Graph Learning Based Financial Market Crash Identification and Prediction

Dayu Qin* and Ercan E. Kuruoglu*
*Tsinghua-Berkley Shenzhen Institute, Institute of Data and Information,
Shenzhen International Graduate School, Tsinghua University, 518055, China

***Abstract*—The recognition and anticipation of financial crises play a pivotal role in safeguarding investor interests and maintaining the normalcy of the market order. Therefore, we aim to learn and predict the market abnormal fluctuation from the perspective of its graph structure. In this study, we simplify the financial complex networks with the Planar Maximally Filtered Graph(PMFG) algorithm and analyze their structural characteristic. Then, we integrate the Ollivier-Ricci curvature (ORC) of each network as a manual feature into a graph learning algorithm. We validate the effectiveness of our proposed methodology and engage in a discussion of the future research directions illuminated by our experimental results.**

***Index Terms*—Ollivier-Ricci curvature, financial complex network, PMFG, graph learning**

I. Introduction

The detrimental repercussions of a financial market crash on society are incalculable. Therefore, it is crucial for the investors to have a profound comprehension of its dynamics and a concerted effort towards early detection. Many studies have approached the examination of financial crises by investigating the structural and topological changes in financial complex networks. The Minimal Spanning Tree (MST) technique has been applied on financial complex network and the MST graph tends to exhibit increased centralization as the financial crisis approaches[1]. The identification of strongly negative Ollivier-Ricci curvature has been observed to be highly associated with abnormal edges within financial complex networks and market crash[2]. However, there is a dearth of research endeavors systematically attempting to predict financial crises from a graph learning perspective.

In our research, we generated financial complex networks and visualized the time series of their ORCs. We discovered some connections between ORC and financial crises. We employed the ORC as a manual feature in graph learning in a novel manner, which means that we choose ORC as our feature rather than running auto feature selection by machine. We aim to predict market fluctuations and provide early warnings before the onset of a financial crisis.

Corresponding author: Ercan E. Kuruoglu. This work is supported by Shenzhen Science and Technology Innovation Commission under Grant JCYJ20220530143002005, Shenzhen Ubiquitous Data Enabling Key Lab under Grant ZDSYS20220527171406015, and Tsinghua Shenzhen International Graduate School Start-up fund under Grant QD2022024C.

II. Approach

A. Building financial networks

We need to transform stock data into graph data to generate a financial complex network. Stock codes and return are chosen as nodes. The selection of edges is crucial, and in this context, we define edges between node i and j based on distance[3].

$$D_{ij} = \sqrt{2\left(1 - C_{ij}\right)} \tag{1}$$

where C_{ij} is the Pearson cross correlations.

This approach yields a fully connected graph with N nodes and N(N-1)/2 edges.

However, analyzing such a graph is often inefficient and meanningless. Because it will contain many irrelevant edges. We aim to retain as much relevant information as possible while still simplifying the structure. Therefore, the Planar Maximally Filtered Graph (PMFG)[4] is applied. It is a type of planar graph that captures the essential structural features of a complex network while minimizing edge crossings. The PMFG is derived from a complete graph by iteratively removing edges based on their weights while preserving the planarity of the resulting graph. It is characterized by being maximally planar, which means it has the maximum number of edges possible while still maintaining planarity. The edges in a PMFG represent strong correlations between nodes, making it a valuable tool for identifying key relationships in complex networks. Besides, the number of edges is reduced from N(N-1)/2 to 3(N-2) and it is easier for us to analyze now.

B. Ollivier-Ricci curvature

For 2 nodes x and y in $G(V, E)$, the Ollivier-Ricci curvature(ORC)[5] is defined as

$$ORC(x, y) := 1 - \frac{W_1\left(m_x, m_y\right)}{d(x, y)} \tag{2}$$

where $W_1\left(m_x, m_y\right)$ is L^1 transportation distance and $d(x, y)$ is the geodesic distance between x and y.

When applying ORC on graphs, $W_1\left(m_x, m_y\right)$ can be expressed as

$$W_1\left(m_x, m_y\right) = \inf_{\xi \in \Pi(m_x, m_y)} \sum_{(x,y) \in V \times V} d(x, y)\xi(x, y) \tag{3}$$

where m_x, m_y is the probability of hopping from node x,y to its neighbor nodes, $\xi(x, y)$ is the mass moves from x to y.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00127

For node x and its neighbor nodes N_x, Ollivier uses a lazy random walk method[5]

$$m_x = \begin{cases} \frac{1}{2}, & \text{if } y = x \\ \frac{1}{2|\mathcal{N}_x|}, & \text{if } y \in \mathcal{N}_x \end{cases} \quad (4)$$

$\xi(x, y)$ is a joint distribution that represents the mapping from node x to y. This mapping satisfies boundary conditions, ensuring that the marginal distributions at each node x and y are equal to the given node distributions.

