Spatio-Temporal Topological Data Analysis of Ethereum's Transaction Data

Moturi Pradeep[1], Cheong Siew Ann[2], and Xia Kelin[2]

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
[2]Complexity Institute, Nanyang Technological University, Singapore

Keywords: Ethereum Blockchain, Transaction Graphs, Topological Data Analysis, Functional Segmentation, Information Propagation

Extended Abstract

Motivation. The Ethereum blockchain and its digital assets have revolutionised the landscape of digital assets and decentralised financial markets. Transaction graphs, representing the flow of value between wallets within Ethereum, have played a crucial role in understanding the system's dynamics[1], such as digital asset transfers and the behaviour of traders. Analysing the transactional graphs sheds light on underlying functional features, providing a better understanding of how interactions affect the system. Complex systems process information at different length and time scales using emergent functional organisations of its microscopic variables at these scales. In this talk, we will demonstrate how these functions can be discovered using topological data analysis[2] on blockchain transaction data from the Ethereum cryptocurrency.

Approach and Methodology. We perform the functional segmentation at various times and length scales on a directed transactional graph. To perform temporal topological data analysis, we vary the timescale and examine the M cryptocurrency transactions between N accounts within a time window of width τ. Within different window lengths τ's, the weighted directed graphs we inferred are different, but the smaller window network is nested within the larger time window ($W_\tau(i,j) \leq W_{\tau'}(i,j)$ for all $1 \leq i,j \leq N$). So, we find the weight matrix W_τ that is averaged over all sliding time windows of scale τ, and the weight matrix W'_τ averaged over all sliding time windows of scale τ'. To perform spatial topological data analysis, we obtain the shortest distance matrix D_τ(i,j) of the reciprocal values of the average weight matrix W_τ(i,j) using the Dijkstra algorithm, since the smaller distance between nodes represents frequent interactions. We deduce an network G(ϵ) of data points that are D(i,j)<ϵ, starting from ϵ=0, increasing the length scale ϵ progressively until $\epsilon = \max\limits_{(i,j)} D(i,j)$. For G($\epsilon$) at different ϵ, we plot barcodes, persistent diagrams, landscapes and compute the various Betti numbers. At scales of ϵ, we find a group of nodes which interact strongly with themselves, and we expect these groups to represent functions. We track how these groups evolve as we vary ϵ. The groups that remain the same as we vary ϵ are said to be persistent and can be thought of as the functional architecture. We compare the functional architecture for various τ 's and are concerned with functions that are equally frequent at different timescales and length scales. We would then try to organise these into information pathways through comparative analysis, such as alluvial diagrams.

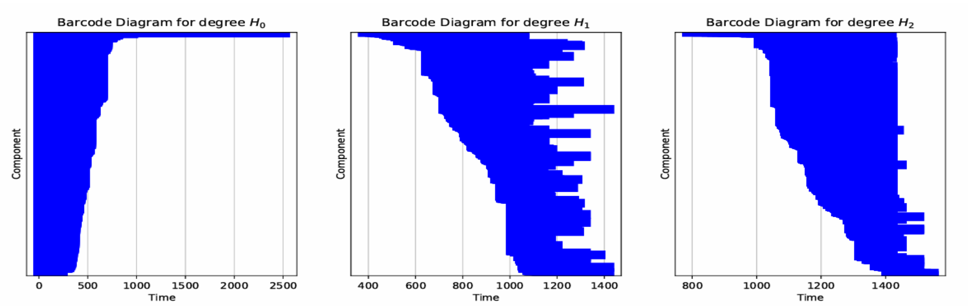
Results. Currently, we have been able to create a directed transaction graph for a particular day and generate barcodes for window length τ = 2 minutes, 20 minutes and 200 minutes, Fig(1d). We are currently working out the persistent spatio-temporal structures (the functions) at these three time scales, and how larger functions are obtained through the composition of smaller functions.

Conclusions and Outlook. With this work, we aim to understand the functional architecture of the Ethereum financial ecosystem, specifically how information propagates inside the system through these functional groups across various time and length scales. By the time of the conference, we believe that we will be able to obtain functional groups for different lengths and time scales and study how information propagates in the financial system for a particular period of time.

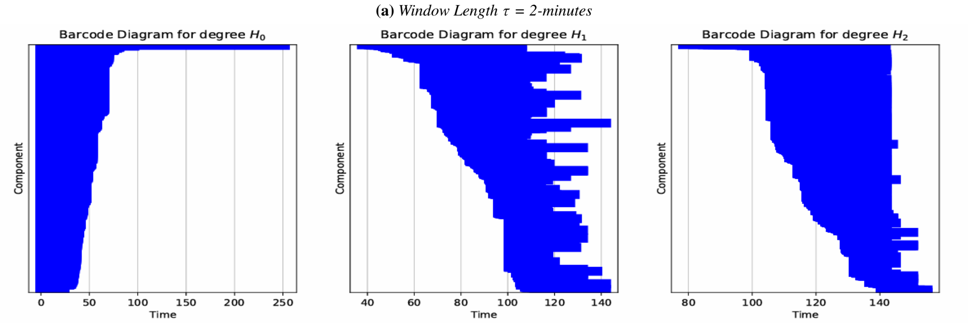
References

[1] G. Gordon A. S. Pentland S. Somin, Y. Altshuler and E. Shmueli. Network dynamics of a financial ecosystem. *Scientific reports*, 2020.

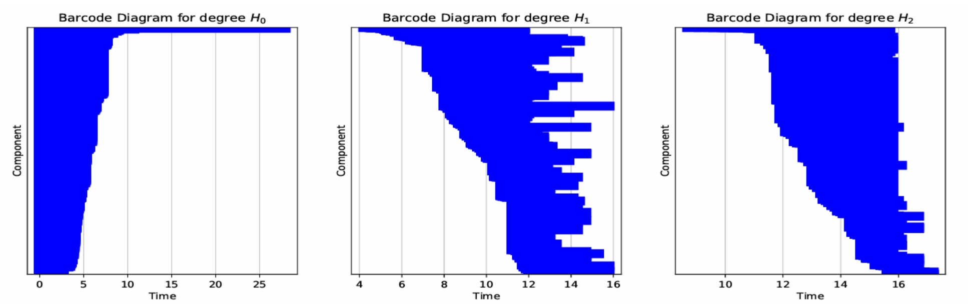
[2] P. T. Yen and S. A. Cheong. Using Topological Data Analysis (TDA) and Persistent Homology to Analyze the Stock Markets in Singapore and Taiwan. *Frontiers in Physics*, 2021.



(a) *Window Length τ = 2-minutes*



(b) *Window Length τ = 20-minutes*



(c) *Window Length τ = 200-minutes*

(d) *Figure1 : Barcodes in 0, 1, and 2 dimensions for different window lengths τ. Each bar for the respective dimensions represents a generator of the homology group. i.e $H_n^{(p,q)}$, where p,q marks a lifetime. The length of p,q signifies the persistence.*