UNIVERSITÀ DI PISA

Department of Computer Science

Bachelor Degree in Computer Science

A cryptocurrency exchange services analysis

Supervisors:

Prof. Laura Emilia Ricci

Dr. Damiano Di Francesco Maesa

Presented by:

Luca Santarella

Academic Year 2019/2020

Contents

Chapter 1

Introduction

In recent years the growth of Bitcoin (or more in general of cryptocurrencies) increased substantially in terms of popularity and volume. Bitcoin is a cryptocurrency, which is designed to work as a medium of exchange that uses cryptography to control its creation and management rather then relying on central authorities, as it happens for regular currency as Euro, US Dollars and so on. The original idea was conceived in a 2008 white paper from a mysterious author Satoshi Nakamoto [1]. His vision was to create a new kind of money which took a decentralized approach, a peer-to-peer version of electronic cash that can be minted with computational power and that had a limited supply. In 2017 Bitcoin received a major media exposure due to its exponential growth in price, reaching in December a price of $\sim$\$20.000 USD.

This expansion in adoption and popularity lead to the establishment of different services connected to Bitcoin, which increased in number over the years. The largest and most popular crypto services are undoubtedly cryptocurrency exchanges, this is due to their essential role in the Bitcoin market. Nowadays their massive trading volumes are a consequence of people using Bitcoin as a speculative digital asset, even though it was originally designed as a network of payment by its creator Satoshi Nakamoto. Further, as discussed in chapter 2, whoever seeks to acquire bitcoins for the first time is forced to obtain them through someone who

already owns coins or more conventionally through a crypto exchange. In fact the mining process conceived by Nakamoto to acquire bitcoins is now performed by professional actors with commercial mining pools. Since most crypto exchanges are centralized this means that a new Bitcoin user will have to rely on a centralized entity, in general terms the introduction of these new crypto services brought centralization in the Bitcoin environment, which is against the principle of a peer-to-peer system conceived by Nakamoto.

This work is focused on dishonest practises in the crypto market, such as wash trading and fake volumes. Both of these illegal activities are a kind of market manipulation, which generate a false perception of the crypto exchange in the market. The reason why an exchange engages in these activities is to attract more customers and increase fame and popularity of the exchange, in fact higher trading volumes are perceived as an indicator of a more successful and popular exchange. Another reason behind these fake volumes is the need to attract more cryptocurrency projects who want to get their coin listed, in fact the more popular a crypto exchange is and the higher listing fees it can impose.

In our analysis we used transparent volume data from trusted exchanges as benchmark to find possible odd patterns in other exchanges that are suspected of faking their trading volumes. We also took a sample of four exchanges that were analyzed more in depth, we choose two exchanges that are considered trusted and two suspected of fake volumes. In this way we conducted a larger analysis by using transaction data that helped us discover new insights about correlation with volumes and distribution. In order to gather and collect this data we used different kind of tools. Daily reported volumes were collected using Scrapy, a framework written in Python, that we used for web scraping from websites that act as data aggregator. Data about the crypto exchanges examined were collected using the REST API that they provided. REST API allowed us to have clean and compact data about transactions that we used to make a complete history of every exchange of cryptos made during March 2020.

We evaluated over 20 different exchanges that we considered relevant for this study, we used daily volume data provided by a data aggregator and for 4 of these exchanges we also used transaction data provided by the exchanges APIs, the results obtained revealed interesting pat-

terns and behaviours. Firstly, we discovered that some of the exchanges that were considered suspicious might have reported false volumes, in fact we found out that 5 out of the 13 suspicious exchanges had significant deviations from the general trend. Furthermore, comparing the number of transactions with reported volumes, we noticed a strong discrepancy between the two. This happened for exchanges suspected of fake volumes whereas trusted exchanges showed a strong correlation, this could be a strong indicator of fraudulent printing. Moreover, we also discovered that LBank had an unusual distribution of number of transactions over the quantity exchanged, which appeared linear in contrast with other exchanges. We also analyzed how the volume reported by the exchanges is partitioned through its markets, demonstrating that the Bitcoin market has a clear dominance with a volume of around 66% of the total volume of the exchange. Additionally, we investigated on when and where the transactions take place, we find out that larger exchanges have a more worldwide spread and that transactions had a uniform distributions during the whole day. Finally we investigated on the wash trade practice examining the top hundred exchanges for the month of February 2020. We tried to estimate the real amount of volume for each exchange by using reported monthly volumes (according to CMC) and number of visitors of their websites (according to SimilarWeb). The result is that 83 exchanges out of the 100 examined reported less then 50% of the real estimated volume.

In chapter 2 we describe Bitcoin and discuss the crypto services in its ecosystem, in chapter 3 we formally give a definition of the crypto exchange, how it works and the services that provides, whereas in chapter 4 we explain how data was collected and which tools were used and finally in chapter 5 we discuss the results obtained by analyzing the data retrieved and some possible reasons behind them.

Chapter 2

Cryptocurrencies: the eco systems

The interest in Bitcoin and other digital currencies like Ethereum[1], Tether[2] and Ripple[3] increased over recent years as well as their value and adoption. The spread of this new kind of money and innovative payment networks made room for new entities such as crypto exchanges, miners and mixers which helped shape the current perspective and vision that the world has about the crypto domain.

2.1 Bitcoin

Currently Bitcoin is the main cryptocurrency in the market with a market capitalization of over $170B USD[4] and a price of ∼$9500 USD, but it is well known that the price of Bitcoin is extremely volatile (see figure 2.1 provided by CoinMarketCap). In fact since its origin the price of Bitcoin has increased slowly over the years until 2017, when it reached its peak in fame and value (so far). At the beginning of 2017 the price for a single bitcoin was around $1000 USD but at the end of the same year it reached an astonishing price of $ 14.500 USD, with a growth

[1] https://ethereum.org/whitepaper/

[2] https://tether.to/wp-content/uploads/2016/06/TetherWhitePaper.pdf

[3] https://ripple.com/files/ripple_consensus_whitepaper.pdf

[4] https://coinmarketcap.com/currencies/bitcoin/

of 1300% in its value [2]. Over the next years its price had spikes but a constant increase in popularity and adoption. Bitcoin transactions are made by all different kind of people ranging from those who are hoping to speculate on its ups and downs in the trading market, to shady characters involved in criminal activity or simply newcomers that wants to join an alternative method of payment and money.

Figure 2.1: Rising value of BTC over the years

Bitcoin is a digital currency created in 2009 by Satoshi Nakamoto, a mysterious Japanese author whose identity is still unknown, the concept for this digital money was published in a white paper [1] in 2008, where he discussed the idea of a peer-to-peer version of electronic cash which would allow online payments between two parties without the need of a financial institution like a bank. He proposed an electronic payment system based on cryptographic proofs instead of trust, where the transactions are verified by its protocol's users using asymmetric cryptography. A Bitcoin transaction is a transfer of amount between two or more Bitcoin wallets that will be included in the blockchain. Bitcoin wallets store a private key, which is used to digitally sign transactions and that should be known only by its owner in order to prove the ownership of coins. The signature also prevents the transaction from being altered by anybody once it has been issued. A public key represents the address is used to obtain an address to send and receive payments (see more in figure 2.2).

Transactions are then processed and gathered in blocks, this activity is called **mining**, which consists in verifying the transactions and then adding the block to the blockchain by demon-

strating a proof-of-work, which consists in scanning for a value that when hashed, such as with SHA-256, the hash begins with a certain number of zero bits. This process is rewarded with an amount of bitcoins that is halved every four years. The **blockchain** acts as a public ledger where every transaction created is available for anyone, in this way practices as double-spending are prevented since it is possible for every user to verify that the transaction proposed is possible (meaning that the creator has enough bitcoins) by accessing the record of public history of transactions.

