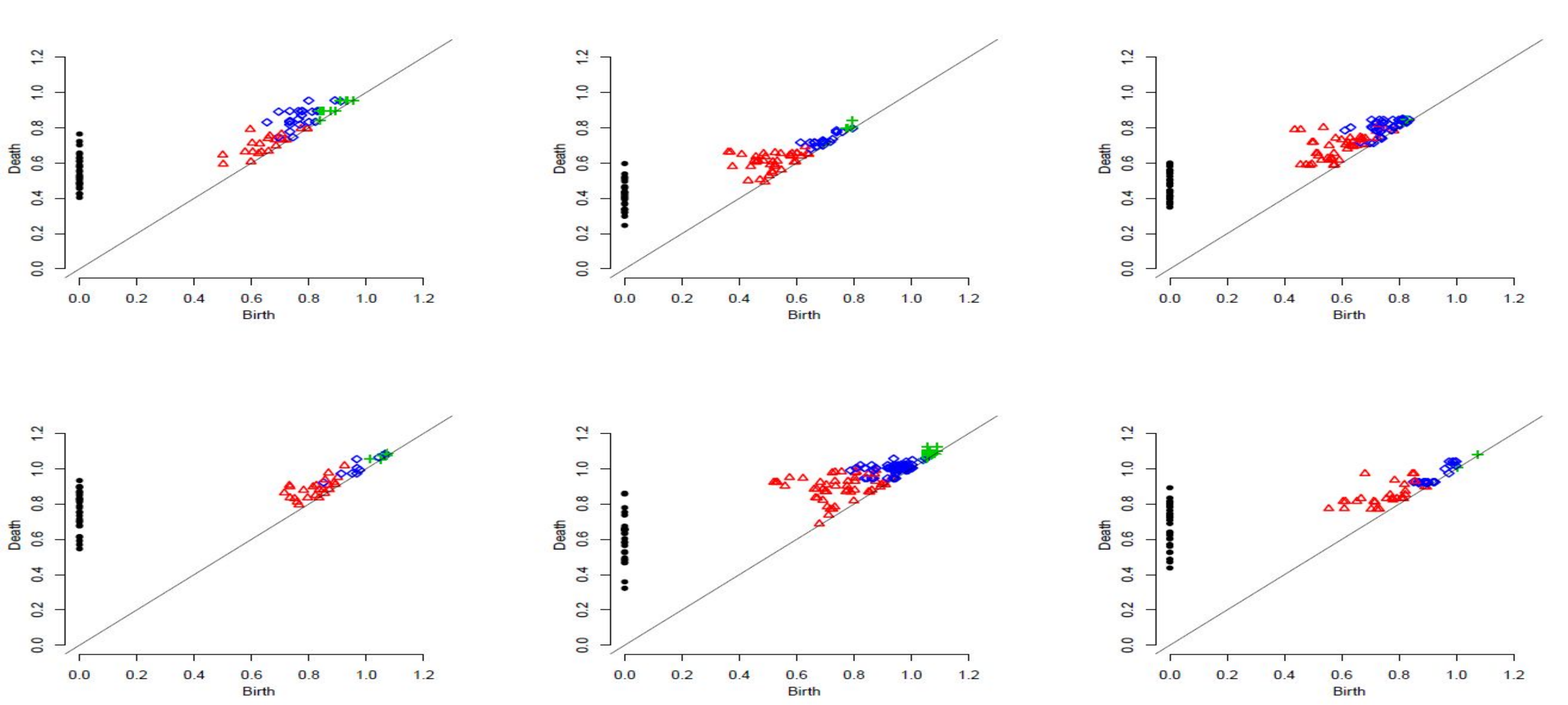
Persistent Diagrams (Sub-level sets)



Analysis

- The 0-dimensional persistent homology provides information on how the network connectivity changes as the value of θ 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 θ 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 θ for which the graph becomes fully 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 θ 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 θ that marks the passage from correlation to anti-correlation is 1.41. Inspecting the diagrams in Fig. 3 we see a concentration of dots with higher vertical coordinates in the first period, and a a concentration of dots with lower vertical coordinates in the second period. There is less correlation in the network in the first period than in the second period.

Persistent Diagrams (Super-level sets)



Analysis

We now compute the super-levels sets of w, which are sub-level set of w'. The resulting persistent diagrams have a different interpretation. The critical value of the threshold θ 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 to lack of correlation). The death of these generators is caused by the addition to correlated edges to the threshold network (in dimension 0, by joining together different connected components, and in dimension 1 by closing the loops). That is, the persistence diagrams capture the loss of normal market conditions.

End