ششمین کنفرانس ملی
۱۳ اسفند ۱۳۹۹
محاسبات توزیعی و پردازش داده های بزرگ
6th National Conference on Distributed Computing and Big Data Processing

Collection and Analysis of Financial Exchange Data on the Stellar Blockchain

Reza Hamidpour Badoe
Electrical and Computer Engineering Faculty
Tarbiat Modares University
Tehran, Iran
reza.hamidpour@modares.ac.ir

Sadegh Dorri Nogoorani
Electrical and Computer Engineering Faculty
Tarbiat Modares University
Tehran, Iran
dorri@modares.ac.ir

Abstract— **One of the most important issues that researchers in various fields such as computers and economics need today is the underlying data related to exchange of money (in cryptocurrency form or fiat money). Analysis of such data enables one to evaluate the growth rate of cryptocurrencies or the impact of various rules and regulatory issues on exchange markets. In this paper, we introduce a new data set collected from the Stellar blockchain. Stellar is an open money transfer network which incorporates a decentralized exchange and is based on blockchain technology. One of the main features of this data set is the co-existence of fiat money and cryptocurrency transfers next to each other. In addition, the exact details of offers, bids, and exchanges can be traced.**

Keywords— cryptocurrency exchange data; distributed exchange market data set; Stellar blockchain; big data; financial data set

I. Introduction

With the increasing development of cryptocurrencies and their acceptance in financial exchanges, issues such as structural analysis of blockchain networks, stabilization of conversion rate exchange and also understanding the behavior of users of these networks have already become important. With the advent of cryptocurrencies, a new generation of digital-native exchanges has emerged which revolutionized financial exchanges and caused a new wave of changes in economic markets. Scientists and economic analysts who study this new phenomena, are in urgent need to have access to this real-world financial exchange data.

Blockchain networks have created a foundation on which financial exchanges do not need a third party to verify and review asset transfers. The blockchain-based Stellar money exchange network allows researchers to access real-time and accurate data about financial transactions without relying on a third party. Based on the before mentioned data, economic analysts, portfolio managers of capital markets, as well as government agencies can understand the users' behavior, detect possible frauds, study the effects of changes and regulations on the market, analyze the correlation between changes in a market and the external effects on the society, and predict changes to exchange rates.

In this paper, we introduce our data set which is collected from the Stellar decentralized exchange during a one-year period, and analyze it from multiple aspects. We have made the data set available to anyone who is interested in financial market analysis and design. The rest of the paper is organized as follows: we review the related work in Section II, introduce the data set in Section III, and analyze it in Section IV. Our paper is concluded in Section V.

II. Related Works

In [1], the data sets created by the popular Bitcoin and Ether cryptocurrencies are introduced as large data sets which are applicable in areas such as fraud detection, illegal transactions, and economic critical point detection. The book also provides a basic description of various features that can be obtained from the data set of a cryptocurrency or the general structure of this type of data.

In [2], it has also been pointed out that blockchains have great potential for use in finance, gaming, financial transfers and asset chains. But one of the serious challenges in this area is the behavioral analysis of these networks' users and their categorization based on their behavior. Here, the authors evaluate this data with the transactional sequences of Bitcoin network which are obtained from the general history of this network and also by using their proposed behavioral analysis method.

In [3], authors have collected data sets from the Chinese future market and analyzed the user behavior in this market by using clustering methods. Due to the nature of this market, there is only one asset on each side of every financial transaction, which can be considered as one of the items that have not been considered in [3]. In contrast, different asset types are exchanged in the new generation of economic markets which are based on blockchain networks. For example, in our data set which is collected from the Stellar network, two users may exchange Ether and Bitcoin in one transaction, while two other users exchange Ripple and Tether in another transaction.

In [4], also, data related to different cryptocurrencies in Binance online exchange is collected. In particular, their data set contained Bitcoin, Ether, Litecoin and Ripple trades. Based on the collected data, the authors in [4] tried to understand these markets better; and studied the effects of the rules applied to the price of these cryptocurrencies. Accordingly, some patterns were recognized especially during the end of year 2017 and the beginning of year 2018.