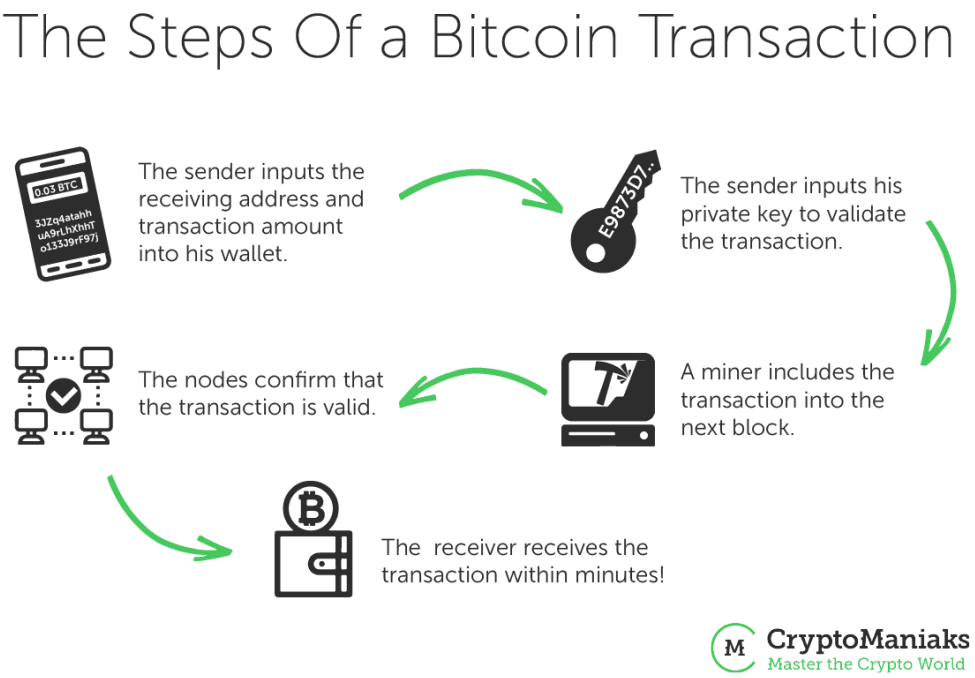


Figure 2.2: A scheme of a Bitcoin transaction (image provided by cryptomaniaks.com).

Nakamoto proposed that the network should run as follows:

1. New transactions are sent to every node in the network in broadcast.
2. Each node collects new transactions into a block.
3. Each node works on finding a proof-of-work for its block, which is deliberately difficult.
4. When a node finds a proof-of-work, it broadcasts the block to all nodes, the node receives a reward for finding the proof-of-work.

5. Nodes accept the block only if all transactions in it are valid and not already spent.

6. Nodes express their acceptance of the block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash.

Bitcoin Core is a free open-source software that manages bitcoins, it allows users to make transactions and provides a Bitcoin wallet which fully verifies payments. Although payments are publicly available to anyone in the network addresses are user alias that are not directly linked to their users information (pseudo-anonymity), also if the private key is lost or stolen that means that bitcoins in that wallet cannot be retrieved in any way, since the private key represents the ownership of the bitcoins.

2.2 Crypto services

Another critical aspect that should be accounted for when discussing Bitcoin are the entities that revolve around it, such as:

- **Miners**

  According to the Bitcoin protocol, there will be only a finite amount of 21 million bitcoins. The only way to obtain bitcoins is to buy them with fiat currency from other users using a crypto exchange (see more in chapter 3) or try to gain the reward for adding a new block, thus participating in the activity of processing transactions, which is known as Bitcoin mining and users who participate in mining as Bitcoin miners. The Bitcoin protocol is designed so that new bitcoins are mined around every ten minutes until all bitcoins are mined around 2140. The rising price of Bitcoin motivates miners to invest on more and more powerful hardware for faster mining. Due to the dramatic growth in both the number of Bitcoin miners and the computation power of their hardware, it has become impossible for a single user to mine Bitcoins on his own. So the majority of miners choose to pool their computation resources to mine Bitcoins together, the so-called

pool mining, in fact pool mining gives individual miners steadier payouts than solo mining. Over time the system became dominated by multi-graphics card systems, then Field-Programmable Gate Arrays (FPGAs) and finally Application-Specific Integrated Circuits (ASICs) in the attempt to find more hashes with less electrical power usage [3].

- **Escrow services**

  When an exchange of physical goods for cryptocurrency happen between two parties that do not trust each other some difficulties may arise. In fact both buyers and sellers wish to perform an atomic exchange, where goods and currency will be exchanged or neither will, yet this situation cannot be guaranteed with a conventional Bitcoin payment as bitcoin transactions are irreversible. To solve this issue buyers rely on a trusted third party which will escrow the funds and transfer them to the seller once the goods arrive. This model has been used for online marketplaces such as the Silk Road [4] and many of its successors but also by many legitimate businesses. Despite the advantages that brings, this method of transaction has two major flaws, it requires that the entity which provides the escrow service is involved in every transaction, even when there is no dispute and also the third party that manages the fund during the transaction may simply pocket the buyer's payment and never transfer it to the seller [5].

  Multisig represents another possible way to create 2-out-3 escrow services without the risk of incurring in a fraud. The intermediary does not manage the funds of the transaction but simply acts as an arbitrator in a possible dispute between buyer and seller. Multisignature (often called multisig) is a form of technology used to add additional security and for Bitcoin transactions, when using a multisignature address it is required that an additional user signs a transaction before it can be broadcast onto the block chain. In this way when Alice wants to pay Bob for a service or goods, she sends a transaction to a multisignature address, which requires at least two signatures from the group "Alice, Bob and Trent" to redeem the money. If Alice and Bob have a dispute, for example Alice wants a refund whereas Bob claim to have fullfilled the order, they can appeal to Trent, who grants his signature to Alice or Bob, so one of them can redeem the funds (see more in figure 2.3).

Figure 2.3: A scheme of a multisig transaction (image provided by l2b.global).

- **Online wallets**

  Many operators offer a wallet service whose purpose is to store private keys and interact with the blockchain making transactions to other users, in many cases these wallet services provide a wallet for more than one cryptocurrency and are free of charge or have an annual fee (see more about wallets in section 3.2).

- **Mixing services**

  A cryptocurrency mixing service (also known as cryptocurrency tumbler) is a service that mix identifiable funds (also referred as "tainted") with other unidentifiable funds in order to ensure private transactions without trace. Mixers will request a standard fee for the process, which consists in submitting the bitcoins you want to "disguise" and then the mixer service will send back random bitcoins gained from a random address to you, this is made by using other user's funds, without the need of any personal information. Since this service allows to obscure the origin, possession and movement of cryptocurrency it resembles traditional money laundering strategies but in a context quite different. As a matter of fact tumblers can protect privacy and keep transactions discreet but can also pose a major risk in covering up activities such as organized crime and terrorism [6].

2.3 Related work

The attention on these exchanges that act as intermediaries between buyers and sellers is growing over the years and it is becoming object of various studies that highlight prominent features and malicious practices that needs to be addressed. Crypto exchanges play a fundamental role in the crypto market because they contribute to the price discovery process, i.e. setting the proper price of an asset, which is determined by various factors such as supply and demand, availability of the asset and many more. In a 2015 article by Brandvold et al. [7], they investigated the role of different exchanges in the price discovery process of Bitcoin. In this article they identified that some exchanges which had high trading volumes acted as leaders, so a higher fraction of the price discovery occurred in these exchanges. On the other hand, the rest of the exchanges were classified as followers, since they correlated more with lagged market returns.

