Probing Persistent Structures in Stock Markets and Citation Networks at Multiple Length Scales

Keywords: Stock Markets; Citation Networks; Persistent Structures; Topological Data Analysis; Multiple Length Scales.

Extended Abstract

Motivation. Complex systems undergo self-organization to produce emergent structures at multiple time and length scales, but a method to simultaneously probe multiple length and time scales, or just multiple time scales is not yet available. To identify the most persistent emergent structures at multiple length scales, topological data analysis (TDA) can be used [1,2]. This was recently applied to stock markets data to identify non-interacting structures [3], and in combination with the Ollivier-Ricci curvature (ORC) to identify weakly interacting structures [4]. In this follow-up study, we combine TDA and ORC analysis to topologically characterize persistent emergent structures not only in stock markets, but also in citation networks.

Approach. From Yahoo! Finance we downloaded the daily closing prices between Jul 2019 and Dec 2020 for 463 component stocks of the S&P 500 index. We then computed the daily returns of each stock i, the cross correlation C_{ij} between stocks i and j over four time windows (one before, two during, and one after the Mar 2020 COVID-19 crash), before computing the ultrametric distances $d_{ij} = \sqrt{2(1 - C_{ij})}$ for TDA filtration. At scale ϵ, we constructed the filtered adjacency matrix such that $A_{ij} = 1$ if $d_{ij} < \epsilon$, and $A_{ij} = 0$ otherwise (**Figure 1a**), before removing filtered links with negative ORCs (weak interactions) (**Figure 1b**). Finally, we identified the connected components for scale ϵ. We also downloaded from the Web of Science bibliographic records between 1985 and 2023 using the keywords 'artificial intelligence' and 'neural networks'. Using these records, we constructed the weighted bibliographic coupling network (BCN, or co-reference network) for each year, drawing an undirected link of weight w_{ij} between papers i and j if they have w_{ij} references in common. We then performed a discrete filtration process by keeping links with $w_{ij} \geq w$, remove filtered links with negative ORCs, before identifying the clusters associated with weight w.

Results. Varying the 'length scale' over $0.2 \leq \epsilon \leq 2.0$, or over $6 \geq w \geq 1$, we found the S&P 500 component stocks or AI+NN papers organizing into many small clusters at small 'length scales', a few larger clusters at intermediate 'length scales', or just one giant cluster at large 'length scales' (**Figures 1d & 1e**). For the US stock market, we find at large 'length scales' networks with very distinct core-shell structures during market crashes, whereas at intermediate 'length scales', we find a large cluster of growth stocks (low B/M ratio) and a few small clusters of value assets (high B/M ratio). For AI research, we find core-shell networks at large 'length scales', and many small, elongated clusters at intermediate 'length scales' during normal years. In breakthrough years, the clusters are larger and more spherical, connected to each other via star centers whose links have negative ORCs. As ϵ goes from intermediate to large, holes and voids proliferate in connected components. This is true for both the US stock market and AI research.

Conclusions. For two very different complex systems (stock market and citation network), we found a similar hierarchy of self-organized persistent structures (many small clusters at small length scales, few large foam-like clusters with holes and voids (**Figure 1c**) at intermediate length scales, and core-shell structures at large length scales.

References

[1] Patania, A., Vaccarino, F. and Petri, G. Topological analysis of data. *EPJ Data Science* 6(1):1-6, 2017.

[2] Salnikov, V., Cassese, D. and Lambiotte, R. Simplicial complexes and complex systems. *European Journal of Physics* 40(1):014001, 2018.

[3] Yen, P.T.W. and Cheong, S.A. Using topological data analysis (TDA) and persistent homology to analyze the stock markets in Singapore and Taiwan. *Frontiers in Physics* 9:572216, 2021.

[4] Yen, P.T.W., Xia, K. and Cheong, S.A., Understanding changes in the topology and geometry of financial market correlations during a market crash. *Entropy* 23(9):1211, 2021.