One of the researches in the field of economics and especially in the field of digital cryptocurrencies is [5]. In this paper, the authors used Bitcoin network data to simulate a Bitcoin exchange market, and compared the performance of two sales agents. One agent used a neural networks-based strategy, and the other used a random strategy. The modeling performed was based on a multi-factor method which is suitable for analyzing complex systems affected by multi-factors. This is one of the first multi-factor simulations in the field of digital cryptocurrencies. However, due to the huge growth in the number of cryptocurrencies, it is necessary for the simulated data to be more diverse. This means that the agents must have the ability to exchange multiple assets, and the number of simulated agents should be increased, as well.

Our data set is collected from the Stellar network. In this network, ledgers have the role of storing transactions, such that in each ledger several transactions can be incorporated, and each transaction includes several operations. Examples of these operations are changing a user account, registering an offer to purchase or sell an asset, editing or deleting an offer, and money transfer between users.

According to the structure of the Stellar network transactions, the data of the sale and purchase of assets are divided into two general modes [6]:

- Offer to purchase an asset
- Offer to sell an asset

Any offer to buy or sell by a market agent can be generally converted into a buy or sell offer. If a user intends to buy BTC and in exchange offers Ether (ETH), it can be seen that the user has registered a request to sell ETH in the network. Similarly, if the same user intends to sell BTC again, while in return he/she intends to buy ETH, this time he/she can be considered as an ETH seller. According to this description, without loss of generality, all requests are seen as a sale request.

III. Our Data set

Our data set includes transactions and financial exchanges between Stellar network users. Using blockchain technology, this network creates a new generation of distributed exchange market, in which the users can exchange money with each other without the need for a third party.

We queried the Application Programming Interface (API) provided by Horizon Stellar servers to collect data. Our data set is related to the time period starting from Dec. 1, 2018 and ending on Dec. 30, 2019, and includes 1,650,718 purchases or sale operations organized into 6,680,451 ledgers. Of this total, 1,165,274 operations are new purchase requests, and 891,449 operations are edits or removal of past requests. In total, 163,38 users in our collected data have performed asset exchange operations, 466 assets have been put up for sale and 444 assets have been purchased. Our data set is freely accessible from Github[1].

A. Description of Data Fields

The data fields of each proposal in the Stellar network are as follows:

- **Source_accoun**t: the identifier (ID) of the user who created the offer in the network.
- **Offer_id:** If this field is equal to zero, it means that this offer has just been created by the user and is included in the sales list.
- **Created_at:** the creation time of this proposal.
- **Ledger_id:** the ID of the ledger in which this offer is placed.
- **Transaction_hash:** the transaction hash value in which this operation belongs to. Each offer in the Stellar network is included in a transaction where each transaction is composed of several operations.
- **Selling_asset_type:** the type of asset that the bidder intends to sell. If this value is equal to the native asset type within the Stellar network, the next two fields have no value.
- **Selling_asset_code:** The code of the asset to be sold.
- **Selling_asset_issuer:** The ID of the user who created this asset on the network. Due to the structure of the Stellar network, any asset can be created by a user and transferred through an anchor that exists in the network. In other words, the anchors have the role of confirming the transfers in the network.
- **Amount:** The amount of the asset that the bidder provides for sale.
- **Price_r {n,d}:** The value of the asset being sold versus the asset being purchased. This property consists of two values, n and d. For example, if a user wants to sell 1 BTC for 100 ETH, the value of this field is {1,100}.
- **Price:** Multiplication of the two n properties in the price_r field by the amount field.
- **Buying_asset_type:** The type of the asset that the bidder intends to receive in return for the asset he/she is selling. If the value is equal to the native asset within the Stellar network, the next two fields have no value.
- **Selling_asset_code:** The code of the asset to be sold.

[1] https://github.com/reza-hamidpour/stellar-dataset

- **Buying_asset_code:** The asset code that the bidder wants to receive.
- **Buying_asset_issue:** The ID of the user who created this asset on the network and is used by the bidder.

B. *An Overview of the Data Set*

Fig. 1 shows the number of asset transactions that are in the sales section, and Fig. 2 shows the number of asset transactions that are in the purchase section. Fig. 1 and Fig. 2 show that in the Stellar network, the largest volume of financial transactions is related to the native assets of the Stellar network (Lumens - XLM), Ether (ETH) and Bitcoin (BTC), respectively. In Fig. 1 the vertical axis represents the number of operations in which each asset is in the sales section of operations, and in Fig. 2, the vertical axis indicates the number of operations in which each asset is in the purchase section.

