Lecture 16: Persistent Homology in Financial Markets

Correlation networks, persistence diagrams, and early-warning signals

Jordi-Lluís Figueras

Matematiska Institutionen, Uppsala Universitet

May 13, 2026

Today: Project 16 and stock-market topology

First, I will explain Project 16.

Then we will study how persistent homology is used to analyze stock-market data.

The main case study is Marian Gidea's paper:

Primary source

M. Gidea, "Topological Data Analysis of Critical Transitions in Financial Networks," Springer Proceedings in Complexity, 47–59, 2017.

The paper is included in the suggested bibliography at the end of the lecture.

Where Lecture 15 stopped

In Lecture 15, persistent homology answered the question:

Which holes survive across scales?

For a point cloud or weighted network, we built a filtration

$$K_{\theta_1} \subseteq K_{\theta_2} \subseteq K_{\theta_3} \subseteq \cdots$$

and recorded births and deaths of homology classes.

Today the data are not sensor positions, but stock-market time series.

The central translation is:

From finance to topology

Stock returns give correlations; correlations give a weighted network; the weighted network gives a filtration.

The financial question

Financial markets are complex systems with many interacting components.

A crash is not only a large movement in one stock or one index.

It is often accompanied by a change in the dependence structure between many assets.

Gidea's question is:

Question

Can persistent homology detect structural changes in the stock-correlation network before a critical transition?

This is not a claim of deterministic crash prediction. It is a way to quantify changes in global market structure.

From prices to returns

Let $S_i(t)$ be the adjusted closing price of stock i at day t.

Gidea uses daily arithmetic returns

$$x_i(t) = \frac{S_i(t + \Delta t) - S_i(t)}{S_i(t)}, \qquad \Delta t = 1 \text{ day}.$$

The raw price level is less useful than the return series because we want to compare co-movement, not absolute price scale.

The data set in the paper is based on DJIA stocks during 2004–2008.

Remarks

- Adjusted closing price means the closing price corrected for dividends, stock splits, and similar corporate actions, so historical returns are comparable.
- DJIA means the Dow Jones Industrial Average, a U.S. index of 30 large companies. Here each DJIA stock is one node in the financial network.

Rolling correlations

At each day t, compute the Pearson correlation between stocks i and j over a recent time window of length T:

$$c_{ij}(t) = \mathrm{corr}\big(x_i, x_j\big) \text{ over } [t - T, t].$$

The paper uses a short window:

$$T = 15 \text{ trading days.}$$

The result is a time-dependent correlation matrix

$$C(t) = \big(c_{ij}(t)\big).$$

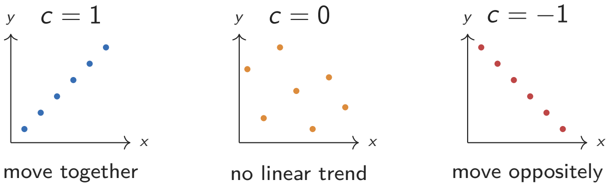
Each entry tells us how similarly two stocks moved during the recent past.

What correlation measures

For two return series x_s and y_s in one time window, Pearson correlation is

$$c = \frac{\sum_s (x_s - \bar{x})(y_s - \bar{y})}{\sqrt{\sum_s (x_s - \bar{x})^2}\sqrt{\sum_s (y_s - \bar{y})^2}}.$$

It satisfies $-1 \leq c \leq 1$ and measures linear co-movement.



Correlation is not causation: it only quantifies how two return series move together in the chosen window.

Correlation distance

Gidea converts correlations into distances by

$$d_{ij}(t) = \sqrt{2(1 - c_{ij}(t))}.$$

This gives

$$0 \leq d_{ij}(t) \leq 2.$$

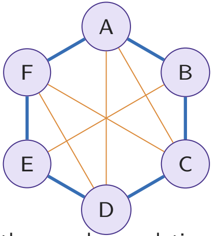
Correlation	Distance	Interpretation
$c_{ij} = 1$	$d_{ij} = 0$	perfectly correlated
$c_{ij} = 0$	$d_{ij} = \sqrt{2}$	uncorrelated
$c_{ij} = -1$	$d_{ij} = 2$	perfectly anti-correlated

Small distance means strong positive co-movement.

For each day t, form a complete weighted graph:

$$G(t) = (V, E, w_t).$$

- vertices V: stocks;
- edges E: all stock pairs;
- edge weight: $w_t(i,j) = d_{ij}(t)$.



edge lengths encode correlation distances

What changes with time?

The vertex set is almost fixed over the period studied.

The edge weights change every day because the rolling correlations change:

$$d_{ij}(t_1) \neq d_{ij}(t_2).$$

Therefore the market produces a time series of weighted networks:

$$G(t_0), G(t_1), G(t_2), \ldots$$

Persistent homology will be used twice:

1. within each day, vary a threshold to get one persistence diagram;
2. across days, compare those diagrams to track structural change.