Since digital assets are quite often considered a financial lucrative business, people can be deceived by fraudsters who could use social engineering techniques such as phishing and trust-trading scams to steal users' fund, leading to huge amounts of financial loss as seen in a 2020 article by Xia et al. [8]. In this work they analyzed as many scams as possible, they investigated on existing known scams and then developed an automated approach to identify both well-known scams and scams that have not been public yet. The results revealed 1595 scam domains, over 60% of them were not publicly known and identified over 300 fake exchange apps. Another major risk is represented by crypto exchange closure, which has been explored in a 2018 article by T. Moore et al. [9] where they discovered that Bitcoin exchange closure rate is very high, 38 out of 80 exchanges analyzed faced closure and 26 experienced at least one security breach. It has also been discovered that experiencing a breach in a given quarter is strongly correlated with the exchange closing that same quarter and that an exchange's trading volume is positively correlated with its continued operation. Despite the interesting results achieved, in this article no correlation was found between the exchange lifetime and the presence of security controls, e.g. two-factor authentication. Krückeberg and Scholz (2019) [10] have also found out that increased spreads are more likely to occur during particular times such as early

UTC hours and after bitcoin heists and hacks, thus presenting arbitrage opportunities and proving that Bitcoin market inefficiency increases over time. Another great effort in "Making Bitcoin Exchanges Transparent" comes from a 2015 publication of the same name, made by Decker et al. [11]. In their work they discussed on a possible solution by proposing a software based audit[5] of Bitcoin exchanges without revealing information that could be of strategic importance or deemed sensible. A proof-of-concept is proposed which explains how to achieve this automatic digit, which is based on an architecture that splits the computation into individual pieces to overcome memory limitations. In this way, it could be possible to determine the solvency of Bitcoin exchanges without revealing any private data and also replace the human financial auditor. Finally we introduce the work of Yue et al. [12] that proposed a software that is capable to visualize transactions between exchanges in a hyper-graph, which helped discover new patterns and gather intelligence about the relationship between exchanges. As a matter of fact, they discovered a possible correlation between the 2016 Bitfinex hack, which resulted in approximately $70 millions USD worth of Bitcoins being stolen and the adoption of the BitGo custodian service to accelerate their customers transactions.

In this thesis we focused more on the wash trading and fake trading volume reports, discussing on how they take place, their effects on the general market and moreover the recognizable patterns that occur. In our work we took an approach similar to the work of Cong (2019) [13] which revealed that unregulated crypto exchanges are highly likely involved in wash trading and that most of them showed anomalous activities compared to the regulated ones. They further showed that wash trading is a dangerous activity since it can distort the attention of investors, the exchanges prominence and subsequent trading activities. Additionally, we also considered the report presented to the U.S. Securities and Exchange Commission by Bitwise (2019) [14] on wash trading. This report revealed that an astonishing "95% of reported trading volume in bitcoin is fake or non-economic in nature" and explored why and how wash trading occurs. They analyzed trade size histograms, volume spike alignment and spread patterns of the exchanges examined, showing that there are only 10 exchanges with 100% real trading volumes which accounted for 5% of the collectively reported trading volume.

[5] an audit is an independent evaluation made by a third party to inspect financial records.

Chapter 3

Cryptocurrency exchanges

3.1 The anatomy of crypto exchanges

Cryptocurrency exchanges or Digital Currency Exchanges (DCE) are online platforms which allow to buy and sell most major cryptocurrencies (also referred as digital/cyber/virtual currencies).

They can be a brick-and-mortar business, that is a company which has an actual physical workplace or they can be a completely online business. These exchanges allow their users to trade digital currencies with fiat[1] currency (and vice versa) or with other cryptocurrencies, applying a charging fee for each transaction. They represent the main "middle man" that allows users to trade the major cryptocurrencies.

A crypto exchange is either a matching platform, which charges a simple fee or a market maker, i.e. "any intermediary who creates a market for a financial obligation" [15], that is a company that offers a buy and sell price for a financial instrument or commodity held in inventory, hoping to make a profit on the bid-offer spread. "Its basic function is to service the public's demand to trade with immediacy by continuously standing ready to buy shares

[1] currency without intrinsic value that has been established as money, often by government regulation.

from customers who wish to sell, and to sell shares to customers who wish to buy. Additionally, the market maker helps to stabilize prices and to facilitate a reasonably accurate price discovery" [16].

Generally speaking, a cryptocurrency exchange establishes a common ground between buyers and sellers (also referred as makers[2] and takers[3] [17] in case of matching platforms) providing the necessary availability. Its main goal is to encourage a substantial number of transactions, maintaining a certain difference between ask and bid value, called ask-bid spread on which they profit. The increase of difference between ask and offer is strictly proportional to the gain generated on each transaction. For instance if the bid price for an exchange for a given cryptocurrency X is hundred units and the offer price is 120 units, the difference between the two will be 20 units. Therefore the exchange will offer to buy the asset for hundred units and sell it back to 120 units, scoring a profit of 20 units, which represents exactly the bid-ask difference that it offers to its users. An exchange must provide a significant amount of liquidity in order to satisfy the great demand of the market. It also provides other services such as lending, trading and wallet services. One of the main goals of a crypto exchange is to keep the market active, by actions of compensation (speculation of the asset value) which makes it more volatile, as a consequence the market will be more fluid allowing a faster and easier exchange of cryptocurrencies.

A crypto exchange is quite similar to a foreign exchange market (also known as forex), which is divided in **spot market**, that is a "market that deals in commodities or foreign exchange for immediate delivery" [15] this is the traditional way to buy or sell a digital currency and **derivatives market**, in which two parties agree on a contract called derivative. A derivative is a financial instrument, the price of which has a strong relationship with an underlying commodity, (crypto)currency, economic variable or financial instrument. A future contract is the most common type of derivative, it is "an agreement to buy or sell a fixed quantity of a

[2] a "maker" is whoever places an order, and it does not trade immediately, so the order stays in the order book and waits for someone else to fill/match with it later.

[3] the "taker" is someone who decides to place an order that is instantly matched with an existing order on the order book.

particular commodity, currency, or a security for delivery at a fixed date in the future at a fixed price" [15], it is often used to speculate on the value of cryptocurrencies on the long term in contrast of spot transactions that are short-term trades.

3.2 Crypto exchange services

In order to start trading cryptocurrencies it is crucial to choose the right crypto exchange. There are several aspects to keep in mind, here are listed some of the criteria that users should know before start buying or selling:

1. **Reputation**

   Since many exchanges are often involved in shady, suspicious or even criminal activities, it is important to inquire about them. One way to go is to search for as much information as possible, for example: possible history of frauds or scams, general policy, terms and conditions listed on their website, social media and general complaints issued by other users.

2. **Number of tradable assets**

   The most famous cryptocurrencies such as Bitcoin, Ethereum and Ripple are generally always listed, but large crypto exchanges often list many more cryptocurrencies or digital tokens that users may be interested in.

3. **Fiat exchange**

   Traders may need to start buying digital currencies somewhere or cash out profits, so checking if the exchange has the option to buy with fiat currencies and which method of payments are accepted (bank transfers or debit/credit card) is one of the key aspects.

4. **Fees**

   Every exchange has a different fee plan, most of them charge a small fee for every transaction and they may also charge for other operations such as withdrawal and deposit.

5. **Trading volume**

   Another criterion that has to be considered when trading or investing is the trading volume reported by the exchange, this value gives a rough idea of the size of an exchange and its liquidity. It is possible to get these trading volumes by accessing the order book for each asset, alternatively, it is possible to rely on websites such as CoinMarketCap (CMC) [4] and CoinGecko[5] that keep track of the reported volume for every major cryptocurrency exchange in the market. These values, however, often do not represent the real traded volume of the exchange since many exchanges tend to inflate the reported traded volume, engaging in practices as wash trading, which is a kind of market manipulation (see more in section 5.3) and reports of inflated or fake volumes (see more in chapter 5).

6. **Security Measures**

   Most exchanges have to comply with laws and regulations which require to implement KYC (Know Your Customer) and AML (Anti-Money Laundering) procedures in order to prevent illegal activities. Such procedures require the user to provide personal information, without this data many exchanges limit the amount of trading or do not allow withdrawals. Furthermore, crypto exchanges also implement security measures to protect the users such as: 2FA (Two-Factor Authentication), OTP (One-Time Password) and general warnings in order to prevent hacking, phishing or theft of funds.