In similar papers that have worked on the financial data of similar markets, the markets have been structured in such a way that the users pay only one type of the listed assets in exchange to sellers. On one hand, it is the exchange of a specific and unique asset, while in the Stellar exchange market, different assets are offered in exchange for a specific asset. For example, a user who intends to buy BTC, can offer US dollars for it, while another user can offer Ripples (XRP) in exchange for BTC.

IV. Analyzing the Data Set

In this section, we focus on Bitcoin (BTC) and Ether (ETH), and provide a better description of their related transactions in our data set. For this reason, we select the operations that at least one side of them is BTC or ETH in the period starting from January 15, 2019 and ending on Dec. 15, 2019. Table 1 shows the number of transactions related to each asset. Fig. 3 and Fig. 4 are the histograms demonstrating the dispersion of the number of users who have used BTC and ETH assets, respectively. The horizontal axis specifies the number of offers and bids that a user has placed on the network, and each bar represents the number of users with the specified number of offers and bids.

According to Fig. 3, Fig. 4, and Table 3, it can be seen that users in the Stellar network are more interested in the ETH cryptocurrency, and ETH is used more than BTC in the exchanges of this network. Moreover, users who have conducted more operations, perform most of their transactions on ETH. An interesting point here is that the users who have done more transactions are far more than the users who have only one or a limited number of transactions in the network and no longer use their account. We can conclude that most users trust the Stellar network for their exchanges.

TABLE I. Number of transactions for each focus asset.

Asset type	Number of offers
Bitcoin (BTC)	169,130
Ether (ETH)	209,727

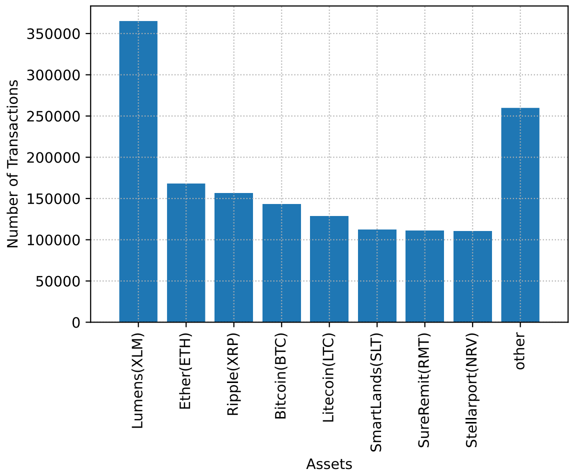

Fig. 1. Number of transactions based on the assets that are in the sales section.

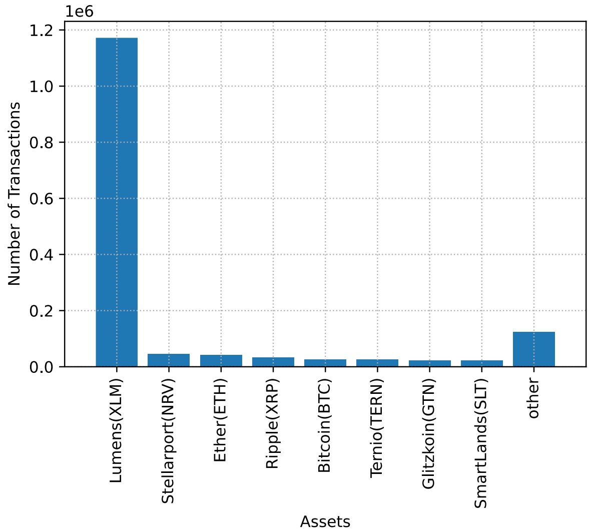

Fig. 2. Number of transactions based on assets in the purchase section.

Fig. 5 demonstrates the number of asset exchange operations between the two assets. By a look at the number of operations by the end of 2019 year, it is evident that with the outbreak of COVID-19, the number of financial exchanges in the network also declines significantly.