Threshold Networks

Turn correlations into filtrations

Sublevel threshold graph

Fix a day t and a threshold $\theta \in [0, 2]$.

The sublevel threshold graph keeps edges with distance at most θ:

$$E_\theta(t) = \{\{i, j\} : d_{ij}(t) \leq \theta\}.$$

Small θ keeps only strongly correlated stock pairs.

As θ increases, more edges are added:

$$\theta_1 \leq \theta_2 \quad \Longrightarrow \quad G_{\theta_1}(t) \subseteq G_{\theta_2}(t).$$

This is the graph-level filtration.

From threshold graph to clique complex

For each threshold graph $G_\theta(t)$, build the clique complex

$$K_\theta(t) = \mathsf{Cl}\big(G_\theta(t)\big).$$

This means:

- vertices are stocks;
- edges are strongly enough correlated pairs;
- triangles are triples whose three pairwise distances are all below θ;
- higher cliques are filled as higher-dimensional simplices.

Because threshold graphs are nested, the clique complexes are nested:

$$K_{\theta_1}(t) \subseteq K_{\theta_2}(t) \quad \text{when} \quad \theta_1 \leq \theta_2.$$

One day gives one persistence diagram

At a fixed day t, we compute persistent homology of

$$K_\theta(t), \qquad 0 \leq \theta \leq 2.$$

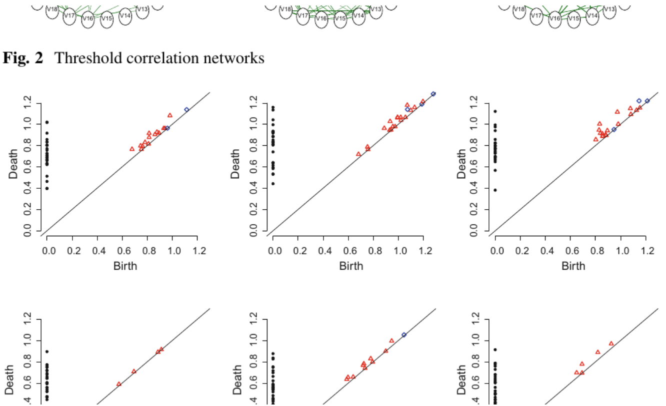
This produces diagrams

$$\mathrm{PD}_0(t), \qquad \mathrm{PD}_1(t).$$

- $\mathrm{PD}_0(t)$ records when connected components merge.
- $\mathrm{PD}_1(t)$ records when loops appear and become filled.

So one market day becomes a topological summary of the correlation network at all thresholds.

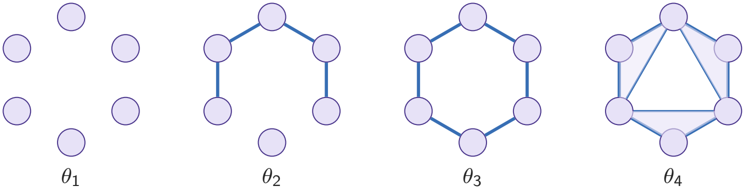
Persistent diagrams in Gidea's paper



The figure shows persistence diagrams for sublevel threshold networks. Black dots represent H_0 information; red triangles represent H_1 information.

Source: cropped from Fig. 3 in M. Gidea, "Topological Data Analysis of Critical Transitions in Financial Networks," Springer Proceedings in Complexity, 47–59, 2017.

Sublevel filtration picture



isolated stocks $\rightarrow$ correlated groups $\rightarrow$ loop $\rightarrow$ loop filled by cliques

Why not choose one threshold?

Choosing one threshold forces a fragile decision.

At one value of θ, a loop may be present.

At a slightly different value, it may disappear because a few extra edges create filled triangles.

Persistent homology avoids this arbitrary choice:

Persistence viewpoint

Do not ask whether a feature exists at one threshold. Ask how long it survives across thresholds.

This is especially important in financial data because correlations are noisy and time-dependent.

Interpreting H_0

A point in $\mathrm{PD}_0(t)$ represents a connected component of the threshold network.

In the sublevel filtration:

- components are born at $\theta = 0$;
- a component dies when it merges into an older component;
- a low death value means merger through strong positive correlations;
- a high death value means merger only after weaker or negative-correlation edges are allowed.

Thus H_0 summarizes how the market separates into correlated clusters and how quickly those clusters merge.

Interpreting H_1

A point in $\mathrm{PD}_1(t)$ represents a loop in the stock-correlation network.

In a clique complex:

- a triangle of three mutually correlated stocks is filled and does not create H_1;
- an H_1 loop needs at least four stocks arranged cyclically;
- the loop dies when added edges create filled triangles that cover it.