In order to buy and sell cryptocurrency it is required to have a crypto wallet, i.e. a tool that stores private and public keys that you can use to interact with a blockchain network. There are various types of crypto wallets which can be divided into three groups: software, hardware, and paper wallets. A crypto wallet may be referred as a hot wallet if it is connected to the Internet or a cold wallet otherwise, generally a cold wallet is used to store large amounts which are not traded frequently, this type of wallet is more secure but also more expensive. A hot wallet is usually used as main interaction between exchanges and external parties, this type of wallet is free and user-friendly but less secure since it is more exposed to cyber attacks.

[4] https://coinmarketcap.com/rankings/exchanges/reported/

[5] https://www.coingecko.com/en/exchanges

The majority of crypto wallet providers offer a software solution, which makes their use more practical than hardware wallets, however since they are usually used as cold storage, hardware wallets tend to be the most secure alternative. Paper wallets, on the other hand, consist of a "wallet" printed out on a piece of paper, but their use is now deemed as obsolete and unreliable[18].

3.3 Security and regulation

As far as security is concerned most crypto exchanges claim to handle and store users' assets safely and to posses high standard defenses systems against hackers, but as a matter of fact cyber-security it's still today a great issue. Over the years numerous exchanges reported to have suffered from thefts and mishandling of their funds, these two represent a major risk of loss for traders. Cyber-attacks involving crypto exchanges are typically phishing, scamming, supply chain attacks[6] and hacking, therefore there is a substantial worry in dealing with the accountability deficits relating to all cryptocurrencies and two examples of this pertain to (1) theft and (2) exchange shutdowns.

One of the most infamous case of shutdown is MtGox, which in 2014 was the largest crypto exchange (handling approximately 70% of all Bitcoin transactions) that suffered from security breach. At that time, Mt. Gox announced that approximately 850,000 bitcoins (valued at the time at about $460 million USD) in addition to $27 million USD in bank deposits were missing, since then, some 200,00 bitcoins have been located, but exactly what happened has never been fully disclosed [19]. In August 2016, the major Hong-Kong-based exchange Bitfinex reported that hackers stole the equivalent of $72 million USD worth of bitcoin [20]. Moreover in July 2019 a cryptocurrency exchange in Tokyo has halted its services after it lost $32 million USD. Remixpoint, which runs the Bitpoint Japan exchange, announced [21] that about ¥3.5bn in various digital currencies were missing. The apparent hack emerged after an error appeared

[6] a supply chain attack involves physically tampering with electronics (computers, ATMs, power systems, factory data networks) in order to install undetectable malware for the purpose of bringing harm to a player further down the supply chain network.

in the outgoing funds transfer system, it said that the cryptocurrency went missing from a hot wallet, but that currency held in cold wallets that are offline was not affected [22].

In some countries a crypto exchange needs to posses a license in order to operate, the regulation of these agents differs from state to state. For instance in Italy crypto exchanges are required to comply with the European directive on the Prevention of the Use of the Financial System for the Purposes of Money Laundering or Terrorist Financing, as stated in the Legislative Decree n. 90 of May 25 2017[7]. Virtual currency providers are subjected to the regulations established for traditional money exchange operators, also in the Legislative Decree n. 125 of October 4 2019[8], it is stated that the definition of virtual currency providers is extended introducing the classification of virtual currency wallet providers. Furthermore, for these providers of wallet services, it is required to be enlisted in a special section managed by the "Organismo degli Agenti in attività finanziaria e dei mediatori creditizi" (Oam), similarly, to what it is already requested for virtual currency operators. Regarding the European vision on crypto currencies, on July 5 2016, the European Commission presented a legislative proposal to amend the Fourth Anti-Money Laundering Directive (AMLD). It suggested, bringing custodian wallet providers and virtual currency exchange platforms within the scope of the AMLD, meaning they would be obligated to fulfill due diligence[9] requirements and have in place policies and procedures to detect, prevent, and report money laundering and terrorist financing [23].

3.4 Decentralized crypto exchanges

Until now, we talked about the most common type of exchange, the centralized type, meaning that a central authority acts as a third party in a transaction involving a buyer and a seller, in

[7]Decreto Legislativo 25 maggio 2017, n. 90 Attuazione della Direttiva (UE) 2015/849 Relativa alla Prevenzione dell'Uso del Sistema Finanziario a Scopo di Riciclaggio dei Proventi di Attivita Criminose e di Finanziamento del Terrorismo, https://www.gazzettaufficiale.it/eli/id/2017/06/19/17G00104/sg

[8]Decreto Legislativo 4 Ottobre 2019, https://www.gazzettaufficiale.it/eli/id/2019/10/26/19G00131/sg

[9]the investigation that a reasonable business is expected to take before entering into an agreement or contract.

this case both parties trust a "middle man" to handle their assets. This approach can lead to potential cyber attacks [24] or theft of funds [25], as already discussed in section 3.3, that can potentially hurt the users of the exchange, as matter of fact this lack of security is considered one of the major flaws in the centralized approach. Another critical aspect is the total absence of anonymity of the transactions, which beats the decentralized nature of cryptocurrency, most CEXs (Centralized EXchanges) has to guarantee total transparency since they are required to implement KYC (Know Your Customer) procedures. These issues motivated the development of decentralized exchanges, the term refers to "distributed ledger protocols and applications that enable users to transact cryptocurrencies without the need to trust a centralized entity to be an intermediary for the trade or a custodian for their cryptocurrencies". They provide a number of important benefits that make them more suitable compared to standard CEXs, such as lower correspondent risk (no need to trust a centralized exchange to secure and manage private keys) and the potential for lower transaction fees [26].

In conclusion DEXs (Decentralized EXchange) are still less popular, less widespread and they lack the great liquidity that CEXs often provide. Furthermore it is not possible to trade fiat to cryptocurrency (or vice versa), as a matter of fact their higher trade latency, lower liquidity, and less intuitive user interfaces make them less attractive for mainstream retail users.

3.5 Summary

The table 3.1 shows a list of some of the many crypto exchanges in the market that will be analyzed in chapter 5, for each one a series of features is listed. As shown in the list, almost every exchange is centralized and provides a wallet service and implements KYC procedures, while not many of them provide a service of margin trading, i.e. "the buying of securities on credit in order to make short-term gains" [15]. This method of trading, which is also known as trading with leverage, is an agreement that involves two parties, where the trader borrows funds from the exchange or other users in return of a small interest, this is done to amplify the trading results. Only two crypto exchanges are based in USA, which is one of the countries

with the most strict laws regarding crypto currencies and exchanges, most of them are based in countries which are considered to be “tax havens” or where the regulation of crypto exchanges is not well defined [27] [23]. The fees listed are the trading fees (they can vary from maker to taker and fidelity programs), for many of them there also deposit and withdrawal fees that are not listed in the table.

	KYC	OFFICE	WALLET	MARGIN	TYPE	YEAR	PAIRS	FEES
Coinbase	✓	U.S.A.	✓	✗	CEX	2012	58	0% - 0.5%
Kraken	✓	U.S.A.	✓	✓	CEX	2011	140	0% - 0.26%
Binance	✓	Malta	✓	✓	CEX	2017	611	0.02% - 0.1%
Bitfinex	✓	B.V.I.	✓	✓	CEX	2012	277	0.1% - 0.2%
LBank	✓	China	✓	✗	CEX	2016	98	0.05% - 0.2%
TAGZ	✓	Australia	✓	✓	CEX	2019	33	0.01%
HitBTC	✓	Chile	✓	✗	CEX	2013	830	0.02% - 0.07%
Coinsbit	✓	Estonia	✓	✗	CEX	2018	149	0.2%
IDEX	✓	Panama	✓	✗	DEX	2019	135	0.1% - 0.2%
Changehero	✗	N/A	✗	✗	DEX	2018	23	0.5%

Table 3.1: A summary of the characteristics analyzed.

Chapter 4

Data collection from crypto exchanges

The majority of crypto exchanges provide an Application Programming Interface (API) that allow users to automate certain operations such as trading, deposits and withdrawals. They offer an API documentation which states the endpoint where the HTTP (HyperText Transfer Protocol) request should be sent and the data that can be queried.