Fig. 6 demonstrates another view of the exchanges. In this figure, the volume of asset exchanges per day are depicted. This can be seen that as the last days of 2019 are approached, the volume of exchanges also decreases. It can also be concluded from this figure that despite the fewer number of BTC exchanges in comparison to ETH, the volume of the exchanges are much higher than ETH. One of the main reasons for this can be attributed to the value of BTC compared to ETH.

In fact, it can be said that due to the sharp fluctuations of BTC in this time period, the owners of this asset are not very

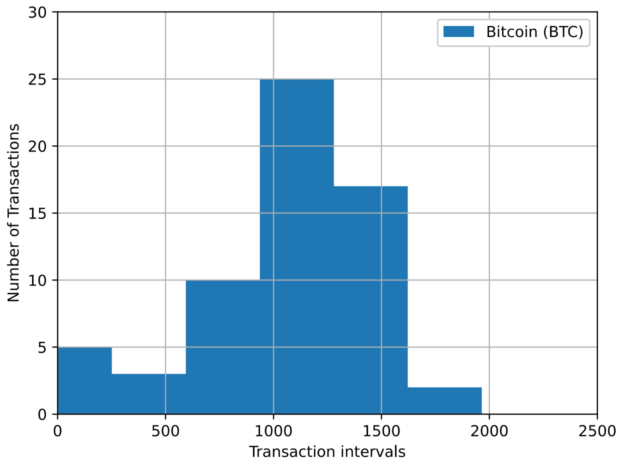

Fig. 3. Histogram of the number of transactions of Bitcoin (BTC) users.

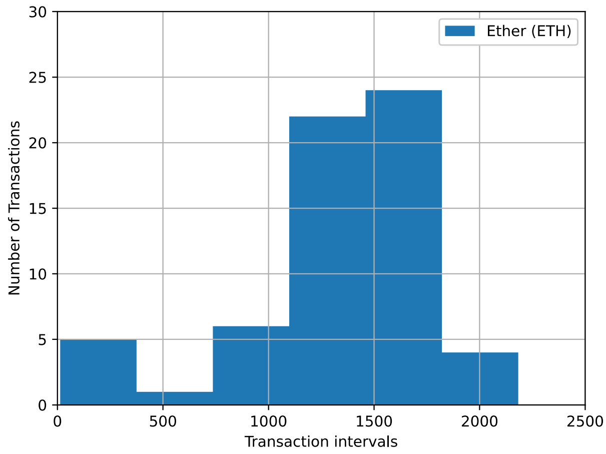

Fig. 4. Histogram of the number of transactions of Ether (ETH) users.

interested in making small amount exchanges on this asset. In other words, BTC is used as a capital asset among the people, and the exchanges on it will be done in high amounts. Despite the fact that ETH assets are somewhat less valuable than BTC and have also seen fewer fluctuations, it can be said that due to the inherent characteristics of this asset, it is more capable for daily usage than BTC.

Fig. 7 depicts the average change in inventory for all users of the network on a daily basis. This value is the difference between the total number of user purchases on a day and the total number of sales on the same day [3]. In Fig. 7, we present the rate of change in inventory of all users who have conducted a BTC or ETH exchange on the network by day. It can be said that in general, the rate of change of this value for the two assets of BTC and ETH is near zero or limited to be between +1 and -1.

Change in inventory is an indicator of the risk taking characteristic of users. In particular, if the amount of user purchase is more than the amount of sales, it can be said that the user is investing in the desired asset, or in other words, the user is taking more risk on that asset. On the reverse side, if he/she is selling an asset, it means that the user is releasing that asset, so the user's risk taking level for that asset is reduced.

V. Conclusion

In this paper, the we introduced our Stellar network exchange data set. Using this data set and the features obtained through it, it can be pointed out that researchers in the field of economics or computer science can use the financial exchange information of this network to simultaneously investigate the financial data of fiat money and cryptocurrencies.

As mentioned earlier, one of the main challenges in existing research in the field of user behavior was that the data diversity of assets was not considered. In addition, we have exact details of offers and bids besides finalized purchases, which needs more investigation, and can be a potential advantage in market analysis. Our data set can be used in simulating new exchange markets, stable cryptocurrency markets, studying user behavior or predicting the conversion rate of cryptocurrencies after changes in market parameters.