Financial interpretation:

Loop interpretation

H_1 records cyclic dependence patterns that are not explained by a single tightly connected clique.

Toy example: four stocks

Suppose the correlation distances on one day are:

d_{AB}	d_{BC}	d_{CD}	d_{DA}	d_{AC}	d_{BD}
0.35	0.36	0.34	0.37	0.55	1.10

At threshold $\theta = 0.40$, the graph contains the square edges but no diagonal:

$$\beta_1 = 1.$$

At threshold $\theta = 0.60$, the diagonal AC appears. The clique complex contains triangles ABC and ACD, so the square loop is filled:

$$\beta_1 = 0.$$

The loop is born near 0.37 and dies near 0.55.

Toy example diagram

This is the same algebra as in Lectures 14–15, but the edges now come from stock correlations.

Tracking Market Change

Compare persistence diagrams over time

A diagram-valued time series

For each market day t, persistent homology gives diagrams

$$\mathrm{PD}_0(t), \qquad \mathrm{PD}_1(t).$$

Thus the market produces a time series of persistence diagrams:

$$\mathrm{PD}_k(t_0), \mathrm{PD}_k(t_1), \mathrm{PD}_k(t_2), \ldots$$

To turn this into a numerical signal, choose a reference day t_0 and compute distances

$$D_p(\mathrm{PD}_k(t), \mathrm{PD}_k(t_0)).$$

Gidea uses Wasserstein distances between persistence diagrams.

Distance between persistence diagrams

The Wasserstein distance matches points in one persistence diagram with points in another.

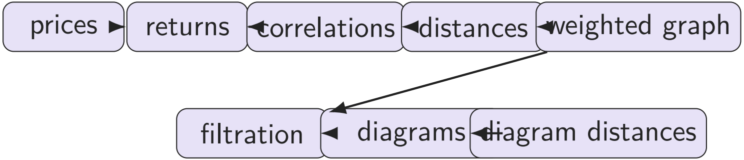
Points may also be matched to the diagonal, which means treating a short-lived feature as noise.

For degree p, the distance has the form

$$D_p(P, Q) = \inf_{\varphi : P \to Q} \left(\sum_{x \in P} \|x - \varphi(x)\|_\infty^p \right)^{1/p}.$$

Gidea uses $p = 2$ in the reported experiment.

Large distance means that the topology of the correlation network has moved far from the reference topology.



The final output is an ordinary time series measuring topological change.

What the H_0 signal measures

The H_0 persistence diagram records how correlated groups of stocks merge as θ increases.

If correlations become stronger across the market, components merge at lower distance thresholds.

In the paper, the time series

$$D_2\big(\mathrm{PD}_0(t), \mathrm{PD}_0(t_0)\big)$$

shows a substantial change in behavior before the peak of the 2007–2008 crisis.

Interpretation:

Connectivity warning

The topology of market connectivity changes as correlations reorganize before the critical transition.

What the H_1 signal measures

The H_1 persistence diagram records loops in the correlation network.

These loops represent cyclic structures that are not immediately filled by highly correlated triples.

Gidea observes that the H_1 diagram-distance time series also changes before the crisis, but its interpretation is more subtle.

The paper describes this as a change in the topology of cliques and loops: the structure of correlated sub-networks stabilizes differently as the transition approaches.

Sublevel sets and the crisis interpretation

For sublevel sets, low thresholds keep strongly positively correlated edges.

Before a financial crisis, correlations between many stocks often increase.

Topologically this can appear as:

- connected components merging at lower thresholds;
- changes in the number and persistence of loops;
- changes in distances between persistence diagrams over time.

Gidea reports significant topological changes starting around February 2007, before the U.S. market peak in October 2007.

Superlevel sets

The paper also studies superlevel sets of the original distance function.

Equivalently, use the dual weight

$$d'_{ij}(t) = 2 - d_{ij}(t)$$

and take sublevel sets of d'.

Small thresholds for d' keep pairs that are weakly correlated or anti-correlated in the original distance.

This gives complementary information:

Complementary view

Sublevel sets emphasize strong co-movement; superlevel sets emphasize the loss of normal non-correlated structure.

What changes near a crash?

The paper's conclusion can be summarized as follows.

- The correlation network changes topology before the crisis peak.
- The changes are visible in distances between persistence diagrams.
- The changes are consistent with increasing cross-correlations and emerging correlated sub-networks.

Main point for this course

Persistent homology turns a time-dependent network of pairwise stock correlations into a measurable signal of global structural change.

What this does not prove

It is important to state the limitations clearly.

- The method does not prove that a crash must occur.
- It does not identify a unique economic mechanism.
- Results depend on the asset universe, time window, return definition, and distance choice.
- Correlation estimates over short windows are noisy.
- Backtesting on historical crises is not the same as real-time prediction.

