RIETI Discussion Paper Series 24-E-085

Verification of Elemental Technologies for Anomaly Detection in Crypto Asset Transactions

IKEDA, Yuichi
Kyoto University

AOYAMA, Hideaki
RIETI

HATSUDA, Tetsuo
RIKEN

HIDAKA, Yoshimasa
Kyoto University

SHIRAI, Tomoyuki
Kyushu University

SOUMA, Wataru
Rissho University

IYETOMI, Hiroshi
Rissho University

CHAKRABORTY, Abhijit
Indian Institutes of Science Education and Research / RIKEN

FUJIHARA, Akihiro
Chiba Institute of Technology

NAKAYAMA, Yasushi
SBI Financial and Economic Research Institute Co. Ltd

ARAI, Yuta
Reitaku University

SANKAEWTONG, Krongtum
Kyoto University

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-085
December 2024

Verification of Elemental Technologies for Anomaly Detection in Crypto Asset Transactions

Yuichi Ikeda [1], *Hideaki Aoyama* [2,3], *Tetsuo Hatsuda* [3], *Yoshimasa Hidaka* [4], *Tomoyuki Shirai* [6], *Wataru Souma* [5], *Hiroshi Iyetomi* [5], *Abhijit Chakraborty* [7,3], *Akihiro Fujihara* [8], *Yasushi Nakayama* [9], *Yuta Arai* [10], *Krongtum Sankaewtong* [1]

[1] *Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University,* [2] *Research Institute of Economy, Trade and Industry,* [3] *RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program,* [4] *Yukawa Institute for Theoretical Physics, Kyoto University,* [5] *Faculty of Data Science, Rissho University,* [6] *Institute of Mathematics for Industry, Kyushu University,* [7] *Department of Humanities and Social Sciences, Indian Institutes of Science Education and Research Tirupati,* [8] *Faculty of Engineering, Chiba Institute of Technology,* [9] *SBI Financial and Economic Research Institute Co. Ltd.,* [10] *Faculty of Economics and Business Administration, Reitaku University*

Abstract

Realizing a cyber-physical economy requires dealing with the problems of the digital society that have arisen with the development of information technology. This study systematizes the mathematical basis for detecting anomalies for a dynamic graph, a network representation of relationships among nodes of crypto asset transactions and changes as time passes, based on graph theory, topology, and high-dimensional statistical analysis, to answer the three research questions: (1) Are there leading indicators of transactions that precede prices? (2) Is there a correlation between the velocity of circulation and prices? (3) Is there a herding phenomenon in the transaction network? Here, we define "anomaly" as large price fluctuations that affect transactions. The multiple methods above are applied to dynamic graphs during higher priced periods of crypto asset transactions to estimate individual anomaly indicators. We verify the effectiveness of the various anomaly detection methods by answering the three research questions for a major crypto asset. Finally, we propose a concept for an anomaly detection AI that estimates a comprehensive anomaly indicator by inputting various features from individual analysis methods.

Keywords: crypto asset, transaction network, anomaly detection, graph theory, topological data analysis
JEL classification: C55, D54, G14

Correspondence: Yuichi Ikeda, email: ikeda.yuichi.2w@kyoto-u.ac.jp
*This study is conducted as a part of the Project "Dynamics of Price in Crypto Assets and Real Economy and Their Underlying Complex Networks" undertaken at the Research Institute of Economy, Trade and Industry (RIETI).

I. Introduction

A. *Background and Objectives*

Social, environmental, and economic issues that have a broad international impact and require robust solutions, such as those symbolized by the Sustainable Development Goals (SDGs), are called global problems. In considering solutions to global problems, it is essential not to turn back from the globalized economy but to search for solutions to challenge the next economic frontier. The next economic frontier includes cyber activities brought about by new information technologies developed by the global economy in physical space. For example, in smart grids and automated driving systems, physical systems have achieved significant functional development by incorporating information systems, providing new economic value to society. In this way, the cyber-physical economy is defined as economic activities created during physical systems incorporating information systems.

In this context, we focus, in particular, on economic activities in cyberspace, which are grounded by blockchain, a technology that enables crypto assets. In cyberspace, a movement has begun to use blockchain to create autonomous decentralized organizations (DAOs), different from conventional enterprises that centralize the distribution of goods and services based on instructions from managers to their subordinates. In a DAO, all participants (members) are involved in decision-making using a governance token, a crypto asset that enables management based on the cooperation and consensus of members instead of conventional management based on top-down instructions. The spread of DAOs, which will replace or complement conventional companies, is expected to enable the construction of an economy based on a new set of human-centered values rather than values that excessively pursue economic growth.

However, today, various criminal acts and other anomalous events (anomalies) are occurring in crypto asset transactions, representing economic activities in cyberspace. They are causing significant damage to the credibility of crypto assets. An anomaly is a general term for a peculiar transaction with characteristics that deviate significantly from those considered normal. Anomalous events are often accompanied by increased trading volume and significant price fluctuations and may be caused by criminal acts such as price manipulation or money laundering. Therefore, analyzing transactions during periods of substantial price fluctuations is essential. In this study, we define an anomaly as a feature of transactions that involve large fluctuations in price. Note that the transaction data are a record of crypto asset transfers on the blockchain, and the price is determined when the crypto asset is exchanged for legal tender on the exchange market. In other words, the market where the price of crypto asset is determined and the blockchain that transfers crypto asset are essentially different. Regulatory authorities such as the Financial Services Agency monitor transaction anomalies and take appropriate countermeasures. Therefore, it is of great social significance to automatically detect criminal acts in crypto asset transactions using mathematical methods. So far, we have been working on mathematical methods to detect anomalies in crypto asset transactions using transaction data recorded on the blockchain, which is called the on-chain data (Ikeda, 2022; Aoyama et al., 2022; Ikeda and Chakraborty, 2023; Chakraborty et al., 2023, 2024).

In data science, anomaly detection generally refers to finding patterns in data that do not conform to expected behavior (Chandola et al., 2009). Anomaly detection on networks began with Noble's work in 2003 (Noble and Cook, 2003). Early research was limited to extracting subgraphs of specific patterns using information theory, etc. Around 2010, reflecting data availability on dynamic graphs, research on anomaly detection focusing on temporal changes

began. Furthermore, around 2020, applying topology and machine learning to network science became active. Developing a comprehensive indicator of anomalous events through the systematic application of graph theory, topology, and high-dimensional statistical analysis to dynamic graphs, which is the subject of this research, is in line with the development of recent trends in network science.

Network analysis is essential for preventing financial crime and money laundering (AML). It can improve efficiency by detecting anomalies in transaction networks and automating fraud detection. Tools such as VISFAN use network indicators to identify suspicious transactions (Didimo et al., 2011). García et al. applied network analysis to a tax investigation study by the Spanish Revenue Agency, using algorithms for rapid fraud detection and community detection techniques for representing economic situations (García and Mateos, 2021). Colladon et al. highlighted social network indicators to identify money laundering through the relationship graph of economic sectors, regions, transaction volumes, and ownership links (Fronzetti Colladon and Remondi, 2017). The CoDetect framework integrates network and feature data for fraud detection (Huang et al., 2018). Previous studies have been limited to detecting subgraphs corresponding to specific transaction patterns corresponding to particular irregularities (Noble and Cook, 2003; Chandola et al., 2007; Ranshous et al., 2015; Novikova and Kotenko, 2014; Huang et al., 2018; Thudumu et al., 2020; Hilal et al., 2022; Pourhabibi et al., 2020).

Machine learning and deep learning, especially graph neural networks, have been studied for fraud detection (Chen et al., 2018). Due to the lack of annotated training data, unsupervised anomaly detection is often used, and innovative approaches such as zero-shot learning are adopted (Chen et al., 2018). Reviews highlight effective anomaly detection strategies in fraud detection (Bolton and Hand, 2002; Phua et al., 2010), and social network analysis is used to uncover organized criminal activities (Šubelj, Štefan Furlan and Bajec, 2011). Recent research has focused on unsupervised and semi-supervised machine learning algorithms. These are based on unsupervised learning methods that classify suspicious transactions using fixed rules and thresholds defined by financial regulations or focus on cluster analysis (Yang et al., 2023).

In this study, we systematize the mathematical basis for detecting anomalous events in the dynamic graphs (directed and weighted graphs) of on-chain crypto asset transactions to answer the three research questions: (1) Are there leading indicators of transactions that precede prices? (2) Is there a correlation between the velocity of circulation and prices? (3) Is there a herding phenomenon in the transaction network? Based on graph theory, topology, and high-dimensional statistical analysis, we estimate multiple anomaly features from the dynamic graph analysis of crypto asset transactions and identify anomalous events related to the transactions. We also estimate price-related anomaly features by studying price time series in the exchange market of crypto assets. We aim to validate individual techniques for anomaly detection by conducting case studies in which we estimate individual indicators of anomalous events for the dynamic graphs of specific crypto asset transactions during the high-price period using multiple mathematical methods. In this study, we do not limit ourselves to specific transaction patterns but use multiple mathematical methods to estimate features of dynamic graphs that are highly correlated with price changes and use AI to synthesize these features to detect anomalous events. This research fundamentally differs from conventional approaches because it does not assume knowledge of trading patterns.

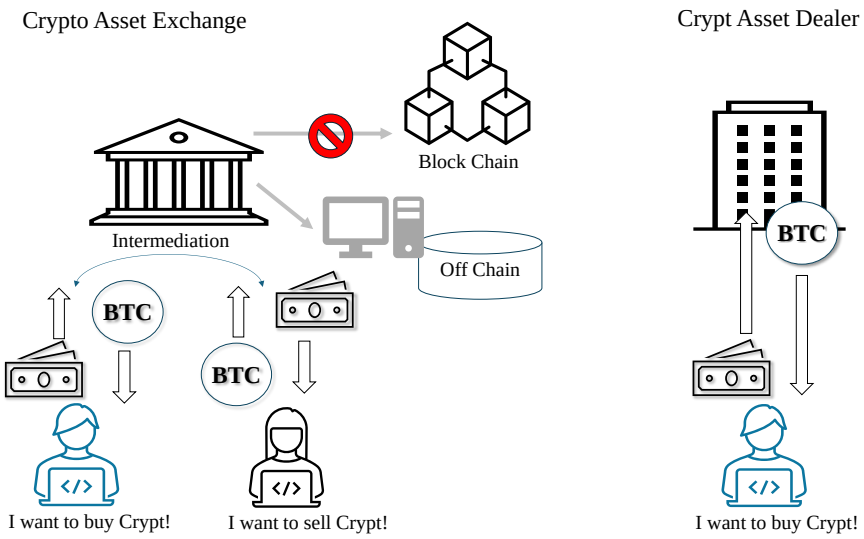

Figure 1. Crypto Asset Exchange and Crypto Asset Dealer

B. *Price formation on exchange markets*

The price of crypto assets such as Bitcoin is determined by market supply and demand. In other words, the exchange rate at which a transaction is concluded between people who want to sell and people who want to buy crypto assets is the price at the time, and it fluctuates in real-time. On platforms called crypto asset exchanges, exchanges between crypto assets and fiat currencies (legal tender) or other crypto assets are mediated, and in particular, the exchange rate between crypto assets and legal tender is the price of the crypto asset denominated in the legal tender in question.

In the exchange market, a seller and a buyer each place an order specifying their desired price and quantity, as shown in Fig. 1. These orders are called the order book, in which sell and buy orders are sorted in order of highest to lowest price, respectively (see Table 1). The difference between the lowest price of a sell order and the highest price of a buy order is called the spread, and the narrower the spread, the more liquid the exchange is. The price formation on the exchange depends on the status of the order book, and when a sell order and a buy order match, the transaction is executed. That price becomes the price on the exchange (in addition to "limit" orders, "market" orders are also used as the actual order method).

In this way, the price of a crypto asset is determined by transactions based on supply and demand, reflecting changes in the external environment, such as the fundamentals of the crypto asset itself and interest rates in traditional finance. However, it may also reflect exchange-specific factors such as trading volume, liquidity, transaction fees, regulations, reliability, and hacking risk on the exchange. The price of an exchange's shares may vary slightly from one exchange to the next. However, even if there is a temporary price divergence between exchanges, the difference is usually only marginal due to arbitrage. In addition to crypto asset exchanges, there are other ways to obtain crypto assets, such as purchasing them from dealers who own them. However, they often charge a higher spread based on the prevailing price on the exchange, which is effectively a commission. For this reason, exchange

Table 1. **Order Book** (As of 7/26/2024 14:20 BTC)

Selling	Rate	Buying
0.59744	10,401,034	
0.00664	10,400,023	
0.00359	10,400,000	
0.00500	10,399,996	
0.30541	10,399,993	
	10,399,992	0.00001
	10,320,001	0.00113
	10,320,000	0.00621
	10,300,003	0.00701
	10,300,002	0.05960

rates or their weighted averages on major crypto asset exchanges with many transactions are usually used as price indices for crypto assets.

Crypto assets are based on blockchain technology, which allows transactions to be linked chronologically by compiling transaction data into a single block of data and linking them together on a chain. In order to record a transaction into the blockchain, an approval process (connecting the correct blocks) such as Proof of Work (PoW) is required, which takes several seconds to several minutes, depending on the type of crypto asset (Bitcoin takes about 10 minutes, Ethereum 15 to 17 seconds). However, transactions on crypto asset exchanges are usually conducted without going through the blockchain (off-chain transactions) because waiting for blockchain approval does not allow for real-time transactions conducted on exchanges and because of the fees involved. In other words, a dedicated system runs on a server provided by the crypto asset trader to allow users to trade freely on the board, and transactions are conducted within that system. The only transactions recorded in the blockchain are when crypto assets are transferred from one exchange to another; for example, crypto assets deposited at an exchange are transferred to a wallet on your own or to an account at another crypto asset exchange.

The price of crypto assets is fundamentally determined by the market's supply and demand dynamics. These dynamics are influenced by the characteristics and environment of the crypto assets themselves. For instance, the characteristics of representative crypto assets like Bitcoin, Ethereum, and XRP play a significant role in price formation. Understanding these characteristics and their influence on supply and demand is crucial for comprehending the price dynamics of crypto assets.

[**Bitcoin**]

- The oldest and most popular crypto asset (the largest market capitalization).
- No initial coin offerings (ICOs) are issued, and all are issued through mining.
- The number of coins issued through mining is halved every four years, and the upper limit of the number of coins issued is set at 21 million. It is said that the aim is to stabilize the value by creating a sense of scarcity.
- On the other hand, it takes time to approve transactions (about 10 minutes), and it is said that it is necessary to wait for six blocks (i.e., 60 minutes) before a transaction

can be considered finalized.

- According to one estimate, the number of transactions that can be processed is limited to about seven per second, and scalability is also low (one block contains an average of 1,900 transactions).
- Compared to Ethereum, it is more challenging to make technical improvements and has only been successfully upgraded twice in the past.
- Both the U.S. Securities and Exchange Commission (SEC) and the U.S. Commodity Futures Trading Commission (CFTC) recognize it as a "commodity" rather than a "security".
- The demand for Bitcoin is increasing due to the approval of a Bitcoin spot ETF (Exchange Traded Fund) in the U.S. in January 2024.
- The cost of mining may affect the price formation, as the cost of mining is considered at least "worth" more than the cost to those willing to mine without purchasing Bitcoin because of the cost of computing resources and electricity.

[Ether]

- Ethereum is not a crypto asset but a platform for running decentralized applications (DApps). It uses a technology called smart contracts that enables automatically executed contracts, and various DApps have been developed in fields such as finance, gaming, real estate, and insurance.
- Ether is the native currency of Ethereum, and Ether is required to use DApps.
- The price of Ether fluctuates depending on the activity of the Ethereum network, and in general, when the demand for DApps increases, the price of Ether also increases.
- On the other hand, Ethereum, like Bitcoin, has scalability issues, and network congestion and high transaction fees may affect the price.
- Ethereum is undergoing regular planned technical improvements (such as the transition from Pow to PoS) and is constantly undergoing significant upgrades to move to a more efficient and secure system.
- Ether is issued by mining (currently staking). However, since the significant upgrade "London" in August 2021, a mechanism has been introduced to burn a portion of Ether that corresponds to gas fees (transaction fees), suppressing the increase in the total amount issued (equivalent to a share buyback in the case of stocks). Ether does not initially have a predetermined total issuance (cap). However, the scale of basic fee burning has expanded, and there have been cases where it has exceeded the issuance amount (mining amount), also affecting its scarcity.
- The U.S. Commodity Futures Trading Commission (CFTC) has ruled that it is a "commodity". At the same time, a senior official at the Securities and Exchange Commission (SEC) said it was a "security" and that the decision was shaky.
- In the U.S., an application for an Ether spot ETF (Exchange Traded Fund) from an exchange was approved.

[XRP]

- Because it is traded on a network operated by a company called Ripple, it is considered one of the centralized crypto assets.
- All 100 billion XRP coins were issued in 2005, and there are no plans to issue new ones. As of March 2020, most of the coins are held by Ripple Inc. and its founders, so not all are in circulation on the market.
- Ripple Inc, which owns a lot of the currency, may release Ripple into the market to balance supply and demand, which could affect the price formation.
- Originally developed to make international remittances more efficient, Ripple Inc. has partnered with banks and financial institutions to develop a remittance service using XRP. If its use increases, the price of XRP is expected to rise.
- It is designed to disappear little by little each time it is used in the international remittance system, and the number of coins gradually decreases, which is said to create scarcity and stabilize value.
- XRP was sued by the U.S. Securities and Exchange Commission (SEC) for violating securities laws, claiming that the issuance of XRP is a "security" (i.e., an investment contract). However, the court ruled that XRP is not a "commodity" concerning its secondary distribution. The ruling that XRP in circulation does not constitute security has led to buyers' widespread sense of relief.