Usually the API offers two types of data, **private user data** which involves user information such as balance, orders, history and reports of the account. In order to access this data an account needs to be created and then an API key and an API signature are required, these two information must be included in the HTTP header whenever making a request. On the other hand **public market data** consists in public data of an exchange which is accessible by anyone, some of its features are:[1]

- **Tickers**

  A symbol ticker is a unique abbreviation of a pair of digital currencies[2], this symbol is used to retrieve the price of the asset pair. This value is updated continuously throughout the day, other information related to the asset pair is often shown such as traded volume, asks, bids, last trade and opening price.

[1] in this list the exchange Kraken was used as an example (https://www.kraken.com/features/api).

[2] e.g. the pair Bitcoin-Ethereum is represented as BTC-ETH.

- **OHLC Data**

  Open High Low Close (OHLC) data shows how the price of an asset pair fluctuates over time, it is used to make a candlestick chart, just like the one in figure 4.1.

Figure 4.1: A OHLC chart of the asset pair BTC-USD (Kraken).

The horizontal axis represents the time, (the granularity chosen is 24 hours, so every candlestick represents one day) and the vertical axis represents the price of the asset.

The following price points are needed to create each candlestick (see figure 4.2): **Open** (1), - the first recorded trading price of the asset, **High** (2), - the highest recorded trading price of the asset, **Low** (3), - the lowest recorded trading price of the asset and **Close** (4), - the last recorded trading price of the asset. Each price point is the price within that particular time frame (in figure 4.1 is 24 hours).

Collectively, this data set is often referred as the OHLC values. The relationship between the open, high, low, and close determines how the candlestick looks. The distance between the open and close is referred to as the body, while the distance between the body and the high/low is referred to as the wick or shadow. The distance between the high and low of the candle is called the range of the candlestick. If the body is green, it means

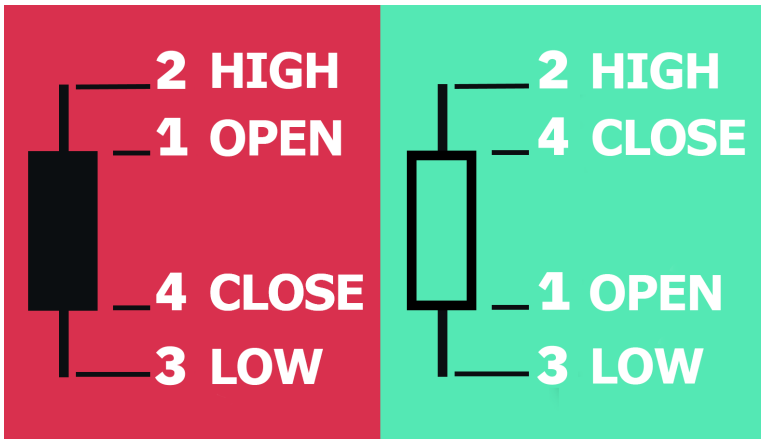


Figure 4.2: A candlestick which contains the price points (image from Binance Academy).

that the asset closed higher than it opened, whereas red means that the price moved down during the measured time frame, so the close was lower than the open [28].

- **Order Book**

  Every ask and bid offer is listed in the order book (see figure 4.3), which changes continuously throughout the day, every user can access this record and have a rough idea of the price of the asset pair. Each entry is made up by type (buy or sell), price and volume.

Order Book (XBT/EUR)

Buying		Selling	
Volume	Price	Price	Volume
0.042	€7,112.4	€7,116.3	0.660
0.048	€7,112.1	€7,116.4	0.556
0.690	€7,112.0	€7,116.5	3.880
0.002	€7,111.9	€7,116.8	0.490
0.241	€7,111.7	€7,116.9	4.000
0.010	€7,111.6	€7,117.0	2.000

Figure 4.3: The order book containing the ask and bid offers for BTC-EUR (Kraken).

- **Recent Trades**

  A list of the most recent trades for a certain pair (see figure 4.4), most exchanges do not provide historical data but just the last 500 - 1000 transactions. In each transaction are listed timestamp, price, volume and type of transaction (either buy or sell).

Recent Trades (XBT/EUR)

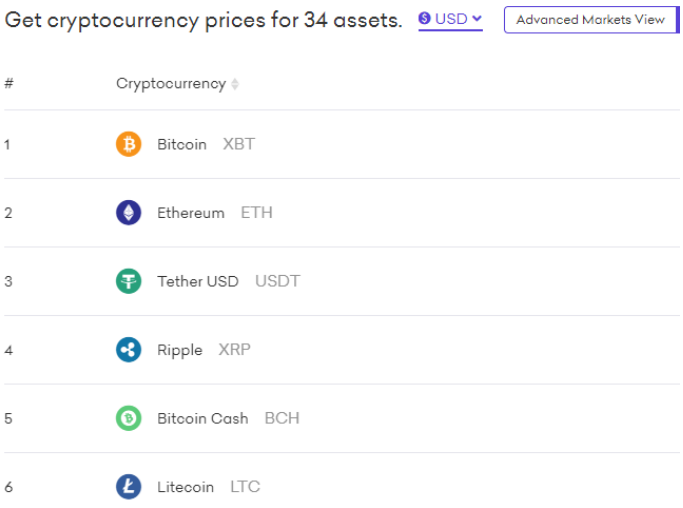
1 - 15 of 250 trades

Time	Order	Price	Volume
15:02:10 +00:00	buy	€7,124.9	0.02105144
15:02:08 +00:00	buy	€7,124.9	0.29378985
15:02:08 +00:00	buy	€7,124.9	0.25000000
15:02:08 +00:00	buy	€7,124.9	0.15615453
15:02:07 +00:00	buy	€7,124.9	0.33384547
15:02:05 +00:00	buy	€7,124.9	0.47589869

Figure 4.4: The most recent trades for BTC-EUR (Kraken).

- **Tradable Pairs**

  A list of pairs of digital currencies that the platform currently supports.



Get cryptocurrency prices for 34 assets. USD Advanced Markets View

#	Cryptocurrency
1	Bitcoin XBT
2	Ethereum ETH
3	Tether USD USDT
4	Ripple XRP
5	Bitcoin Cash BCH
6	Litecoin LTC

Figure 4.5: Some of the cryptocurrencies supported by Kraken

- **Spread Data** shows the bid-ask spread trend for a pair throughout the time (see figure 4.6), the horizontal axis represents time and the vertical axis represents the price of the asset pair. The black line represent the price of the asks and the red one of the bids, at the bottom the spread value.

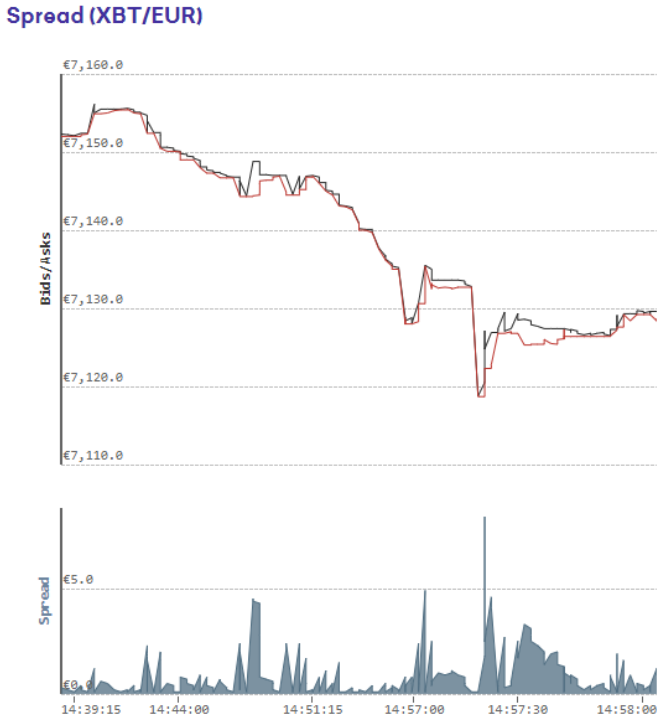


Figure 4.6: Spread chart for the asset pair BTC-EUR (Kraken)

In our analysis we collected data using APIs from four different crypto exchanges, two exchanges that are considered trustworthy and well established, which are Kraken and Binance[3] and two exchanges that are suspected to report a false amount of trading volume, which are HitBTC[4] [29] and LBank[5] [30]. In order to make a better comparison between them, Kraken and HitBTC were selected to represent a medium sized[6] exchange and Binance and LBank to represent a large sized exchange. The data collected involves the trades and daily volumes for each pair listed on the exchanges during the whole month of March 2020. In table 4.1 are listed

[3] https://www.binance.com/en

[4] https://hitbtc.com/

[5] https://www.lbank.info/

[6] size in terms of reported volume and number of trades.

the features of each exchange; the number of volume and trades reported are an average for the month, the data for the website visits are provided by SimilarWeb[7].