The safer claim is that TDA gives a robust way to quantify structural changes that may serve as early-warning indicators.

A Related Extension

Persistence landscapes for financial time series

Gidea–Katz: landscapes of crashes

A later paper studies a related but different construction:

Related source

M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A* 491, 820–834, 2018.

Instead of building a stock-correlation network, they build point clouds from multiple index-return time series.

They use:

- S&P 500;
- DJIA;
- NASDAQ;
- Russell 2000.

Each sliding window gives a point cloud in $\mathbb{R}^4$.

Point clouds from index returns

At day n, form the return vector

$$x_n = \left(x_n^{(1)}, x_n^{(2)}, x_n^{(3)}, x_n^{(4)}\right) \in \mathbb{R}^4.$$

A sliding window of size w gives a point cloud

$$X_n = \{x_n, x_{n+1}, \ldots, x_{n+w-1}\} \subset \mathbb{R}^4.$$

Then compute H_1 persistence of the Vietoris–Rips filtration of X_n.

The financial time series becomes a time series of point clouds, then a time series of persistence summaries.

Persistence landscapes

Persistence diagrams can be hard to treat statistically because they are multisets of points.

A persistence landscape turns each birth–death pair (b, d) into a tent function:

$$f_{(b,d)}(x) = \begin{cases} x - b, & b < x \leq (b+d)/2, \\ -x + d, & (b+d)/2 < x < d, \\ 0, & \text{otherwise.} \end{cases}$$

The landscape is a sequence of functions obtained by taking the largest, second largest, third largest, and so on.

This embeds persistence information into a function space.

Landscape norms as warning signals

Gidea–Katz compute norms of persistence landscapes:

$$\|\lambda(X_n)\|_p, \qquad p = 1, 2.$$

These numbers measure the total strength of persistent loops in each sliding-window point cloud.

They report strong growth in these norms near major crashes, including:

- the dot-com crash of 2000;
- the Lehman bankruptcy period in 2008.

This is another way to turn persistent H_1 into a financial time-series diagnostic.

Network method versus landscape method

Method	Data object	Topological summary
Gidea 2017	stock-correlation network	diagram distances
Gidea–Katz 2018	sliding-window point clouds	landscape norms

Both methods use persistent homology to detect structural changes.

The difference is the object being filtered:

- a weighted network of pairwise correlations;
- or a point cloud of recent multidimensional returns.

Both are consistent with the course theme: local measurements can reveal global organization.

Summary

The complete financial-network pipeline

1. Start from adjusted closing prices $S_i(t)$.
2. Compute daily returns $x_i(t)$.
3. Estimate rolling correlations $c_{ij}(t)$.
4. Convert correlations to distances $d_{ij}(t) = \sqrt{2(1 - c_{ij}(t))}$.
5. Build a weighted stock network for each day.
6. Threshold the network over all θ.
7. Fill cliques to obtain a filtration of simplicial complexes.
8. Compute $\mathrm{PD}_0(t)$ and $\mathrm{PD}_1(t)$.
9. Compare diagrams over time using Wasserstein distances.

The result is a numerical signal describing topological change in the market.

What to remember

- Correlations can be converted into distances between stocks.
- A weighted stock network gives a threshold filtration.
- Clique complexes fill mutually correlated groups.
- H_0 tracks how correlated clusters merge.
- H_1 tracks loops not explained by filled cliques.
- Distances between persistence diagrams quantify changes in market topology.
- In Gidea's case study, these topological distances change before the 2007–2008 crisis peak.
- Persistent homology is an early-warning diagnostic, not a deterministic prediction tool.

Discussion questions

1. Why is it useful to analyze all thresholds instead of choosing one correlation cutoff?
2. What does $d_{ij} = \sqrt{2}$ mean in terms of correlation?
3. What financial behavior could make many H_0 deaths move to lower thresholds?
4. Why does a triangle of three strongly correlated stocks not produce an H_1 loop?
5. What would you need to check before using this as a real-time warning indicator?

Suggested bibliography

For students who want to read more:

- M. Gidea, "Topological Data Analysis of Critical Transitions in Financial Networks," Springer Proceedings in Complexity, 47–59, 2017.
- M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A* 491, 820–834, 2018.
- R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, 2000.
- H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, American Mathematical Society, 2010.
- P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, 16, 77–102, 2015.

Acknowledgment

These slides were prepared with the assistance of OpenAI Codex / ChatGPT 5.5 as a drafting and editing tool.
All mathematical, pedagogical, and course-specific content remains under the responsibility of the lecturer.

Figure sources: all schematic figures in this lecture are original TikZ diagrams created for these slides. The persistence-diagram slide reproduces a cropped figure from Gidea (2017), acknowledged on that slide.