C. *Reality of Fraud and the government response*

In Japan, the "Act on Prevention of Transfer of Criminal Proceeds" stipulates regulations to prevent money laundering, terrorist financing, and proliferation financing. This law mandates financial institutions and other entities to conduct transaction verification, maintain transaction records, and report suspicious transactions, among other obligations (HoureiRead, 2019). Money laundering refers to activities that make it difficult for investigative authorities to trace the origins of funds by making illicit proceeds from crimes or improper transactions appear to be from legitimate sources. This can involve transferring money into accounts under other names, selling assets using aliases, or moving funds through multiple financial institutions. The total amount of money laundering worldwide is estimated to be approximately 2-5% of the global GDP. Terrorist financing is the act of providing funds or resources to terrorists to support terrorist activities or the operations of terrorist organizations. Proliferation financing refers to the provision of funds or financial services to individuals or entities involved in the development, possession, or export of weapons of mass destruction (nuclear, chemical, and biological weapons), who are subject to measures such as asset freezes (Ministry-Of-Finance, 2024). The background for the establishment of this law includes the revised FATF Recommendations (40 Recommendations) released by the Financial Action Task Force on Money Laundering (FATF) in February 2012 and the subsequent public statement on Japan by the FATF in June 2014, which called for a swift response to deficiencies in anti-money laundering measures.

The 2016 amendment to the Act on Prevention of Transfer of Criminal Proceeds (APTCP) included crypto asset exchange providers as entities subject to compliance with the law. As a result, crypto asset exchanges are now required, like financial institutions, to conduct "Know Your Customer" processes, maintain records, and report suspicious transactions to authorities, such as the Financial Services Agency, to prevent money laundering and financial crimes. Moreover, a recent topic of interest is the amendment to the same law in December

2022, which came into effect in June 2023, mandating the implementation of the Travel Rule for crypto asset exchanges. The Travel Rule stipulates that "crypto asset exchange providers facilitating transfers of crypto assets on behalf of users must notify the recipient' s crypto asset exchange provider of specific information regarding both the sender and the recipient". The FATF has recommended this rule as part of its international standards (FATF Standards) for combating money laundering and terrorist financing, which it urges national regulators to adopt. Under the Travel Rule, crypto asset exchange providers must obtain information on the origin and destination of a crypto asset transfer and notify the recipient exchange provider. This is expected to enhance transparency regarding the parties involved in crypto asset transfers, thereby mitigating the risk of illicit use. However, it is important to note that, in crypto asset trading, there are methods other than using crypto asset exchanges, such as peer-to-peer (P2P) transactions between users, which include Decentralized Exchanges (DEXs) and unhosted wallets. These forms of P2P trading fall outside the scope of such regulatory efforts.

Under these laws, Japanese financial institutions and other entities must detect id, entify, and report suspicious transactions that are believed to be related to the transfer of proceeds from crimes, such as money laundering, terrorist financing, and proliferation financing. The Financial Services Agency has compiled reference examples of suspicious transactions, categorizing them as illustrative cases (Financial-Services-Agency, 2024). Among these examples, reference cases for Crypto Asset Service Providers are also provided. The three main types of transactions that are subject to reporting are (1) Large transactions involving crypto assets, (2) Transactions conducted frequently within a short period, and (3) Transactions where the account holder's name is fictitious or customer information is anonymized using anonymization techniques. On the dark web, Bitcoin addresses are often displayed. Bitcoin is the most valuable among crypto assets, and its price is highly volatile, which can result in large transaction amounts. One of the anonymization techniques for crypto assets is mixing (Une, 2018). This technique involves adding many unrelated addresses to the inputs and outputs in the transaction data of crypto assets, mixing them with the original addresses, thereby making it more difficult for third parties to trace the transactions or link addresses. Research has also shown that mixing-related transactions are conducted regularly and frequently (Hirosawa and Uehara, 2018). Advanced anonymization techniques using cryptographic technologies such as ring signatures and zero-knowledge proofs are also known (Une, 2018). The off-chain technology known as the Lightning Network allows for transactions between any parties, even those who have not directly opened a payment channel (a mechanism enabling off-chain transactions between two parties), making anonymous transfers possible through this method (Financial-Services-Agency, 2019).

With the future development of Web3 and blockchain technology, as well as the general public's increased literacy regarding crypto assets, it is anticipated that there will be an increase in opportunities and cases for detecting, identifying, and reporting suspicious transactions that are believed to be related to the transfer of proceeds from crimes such as money laundering, terrorist financing, and proliferation financing. Since some crypto asset transactions are automated, like mixing, it is expected that humans will become increasingly challenging to handle every case individually. Therefore, it is necessary to consider methods that utilize automation technologies, such as AI based on machine learning, wherever possible. Developing and implementing technology that can automatically report suspicious transactions to humans is essential.

II. Theory of evaluation of transaction and price features

We explain the theory of various anomaly detection methods. To answer research question 1, we use transaction features, Indicators 1 to 10, and price features, Indicator 11. To answer research question 2, we use transaction features, Indicators 3, 4, 5, and 7, and price features, Indicator 11. Fisher's equation of exchange suggests research question 2. The velocity of money indicates how often money is used in transactions over a certain period; in other words, how much money circulates in the economy. In this analysis, the velocity of money is represented by the number of loops. To answer research question 3, we use transaction features, Indicators 6, 8, 9, and 10. In this analysis, the herding phenomenon means most nodes that make up the network change similarly when prices change significantly.

A. *Transaction Features*

a. **[Indicator 1: Graph Theory] Clustering coefficient** A clustering coefficient measures the ratio to which a specific node's neighboring nodes in a graph are linked. A graph (network) $G = (V, E)$ consists of a set of vertices (nodes) V and a set of edges (links) E. A link e_{ij} connects vertex i with vertex j. The neighbor nodes N_i for a node i are defined as its immediately connected nodes:

$$N_i = \{v_i : e_{ij} \in E \vee e_{ji} \in E\} \tag{1}$$

The clustering coefficient C_i of a node i is a proportion of the number of links between the nodes within its neighborhood divided by the number of links that could exist between them. The clustering coefficient C_i is defined for the directed binary graphs:

$$C_i = \frac{|e_{jk} : v_j, v_k \in N_i, e_{jk} \in E|}{k_i(k_i - 1)}, \tag{2}$$

where k_i is the degree of node i. For the un-directed binary graphs, C_i must be multiplied by factor 2.

b. **[Indicator 2: Graph Theory] Degree Entropy** Entropy is a measure of network complexity, where low entropy means low complexity. Degree entropy S of network is defined as follows using the degree distribution $f(k)$ of network:

$$S = -\sum_{k=1}^{k_{max}} p(k) \log p(k) \tag{3}$$

$$p(k) = \frac{f(k)}{\sum_{k'=1}^{k_{max}} f(k')} \tag{4}$$

where degree k is the sum of in-degree k_{in} and out-degree k_{out}.

c. **[Indicator 3: Graph Theory] Triangular motif analysis** A motif is a small pattern contained in a network. For three nodes connected by directed links, there are 16 motifs, including three motifs (motifs 1, 2, and 4) that are only partially connected.

First, we count the number of motifs $N_k^{real}(k = 1, \cdots, 16)$ of the actual network. We assume the null-hypothesis network (directed) to be a randomized graph without changing

the in-degree and out-degree of each node. Next, we generate 1000 null-hypothesis networks (directed) and find the number of motifs N_k^{rand}. Then, we calculate the Z-score Z_k of each motif k:

$$Z_k = (N_k^{real} - E[N_k^{rand}])/sd[N_k^{rand}], \tag{5}$$

where $E[\cdot]$ and $sd[\cdot]$ are expectation value and standard deviation, respectively. If Z_k is larger than n (or smaller than $-n$), the increase (or decrease) in motif k is statistically significant because the number of motifs exceeds n times the standard deviation.

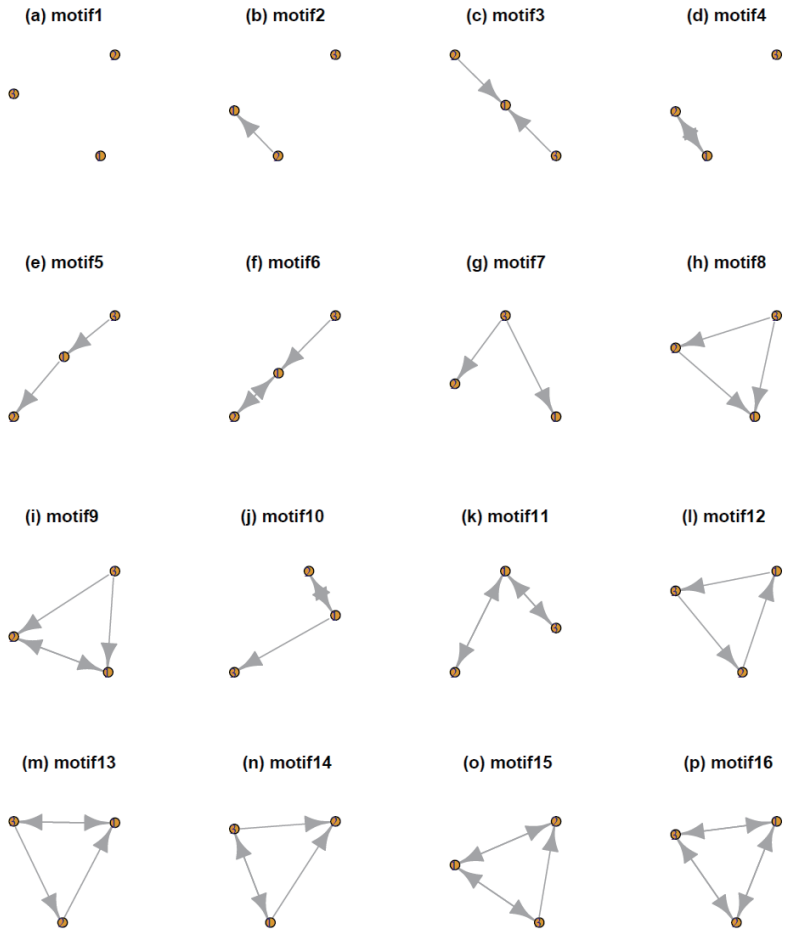

Figure 2. Sixteen Triangular Motifs

d. **[Indicator 4: Graph Theory] Transaction loop analysis considering the time of edge occurrence** Focusing on circular transactions in a financial network is an interesting analysis area, particularly for identifying potential fraudulent activity or money laundering. Circular transactions, in which funds move through a series of accounts only to return to the original source, can often indicate an attempt to disguise the origin or destination of funds, evade taxes, or conduct illicit activities. Circular transactions can also

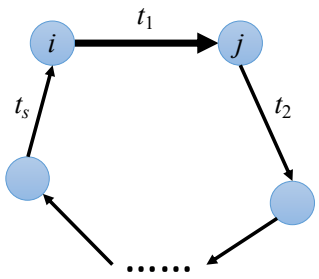

Figure 3. Definition of a loop associated with a directed link from node i to node j

help fuel high price by creating a false sense of value and stability. Of course, the high-price periods are typically the result of multiple factors, including speculative behavior, low interest rates, and overconfidence in asset markets. Circular transactions are one mechanism among many that can amplify these dynamics. Therefore, the study should emphasize the circular pattern of transactions and the circular flow of money within the networks under consideration.

To this end, we develop a methodology to identify transaction loops in a network and determine to what extent these loops are causal. We first define a loop associated with a directed link from node i to node j by connecting the two nodes via the shortest path in the backward direction, from node j to node i; the loop is *irreducible* in this sense. Figure 3 illustrates the loop for the directed link as defined above. It comprises s links with timestamps, $t_1, t_2, \ldots, t_s$, and is referred to as a loop of size s. The size distribution of the loops provides important information on the topological properties of transaction networks.

The causality of each extracted loop is then evaluated by examining timestamps associated with the constituent links and ensuring that they are aligned in chronological order. We can carry out such an analysis as the available dataset contains information about the occurrence of transactions over time. If transactions occurred randomly, the probability of finding causal loops out of all loops of size s is given by $1/(s-1)!$. We thereby propose the following indicator for anomalous transactions:

$$\xi_{\rm cl}(s) = \frac{n_{\rm cl}(s)}{n(s)/(s-1)!} \ , \tag{6}$$

where $n(s)$ and $n_{\rm cl}(s)$ are the numbers of all and causal loops of size s, respectively. If the indicator $\xi_{\rm cl}(s)$ exceeds the threshold $\eta_\alpha(s)(>1)$ determined by a given significance level α of the hypothesis test for abnormal states, a warning message should be sent out.

In passing, it is enough for us to pay attention to strongly connected components of networks for this loop statistics analysis because loops are only embedded in them.

e. **[Indicator 5: Topology] Transaction loop component by Hodge decomposition** We explain the Hodge decomposition Kichikawa et al. (2019); Fujiwara and Islam (2020); Ikeda and Chakraborty (2023) to estimate the "potential flow" and "loop flow" in the international remittance of crypto assets during the high-price period. The higher-order interaction in the remittance network disintegrates into the two subsequent two-body interactions. In this approximation, we obtained a weighted directed network for

international remittance. A weighted directed network consisting of N nodes is defined by adjacency matrix A and weighted adjacency matrix B:

$$A = [a_{ij}] = \begin{cases} 1 & \text{(if directed edge from } i \text{ to } j\text{)} \\ 0 & \text{(otherwise)}, \end{cases} \tag{7}$$

$$B = [b_{ij}] = \begin{cases} b_{ij} & \text{(if directed edge from } i \text{ to } j \text{ has a weight)} \\ 0 & \text{(otherwise)}, \end{cases} \tag{8}$$

where $i = 1, \cdots, N$ and $j = 1, \cdots, N$.

We define total flow matrix F and weight matrix W using adjacency matrix A and weighted adjacency matrix B as follows:

$$F_{ij} = B_{ij} - B_{ji}, \tag{9}$$

$$W_{ij} = A_{ij} + A_{ji}. \tag{10}$$

Graph Laplacian L is written as follows using weight matrix W:

$$L_{ij} = D_{ij} - W_{ij}, \tag{11}$$

where $D_{ij} = \delta_{ij} \left(\sum_k W_{ik} \right)$ is the degree matrix. We obtain flow potential ϕ by solving the following Laplace-like equation numerically on the network:

$$\sum_j L_{ij} \cdot \phi_j = \sum_j F_{ij}. \tag{12}$$

Flow potential ϕ is arbitrary for the shift from the origin. We shift the origin to get $\sum_j \phi_j = 0$. By taking numerical derivative of potential ϕ, we obtain potential flow matrix F^{pot}:

$$F_{ij}^{pot} = W_{ij} \cdot (\phi_i - \phi_j) \,. \tag{13}$$

Finally, we obtain loop flow matrix F^{loop} by subtracting F^{pot} from total flow matrix F as follows:

$$F_{ij}^{loop} = F_{ij} - F_{ij}^{pot} = F_{ij} - W_{ij} \cdot (\phi_i - \phi_j) \,. \tag{14}$$

This procedure is called Hodge decomposition on a weighted directed network. We define the potential flow ratio f^{pot} and the loop flow ratio f^{loop} as follows:

$$f^{pot} = \frac{\sum_{ij} F_{ij}^{pot}}{\sum_{ij} F_{ij}}, \tag{15}$$

$$f^{loop} = \frac{\sum_{ij} F_{ij}^{loop}}{\sum_{ij} F_{ij}}. \tag{16}$$

f. **[Indicator 6: Topology] Classification by graph Laplacian eigenvalue distance** We introduce states into dynamic networks and analyze the temporal changes in these states. For this purpose, we consider the distance between different graphs and perform clustering. Following Masuda and Holme (2019), we define the distance between graphs and the transitions of states. While distances between graphs can be defined in various ways, we focus here on a distance based on the graph Laplacian matrix. It can

be expressed as $L = BB^T$ using an incidence matrix B, where B^T denotes the transpose of B. This implies that the graph Laplacian is symmetric and positive-semidefinite; the eigenvalues are nonnegative real values. In the following, the eigenvalues are assumed to be arranged in ascending order, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$, where N denotes the number of vertices. The eigenvalues of the Laplacian matrix are of significant importance. The number of zero eigenvalues, i.e., the dimension of the kernel of L, corresponds to the number of connected components in the graph. In contrast, the smallest non-zero eigenvalue is known as the spectral gap. The second smallest eigenvalue, algebraic connectivity, offers insight into the graph's overall connectivity. The eigenvalues can take a wide range of values, so it would be convenient to normalize them. The symmetrically normalized Laplacian matrix is defined by using the degree matrix D as

$$\tilde{L} = D^{-\frac{1}{2}} L D^{\frac{1}{2}}, \tag{17}$$

where any eigenvalue λ of $\tilde{L}$ satisfies $0 \leq \lambda \leq 2$. One way to introduce distances between graphs is to define them using these eigenvalues. For example, the distance between graphs G_1 and G_2 with an equal number of vertices can be defined as

$$d(G_1, G_2) = \sqrt{\sum_{i=1}^{N} \left(\lambda_{N+1-i}(G_1) - \lambda_{N+1-i}(G_2)\right)^2}. \tag{18}$$

Here, $\lambda_i(G)$ represents the i-th eigenvalue of (normalized) graph Laplacian of G.

With the distance between graphs now defined, it enables the classification of states using topological data analysis and clustering algorithms. For example, hierarchical clustering can be used to classify the states of a graph.

g. **[Indicator 7: Topology] Topological data analysis** Let V be a vertex set, A a set of directed edges connecting the vertices of V, and $w : A \to \mathbb{R}$ the weight function. The triplet $G = (V, A, w)$ is called a weighted directed graph. The specification of a weighted directed graph G is equivalent to that of a weight matrix $W = (w(x, y))_{x,y \in V}$ indexed by V, where we understand $(x, y) \notin A$ if $w(x, y) = 0$. We consider a time series $\mathbb{G} := (G_n)_{n=0,1,2,\ldots,T}$ of weighted graphs as data, or equivalently, a time series of weight matrices $\mathbb{W} := (W_n)_{n=0,1,2,\ldots,T}$. Here we will consider two topological features for this $\mathbb{G}$ or $\mathbb{W}$.