DCE (SIZE)	VOLUME	TRADES	WEBSITE VISITS	YEAR	OFFICE
Kraken (M)	$246.491.775 USD	284.897	4,28 M	2011	U.S.A.
HitBTC (M)	$1.012.617.424 USD	839.667	1,09 M	2013	Chile
Binance (L)	$2.150.701.021 USD	3.016.786	23,67 M	2017	Malta
LBank (L)	$2.525.777.644 USD	2.163.133	1,64 M	2016	China

Table 4.1: Some of the features of the four exchanges analyzed.

From table 4.1 it is easy to notice that even though Binance and LBank have a similar amount of trades and volume, the number of website visits for Binance is 23,67 million whereas LBank has only 1,64 million of visits. This comparison brings the attention on the possibility of some kind of deception with these exchanges, also with the uncovering of other aspects, that will be analyzed later on in chapter 5, the suspicion of some kind of dishonesty raises greatly.

The API documentation provide an accurate description of how the API works, the limitations imposed on the call rate and how to make a HTTP request for every kind of data. While describing the API documentation the exchange Kraken will be used as an example, the focus will be on transaction data, which gives a great sense of how well an exchange is doing in the market, also they provide a valuable insight of how real world events affect the exchange by values such as price of the asset or number of transaction per day.

Every crypto exchange has security measures in place to protect against abuse/DoS (Denial of Service) attacks and manipulations of the book order. For this reason from one request to another there is going to be a waiting time which will be tuned according to the API call rate limit provided by the documentation. In table 4.2 is possible to see the limit of the number of API calls that can be made in a single second and also the period of suspension in case of the number of calls made exceeds the limitations imposed.

[7] https://www.similarweb.com/

EXCHANGE	ORDER LIMIT	MARKET DATA LIMIT	SUSPENSION TIME
Kraken	15	~ 1	15s - 5m
HitBTC	300	100	N/A
Binance	10	20	2m - 3d
LBank	50	20	N/A

Table 4.2: Number of API calls per second for the examined exchanges.

API documentation usually provide more than one way to access data, in our case we will use REST (REpresentational State Transfer) API which is more suitable for our purpose of gathering public market data and make a history of the recent trades. Almost every exchange will also offer WebSocket API, i.e. using a WebSocket protocol which achieves full-duplex communications over a single TCP connection between client and server. In this case the server sends information to the client, reducing unnecessary overhead such as frequent authentication, this method of collecting data will require an authentication by using an API key and API signature. We will not use this method because is more suitable to gather real-time information instead of a complete history.

When making a HTTP request to the endpoint it is possible to provide input parameters for the data requested, once the request is elaborated the response is returned as a JSON (JavaScript Object Notation) object or array (some APIs also provide data in other formats). Each exchange has a different model for how data is shown and different kind of input parameters accepted. For Kraken, responses are JSON encoded in the form of: **error**, which is an array of error messages and **result**, that is the result of the API call (may not be present if errors occur). The result returned has a different form based on the data requested, for instance, in order to get recent trades data it is required to have the exact URL (Uniform Resource Locator) and the values "pair" (the asset pair specified) and "since" (timestamp that specifies the date since the data should be returned). The response is an aggregation of a certain number of transactions. Every transaction has different features such as price, volume and time for each transaction (see figure 4.7).

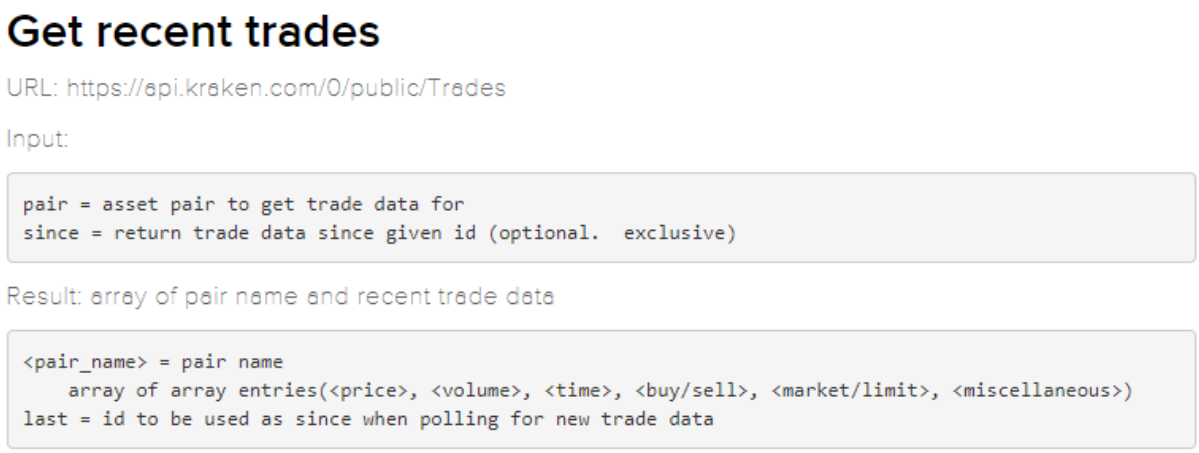


Figure 4.7: API documentation for the recent trades data (Kraken).

In order to make a complete history of trades (of every pair supported by the exchange) for the whole month of March, a script[8] has been used to collect data using the API provided by each exchange. This program was developed using the programming language Java 8[9] and the external library JSON Simple[10] for parsing the JSON objects that the API returns. The flow of the program is structured in this way: first the variables that define the initial day and the final day for the range of time desired must be set, these values will be then converted in nanoseconds using the epoch Time[11] format. The first API call made will request the symbol tickers that the exchange supports, once the response is returned the parser will extract the pair names that will be added to a list of strings.

Then the list will be used to request the trades for every pair name, every request made will contain a maximum of 1000 transactions. Multiple requests will be necessary to cover 24 hours, once the data for the current day has been aggregated the whole process will be repeated until the timestamp of midnight of the final day is reached. Once the JSON object or array for a single request is returned, it will be parsed in order to obtain the new timestamp value that is required for the next API call, further the transactions collected are written in a new JSON

[8] Kraken will be used as an example, the programs for the other exchanges work in a similar way with slight differences.

[9] https://www.java.com/en/

[10] http://www.java2s.com/Code/Jar/j/Downloadjsonsimple111jar.htm

[11] the number of seconds that have elapsed since the Unix epoch, that is 1 January 1970.

file that will aggregate the trades for a single pair in a single day. For istance, if trades for a certain day are requested from an exchange that has hundred pairs, the script will generate hundred JSON files, one for each pair.

In figure 4.8 a flow diagram shows how the code works, the rectangles represent entities and the arrows represent actions or objects, the central oval represent the current state that is repeated during the whole process. The first phase (1) is required in order to have a complete list of strings containing the pair names. Once this list is available, the second phase (2) consists in a constant loop that allow to obtain the trades data for the current day, the process is completed when the final day is reached.