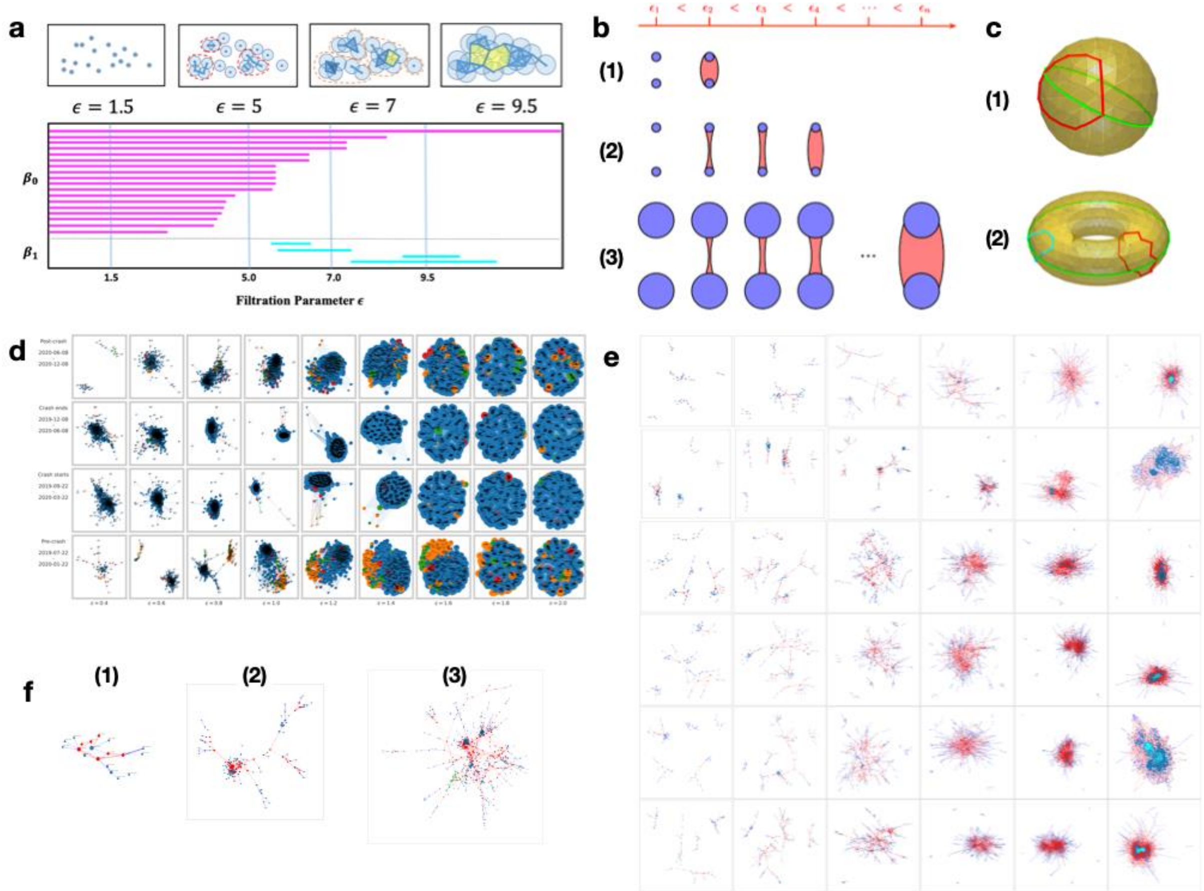

Figure 1. Using topological data analysis to identify persistent structures in stock markets and citation networks at multiple length scales. (a) The TDA filtration procedure for a data set, showing different connected structures (isolated by closed red dashed curves) emerging at different scales ϵ. In this figure, triangles and tetrahedrons are shown with a darker shade of blue, whereas 'holes' are shaded yellow. The barcode summary of this filtration procedure. **(b)(1)** A non-persistent pair of small structures is not distinct when the 'neck' connecting them is thick. **(2)** A persistent pair of small structures remains distinct while the 'neck' remains thin. **(3)** An extremely persistent pair of large structures whose 'neck' remains thin over a very wide range of ϵ. **(c)(1)** The homological quotient group H_1 of a sphere is trivial, because all closed loops can be shrunk to a point. **(2)** H_1 of a torus is non-trivial, because of two equivalent classes of closed loops that cannot be shrunk to a point. **(d)** Persistent connected structures of the US stock market at different scales. **(e)** Persistent connected structures of AI research at different scales. **(f)(1)** The cluster of AI + Medicine research at $w = 4$. **(2)** The largest cluster at $w = 4$. **(3)** The non-trivial 'docking' of AI + Medicine onto the largest cluster at $w = 3$.