(i) Traces of Powers of Weighted Adjacency Matrices. The trace of a square matrix $B = (b(x, y))_{x,y \in V}$ is defined by

$$\mathrm{Tr}(B) = \sum_{x \in V} b(x, x). \tag{19}$$

The trace of the n-th power of W has the following expression:

$$\mathrm{Tr}(W^n) = \sum_{(x_1, x_2, \ldots, x_n) \in V^n} w(x_1, x_2) w(x_2, x_3) \cdots w(x_{n-1}, x_n) w(x_n, x_1), \tag{20}$$

where $w(x_1, x_2) w(x_2, x_3) \cdots w(x_{n-1}, x_n) w(x_n, x_1) \neq 0$ if and only if there exists a sequence of directed edges $(x_1, x_2), \ldots, (x_n, x_1)$ with non-zero weight. For example, when $n = 2$,

$$\mathrm{Tr}(W^2) = \sum_{(x_1, x_2) \in V^2} w(x_1, x_2) w(x_2, x_1) \tag{21}$$

Since the sum is taken treating $w(x,y)w(y,x)$ and $w(y,x)w(x,y)$ as distinct terms, it is equivalent to counting the product of weights for all pairs of vertices twice. Of course, if there is no directed edge between two vertices x and y or if there is a directed edge in only one direction, then $w(x,y)w(y,x) = 0$. Therefore, $\mathrm{Tr}(W^2)$ is the total sum of the size of mutual transactions. When $n = 3$, we see that

$$\mathrm{Tr}(W^3) = \sum_{(x_1,x_2,x_3)\in V^3} w(x_1,x_2)w(x_2,x_3)w(x_3,x_1) \tag{22}$$

For three vertices x, y, z, there are six possible contributions:

$$\begin{aligned} &x \to y \to z \to x, \quad y \to z \to x \to y, \quad z \to x \to y \to z \\ &x \to z \to y \to x, \quad y \to x \to z \to y, \quad z \to y \to x \to z \end{aligned} \tag{23}$$

and thus the contributions of

$$w(x,y)w(y,z)w(z,x),\ w(x,z)w(z,y)w(y,x) \tag{24}$$

are counted three times each. This is equivalent to examining the total sum of transactions along the edges of triangles. For $k \geq 4$, there are combinations of multiple primitive cycles such as, for $k = 4$,

$$x \to y \to z \to w \to x, \quad x \to y \to x \to z \to x. \tag{25}$$

The former is a primitive cycle of length 4, while the latter is a concatinated two primitive cycles of length 2. If we focus on the contribution of transactions along primitive cycles, we need to separate them.

(ii) Betti numbers of a flag complex defined by a directed adjacency matrix. A directed k-simplex is an ordered sequence of $(k+1)$ vertices $(v_0, v_1, \ldots, v_k)$ such that for all i and j with $0 \leq i < j \leq k$, $(v_i, v_j) \in A$. We denote the set of all directed k-simplices by K_k. The collection of all such directed k-simplices, denoted by $\mathbb{K} = (K_k)_{k=0,1,2,\ldots}$, is called a directed flag complex for the directed graph G. Suppose a directed graph $G = (V, A)$ is given as follows:

$$V = \{0,1,2,3,4\}, \quad A = \{(0,1),(0,2),(1,2),(2,1),(2,3),(4,2),(4,3)\} \tag{26}$$

For this G, the directed flag complex $\mathbb{K} = (K_i)_{i=0}^2$ is given by

- $K_0 = V = \{(0),(1),(2),(3),(4)\}$
- $K_1 = A = \{(0,1),(0,2),(1,2),(2,1),(2,3),(4,2),(4,3)\}$
- $K_2 = \{(0,1,2),(0,2,1)\}$

For the ordered triple $(0,1,2)$, all the ordered pairs $(0,1),(0,2),(1,2)$ belong to the set A so that $(0,1,2) \in K_2$. Similarly, we can verify that $(0,2,1) \in K_2$. However, for the ordered triple $(1,2,3)$, the ordered pairs $(1,2),(1,3),(2,3)$ do not all belong to A since $(1,2),(2,3) \in A$ but $(1,3) \notin A$. Therefore, $(1,2,3) \notin K_2$. Similarly we can verify other ordered triples do not belong to K_2.

Let $C_k(\mathbb{K}) = \{\sum_{\sigma\in K_k} a_\sigma \sigma : a_\sigma \in \mathbb{R}\}$ be the real vector space with the elements of K_k as a basis. For $(v_0, v_1, \ldots, v_k) \in K_k$, we define the boundary homomorphism by

$$\partial_k(v_0, v_1, \ldots, v_k) = \sum_{j=0}^{k} (-1)^j (v_0, v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k) \tag{27}$$

and extend it linearly. For example,

$$\partial_2(0,2,1) = (2,1) - (0,1) + (0,2), \quad \partial_1(2,1) = (1) - (2), \tag{28}$$

from which it is easy to see that $\partial_1(\partial_2(0,1,2)) = 0$. Moreover, $\partial_{k+1} \circ \partial_k = 0$ holds for $k = 0, 1, \ldots$ in general, which implies $\mathrm{Im}\partial_{k+1} \subset \ker \partial_k$. The k-th homology group of the directed flag complex $\mathbb{K}$ is defined by $H_k(\mathbb{K}) := \ker \partial_k / \mathrm{Im}\partial_{k+1}$, and its dimension is called the k-th Betti number of the directed flag complex $\mathbb{K}$. Flagser can efficiently compute these Betti numbers and persistent homology (Luetgehetmann et al., 2019; Tauzin et al., 2020; Tauzin, 2021a,b). For Betti numbers and persistent homology for undirected graphs, see e.g. (Shirai, 2023) and reference therein.

h. **[Indicator 8: Topology] Ricci curvature based on optimal transport theory** The local curvature of a graph reflects the properties of specific nodes and links, and its temporal changes and differences from neighboring links can be expected to indicate its anomaly. Consider the undirected-weighted graph consisting of n nodes with the weight of an edge w_{ij} connecting node v_i and node v_j. The Ricci curvature $\kappa(v_\alpha, v_\beta)$ along the link connecting nodes v_α and v_β is defined by

$$\kappa(v_\alpha, v_\beta) = 1 - \frac{W_1(\mu_\alpha, \mu_\beta)}{d(v_\alpha, v_\beta)}. \tag{29}$$

Here $W_1(\mu_\alpha, \mu_\beta)$ is the Wasserstein-1 distance and $d(v_\alpha, v_\beta)$ is the hop distance, which is equal to #edge between node v_α and node v_β. $W_1(\mu_\alpha, \mu_\beta)$ is obtained by minimizing the objective function:

$$W_1(\mu_\alpha, \mu_\beta) = \min_{\mu} \sum_{i,j=1}^{n} d(v_i, v_j)\mu(v_i, v_j), \tag{30}$$

where $\mu(v_i, v_j) \geq 0$ are variables for $v_i, v_j \in V$, with the following constraints:

$$\sum_{j=1}^{n} \mu(v_i, v_j) = \mu_\alpha(v_i) \tag{31}$$

$$\sum_{j=1}^{n} \mu(v_j, v_i) = \mu_\beta(v_i) \tag{32}$$

for $v_i \in V$. Here we note that $\mu_\alpha(v_i) = w_{\alpha i}/s_\alpha$, $s_\alpha = \sum_{i=1}^{n} w_{\alpha i}$, $\mu_\beta(v_i) = w_{i\beta}/s_\beta$, and $s_\beta = \sum_{i=1}^{n} w_{i\beta}$.

i. **[Indicator 9: High-dimensional statistical analysis] Correlation tensor analysis** We utilized the node2vec algorithm (Grover and Leskovec, 2016) to embed weekly weighted directed networks into a D-dimensional space. We employed unbiased random walks by setting the parameters $p = 1$ and $q = 1$, effectively reducing node2vec to the DeepWalk method (Perozzi et al., 2014). This approach, inspired by natural language models, captures the structural regularities of the network, particularly community structure, by encoding it into the vector representations of the nodes. The method generates sequences of nodes, $S = V_1, V_2, V_3, ...V_S$, through truncated random walks, akin to sentences in natural language. These sequences are then processed using the SkipGram algorithm (Mikolov et

al., 2013), mapping each node V_j to its vector representation $\Phi(V_j) \in \mathcal{R}^D$, while maximizing the co-occurrence probability of neighboring nodes within the random walks.

We introduced the correlation tensor method and its diagonalization via double SVD (Chakraborty et al., 2023). Here, we provide a concise overview of this approach. In the weekly XRP transaction networks, we identify N nodes that do at least one transaction every week throughout the study period. We define these nodes as regular nodes. Each regular node is represented in the embedding space by a time series of D-dimensional vectors, denoted as $V_i^\alpha(t)$, where i ranges from 1 to N, t from 1 to T, and α from 1 to D. The correlation tensor between the components of these regular nodes is then defined as follows:

$$M_{ij}^{\alpha\beta}(t) = \frac{1}{2\Delta T} \sum_{t'=t-\Delta T}^{t+\Delta T} \frac{[V_i^\alpha(t') - \overline{V_i^\alpha}][V_j^\beta(t') - \overline{V_j^\beta}]}{\sigma_{V_i^\alpha}\sigma_{V_j^\beta}}, \tag{33}$$

In this equation, the summation is taken over five weekly networks at times $t' = \{t-2, t-1, t, t+1, t+2\}$, corresponding to a time window of $(2\Delta T+1)$, with $\Delta T = 2$ for our analysis. The terms $\overline{V_i^\alpha}$ and $\sigma_{V_i^\alpha}$ represent the mean and standard deviation of V_i^α over the same time window of five weekly networks, covering the times $t-2, t-1, t, t+1, t+2$. Notably, a smaller ΔT value introduces more noise into the correlation tensor. However, selecting a large ΔT is also impractical, as we aim to capture the detailed temporal evolution of the networks. The influence of the window size $(2\Delta T + 1)$ on the correlation tensor is discussed in Chakraborty et al. (2023). While we use a dimension of $D = 32$ for the correlation tensor $M_{ij}^{\alpha\beta}(t)$ in our analysis, the results remain qualitatively consistent across other values of D. The quantitative dependence of the largest singular value of the correlation tensor, assuming normally distributed elements, on D is presented in Chakraborty et al. (2024).

To determine the spectrum of the correlation tensor, we employ a double SVD approach. This involves successively diagonalizing $M_{ij}^{\alpha\beta}$ using a bi-unitary transformation, also known as singular value decomposition (SVD), first along the (ij)-index and then along the $(\alpha\beta)$-index. The initial step expresses $M_{ij}^{\alpha\beta}$ as a sum of matrices through the SVD method:

$$M_{ij}^{\alpha\beta} = \sum_{k=1}^{N} L_{ik}\sigma_k^{\alpha\beta}R_{kj}. \tag{34}$$

The second step involves further decomposing each singular value, $\sigma_k^{\alpha\beta}$, as a sum of matrices using SVD:

$$\sigma_k^{\alpha\beta} = \sum_{\gamma=1}^{D} \mathcal{L}^{\alpha\gamma}\rho_k^\gamma\mathcal{R}^{\gamma\beta}. \tag{35}$$

Finally, we put together these steps to arrive at the following expression for $M_{ij}^{\alpha\beta}$:

$$M_{ij}^{\alpha\beta} = \sum_{k=1}^{N}\sum_{\gamma=1}^{D} \rho_k^\gamma(L_{ik}R_{kj})(\mathcal{L}^{\alpha\gamma}\mathcal{R}^{\gamma\beta}). \tag{36}$$

In this expression, ρ_k^γ represents the generalized singular values, which form an $N \times D$ matrix. These singular values are real and positive since M is a real correlation tensor.

j. **[Indicator 10: High-dimensional statistical analysis] Feature extraction of transaction frequency statistics** In evaluating the behavior of individual nodes, the number of their transactions and the amount of each transaction play crucial roles. Both are important and complementary to each other: If one node makes millions of transactions a day and another node makes only a couple of transactions a day, does it make the first node more important than the latter? The answer is No, if the first node's transactions are mostly worth a small amount, say worth 1 USD, and the latter's transactions are worth millions of US dollars.

The concept of "Flow-weighted Frequency (*"F-frequency"* for short) was invented to deal with this aspect of the importance of both the frequency and the transaction amount (Aoyama et al., 2022). It defines a measure of the importance of the activity of a node. In general, this concept is useful for analyzing activities in a directed network, such as the transactions of account holders. Here, we concentrate on using the F-frequency in the crypto asset transaction network.

Before defining F-Frequency, let us study how to count the number of "effective nonzero elements" in a set of numbers. For example, if the set is made of the same numbers, such as $\ell_1 = (10^2, 10^2, 10^2, 0, 0, 0)$, we just assume that they are and the answer is 3. But if the set is made with numbers that differ from each other on a large scale, such as $\ell_2 = (10^3, 10^5, 10^5, 0, 0, 0)$, what is the appropriate counting? A useful way of counting such an "effective" number is to use the following:

$$\mathrm{Rat}(\ell) = \frac{\mathrm{Total}(\ell)}{\mathrm{Max}(\ell)}. \tag{37}$$

For the first case of ℓ_1, this gives you 3, an apparent result. The second one, ℓ_2, yields 2.01, which is a reasonable value.

In the transaction network we are dealing with now, we need a measure for the inflow and outflow. For this discussion, let us denote the time series of the daily outflow by f_{out} and the daily inflow by f_{in}. All components of f_{out} and f_{in} are positive. In a case with no flow, say $f_{\mathrm{out}} = \{\}$ (an empty set), we define $\bar{f}_{\mathrm{out}} = \{\}$ and $M_n(\bar{f}_{\mathrm{out}}) = 0$ and so on.

Here, we are dealing with aggregated flows by day. Alternatively, one may deal with tick data from the flows. The difference is that if a node made several large transactions quickly, treating them as one transaction is most appropriate. Daily aggregation would take care of them unless several transactions were made in a time window, including 0:00 UTC. For this reason, the daily aggregation is chosen for this article.

Now, we need to deal with both inflow and outflow. As we saw above, the effective number of elements, or the number of transactions, is gven by $\mathrm{Rat}(\cdot)$ for each of them. The key is the difference in scale of the flows between them. To account for it, we multiply the ratio between them;

$$\boldsymbol{A} = \left\{ \mathrm{Rat}(f_{\mathrm{in}}) \times \frac{\mathrm{Max}(f_{\mathrm{in}})}{\mathrm{Max}(f_{\mathrm{in}}, f_{\mathrm{out}})}, \mathrm{Rat}(f_{\mathrm{out}}) \times \frac{\mathrm{Max}(f_{\mathrm{out}})}{\mathrm{Max}(f_{\mathrm{in}}, f_{\mathrm{out}})} \right\} \tag{38}$$

$$= \left\{ \frac{\mathrm{Total}(f_{\mathrm{in}})}{\mathrm{Max}(f_{\mathrm{in}}, f_{\mathrm{out}})}, \frac{\mathrm{Total}(f_{\mathrm{out}})}{\mathrm{Max}(f_{\mathrm{in}}, f_{\mathrm{out}})} \right\} \tag{39}$$

This is the definition of the F-frequency.

B. *Price Features*

a. **[Indicators 11: The maximum price fluctuation in the week, day, hour, minute window** When we analyze time series data, it is crucial to consider whether the time series is stationary. If we adopt the return or the logarithmic return, we can usually transform the non-stationary process into stationary. Although we use the logarithmic return in this article, there are two types. If we denote the open price at t-th time interval as $O(t)$ and the closed price at time t as $C(t)$, by using natural logarithm, we define the logarithmic return:

$$r_{\mathrm{OC}}(t) = \log C(t) - \log O(t) \ , \tag{40}$$

where the subscript "OC" stands for "open to close". On the other hand, if we denote the highest price at t-th time interval as H_t and the lowest price at time t as L_t, by using natural logarithm, we define the logarithmic return:

$$r_{\mathrm{LH}}(t) = \log H(t) - \log L(t) \ , \tag{41}$$

where the subscript "LH" stands for "low to high".

Methods like moving averages and the z-score approach identify anomalies by comparing data points to established statistical baselines, such as mean and standard deviation. Machine learning-based methods, including supervised learning and unsupervised techniques like clustering, learn the patterns within the data to spot outliers. Deep learning methods, such as Long Short-Term Memory (LSTM) networks and autoencoders, are particularly effective for detecting complex and long-term dependencies in time series.

Another approach involves time series models like ARIMA, which forecasts future data based on past trends and detects anomalies when the observed values significantly deviate from predictions. Each method has its strengths, and the choice of technique depends on the data characteristics, complexity, and the specific requirements of the anomaly detection task. Combining these methods can improve the accuracy and robustness of anomaly detection in time series data.

The z-score is defined by

$$z_i(t) = \frac{r_i(t) - \mu_i}{\sigma_i} \quad i \in \{\mathrm{OC}, \mathrm{LH}\} \ . \tag{42}$$

Here, μ_i and is the average value of r_i and σ_i is the standard deviation of r_i:

$$\mu_i = \frac{1}{T}\sum_{j=1}^{T} r_i(t) \ , \quad \sigma_i = \frac{1}{T}\sum_{j=1}^{T} \left(t_i(t) - \mu_i\right) \ . \tag{43}$$

The z-score, or standard score, measures a data point's distance from the mean regarding standard deviations. The z-score of 0 means the data point is strictly at the mean, while positive or negative values indicate how many standard deviations the point is above or below the mean. Z-scores greater than 2 or 3 typically signify outliers or anomalies, as most data in a normal distribution falls within two standard deviations of the mean. This fact makes z-scores a simple and effective tool for identifying anomalies in a dataset.

However, in the case of financial phenomena, we encounter time series following non-normal distributions, such as those with fat tails. Therefore, we need to use the robust z-score defined by

$$Z_i(t) = \frac{0.6745 \times \{r_i(t) - \mathrm{Med}[r_i(t)]\}}{\mathrm{MAD}}, \tag{44}$$

where $\text{Med}[r_i(t)]$ corresponds to the median of $r_i(t)$, and MAD is the median absolute deviation defined by

$$\text{MAD} = \text{Med}\left[|X_i - \text{Med}(X)|\right]. \tag{45}$$

In dis paper, we use the robust z-score to detect the anomalies.

III. Description of Data

A. *BTC*

We utilize the dataset containing all transactions recorded on the Bitcoin blockchain, starting from the genesis block (the first block issued on January 9, 2009) up to and including block number to the block height of 693999 (issued on August 3, 2021). A typical transaction data contains multiple input and output addresses. Each transaction represents a transfer of a specific amount of BTC (the monetary unit of Bitcoin) between one or more addresses, as will be explained further. We refer to these BTC transfers as crypto flows. An address functions like a wallet owned by a user, who could be an individual or, more commonly nowadays, an entity involved in exchanges, services, gambling, and similar activities. We will use a straightforward yet effective method to identify users from addresses, allowing us to create a large graph where the nodes represent users and the edges represent crypto flows.