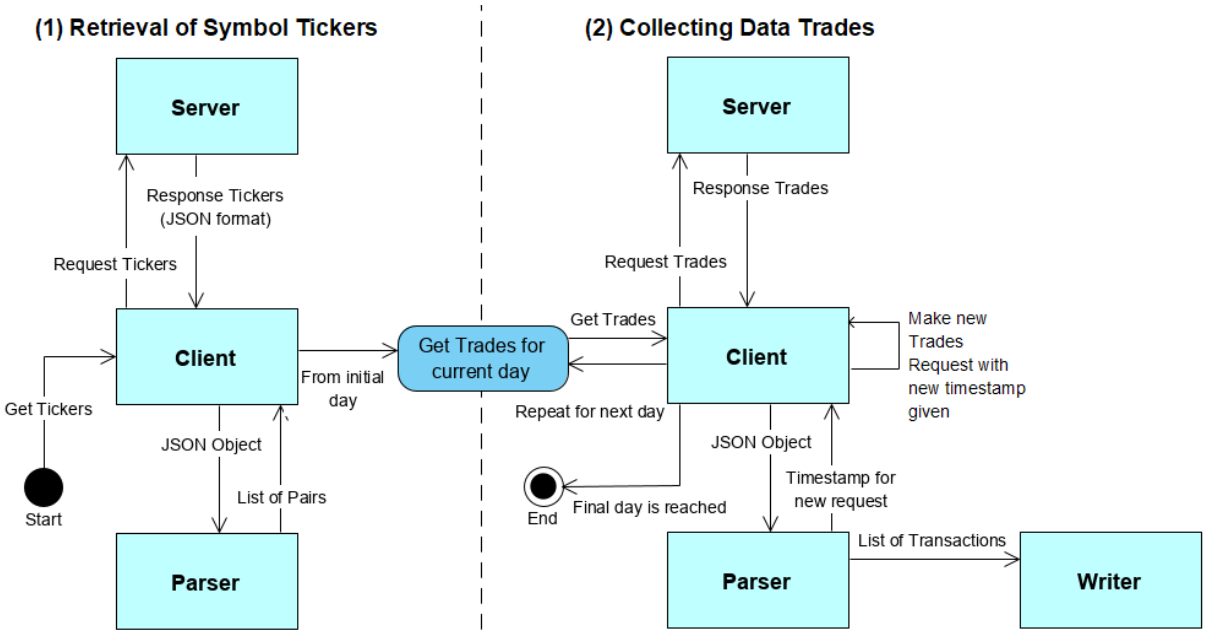


Figure 4.8: Flow diagram showing how trades data is collected.

4.1 Scraping using Scrapy

Web scraping (also referred as web harvesting) is defined as "the practice of gathering data through any means other then a program interacting with an API (or, obviously, through a human using a web browser)" [31]. This is usually done by writing a software program that

automatically extracts data from one or more websites, first it queries a web server, requests data (fetching an HTML file) and then parses that data to extract the desired information.

Web scraping is often used in many fields such as: **web indexing**, i.e. the process of indexing the contents of a website or of the Internet as a whole done by search engines (such as Google and Bing) to organise information before a search, **data mining**, which is the science of extracting useful knowledge from such huge data, **online price change monitoring** and **price comparison** (we will use it in a similar way but for volume data), **market analysis**, **research** and for **collecting training and testing data set** for Machine Learning applications. A Web crawler (also known as a spider) is an Internet bot that systematically browses the World Wide Web, it is usually used for web scraping and in particular for web indexing, in our case we used one to automate data retrieval operations in the context of daily volume data from market data aggregators websites such as CMC and CoinGecko.

The framework used was Scrapy[12], which is an open source and collaborative framework for creating spiders to extract data from websites, it is written in Python[13] and runs on Linux, Windows, Mac and BSD. Every implementation of a spider is a class that must subclasses `Spider`, which is a general class that provides a default implementation and define the initial requests to make, optionally how to follow links in the pages and how to parse the downloaded page content to extract data. The subclass (our spider implementation) can define some attributes, for instance `name` that uniquely identifies the Spider and override default methods such as `startrequests()`, which returns a list of requests defined by the user that the spider will start to crawl from and `parse()`, which will be called to handle the response downloaded for each one of the requests made, it usually parses the response, extracting the scraped data as dicts and also finding new URLs to follow and creating new requests from them. In our implementation the requests were made to CMC and CoinGecko to extract volume data about the top hundred crypto exchanges by reported volume, then the responses were parsed using XPath (XML Path Language) [14], which is a W3C (World Wide Web Consortium)[15] standard

[12] https://scrapy.org/

[13] https://www.python.org/

[14] https://www.w3.org/TR/xpath-3/

[15] https://www.w3.org/

used to navigate through elements and attributes in an XML document (or documents with a similar structure like HTML). Thus once the response was downloaded, we addressed the names of the exchanges and volume values and collected them in a JSON object.

The volume data gathered consists in the amounts of reported volume for the top hundred exchanges for CMC and for CoinGecko we also collected the "normalized volumes", which are an estimate of real volumes based on traffic web data [32], the process was repeated on a daily basis for the month of March 2020.

Chapter 5

Crypto exchanges data analysis

Once the data set was created we proceeded to analyze some of its features to discover possible patterns and reveal significant insights on the exchange market. The data that will be described involves transactions/orders data collected using the exchange APIs of Kraken, HitBTC, Binance and LBank, whereas volume data are scraped from CMC that acts as a data aggregator.

For reported volumes we created a crawler that gathered the desired data instead of using the APIs of each exchange. We did this because websites as CMC provide a list of the top hundred exchanges by reported volume, scraping from this list every day for an entire month gave us an idea of which one of the exchanges are more consistent throughout the time. Further, we compared data provided by CMC with volumes reported by the exchanges. We did this to better understand if the values returned were valid, so we calculated an accuracy score by dividing the real volume provided by the exchange (API data) to the volume provided by the aggregator (CMC DATA), revealing that the data that CMC provided is reliable and so can be taken as valid data[1] (see table 5.1).

Once volume data was scraped, we selected a subset of twenty exchanges that remained on the top hundred list for the whole month of March. We divided this set into two groups: the **trusted** exchanges, those which are well established, have a great reputation and a trans-

[1] values reported in table 5.1 refer to April 28

EXCHANGE	MARKET	CMC DATA	API DATA	ACCURACY
Kraken	BTC-EUR	₿4.212	₿4.216	99,9%
Kraken	BTC-USD	₿5.613	₿5.615	99,9%
HitBTC	BTC-USDT	₿28.385	₿28.420	99,8%
HitBTC	ETH-USDT	333.520 ETH	333.726 ETH	99,9%
Binance	BTC-USDT	₿51.149	₿51.242	99,8%
Binance	ETH-USDT	784.281 ETH	785.610 ETH	99,8%
LBank	BTC-USDT	₿104.712	₿105015	99,7%
LBank	ETH-USDT	2.593.220 ETH	2.599.924 ETH	99,7%

Table 5.1: Accuracy percentage for the main markets of Kraken, HitBTC, Binance and LBank.

parency policy and the **suspicious** exchanges, i.e. exchanges considered shady and suspected of wash trade and false volume reports. In order to achieve a better classification of level of trust, we used a transparency rating, provided by the website Nomics[2] which measures the transparency of these exchanges in terms of their willingness to provide auditable history, these ratings go from A (full transparency) to D (poor) or even F (fail) for those exchanges which do not provide a public market data API [33]. We also used another type of scoring provided by CoinGecko[3], this one is a trust score that goes from 1 (worst exchange) to 10 (best exchange) that gives an evaluation based on liquidity, scale, API coverage and regulatory compliance [32]. In our classification we defined as "trustworthy", exchanges that achieved at least a score of 7 in the trust score and at least a B score in the transparency rating, whereas the exchanges that did not meet this requirements are defined as "suspicious".