References

[1] S. S. Chawathe, "Clustering Blockchain Data," in Clustering Methods for Big Data Analytics: Techniques, Toolboxes and Applications, O. Nasraoui and C.-E. Ben N'Cir, Eds. Cham: Springer International Publishing, 2019, pp. 43–72.

[2] B. Huang, Z. Liu, J. Chen, A. Liu, Q. Liu, and Q. He, "Behavior pattern clustering in blockchain networks," Multimed Tools Appl, vol. 76, no. 19, pp. 20099–20110, Oct. 2017.

[3] G. Shi, L. Ren, Z. Miao, J. Gao, Y. Che, and J. Lu, "Discovering the Trading Pattern of Financial Market Participants: Comparison of Two Co-Clustering Methods," *IEEE Access*, vol. 6, pp. 14431–14438, Feb. 2018.

[4] M. Gidea, D. Goldsmith, Y. Katz, P. Roldan, and Y. Shmalo, "Topological recognition of critical transitions in time series of cryptocurrencies," *Physica A: Statistical Mechanics and its Applications*, p. 123843, Jan. 2020.

[5] L. Cocco, G. Concas, and M. Marchesi, "Using an artificial financial market for studying a cryptocurrency market," J Econ Interact Coord, vol. 12, no. 2, pp. 345–365, Jul. 2017.

[6] Stellar Development Guides, Stellar Foundation, 2020.

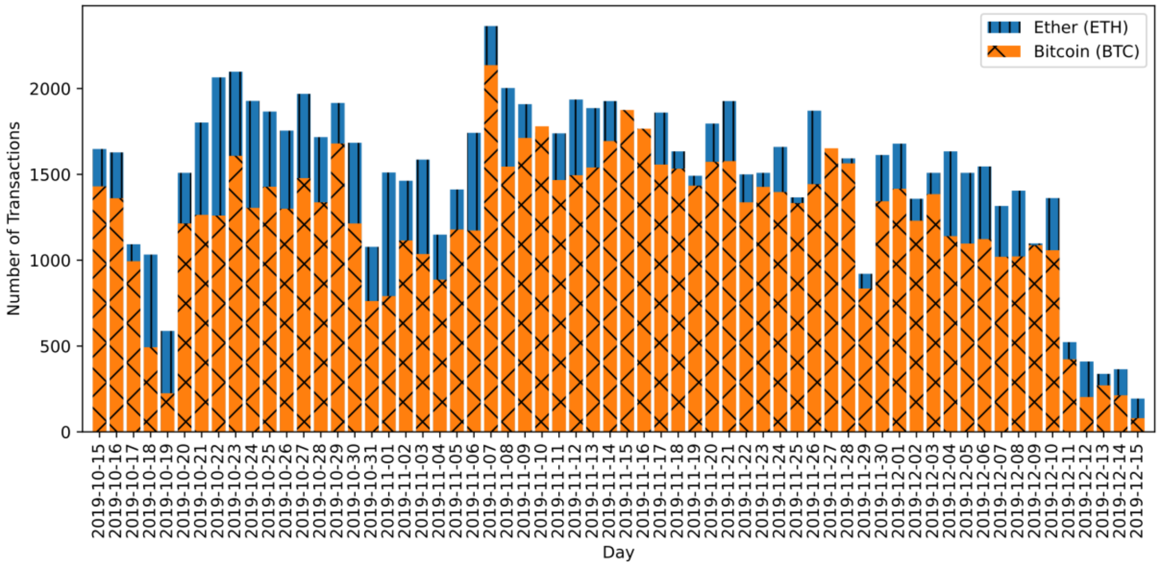

Fig. 5. Number of transactions of Bitcoin (BTC) and Ether (ETH) users on a day-by-day basis.