Let us consider an example of a transaction (TX) where Alice transferred 1 BTC to Bob on a given day:

$$\text{TX1} : \{a_1, a_2\} \to \{a_{123}, a_1\}. \tag{46}$$

Here, a_1 and a_2 are Alice's addresses, while a_{123} belongs to Bob. Alice needed multiple addresses as inputs for TX1 because one address alone did not have enough BTC to cover the 1 BTC transfer. The output includes a1 as a change address. On another day, Alice made another transaction:

$$\text{TX2} : \{a_1, a_3\} \to \{a_{45}, a_3\}. \tag{47}$$

In this case, a_3 is also Alice's address. It is clear that when multiple addresses appear as inputs in a transaction, they all belong to the same user, which in this case is Alice. Based on TX1 and TX2, we can deduce that a_1, a_2, and a_3 are all owned by Alice. Even though a_2 and a_3 did not show up in both transactions, examining the entire transaction history shows that many addresses can be linked to specific users.

B. *ETH*

There are two types of accounts on the Ethereum blockchain: Externally Owned Accounts (EOAs) and Contract Accounts (CAs). EOAs are controlled by private keys held by individuals or organizations, analogous to user accounts in traditional systems. These accounts can initiate transactions and are managed by human users. In contrast, CAs are governed by smart contract codes embedded within the blockchain. Unlike EOAs, CAs do not have private keys, and their behavior is entirely dictated by predefined rules set in the smart contract code. These contracts can execute autonomously in response to specific transactions or events, acting as decentralized agents that can enforce agreements, manage assets, or interact with other contracts and accounts on the Ethereum network without human intervention.

It is important to note a fundamental difference between the Ethereum and Bitcoin blockchains regarding how addresses function. In Ethereum, an address represents an actual

account that holds a balance of Ether (ETH) and can directly interact with the network by sending transactions or deploying smart contracts. In contrast, on the Bitcoin blockchain, addresses do not operate as accounts. Instead, Bitcoin utilizes the Unspent Transaction Output (UTXO) model, where addresses serve as references to unspent outputs from previous transactions rather than holding a balance directly. Consequently, Bitcoin addresses are not inherently linked to specific users or accounts. To associate Bitcoin addresses with individual users, further processing—often through clustering and de-anonymization techniques—is required to group related addresses.

Apart from rewarding miners with ETH, similar to Bitcoin's BTC rewards, the Ethereum blockchain introduces gas to measure the computational effort required for transactions and smart contracts. Gas fees, paid in ETH, incentivize miners to process transactions and depend on the complexity of the operation. This mechanism compensates miners and prevents network abuse by limiting computational resources available for each transaction.

We obtained the full dataset of Ethereum blockchain transactions from block number 4,331,764 (October 2, 2017) to block number 6,345,198 (September 16, 2018) using the open-source Cryo library (https://github.com/paradigmxyz/cryo). This library interfaces with the Ethereum blockchain via JSON-RPC requests, eliminating the need for a local full archive node, which would require extensive storage resources. The data is stored in Parquet format, capturing key transaction details, including block number, transaction hash, sender account (the origin of the ETH transfer), recipient account, transaction value (ETH transferred), and timestamp (UTC of block mining). Additionally, the dataset includes information on gas usage and the input code associated with each transaction, which is later utilized to classify transaction types. To ensure data quality, we excluded failed transactions, e.g. those with insufficient gas, transactions with a zero transfer value, and contract creation transactions, identifiable by null values in the recipient column. The total number of transactions is around 240 million.

We focused on transactions between Externally Owned Accounts (EOA-to-EOA), which required a method to filter out irrelevant data. We leveraged that a standard ETH transfer consumes exactly 21,000 gas units (https://ethereum.org/en/developers /docs/gas/), and, in the absence of smart contract execution, the input length (representing code execution) is zero. However, some simple transactions include a small priority fee to incentivize miners to process them faster, leading to slightly higher gas consumption than 21,000 units. We differentiated transaction types for transactions involving smart contracts by analyzing the input length. Fungible token transfers (ERC-20) have a fixed input length of 68 bytes, transfer zero ETH, and require more than 21,000 gas units. The remaining data includes internal transactions, non-fungible token (NFT) transfers, and other tokens such as ERC-721 and ERC-1155. Figure4(a) shows a dendrogram illustrating this heuristic classification, and Fig.4(b) presents the pie chart of the transaction type distribution. The total number of transactions will be 123 million from block number 4,331,764 (October 2, 2017) to block number 6,345,198 (September 16, 2018).

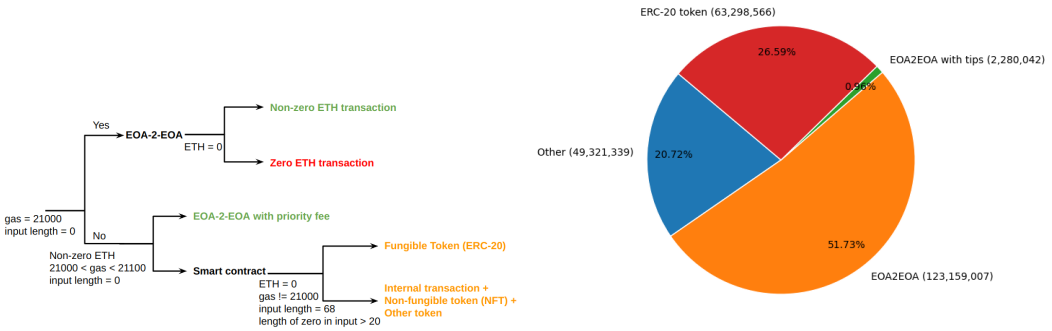

Figure 4. (a) Heuristic classification of transaction on Ethereum blockchain (b) Pie chart illustrating the ratio of each type of transaction on Ethereum blockchain

C. *XRP*

The XRP ledger records two different types of data: direct XRP Transactions and settlement transactions that transfer any type of credit, such as fiat currencies and crypto assets.

a. **[Direct XRP Transaction]** Individual users own their wallets on the XRP ledger. Different wallets may belong to the same users. A type of credit, e.g., fiat currencies such as USD, EUR, and JPY, and crypto assets such as XRP and BTC, is specified for a wallet. We note that XRP is a crypto asset that should be distinguished from the XRP ledger. The hash public key identifies a wallet. Direct XRP transactions, the most usual form of XRP transaction, allow the exchange of XRP between two wallets. For instance, user u wants to pay β XRP to user v, and u has at least β XRP in u's XRP balance. Then, β XRP is removed from u's XRP balance and added to v's XRP balance on the XRP ledger.

b. **[Settlement Transactions]** Settlement transactions transfer any credit (fiat currencies, crypto assets, and user-defined currencies) between two wallets with suitable credit paths on the Ripple network (Moreno-Sanchez et al., 2016). On the Ripple network, remittance is done as settlement transactions that transfer a type of credit, such as fiat currencies and crypto assets, between two wallets with suitable credit paths. The settlement transaction can only be performed by registered users on the Ripple network, but all transactions are recorded in the XRP Ledger and made available to the public. Figure 5 explains a remittance transaction of ¥100 on the Ripple network from user A to user B via a gateway (GW). User A makes ¥100 deposit to GW, and GW issues an IOU to A for ¥100. This IOU is sent from user A to user B. User B sends this IOU to GW and withdraws ¥100. At this time, GW's IOU disappears. Gateway is a well-known, reputed wallet on the Ripple Network that can be trusted to create and maintain an IOU credit correctly. Here, IOU credit guarantees a claim for the amount borrowed. Gateway plays an essential role in remittance transactions on the Ripple Network. We note that IOU issuers are often used instead of a gateway. IOU issuers can issue IOC credits, although they are less reputable wallets than gateways. In this study, we do not distinguish between gateway and IOU issuer as having equivalent functions.

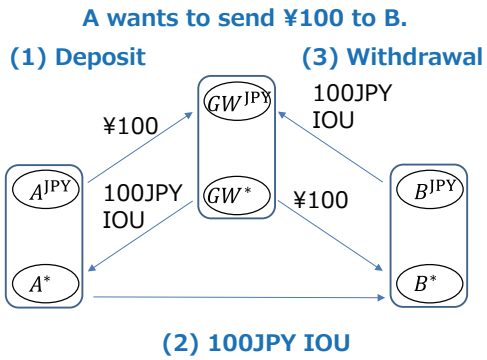

Figure 5. Settlement Transaction of ¥100 from user A to user B Gateway (GW) or IOU issuer is a well-known reputed wallet on the Ripple Network that can trust to create and maintain an IOU credit correctly.

D. *Crypto assets to be analyzed and analysis period*

This paper will analyze the period that includes the two high-price periods (period A–B and period E–F), assuming that the crypto asset to be analyzed is XRP. The analysis period is as shown in Fig. 6.

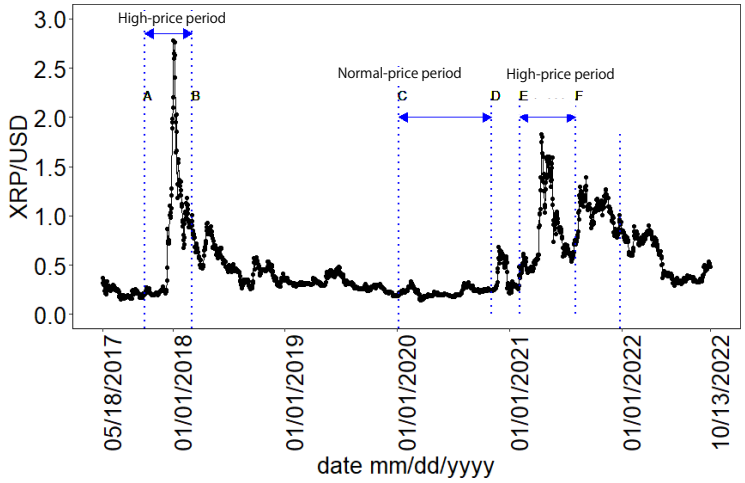

Figure 6. The XRP daily closing price in USD is recorded from May 05, 2017 to October 13, 2022. The blue horizontal line segments between different pair of blue vertical lines represent different periods, which are explained in the main text.

IV. Analysis of XRP transaction network and price

We systematize the mathematical basis for detecting anomalous events in the dynamic graphs of on-chain crypto asset transactions. Based on graph theory, topology, and high-dimensional statistical analysis, we estimate multiple anomaly features from the dynamic graph analysis

of crypto asset transactions and identify anomalous events related to the transactions. We also estimate price-related anomaly features by studying price time series in the exchange market of crypto assets.

A. *Transaction feature vector*

a. **[Indicator 1: Graph Theory] Clustering coefficient** We considered the weekly network consisting of the regular nodes that carry out at least one transaction every week from October 2, 2017, to March 4, 2018. We calculated the clustering coefficients defined by Eq. (2) for each regular node and averaged them to obtain the clustering coefficient for the regular network. Figure 7 (a) shows the clustering coefficient calculated for the high-price period A-B indicated in Fig. 6, where the price is drawn in red curve. The clustering coefficient increased during the rapid price increase period and then decreased with the price collapse. Cluster coefficients were appropriate as a feature to capture abrupt price increases.

b. **[Indicator 2: Graph Theory] Degree Entropy** We calculated the entropy defined by Eq. (3) for the regular network. Figure 7 (b) shows the entropy calculated for the high-price period A-B indicated in Fig. 6, where the price is drawn in red curve. Entropy increased during both the formation and collapse of the high-price period and decreased during the normal-price period. Entropy was an appropriate feature to capture sharp rises and falls in prices. However, it is difficult to distinguish whether prices are rising or falling from entropy alone.

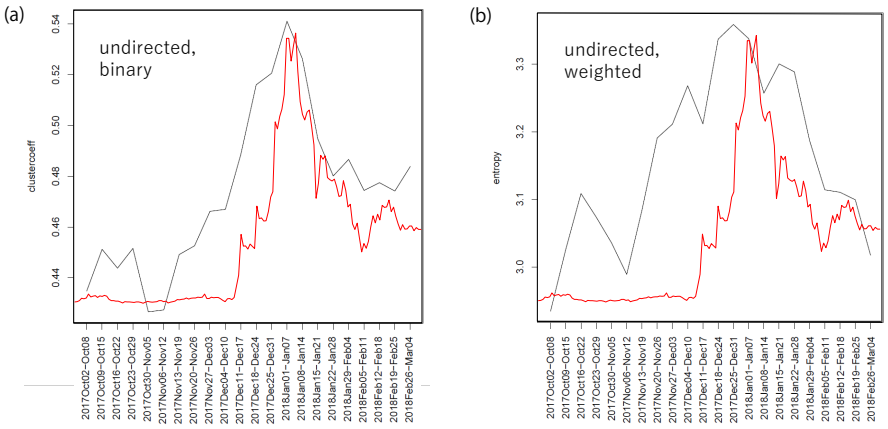

Figure 7. (a) Clustering Coefficient (b) Entropy

c. **[Indicator 3: Graph Theory] Z-score of triangular motifs** We considered the weekly network consisting of regular nodes that carry out more than one transaction every week from October 2, 2017, to March 4, 2018. The networks were treated as the directed binary network. We calculated the Z-score Z_k of each motif $k(k = 1, \cdots, 16)$ for regular networks using Eq. (5). Figure 8 shows the temporal change of Z_k for statistically significant motifs. While motifs 8, 9, 11, 14, 15 and 16 increased, motifs 12 and 13 decreased. Among the increased motifs, motifs 9, 11, 14, and 16 significantly increased. The significantly increased motifs are indicators that accurately capture the rapid price increase. However,

except for Motif 16, none of the transactions circulate among the three nodes. This may suggest the existence of circulation in larger loops with more than three nodes.

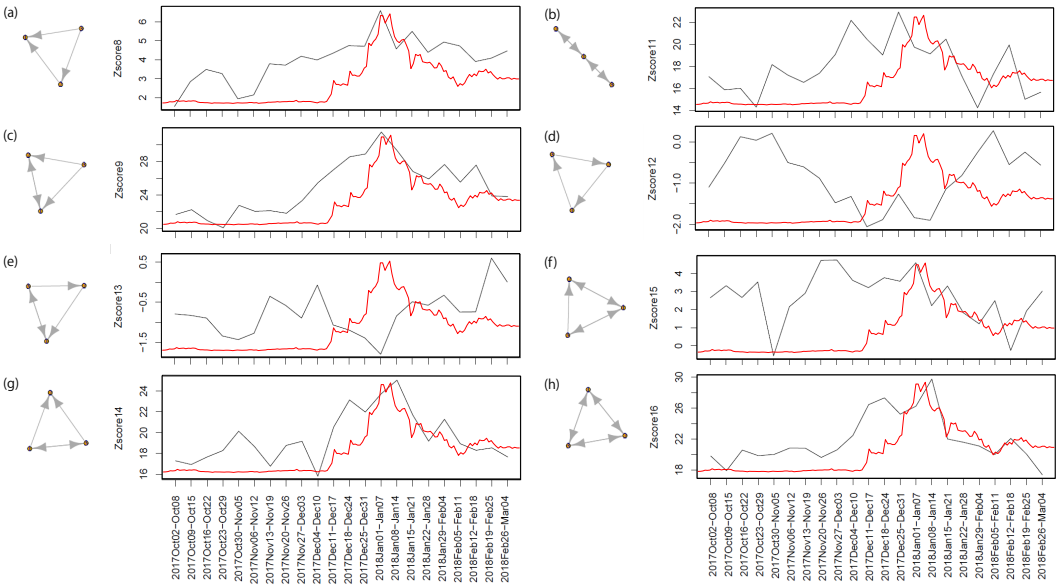

Figure 8. Temporal change of the Z-scores of statistically significant motifs: (a) Motif 8, (b) Motif 11, (c) Motif 9, (d) Motif 12, (e) Motif 13, (f) Motif 15, (g) Motif 14, and (h) Motif 16

d. **[Indicator 4: Graph Theory] Number of transaction loops considering the time of edge occurrence** We examine the indicator introduced in Sec. II.A.d by applying it to the XRP transaction networks constructed every week.

The networks exhibit the existence of multiple connections between nodes. In order to identify loops in such networks, it is helpful to employ the concept of *link network* (Luo et al., 2012; Sato et al., 2021). The link network, which is complementary to an original network, is a representation in which its nodes correspond to the original network's links, and links between these nodes are created when the corresponding links in the original network are directly connected with a common node. Then, we apply the definition of irreducible loops to link networks. We note that a particular portion of loops constructed this way is non-elementary in the original networks; a cycle is called elementary if no node is visited more than once. In financial networks, understanding elementary cycles can help identify patterns like circular transactions that contribute to anomalous behaviors such as market manipulations. We have removed such loops from the calculations given here. Additionally, we note that some traders engage in many transactions simultaneously, both incoming and outgoing. These super traders significantly increase the number of loops. In order to reduce the computational burden, nodes with both more than 500 in-degrees and out-degrees were removed from the original networks when constructing the corresponding link networks.

Figure 9 illustrates the striking variability in the total number of loops during the observation period, with values ranging from tens to millions. Upon examination, it becomes evident that a definitive correlation between the increase in the number of loops and the corresponding rise in the price of XRP, as measured by US dollars, is not feasible to ascertain.

Figure 10 disaggregates the results presented in Fig. 9, displaying the temporal variation in the share of binary and triangle loops, as well as the share of hexagonal and larger loops. The comparison indicates that the network transitions from a shorter-loop dominant structure to a larger-loop dominant structure.

In light of the structural evolution of the XRP networks over time, we calculated the indicator $\xi_{\rm cl}(s)$ defined by Eq. (6) for an intermediate size of loops, specifically rectangular loops ($s = 4$). The results are shown in Fig. 11, where we generated 1000 samples by randomly shuffling the timestamps of links to determine a threshold $\eta_{0.05}(4)$ corresponding to the 5% significance level week by week. The indicator frequently exceeds the threshold $\eta_{0.05}(4)$ during the two high-price periods (shaded in light red); in fact, such anomalous weeks occurred 31 times out of 45 weeks in total. On the other hand, $\xi_{\rm cl}(s)$ typically exhibits lower values, often even close to the expected values for the corresponding random networks, during the normal-price period (shaded in light blue). The probability that $\xi_{\rm cl}(s)$ exceeds $\eta_{0.05}(4)$ is empirically determined as 0.363. If we assume that the abnormal weeks occur randomly, the binomial distribution predicts that we have 16.3 abnormal weeks with the standard deviation σ of 3.23. Remarkably, the actual number of abnormal weeks in the two high-price periods is 4.55σ above the expected value! In contrast, during the normal-price period of 43 weeks, the abnormal weeks is observed to occur 11 times, which falls within the 95% confidence interval $[9.29, 21.9]$ expected for the corresponding random networks.