III. Results

In our experiments, we selected the time frame from July 1, 2013, to June 30, 2017. During this period, the Chinese stock market underwent significant fluctuations. Considering the CSI 300 Index as a reliable indicator of the overall market conditions, we focused on the daily data of the constituent stocks of CSI 300 that existed consistently throughout this period, resulting in a dataset of 308 stocks.

We employed a sliding window of 20 days with a 5-day shift. In order to generate graph data, we used stock daily return as nodes and the distance between node i and j as edges. Within each window, we transformed the data into graph structures, employed the PMFG algorithm to remove some edges and derived the ORC from the PMFG graph.

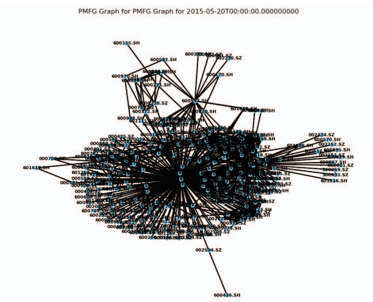

Fig. 1. An example of PMFG results

Next, we compared the variations in ORCs below -0.30, which we considered as strongly negative ORC, with the changes in the CSI 300 Index. It is noticeable that when the index exhibits a fluctuation exceeding 10%, strongly negative ORC tends to manifest in the preceding one to two weeks. Conversely, when ORC does not prominently display negative values, the index changes generally remain within a stable range below 10%, indicating a relatively steady market condition during such periods. Moreover, we

However, the peak number of strongly negative ORCs occurrences doesn't coincide with the stock market crash on June 19, 2015. Instead, there are high peaks in March and April 2015. This suggests that the peak in the number of strongly negative ORC instances occurs slightly earlier than the actual financial crisis, offering potential as an early warning signal. Following the stock market crash in June, although there are some instances of strongly negative ORC slightly higher than normal periods, they were noticeably fewer during the bear market in the year after June 2016 compared to the bull market from 2014 to 2015. Despite the occurrence of several events with index changes exceeding 20% around January 2016, no unusually high number of strongly negative ORCs was observed. In other words, the use of ORC for detecting financial crises may be less effective during bear markets.

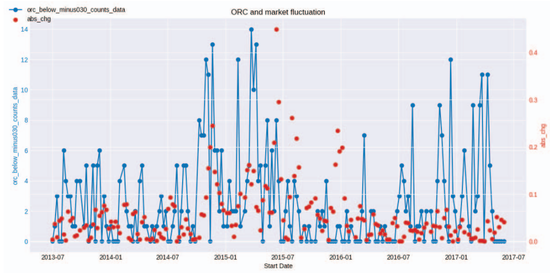

Fig. 2. comparison of ORC with index change

Then, we incorporate Ollivier-Ricci curvature (ORC) as a manually set feature into a simple 2 layers Graph Convolutional Network (GCN) model with the aim of predicting index changes exceeding 10%. The final accuracy achieved is 79.5%. Nevertheless, it is important to note that due to the relatively low stride and the limited dataset size (192 column), this result may lack precision.

IV. Conclusion

In our work, we initially constructed financial complex networks and subsequently investigated the relationship between Ollivier-Ricci curvature and abnormal financial volatility. Then we applied ORC as a manual feature in Graph learning. The results confirmed the predictive capability of ORC for anticipating financial market crashes and highlight its potential as a quantitative indicator.

In future research endeavors, We plan to conduct a more in-depth exploration of PMFG and ORC, examining whether there are additional patterns associated with each change. For instance, during the onset of a financial crisis, we intend to observe if there is a rapid decrease in ORC for major technology companies. Moreover,we aim to expand the sample size, include testing with financial derivatives, and incorporate higher-frequency trading data. We believe that the deeper application of ORC in the financial domain will make even more significant contributions.

References

[1] M. Wiliński, A. Sienkiewicz, T. Gubiec, R. Kutner, and Z. Struzik, "Structural and topological phase transitions on the German Stock Exchange," Physica A: Statistical Mechanics and its Applications, vol. 392, no. 23, pp. 5963-5973, 2013.

[2] P. T.-W. Yen, K. Xia, and S. A. Cheong, "Understanding changes in the topology and geometry of financial market correlations during a market crash," Entropy, vol. 23, no. 9, p. 1211, 2021.

[3] R. N. Mantegna, "Hierarchical structure in financial markets," The European Physical Journal B-Condensed Matter and Complex Systems, vol. 11, pp. 193-197, 1999.

[4] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems," Proceedings of the National Academy of Sciences, vol. 102, no. 30, pp. 10421-10426, 2005.

[5] Y. Ollivier, "Ricci curvature of Markov chains on metric spaces," Journal of Functional Analysis, vol. 256, no. 3, pp. 810-864, 2009.