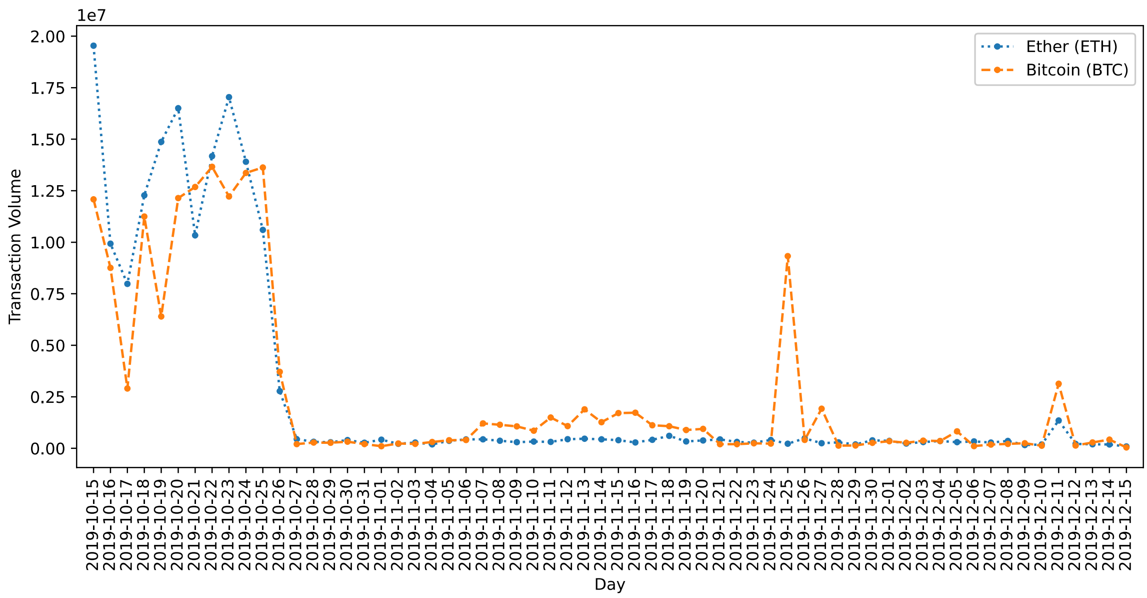

Fig. 6. The volume of transactions of Ether (ETH) users on a day-by-day basis.

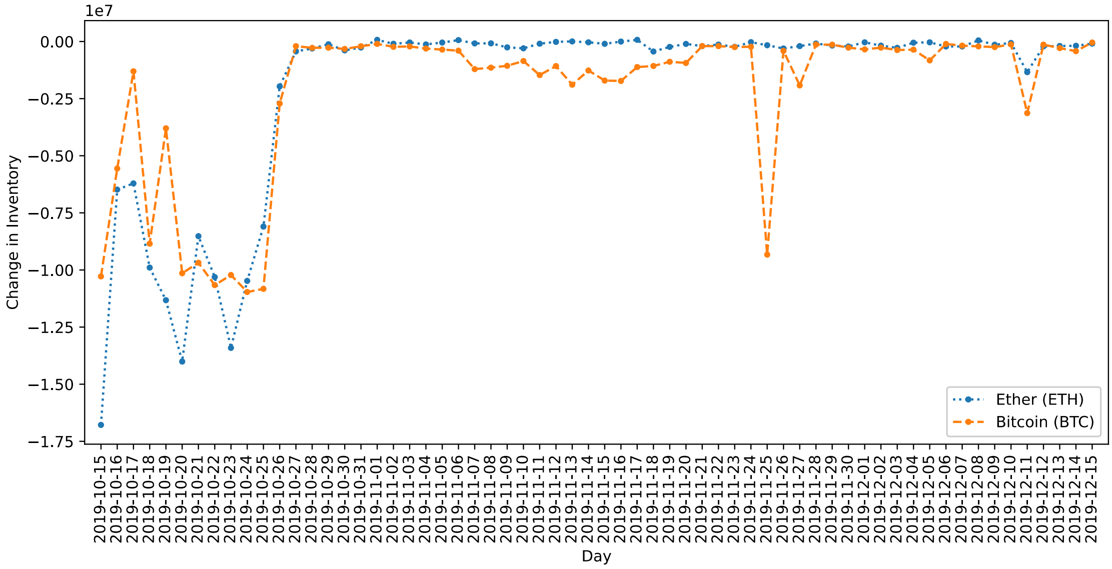

Fig. 7. The amount of changes in the assets of Bitcoin (BTC) and Ether (ETH) users on a day-by-day basis.