We thus see that the indicator $\xi_{\rm cl}(s)$ provides a valuable device for detecting anomalous transactions based on XRP.

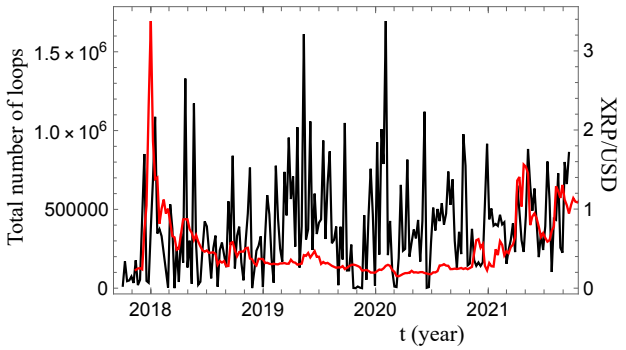

Figure 9. Temporal change in the total number of loops (black line) in relation to the price of XRP in US dollars (red line)

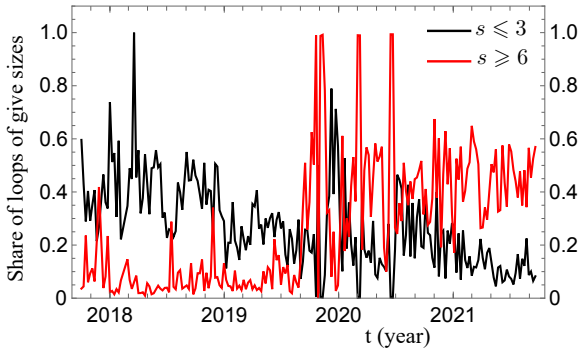

Figure 10. Temporal variation of the share of binary and triangle loops (black line) compared with that of hexagonal and larger loops (red line)

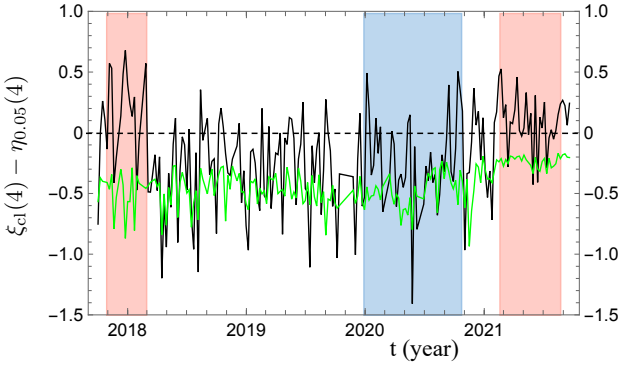

Figure 11. Excess of the indicator $\xi_{\rm cl}(s)$ for causal rectangular loops over the 5% significance level threshold together with the results, $1-\eta_{0.05}(4)$, for the expected value of $\xi_{\rm cl}(4)$ in the corresponding random networks (distinguished by green line).

e. **[Indicator 5: Topology] Ratio of trading loop components by Hodge decomposition** We calculated the potential flow component f^{pot} using Eq. (15) and the loop flow component f^{loop} using Eq. (16). Figure 12 shows the temporal change of the potential flow ratio f^{pot} in dark blue and the loop flow ratio f^{loop} in light blue. Here, the price is drawn in the dotted curve. Loop flows were larger than potential flows throughout the rapid price increase and collapse periods. The potential flow ratio increased during the rapid price increase period and decreased during the price collapse period. The loop flow ratio showed the opposite trend to potential flows, decreasing during the rapid price increase period and increasing during the price collapse period. Motif analysis shows that several types of triangular motifs increase during the high-price period, and we can expect that loop-forming transactions contribute to price appreciation. However, the Hodge decomposition results show a decrease in the proportion of loop flows. This seemingly contradictory result is consistent with interpreted as a relative increase in potential flows because of the increase in transactions due to the participation of many new users during the rapid price increase period.

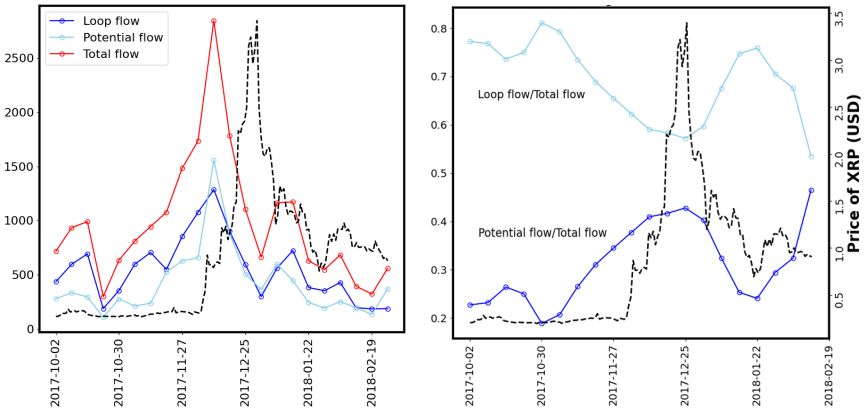

Figure 12. **Hodge Decomposition** (a) the potential flow (dark blue), the loop flow (light blue), and the price (dotted curve) (b) the potential flow ratio (dark blue), the loop flow ratio (light blue), and the price (dotted curve)

f. **[Indicator 6: Topology] Classification by graph Laplacian eigenvalue distance** We analyzed the distance from the average eigenvalue of clusters in the high-price period. We applied the price data taken from `https://coinmarketcap.com/currencies/xrp/historical-data/`, and used the closing price. We focused on two high-price periods: from October 2, 2017, to March 4, 2018, and from February 1, 2021, to August 1, 2021. We refer to these periods as the high-price period in 2017 and the high-price period in 2021, respectively. Regular nodes are defined as those that transact at least once a week during each high-price period. The number of regular nodes for the high-price period in 2017 and the high-price period in 2021 is 71 and 735, respectively. We construct networks where the nodes represent regular nodes, and the edges represent transaction relationships between them. The distance between two networks G_i and G_j is given in Eq. (18), and we define the distance matrix D whose component is

$$[D]_{ij} \coloneqq d(G_i, G_j), \tag{48}$$

where i, j represents the index of the periods. In this paper, we employ the Laplacian matrix (11). Figure 13 shows heat maps of the distance matrix for two the high-price periods. Both graphs have a three-block structure, with the middle block corresponding to periods of rapid price changes.

Next, we classify the states using a hierarchical clustering algorithm. There are several variations in defining cluster distances; here, we employ the Ward method, which minimizes the increase in within-cluster variance at each step of clustering.

Figure 14 shows the dendrogram of clustering results. The cophenetic correlation coefficients are 0.823 and 0.837 for the high-price period in 2017 and the high-price period in 2021, respectively, indicating that the clustering results are reasonably accurate. To obtain the time evolution of states, we set a threshold in the dendrogram so that there are three states. The time evolution of these three states is shown in Fig. 15. For reference, the price time series is indicated by a red line. Both graphs show a change in status during periods of rapid price change. Note that the labels of the states are for convenience only, and their order has no particular significance. These results suggest that distance-based indicators may be useful for capturing such high-price periods.

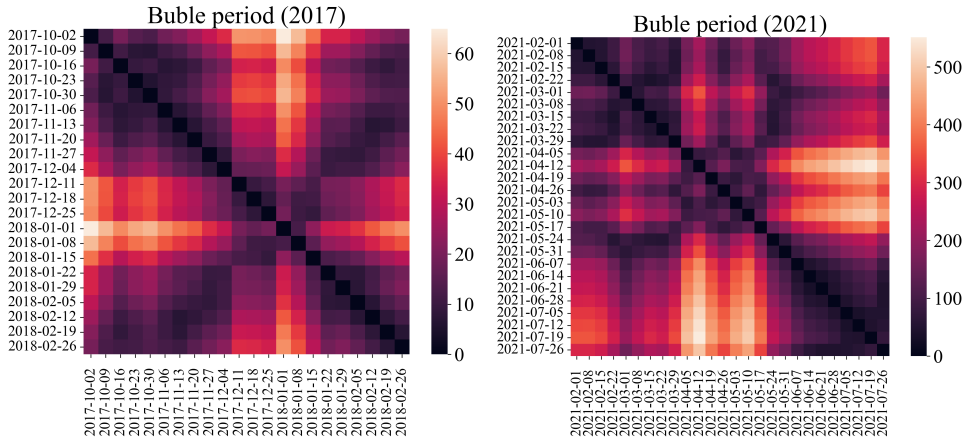

Figure 13. Heat maps of the distance matrix (48) for regular nodes during the high-price periods of 2017 (left) and 2021 (right).

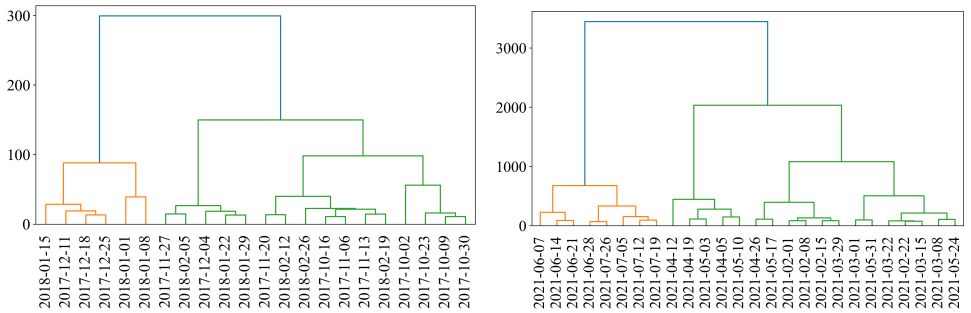

Figure 14. Dendrogram representing the classification of the network of regular nodes using hierarchical clustering during the high-price periods of 2017 (left) and 2021 (right).

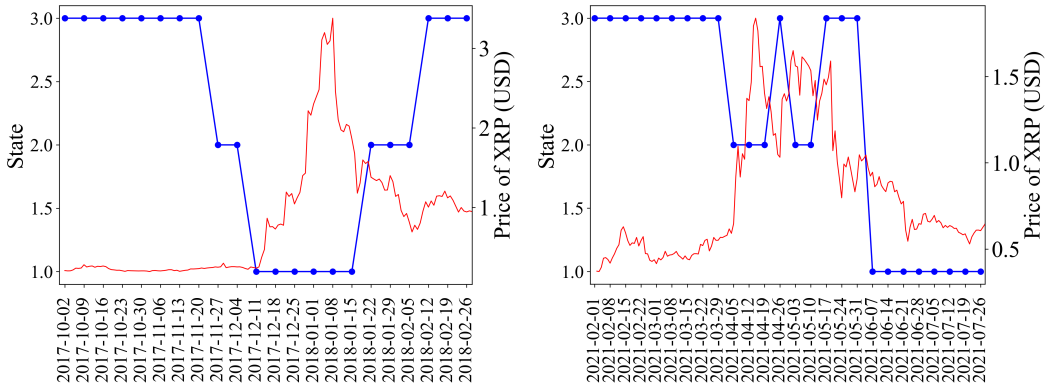

Figure 15. Time evolution of the state during the high-price periods of 2017 (left) and 2021 (right).

g. **[Indicator 7: Topology] Topological data analysis, number of transaction loops, Betti number** The analysis focuses on weekly transaction data for XRP spanning from October 2, 2017, to December 26, 2021, covering 221 weeks. Each week's dataset includes the source node ID (sender's XRP address), the destination node ID (recipient's XRP address), and the total amount of XRP transactions from the source to the destination node, which serves as the weight of the transaction. From this data, a directed weighted graph is constructed for each week, where the nodes represent XRP addresses and the directed edges represent transactions from one address to another. This graph construction process is repeated for all 221 weeks, generating a time series of directed weighted graphs. For each graph in this time series, we compute the trace of powers of the weighted adjacency matrix A and the Betti numbers. These calculations provide insights into the network's evolving transactional relationships and topological features over time. For each weekly weighted adjacency matrix, the column sum quantifies the total amount of XRP a node receives, while the row sum indicates the total amount sent. We construct subgraphs by selecting nodes with column or row sums exceeding certain thresholds such as $10^4, 10^5, 10^6$, and 10^7. Subsequently, we use z-score to detect outliers for these subgraphs. As a result of the calculations, since the results obtained using a threshold of 10^7 were nearly identical to those obtained with thresholds ranging from 10^4 to 10^6, we chose 10^7 as it yields the smallest graph size.

Figure 16 shows the traces of powers of each weighted adjacency matrix (see Indicator 7 (i)). In all powers more than or equal to 3, red points, indicating a z-score exceeding 3, can be observed during the second high-price period, spanning from week 175 to week 200. Additionally, in the trace of A^2 and A^3, which are equivalent to examining the total sum of transactions along the mutual edges and the edges of triangles, respectively, red points are observed preceding the high-price period. Of course, the validity of such an observation must be verified with a broader variety of data.

Figure 17 shows the time series of the 0th, 1st, and 2nd Betti numbers for weighted directed graphs (see Indicator 7 (ii)) and the price of XRP, respectively. As the correlation between the 1st Betti number and the price of XRP is observed, we compare the 10-week moving average of the 1st Betti number and the price of XRP in Fig. 18. The red-colored 10-week moving average curve reaches its peak shortly before the second high-price period.

h. **[Indicator 8: Topology] Average Ricci curvature** We calculated the Ricci curvature using Eq. (29) for undirected binary simple graphs to show the meaning of the graph's curvature. Figure 19 shows the Ricci curvature of undirected binary simple graphs: (a) Tree, (b) Grid, and (c) Clique. For a tree shown in Fig. 19 (a), the curvature along nodes A and B shows negative value $\kappa(A, B) = -0.5$, while the curvatures along edges in an orthogonal direction to edge AB are positive. This means that the tree is embedded in the saddle point. The average Ricci curvature of the tree is 0.091. For a grid shown in Fig. 19 (b), the curvature along nodes A and B shows $\kappa(A, B) = 0$. All curvatures along edges in other edges show small values. This means that the grid is embedded in a plane. The average Ricci curvature of the grid is 0.108. For a clique shown in Fig. 19 (c), the curvature along nodes A and B shows $\kappa(A, B) = 0.625$. All curvatures along edges in other edges show the same positive values. This means the creek is embedded in a sphere with constant positive curvature. The average Ricci curvature of the clique is 0.625.

Next, we computed the Ricci curvature for the actual XRP transaction network consisting of regular nodes during the normal-price period (Oct. 2nd–8th, 2017) and the peak of the high-price period (Jan. 1st–7th, 2018). Although this actual XRP transaction network is a

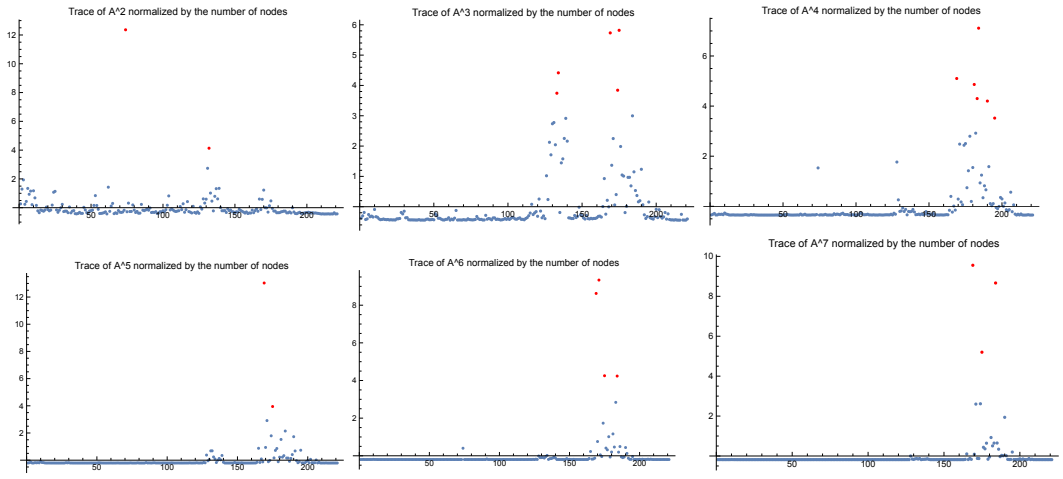

Figure 16. Traces of powers of the weighted adjacency matrix. Red points indicate z-score is more than 3.

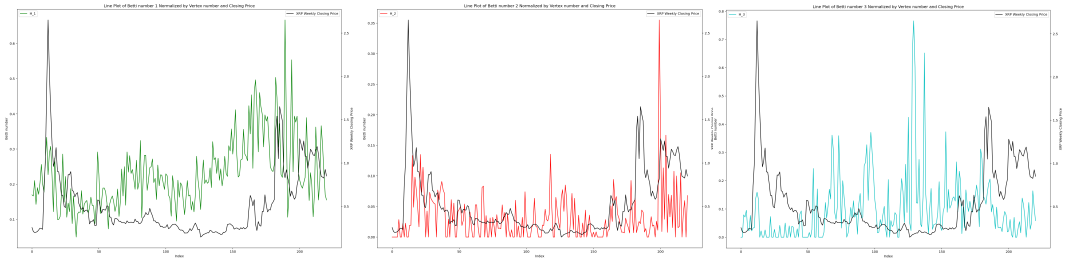

Figure 17. Price of XRP (black) and Betti numbers (from the left, the 0th(green), 1st(red), and 2nd(blue) Betti number.

Figure 18. Price of XRP (black) and 10-week moving average of the 1st Betti number (red).

directed weighted network, link weights were ignored in this curvature calculation. Figure 20 shows the Ricci curvature distribution in the normal-price period (Oct. 2nd–8th, 2017) and the peak of the high-price period (Jan. 1st–7th, 2018). Panels (a) and (b) show the normal-price period, while panels (c) and (d) show the peak of the high-price period.

Both the normal-price period and the peak of the high-price period, the curvature distributions have positive mean curvature. A small fraction of transaction links exhibit negative curvature. The network consisting of regular nodes has a higher density of transaction links, which can exhibited as positive curvature. A comparison of the distributions (a) and (b) with (c) and (d) shows that the curvature distribution shifts toward a larger mean curvature at the highest price than at the normal-price period. In other words, it corresponds to a higher density of trading links at the highest price than at the normal -price period.

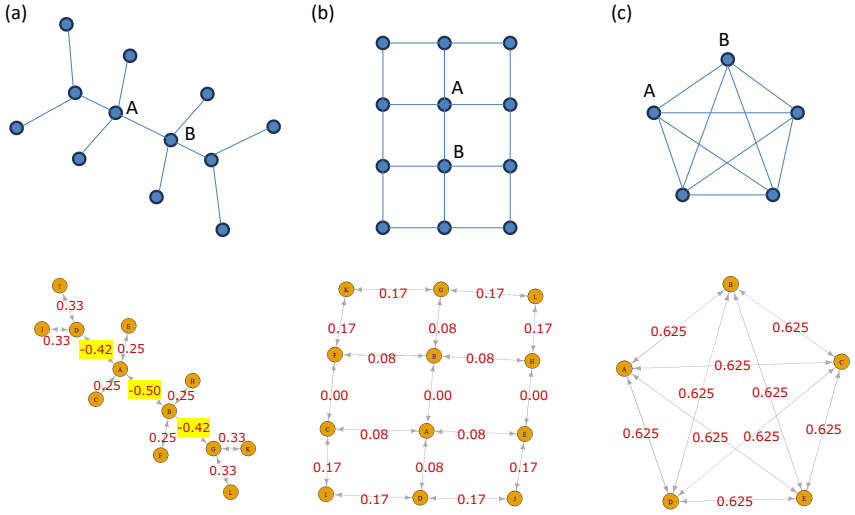

Figure 19. **Ricci curvature of undirected binary simple graphs** (a) Tree, (b) Grid, and (c) Clique

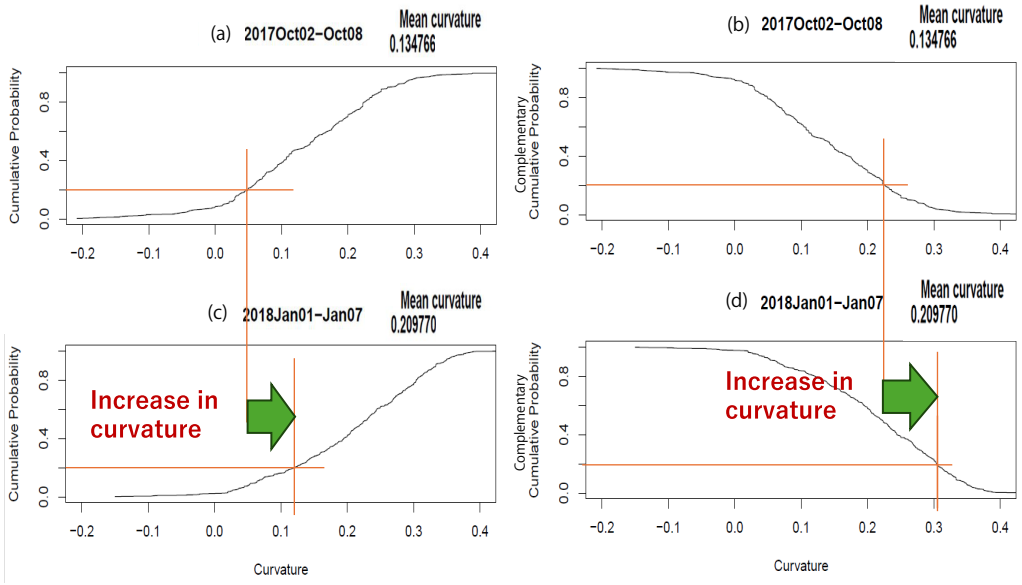

Figure 20. Ricci Curvature Distribution in the normal-price period (Oct. 2nd–8th, 2017) and the peak of the high-price period (Jan. 1st–7th, 2018) (a) Cumulative distribution in the normal time, (b) Complementary cumulative distribution in the normal time, (c) Cumulative distribution in the highest price, and (d) Complementary cumulative distribution in the highest price

i. **[Indicator 9: High-dimensional statistical analysis] Maximum singular value of correlation tensor** The network formed by XRP transactions between wallets changes every week over time. Our analysis focuses on the period from October 2, 2017, to March 2018, which includes the high-price period in XRP prices. This time frame includes 22 weekly networks. Following Eq. (33) we compute the correlation tensor between the components of regular nodes for each week. With 22 weekly networks, we obtain 18 weekly correlation tensors by Eq. (33). Each weekly correlation tensor consists of $N \times N \times D \times D$ elements. To extract key insights from these tensors, we diagonalize them using a double singular value decomposition (SVD).The double SVD is an extension of the standard SVD applied to matrices. By applying the double SVD to the weekly correlation tensor $M_{ij}^{\alpha,\beta}(t)$, we obtain the singular values $\rho_k^\gamma(t)$.

The relevance of the empirical correlation tensor is assessed by comparing it with the reshuffled correlation tensor. To calculate the reshuffled correlation tensor, we randomize the components of the embedded regular node vector v_i^α within the time window $(2\Delta T+1)$. We then compute the reshuffled correlation tensor, following Eq. (33), using these reshuffled embedded regular node vectors. We compare the singular values ρ_k^γ of the empirical correlation tensor with those of the reshuffled correlation tensor ρ_k^γ (reshuffled). The comparison is presented in Fig. 21 for the week of November 13 to November 19, 2017. The results indicate that the largest singular value of the empirical correlation tensor exceeds the largest singular value of the randomized correlation tensor.

To explore the connection between the largest singular value ρ_1^1 and the XRP/USD price, we compare the variation in the daily XRP/USD price with the largest singular value ρ_1^1

in Fig. 22. We calculate their correlation to quantify the relationship between the weekly XRP/USD price and the largest singular values ρ_1^1. The weekly XRP/USD price is defined as the average daily closing price of XRP/USD for each week, denoted as $\overline{\text{XRP/USD}}$. We then compute the Pearson correlation between $\rho_1^1(t)$ and the weekly XRP/USD price for the following week, $\overline{\text{XRP/USD}}(t+1)$. The result is a correlation coefficient of $r = -0.908$ with a p-value of 1.912×10^{-7}, indicating a strong and statistically significant negative correlation.

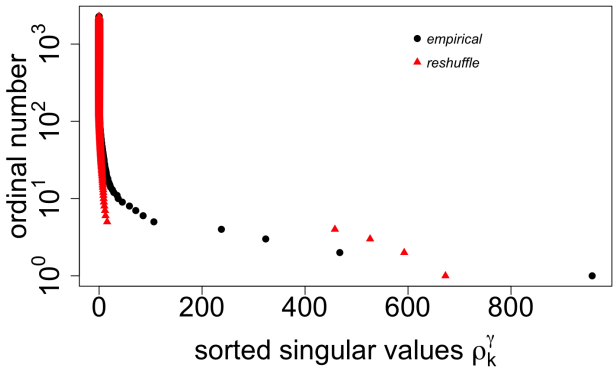

Figure 21. The sorted singular values of the empirical, reshuffled, and randomized correlation tensors for the week of November 13 to November 19, 2017. These values reflect the average computed from 20 distinct, uncorrelated network embeddings.

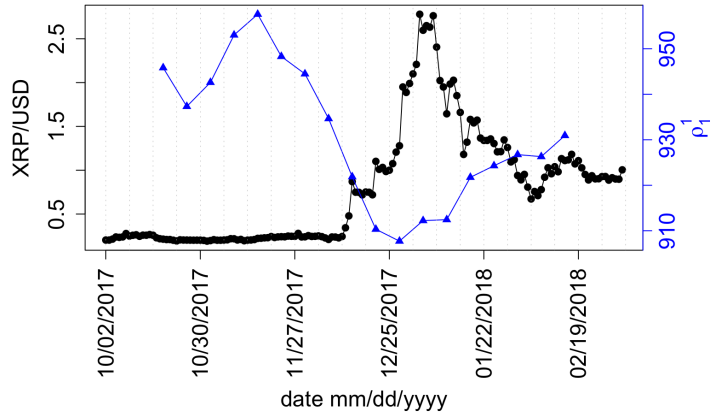

Figure 22. The comparison between the daily XRP/USD price (black curve) and the largest singular value ρ_1^1 (blue curve). The dotted grey vertical lines indicate the boundaries of the weekly windows.

j. **[Indicator 10: High-dimensional statistical analysis] Feature extraction of transaction frequency statistics** We have examined the nodes that were active

during the normal-price period, 7/2/201 (Mon) - 12/30/2017 (Sat), which interval is 182 days, or 26 weeks, using the F-Frequency explained in a previous section. Limiting the data to direct XRP to XRP transactions, we find 5,001,431 transactions altogether. In total, 512,879 nodes were active during this period, either as senders or receivers.

For example, a node "rwW.....Q3C" is among them, The daily activity is shown in Fig. 23. From this, we find

$$\begin{aligned} \mathrm{Total}(f_{\mathrm{in}}) &= 7.644 \times 10^6, \quad \mathrm{Max}(f_{\mathrm{in}}) = 5.000 \times 10^5, \\ \mathrm{Total}(f_{\mathrm{out}}) &= 1.444 \times 10^7, \quad \mathrm{Max}(f_{\mathrm{out}}) = 5.000 \times 10^6, \end{aligned} \tag{49}$$

which yield;

$$\boldsymbol{A} = \{1, 529, 2, 887\} \tag{50}$$

The F-Frequency of the top 400 nodes in this list covers nodes active for more than 120 days are plotted in Fig. 24. Further analyses using F-frequency was detailed in Aoyama et al. (2022). During the high-price period from the winter of 2017, in the case of Bitcoin, we discovered the structure of three groups of players, namely the users balancing surplus and deficit of cryptoassets (Bal-branch), those accumulating the cryptoassets (In-branch), and those reducing it (Out-branch) in the diagram of flow-weighted frequency. We found that the regime switching among Bal-, In-, Out-branches was brought about by the regular users in the case of Bitcoin, while such users are simply absent in the case of XRP.

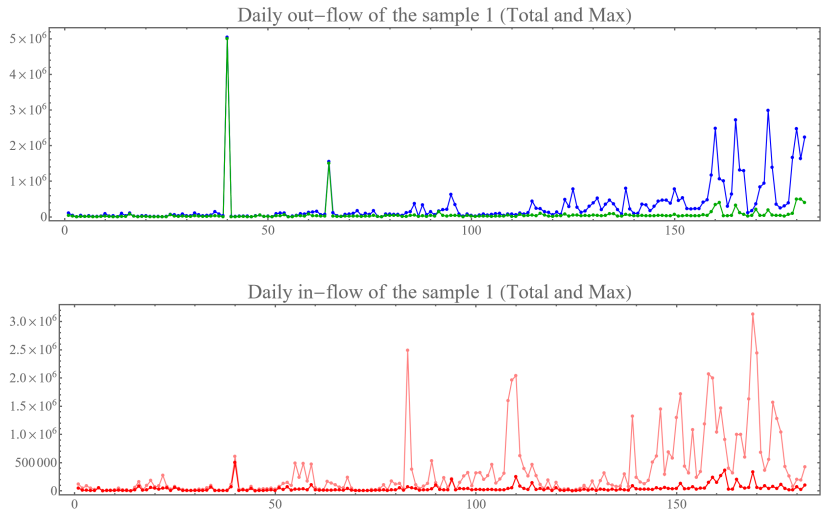

Figure 23. Daily characteristic of the sample node

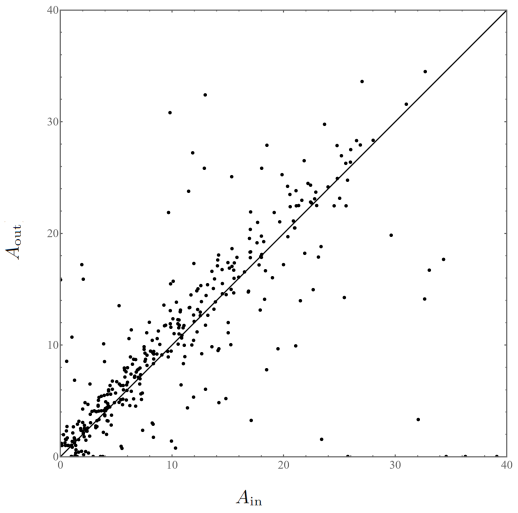

Figure 24. F-frequency $\boldsymbol{A} = (A_{\text{in}}, A_{\text{out}})$ of the top 400 nodes.

B. *Price feature vector*

a. **[Indicator 11: Time Series Analysis] The maximum price fluctuation in the week, day, hour, minute window** As explained in Subsection B, we calculate the robust z-score, i.e., Z_i, $i \in \{\text{OC}, \text{LH}\}$, in Bitfinex for different time intervals: 1 week, 1 day, 1 hour, and 1 minute.

Figures 25 present the results of calculating the robust z-score Z_{OC}. The dotted lines in these figures represent reference lines at $Z_{\text{OC}} = \pm 3$, with points outside these lines considered anomalous values. Figure 25(a) shows the analysis results with a one-week time interval, revealing anomalous values during the high-price period but none during the normal-price period. However, anomalous values are observed between the normal-price and the second high-price periods. Figures 25(b) presents results with a one-day interval, following the same general trend as Figures 25(a), but with an increased number of anomalous values due to the shorter interval. Figures 25(c) and (d) depict analyses with time intervals of one hour and one minute, respectively, revealing many points that qualify as anomalous values.

Figures 26 illustrate the results of calculating the robust z-score Z_{LH}. The broken lines in these figures are reference lines at $Z_{\text{LH}} = \pm 3$, and the figures are arranged in the same order as in Figures 25. A similar phenomenon is observed across Figures 26(a)–(d) as in Figures 25(a)–(d).

This paper proposes a method for detecting anomalies by analyzing money transfer networks. Since the formation of such networks requires a certain time frame—typically a day or a week—rather than shorter intervals like a minute or an hour, future analyses should focus on the topology of money transfer networks when anomalous values detected over a week or a day are observed. We plan to discuss this analysis in more detail in a forthcoming paper.

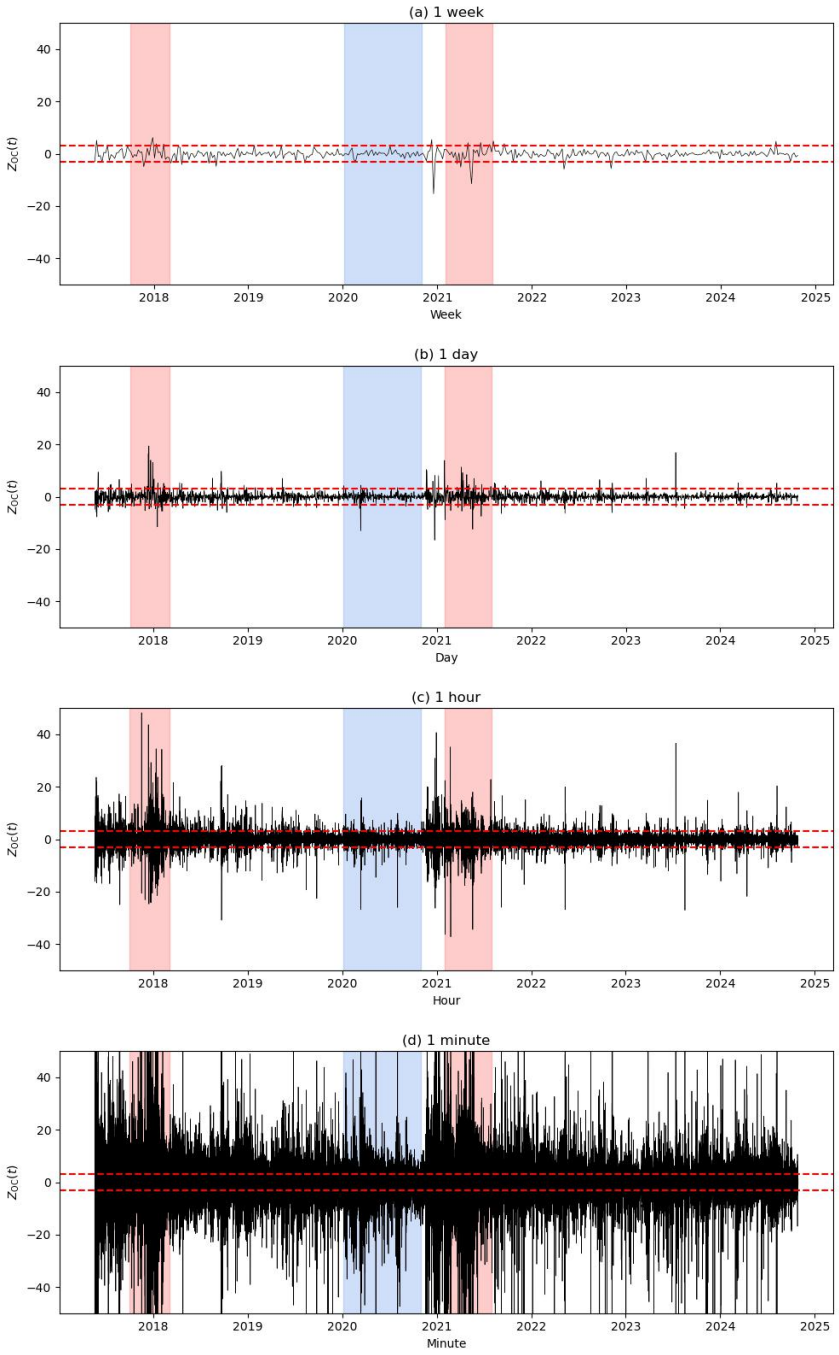

Figure 25. The robust z-score Z_{OC} in time intervals: 1 week, 1 day, 1 hour, and 1 minute. Red dashed lines show $Z = \pm 3$. Therefore, we regard the points up or down the range surrounded by this red dashed line as anomalous events.

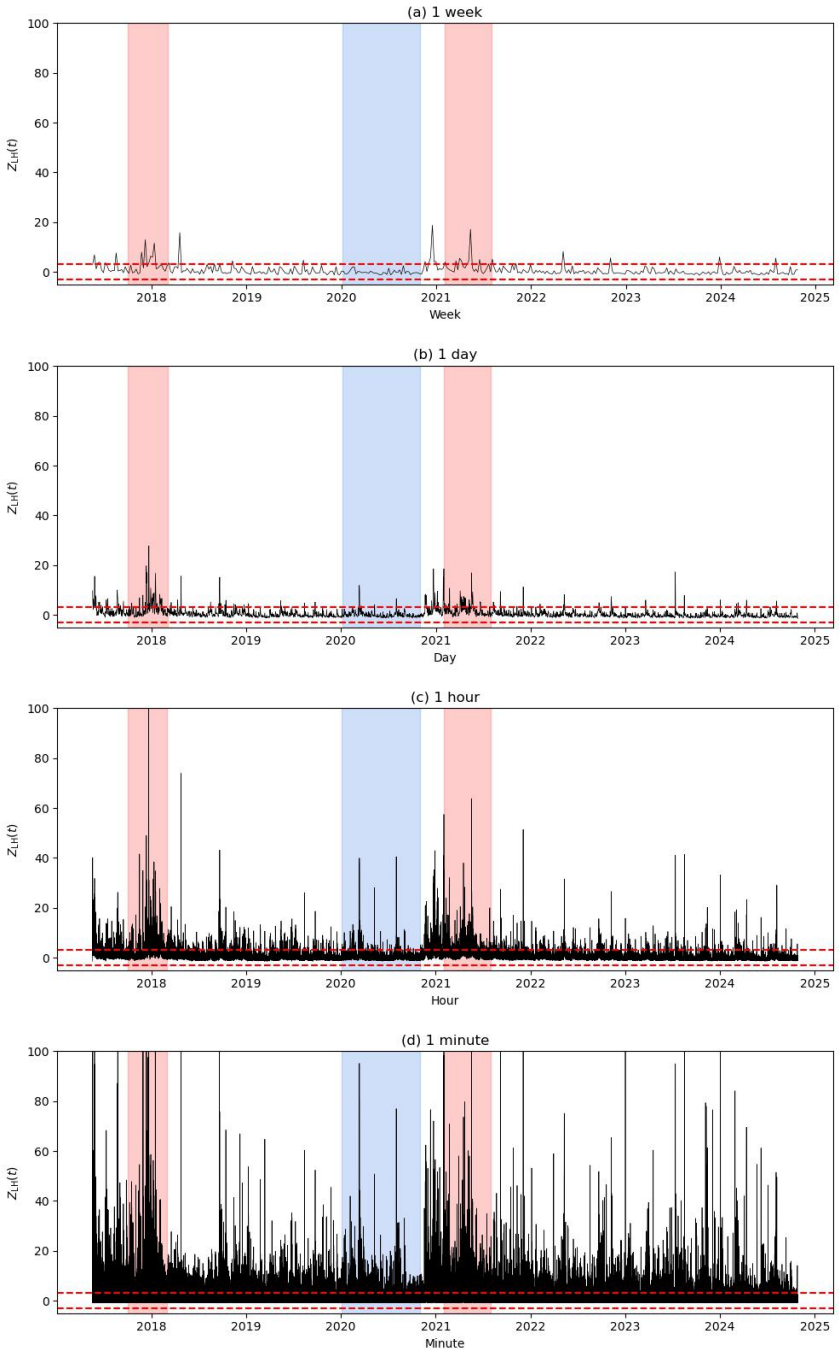

Figure 26. The robust z-score Z_{LH} in time intervals: 1 week, 1 day, 1 hour, and 1 minute. Red dashed lines show $Z = \pm 3$. Therefore, we regard the points up or down the range surrounded by this red dashed line as anomalous events.

V. Implication

Today, various criminal acts and other anomalous events are occurring in trading crypto assets, representing economic activities in cyberspace. They are causing significant damage to the credibility of crypto assets. Therefore, mathematical methods are of great social significance in automatically detecting criminal activities in crypto asset transactions. In this study, we examined the detection of anomalous events based on graph theory, topology, and high-dimensional statistics for a network that indicates the crypto asset transaction changes with time between nodes, i.e. a dynamic graph.

The validation results imply that it would be beneficial to develop an anomaly detection AI system that estimates individual indicators of anomalous events using multiple validated mathematical methods and then estimates a comprehensive indicator of anomalous events by inputting these indicators. In this study, we define anomaly as a feature of transactions that involve large fluctuations in price. Note that the transaction data is a record of crypto asset transfers on the blockchain, and the price is determined when the crypto asset is exchanged for legal tender on the exchange market. In other words, the market where the price of crypto asset is determined and the blockchain that transfers crypto asset are essentially different. However, this study has demonstrated that multiple anomaly features obtained from mathematical analysis can detect large fluctuations in prices in advance. Figure 27 summarizes the concept of anomaly detection AI implied by this validation study. This study corresponds to "Step 1 Elemental Technologies for Anomaly Detection".

Among the components of a feature vector, one feature may show an anomaly, but another feature may not. An AI is needed to determine whether a feature vector is anomalous or not from a comprehensive viewpoint based on a feature vector consisting of multiple anomaly feature values. As such, an AI, for example, the Boltzmann machine, is promising (Moro and Prati, 2023; Stein et al., 2024). The Boltzmann machine estimates (outputs) a comprehensive indicator of anomalous events by inputting a feature vector consisting of multiple anomaly features binarized to $\{0,1\}$. The parameters are learned with inputting a feature vector of normal periods using a probabilistic sampling method. By inputing a feature vector consisting of multiple anomaly features, a Boltzmann machine reconstructs the feature vector using the parameters leaned in the normal period. If the reconstructed feature vector does not match the input feature vector, the system outputs a comprehensive indicator to warn there are signs of anomaly.

If the comprehensive indicator warns of an anomaly, the graph Laplacian eigenvectors of the transaction network and the Ricci curvature of each edge based on the optimal transport problem are used to identify the node (trader) and edge (transaction) responsible for the anomalous event. Random graphs with curvatures equal to the average curvature of the actual dynamic graphs are generated, and edges with significantly larger curvatures can be identified by comparing them. Edges with large curvature are expected to be more likely to be anomalous transactions. The right part of Fig. 27 "Step 2 Anomaly Detection AI System" will be left as a future study.

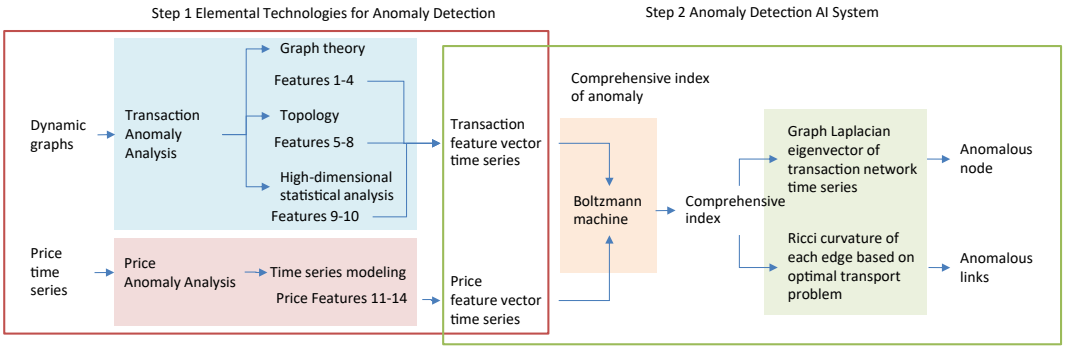

Figure 27. Concept of Anomaly Detection AI

VI. Summary

A. *Verification of elemental technologies*

We summarize the results of each analysis below. The analysis compared the transaction features, Indicators 1, 2, 3, 5, 7, and 9 with the price features; Indicator 11 and the comparison exhibited that these indicators preceded prices. Therefore, the effectiveness of Indicators 1, 2, 3, 5, 7, and 9 is confirmed to verify research question 1. The transaction features, Indicators 3, 4, 5, and 7 exhibited that the loop increased; therefore, the velocity of circulation increased during the high-price periods. Thus, the effectiveness of Indicators 3, 4, 5, and 7 is confirmed to verify research question 2. The transaction features, Indicators 6, 7 (Betti number), 8, and 9 exhibited a herding phenomenon in which most nodes that make up the network change similarly when prices change significantly. Therefore, the effectiveness of Indicators 6, 7 (Betti number), 8, and 9 is confirmed to verify research question 3.

[Transaction Feature Vector]

- **[Indicator 1: Graph Theory] Clustering coefficient** The clustering coefficient increased during the rapid price increase period and then decreased with the price collapse. Cluster coefficients were appropriate as a feature to capture abrupt price increases.
- **[Indicator 2: Graph Theory] Degree Entropy** Entropy increased during both the formation and collapse of the high-price period and decreased during the the normal-price period. Entropy was an appropriate feature to capture sharp rises and falls in prices. However, it is difficult to distinguish whether prices are rising or falling from entropy alone.
- **[Indicator 3: Graph Theory] Z-score of triangular motif** Temporal change of the Z-scores increased for motifs 8, 9, 11, 14, 15 and 16, and decreased for motifs 12 and 13. Among the increased motifs, motifs 9, 11, 14, and 16 significantly increased. The significantly increased motifs are indicators that accurately capture the rapid price increase. However, except for Motif 16, none of the transactions circulate among the three nodes. This may suggest the existence of circulation in larger loops with more than three nodes.

- **[Indicator 4: Graph Theory] Number of transaction loops considering the time of edge occurrence** During the two high-price periods, the indicator $\xi_{\rm cl}(s)$ defined by Eq. (6) for rectangular loops ($s = 4$) frequently exceeds the threshold $\eta_{0.05}(4)$. On the other hand, during the normal-price period, $\xi_{\rm cl}(s)$ typically exhibits lower values. The statistical test using the corresponding random networks shows the statistical significance of the indicator $\xi_{\rm cl}(s)$, which suggests that the indicator captures such high-price periods.
- **[Indicator 5: Topology] Ratio of trading loop components by Hodge decomposition** Loop flows f^{loop} were larger than potential flows f^{pot} throughout the rapid price increase and collapse periods. The potential flow ratio increased during the rapid price increase period and decreased during the price collapse period. The loop flow ratio showed the opposite trend to potential flows, decreasing during the rapid price increase period and increasing during the price collapse period. Motif analysis shows that several types of triangular motifs increase during the high-price period, and we can expect that loop-forming transactions contribute to price appreciation. However, the Hodge decomposition results show a decrease in the proportion of loop flows. This seemingly contradictory result is consistent with interpreted as a relative increase in potential flows because of the increase in transactions due to the participation of many new users during the rapid price increase period.
- **[Indicator 6: Topology] Classification by graph Laplacian eigenvalue distance** The distance calculated from the average eigenvalue of clusters in the high-price period for a weekly network consisting of Regular nodes classifies the three states using a hierarchical clustering analysis. The time evolution of these three states shows a correlation with the rapid price change during the high-price periods. The results suggest that distance-based indicators capture such high-price periods.
- **[Indicator 7: Topology] Topological data analysis, number of transaction loops, Betti number** The z-score exceeding 3 for the traces of powers of weighted adjacency matrix A shows a significant increase during the second high-price period. However, the trace of A^2 and A^3, which are equivalent to examining the total sum of transactions along the mutual edges and the edges of triangles, respectively, increased preceding the second high-price period. The time series of the 1st Betti numbers for weighted directed graphs shows a significant correlation with the XRP price during the first and second high-price periods.
- **[Indicator 8: Topology] Average Ricci curvature** The Ricci curvatures calculated for the actual XRP transaction network consisting of regular nodes during the normal-price period (Oct. 2nd–8th, 2017) and the peak of the high-price period (Jan. 1st–7th, 2018) show a distribution with positive mean curvature. A small fraction of transaction links exhibit negative curvature. The network consisting of regular nodes has a higher density of transaction links, which can be exhibited as positive curvature. The curvature distribution shifts toward a larger mean curvature at the highest price than at the normal-price period. The Ricci curvature detected a higher density of trading links at the highest price than at the normal-price period.
- **[Indicator 9: High-dimensional statistical analysis] Maximum singular value of correlation tensor** The largest singular value of the empirical correlation tensor exceeds the largest singular value of the randomized correlation tensor. The comparison between the daily XRP price and the largest singular value ρ_1^1 shows a strong

negative correlation from October 2, 2017, to March 2018, which includes the high-price period in XRP prices. A correlation coefficient between $\rho_1^1(t)$ and the weekly XRP price was $r = -0.908$ with a p-value of 1.912×10^{-7}, indicating a statistically significant negative correlation. The correlation tensor's largest singular value is an indicator that accurately captures the rapid price increase.

- **[Indicator 10: High-dimensional statistical analysis] Feature extraction of transaction frequency statistics** In the case of Bitcoin, we discovered the structure of three groups of players, namely the users balancing surplus and deficit of cryptoassets (Bal-branch), those accumulating the cryptoassets (In-branch), and those reducing it (Out-branch) in the diagram of flow-weighted frequency. In the case of XRP, the F-frequency $\boldsymbol{A} = (A_{\rm in}, A_{\rm out})$ of the top 400 nodes, which covers nodes active more than 120 day in total, does not show the regime switching among Bal-, In-, Out-branches, as shown in Fig. 24. The additional analysis for the second high-price period might be desired.

[Price Feature Vector]

- **[Indicator 11: Time Series Analysis] The maximum value of the high-low ratio calculated in the week, day, hour, minute window** Detecting anomalous behavior using the Z-score is very sensitive to the time interval selection: 1 week, 1 day, 1 hour, and 1 minute. Therefore, to detect reliable anomalous behavior, we need to analyze not only the z-score for the logarithmic returns of prices but also in conjunction with trading volume.

B. *Future Study*

We will also perform anomaly detection analysis for BTC and ETH, as well as XRP. We will verify that the methods of graph theory, topological geometry, and high-dimensional statistical analysis presented in this study are broadly effective to clarify the transaction feature for all major cryptocurrencies. Furthermore, as shown in the right part; "Step 2 Anomaly Detection AI System" in Fig. 27, we will research concept for estimating comprehensive indices using Boltzmann machines. By conducting these studies, we will establish methods for detecting anomalies in crypto asset transactions, thereby improving the social reliability of cryptocurrencies and contributing to the realization of a new cyber-physical economy.

In this study, we define large price fluctuations as anomalous events, and therefore focus on identifying the senders and transactions that caused these large price fluctuations. On the other hand, financial institutions and exchange market operators identify senders and transactions that are highly likely to be linked to criminal activity in order to report various anomalous transactions to regulatory authorities such as the Financial Services Agency. We believe that systematically investigating the correspondence between senders and transactions identified by financial institutions and exchange market operators and senders and transactions that caused large price fluctuations identified by our anomaly detection AI system has great significance as a crisis management issue. We will continue to conduct research in the future to enable the automation of the anomaly detection process at financial institutions and exchange market operators, and to standardize and improve the quality of reports, as well as to enable the effective use of reports at the Financial Services Agency, through the realization of an anomaly detection AI system with such functions.

Acknowledgements

This study has been conducted as a part of the Project "Dynamics of Price in Crypto Assets and Real Economy and Their Underlying Complex Network" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This work was partially supported by JSPS KAKENHI Grant Numbers 21K03385, 23K11086, 22H05105. It was also partially supported by Ripple Impact Fund 2022-247584 (5855).

The author (Y.I.) is grateful to the following researchers for useful discussions: Prof. Dr. A. Taudes (Vienna University of Economics and Business), Dr. C. Siebenbrunner (Vienna University of Economics and Business), Prof. Dr. S. Thurner (Complexity Science Hub), Dr. B. Haslhofer (Complexity Science Hub), Dr. C. Diem (Complexity Science Hub), Prof. Dr. C. Tessone (University of Zurich), Dr. T. Kim (University of Zurich), Prof. Dr. U. Meyer (Johann Wolfgang Goethe-Universität Frankfurt am Main), and Prof. Dr. K. Ueda (University of Tokyo). The author (T.S.) is grateful to Dr. Taro Hasui (Institute of Mathematics for Industry, Kyushu University) for his generous assistance with the simulations, which greatly contributed to this paper.

Appendix A. Reference examples of suspicious transactions

Financial Services Agency reported the reference examples of suspicious transactions, categorizing them as illustrative cases (Financial-Services-Agency, 2024). The content is summarized below into six categorized items:

- Examples focusing on the patterns of cash usage: Transactions involving large amounts of cash for buying and selling crypto assets, transactions conducted frequently within a short period, and transactions where crypto assets are purchased using a large quantity of small-denomination currency.
- Examples focusing on the potential concealment of the actual account holder: Transactions involving money or crypto assets using accounts suspected to be under fictitious or borrowed names, Transactions using accounts of corporations suspected of having no natural substance, Transactions involving crypto assets using accounts where customers request transaction-related documents to be sent to an address different from the registered one or wish to avoid notifications, Transactions using accounts held by customers found to possess multiple accounts, Transactions involving money or crypto assets by customers who have no apparent reason to conduct face-to-face transactions with the service provider or use the crypto asset automatic exchange machine, Transactions where customers use anonymization techniques when depositing crypto assets into accounts, Transactions accessed from the same IP address despite being conducted by customers with different names and addresses, Transactions where there is no reasonable explanation for the login IP address being located outside the country or the browser language being foreign, even though the customer is a domestic resident, Transactions that make IP address tracking difficult, Account opening transactions where the address obtained during transaction verification differs from the IP address of the computer used, Cases where the same mobile phone number is registered as the contact information for multiple accounts or customers.
- Examples focusing on the usage patterns of accounts: Transactions involving accounts where, after opening the account, a large or frequent deposit and withdrawal of money or crypto assets occurs in a short period, followed by account closure or suspension of

transactions, Transactions involving accounts with frequent deposits and withdrawals of large amounts of crypto assets, Transactions involving accounts that frequently send crypto assets to a large number of addresses, Transactions involving accounts that frequently receive crypto assets from a large number of addresses, Transactions involving accounts that receive funds or crypto assets from names believed to be anonymous or fictitious, Transactions involving accounts that suddenly have large deposits or withdrawals of money or crypto assets, Transactions that appear unnatural in terms of manner or frequency when compared to the purpose of transactions, occupation, or business details confirmed at the time of account opening.

- Examples focusing on the form of transactions: Transactions where a large amount of crypto assets is suddenly bought, sold, or exchanged for other crypto assets, Transactions involving the sale of a large quantity of crypto assets with the condition of cash delivery, Transactions involving a suspiciously large amount of crypto assets that raise doubts about whether the individual owns them, Transactions involving the frequent sale of crypto assets for cash delivery within a short period, Transactions involving customers who attempt to specify a third party's bank account for the deposit of funds or the transfer of proceeds from a sale.
- Examples focusing on transactions with foreign entities: Transactions involving customers based in countries or regions that are non-cooperative in anti-money laundering and counter-terrorist financing measures or are known as sources of illegal drugs.
- Other examples: When a public servant or company employee conducts high-value transactions that do not match their income, When multiple individuals visit at the same time and split the buying or selling of crypto assets so that each amount is just below the threshold requiring transaction verification (as per legal or internal rules), When the same customer visits the same or nearby branches or crypto asset automatic exchange machines several times on the same day or on consecutive days, splitting transactions to keep each below the threshold requiring verification, Transactions where the customer remains uncooperative and transaction verification cannot be completed even though the transaction was conducted before verification was completed, Transactions involving customers who refuse to provide explanations or submit documents when asked to verify the ultimate beneficial owner or the actual controller due to suspicion that they are not acting on their own behalf, Transactions where the ultimate beneficial owner or actual controller of a corporate customer is possibly involved in proceeds of crime, Transactions conducted by internal employees or related parties where the beneficiary of the transaction is unknown, Transactions where there is suspicion that an internal employee has committed crimes under Article 10 (Concealment of Criminal Proceeds, etc.) or Article 11 (Receipt of Criminal Proceeds, etc.) of the Act on Punishment of Organized Crimes and Control of Crime Proceeds, Transactions involving deposits made with counterfeit or stolen currency or stolen crypto assets where there is suspicion that the currency was forged or stolen, or the crypto assets were stolen, Transactions involving customers who unnaturally emphasize the secrecy of the transaction and attempt to prevent reporting, Transactions involving members or associates of organized crime groups, Transactions that appear unnatural in manner or involve customers exhibiting unnatural behavior or attitudes based on the knowledge and experience of staff, Transactions with non-profit organizations where there is no reasonable explanation for the source or ultimate use of the funds, Transactions involving countries, regions, or third parties that have no reasonable relationship

with the activities of the non-profit organization verified at the time of account opening, Transactions with foreign Politically Exposed Persons (PEPs) where there is no reasonable explanation for the purpose of the transaction, Transactions with foreign PEPs where there is no reasonable explanation for the source of wealth or funds used in the transaction, Transactions with foreign PEPs from countries or regions known to have high levels of corruption, Transactions with foreign PEPs based in countries or regions that have not signed or ratified international anti-corruption treaties such as the UN Convention against Corruption or the OECD Anti-Bribery Convention, or are non-cooperative in activities based on these treaties, Transactions where the customer's address used for depositing or withdrawing crypto assets anonymizes part or all of the crypto asset transactions on the blockchain, Transactions with customers who have addresses that receive deposits from or make withdrawals to a large number of addresses on the blockchain, Transactions flagged or inquired about by public agencies or other external entities as potentially involving proceeds of crime.

Appendix B. Collecting Method of Price Time Series

Collecting price information from multiple markets is crucial when discussing anomalous market behaviors. The CryptoCurrency eXchange Trading (CCXT) module in Python is pivotal in this process. As depicted in Figure B.1, the upper panel showcases a sample code used to gather 100 daily data sets of OHLCV for XRP/USD in Bitfinex from January 1, 2020. The lower panel displays the truncated execution results, highlighting the efficiency and reliability of the CCXT module.

If we apply the code shown in the upper panel of Figure B.1 for different market codes and terms, we obtain the bar chart shown in Figure B.2. In this figure, the abscissa represents time, and the ordinate represents the list of market codes. The two red shading ranges correspond to the high-price periods; on the other hand, the blue shaded range corresponds to the stable period.

This paper meticulously focuses on two high-price periods and one stable period, employing data from Bitfinex, Bitstamp, Currencycom, and Exmo. The solid lines in Figure B.3 represent the daily closed values for these four markets. Figure B.3 shows that the data is not stationary. It is evident from this figure that these values align closely, with only a few deviations on certain days, reinforcing the thoroughness and reliability of our analysis. Hence, hereafter, we only use OHLCV in Bitfinex to detect anomalous price behavior. Bitfinex is the most popular in those four markets.

[1]:
```python
import ccxt
import pandas as pd
from datetime import datetime

exchange = getattr(ccxt, 'bitfinex')()
symbol = 'XRP/USD'
timeframe = '1d' #1m, 5m, 15m, 30m, 1h, 3h, 6h, 12h, 1D, 7D, 14D, 1M
since = exchange.parse8601('2020-01-01T00:00:00Z')
limit = 1000
ohlcv = exchange.fetch_ohlcv(symbol, timeframe, since, limit)

df = pd.DataFrame(ohlcv, columns=['timestamp', 'open', 'high', 'low', 'close', 'volume'])
df['timestamp'] = pd.to_datetime(df['timestamp'], unit='ms')
df = df.set_index('timestamp', drop=True)
df
```

[1]:

timestamp	open	high	low	close	volume
2020-01-01	0.19313	0.19658	0.19242	0.19333	7.264490e+06
2020-01-02	0.19314	0.19337	0.18600	0.18817	1.345776e+07
2020-01-03	0.18818	0.19481	0.18519	0.19413	1.338380e+07
2020-01-04	0.19403	0.19464	0.19167	0.19352	3.970285e+06
2020-01-05	0.19392	0.19773	0.19351	0.19468	1.100008e+07
...	...	...	...	...	...
2022-09-22	0.39353	0.50064	0.39353	0.48616	2.339337e+07
2022-09-23	0.48616	0.55970	0.45683	0.50693	3.515667e+07
2022-09-24	0.50782	0.51974	0.47200	0.48807	1.771077e+07
2022-09-25	0.48795	0.52179	0.47562	0.49212	1.332911e+07
2022-09-26	0.49246	0.50245	0.45790	0.46847	1.226704e+07

1000 rows × 5 columns

Figure B.1. Python code

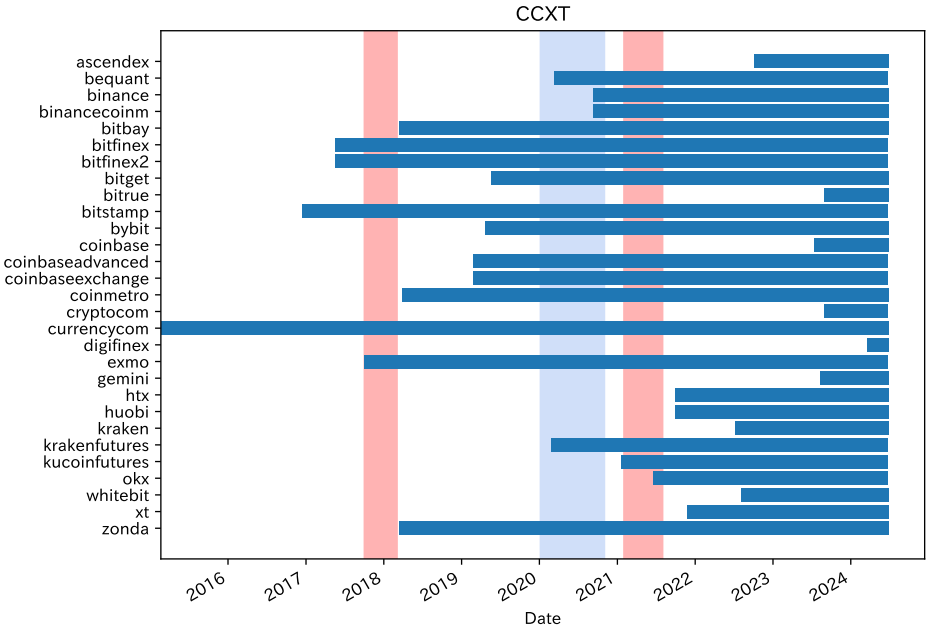

Figure B.2. The daily closed values for Bitfinex, Bitstamp, Currencycom, and Exmo.

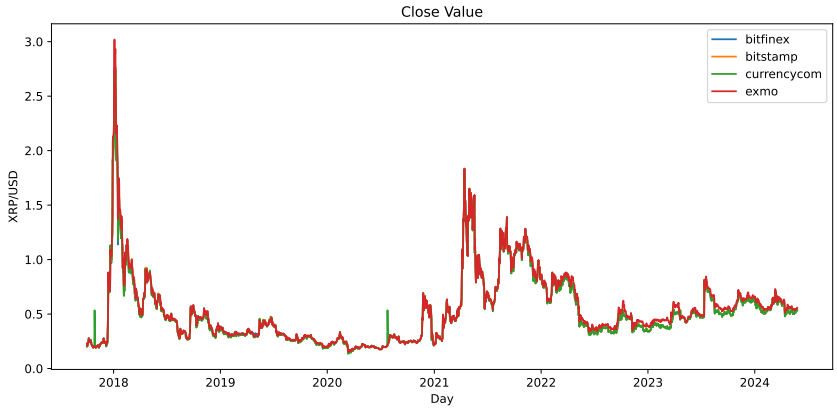

Figure B.3. The timeframe of XRP/USD obtained from several markets using the CCXT module.

References

Aoyama, Hideaki, Yoshi Fujiwara, Yoshimasa Hidaka, and Yuichi Ikeda, "Cryptoasset networks: Flows and regular players in Bitcoin and XRP," *PLoS ONE*, 2022, *17* (8), e0273068.

Bolton, Richard J. and David J. Hand, "Statistical Fraud Detection: A Review," *Statistical Science*, 2002, *17* (3), 235 – 255.

Chakraborty, Abhijit, Tetsuo Hatsuda, and Yuichi Ikeda, "Projecting XRP price burst by correlation tensor spectra of transaction networks," *Sci Rep*, 2023, *13*, 4718.

— , — , and — , "Dynamic relationship between the XRP price and correlation tensor spectra of transaction networks," *Physica A: Statistical Mechanics and its Applications*, 2024, *639*, 129686.

Chandola, Varun, Arindam Banerjee, and Vipin Kumar, "Anomaly Detection: A Survey," Technical Report TR 07-017, Department of Computer Science and Engineering, University of Minnesota August 2007.

— , — , and — , "Anomaly detection: A survey," *ACM Computing Surveys*, 2009, *41* (3), 1 – 58.

Chen, Zhiyuan, Le Dinh Van Khoa, Ee Na Teoh, Amril Nazir, Ettikan Kandasamy Karuppiah, and Kim Sim Lam, "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review," *Knowledge and Information Systems*, 2018, *57*, 245 – 285.

Didimo, Walter, Giuseppe Liotta, Fabrizio Montecchiani, and Pietro Palladino, "An advanced network visualization system for financial crime detection," *2011 IEEE Pacific Visualization Symposium*, 2011, pp. 203–210.

Financial-Services-Agency, "[A Study on Privacy Protection and Traceability in Financial Transactions Using Blockchain] Burokkutyeen wo motiita kinyuu torihiki no praibashi hogo to tsuiseki kanousei ni kansuru tyuosakenkyuu (in Japanese)," 2019.

— , "[Reference Cases of Suspicious Transactions] Utagawashii torihiki no sankou jirei (in Japanese)," 2024.

Fronzetti Colladon, Andrea and Elisa Remondi, "Using social network analysis to prevent money laundering," *Expert Systems with Applications*, 2017, *67*, 49–58.

Fujiwara, Yoshi and Rubaiyat Islam, "Hodge Decomposition of Bitcoin Money Flow," in Lukáš Pichl, Cheoljun Eom, Enrico Scalas, and Taisei Kaizoji, eds., *Advanced Studies of Financial Technologies and Cryptocurrency Markets*, Singapore: Springer, 2020, p. 117–137.

García, Ignacio González and Alfonso Mateos, "Use of Social Network Analysis for Tax Control in Spain," *Hacienda Pública Española / Review of Public Economics*, 2021, *239* (4), 159–197.

Grover, Aditya and Jure Leskovec, "node2vec: Scalable feature learning for networks," in "Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining" 2016, pp. 855–864.

Hilal, Waleed, S. Andrew Gadsden, and John Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," *Expert Systems with Applications*, 2022, *193*, 116429.

Hirosawa, Tatsunori and Tetsutaro Uehara, "[Analysis of fund transfers in Bitcoin mixing] Bittokoin no mikishingu ni okeru shikinidou no bunseki," *IPSJ Technical Report*, 2018, *2018-IOT-41*, pp. 1–8.

HoureiRead, "[Act on Prevention of Transfer of Criminal Proceeds] Hanzai niyoru syuueki no itenboushi ni kansuru houritsu (in Japanese)," 2019.

Huang, Dongxu, Dejun Mu, Libin Yang, and Xiaoyan Cai, "CoDetect: Financial Fraud Detection With Anomaly Feature Detection," *IEEE Access*, 2018, *6*, 19161–19174.

Ikeda, Yuichi, "Characterization of XRP Crypto-Asset Transactions from Networks Scientific Approach," in Yuji Aruka, ed., *Digital Designs for Money, Markets, and Social Dilemmas*, Singapore: Springer Nature, 2022, pp. 203–220.

— **and Abhijit Chakraborty**, "Hodge Decomposition of the Remittance Network on the XRP ledger in the Price Hike of January 2018," *JPS Conf. Proc. , Proceedings of Blockchain Kaigi 2022 (BCK22)*, 2023, *40*, 011004.

Kichikawa, Yuichi, Takashi Iino, Hiroshi Iyetomi, and Hiroyasu Inoue, "Visualization of a directed network with focus on its hierarchy and circularity," *Journal of Computational Social Science*, 2019, *2* (1), 15–23.

Luetgehetmann, Daniel, Dejan Govc, Jason Smith, and Ran Levi, "Computing persistent homology of directed flag complexes," *arXiv preprint arXiv:1906.10458*, 2019.

Luo, Jianxi, Carliss Y Baldwin, Daniel E Whitney, and Christopher L Magee, "The architecture of transaction networks: a comparative analysis of hierarchy in two sectors," *Industrial and Corporate Change*, 2012, *21* (6), 1307–1335.

Masuda, Naoki and Petter Holme, "Detecting sequences of system states in temporal networks," *Scientific Reports*, January 2019, *9* (1), 795.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

Ministry-Of-Finance, "[Teach me! Measures against Money Laundering, Terrorist Financing, and Proliferation Financing] Oshiete! Maneron teroshikinkyouryo kakusankinyuu taisaku (in Japanese)," 2024.

Moreno-Sanchez, Pedro, Muhammad Bilal Zafar, and Aniket Kate, "Listening to Whispers of Ripple: Linking Wallets and Deanonymizing Transactions in the Ripple Network," *Proceedings on Privacy Enhancing Technologies*, 2016, *2016* (4), 436–453.

Moro, Lorenzo and Enrico Prati, "Anomaly detection speed-up by quantum restricted Boltzmann machines.," *Commun Phys*, 2023, *6*, 269.

Noble, Caleb C. and Diane J. Cook, "Graph-based anomaly detection," *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 631 – 636.

Novikova, Evgenia and Igor Kotenko, "Visual Analytics for Detecting Anomalous Activity in Mobile Money Transfer Services," in Stephanie Teufel, Tjoa A. Min, Ilsun You, and Edgar Weippl, eds., *Availability, Reliability, and Security in Information Systems*, Springer International Publishing Cham 2014, pp. 63–78.

Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena, "Deepwalk: Online learning of social representations," in "Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining" 2014, pp. 701–710.

Phua, Clifton, Vincent Cheng-Siong Lee, Kate Smith-Miles, and Ross W. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," *ArXiv*, 2010, *abs/1009.6119.*

Pourhabibi, Tahereh, Kok-Leong Ong, Booi H. Kam, and Yee Ling Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," *Decision Support Systems*, 2020, *133*, 113303.

Ranshous, Stephen, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F. Samatova, "Anomaly detection in dynamic networks: a survey," *WIREs Comput Stat*, 2015, *7*, 223–247.

Sato, Hitomi, Yuichi Kichikawa, Hiroshi Iyetomi, and Tsutomu Watanabe, "Multilayer Network Approach to Dynamics of Japanese Interfirm Transaction Relations," in "Big Data Analysis on Global Community Formation and Isolation: Sustainability and Flow of Commodities, Money, and Humans," Springer, 2021, pp. 63–92.

Shirai, Tomoyuki, "Persistent homology and its application to chainlets in the Bitcoin graph," *JPS Conf. Proc., Proceedings of Blockchain Kaigi 2022 (BCK22)*, 2023, *40*, 011001.

Stein, Jonas, Daniëlle Schuman, Magdalena Benkard, Thomas Holger, Wanja Sajko, Michael Kölle, Jonas Nüßlein, Leo Sünkel, Olivier Salomon, and Claudia Linnhoff-Popien, "Exploring Unsupervised Anomaly Detection with Quantum Boltzmann Machines in Fraud Detection," in "Proceedings of the 16th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART" INSTICC SciTePress 2024, pp. 177–185.

Tauzin, Guillaume, "Flagser Live," 2021. Accessed on Oct 23, 2024.

— , "FlagserPersistence," 2021. Accessed on Oct 23, 2024.

— **, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Wojciech Reise, Anibal Medina-Mardones, Alberto Dassatti, and Kathryn Hess**, "giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration," *arXiv preprint arXiv:2004.02551*, 2020.

Thudumu, Srikanth, Philip Branch, Jiong Jin, and Jugdutt Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," *J Big Data*, 2020, *7*, 42.

Une, Masashi, "[Difficulty of Tracking Transactions and Anonymity in Cryptoassets: Research Trends and Challenges] Angoushisan ni okeru torihiki no tsuisekikonnansei to tokumeisei: kenkyuudoukou to kadai," *[Financial Research] Kinyuu Kenkyuu, Institute for Monetary and Economic Studies, Bank of Japan*, 2018, *2018-J-20*, pp. 1–25.

Yang, Guangyi, Xiaoxing Liu, and Beixin Li, "Anti-money laundering supervision by intelligent algorithm," *Computers Security*, 2023, *132*, 103344.
ΩŠubelj et al.

Šubelj, Lovro, Štefan Furlan, and Marko Bajec, "An expert system for detecting automobile insurance fraud using social network analysis," *Expert Systems with Applications*, 2011, *38* (1), 1039–1052.