[2]https://nomics.com/exchanges

[3]https://www.coingecko.com/en/exchanges

5.1 Fake volumes

One of the first results may indicate the possibility of inflated or fake reported volumes. The table 5.2 shows a summary of the exchanges examined with their daily average volume, the monthly reported volume and whether they are considered (using the trust score and transparency rating) as trustworthy or not. The first noticeable pattern is the significant difference in trading volumes between the group of "trustworthy" exchanges, with an average of $482,4 million USD (Binance is the only one that "escapes" from this average having a volume of $2,1 billion USD) and the group of suspicious exchanges, with an average of $2,3 billion USD [4].

DCE	DAILY	MONTHLY	TRUST
Poloniex	$59,7M	$1,8B	✓
KuCoin	$66,7M	$2B	✓
BitStamp	$189,9M	$5,8B	✓
Kraken	$246,4M	$7,6B	✓
BitFinex	$267,3M	$8,2B	✓
Coinbase	$447,1M	$13,8B	✓
HitBTC	$1B	$31,3B	✗
Folgory	$1,6B	$52B	✗
BiKi	$1,8B	$57,7B	✗
FatBTC	$1,8B	$57,9B	✗

DCE	DAILY	MONTHLY	TRUST
Bit-Z	$1,9B	$61,4B	✗
Binance	$2,1B	$66,6B	✓
Coineal	$2,2B	$69,8B	✗
P2PB2B	$2,3B	$73,5B	✗
LBank	$2,5B	$78,2B	✗
TAGZ	$2,6B	$80,9B	✗
Coinsbit	$2,8B	$87,7B	✗
Hotbit	$2,9B	$90B	✗
MXC	$3,3B	$102,6B	✗
BKEX	$3,4B	$105,3B	✗

Table 5.2: The twenty exchanges examined with the daily average and monthly volume.

The next step was to attempt to set a general trend that could help us trace those exchanges that behaved in a peculiarly way. So we recreated this general trend by averaging the volumes of the group of "trustworthy" exchanges. We discovered that most exchanges that we examined followed this tendency, however some of them showed an opposite or random behaviour. We wanted to better quantify this deviation, so we analyzed the trading volumes and calculated a factor of deviation α for each exchange examined. Thus we determined the volume variation

[4] volume values in table 5.2 refer to the month of March 2020.

of the exchange in exam for two consecutive days as $v.ex._i = \frac{vol_i - vol_{i-1}}{vol_{i-1}}, \forall i \in [2, n]$ where i represents the i-th day, n defines the number of days examined (in our case $n = 31$) and vol is the trading volume for the selected exchange. Then we calculated the volume variation for the general trend (note that we defined the general trend as the arithmetic mean of the "trustworthy" group) as $v.t._i = \frac{vol.t._i - vol.t._{i-1}}{vol.t._{i-1}}, \forall i \in [2, n]$ where this time $vol.t.$ is the trading volume for the trend. Finally we determined the factor as $\alpha = \frac{1}{n-1} \sum_{i=2}^{n} |v.ex._i - v.t._i|$ where the highest and lowest ends were discarded (obtaining a trimmed mean) to better evaluate the central tendency.

At this point we selected a threshold value to classify the exchanges that followed the trend from those which ignored it, with the aid of chart visualization, we set this value at 0.3 as it appeared clear that factors that surpassed this threshold had irregular or opposite patterns. We took a closer inspection at all the exchanges which had a factor greater then the threshold value, which are Hotbit ($\alpha = 0.33$), HitBTC ($\alpha = 0.34$), Folgory ($\alpha = 0.35$), FatBTC ($\alpha = 0.37$) and TAGZ ($\alpha = 0.4$). Analyzing the figure 5.1, which depicts these exchanges, it is clear that their volumes show some significant differences in comparison with the general trend (in green). In many cases when the trend had a climb (a positive variation) these exchanges had a downhill (a negative variation) or no variation at all. One great example of a notable deviation from the general trend is the volume reported by TAGZ for March 13, which had a substantial downhill compared to the previous day, whereas the trend showed an opposite variation.

The totality of the exchanges that exceeded the threshold value were previously considered as "suspicious" by the trust and transparency rating, this odd behaviour discovered can only increase the doubt that there must be something behind these volumes. Perhaps this supposedly practice of reporting fake volumes is coming from two powerful reasons, as stated in a recent report from Bitwise [14], the first one is that **popular exchanges are better perceived**, CMC and other data aggregators such as CoinGecko and Nomics rank exchanges based on their traded volume, since this has a major influence over traders, exchanges may inflate their volumes to increase their popularity and attract new users. The other significant reason are **listing fees from ICOs**, exchanges inflate their volume to attract listing fees from initial coin offerings, the more popular they are the easier is to attract much higher listing fees from

cryptocurrencies that want to be listed on their platform.

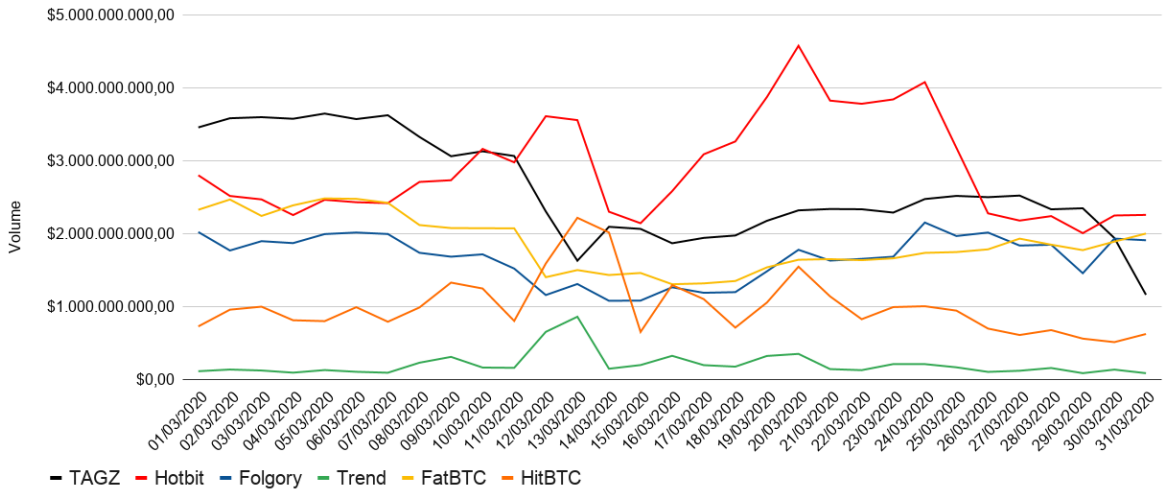


Figure 5.1: A possible scenario of fake volumes.

On the other hand the figure 5.2 shows the amount of reported traded volume during March 2020 with a 24 hours granularity for the exchanges that appear to go along the general trend. We used a scale of red colors (a lighter color indicates a less average traded volume per day) to represent the group of "suspicious" exchanges whereas the scale of green colors was used to represent the group of "trustworthy" exchanges. It is clear that the all the green lines (Binance is the only exception) are situated in the same area, meaning that they have a similar amount of traded volume, while the red lines maintain a significant distance from them.

This is also corroborated by normalized volumes provided by CoinGecko, (which should be an estimate of real volumes based on web traffic data) which portray a much different situation. Using this data we determined a ratio r defined as $r = \frac{\mu_n}{\mu_r} \times 100$ where μ_n and μ_r are the arithmetic mean of the normalized and reported volumes respectively, this percentage should give us an idea of how much of the reported traded volume is real and how much is fake. From the 12 exchanges that CoinGecko covered (out of the total 20) every one of the 5 exchanges that were defined as "suspicious" had a ratio that did not exceed 13%. These exchanges were Hotbit with $r = 2,15\%$, BKEX ($r = 2,98\%$), LBank ($r = 5\%$), Coinsbit ($r = 8,1\%$), HitBTC ($r = 12,18\%$). Whereas the group of "trustworthy" exchanges had the lowest ratio for Binance

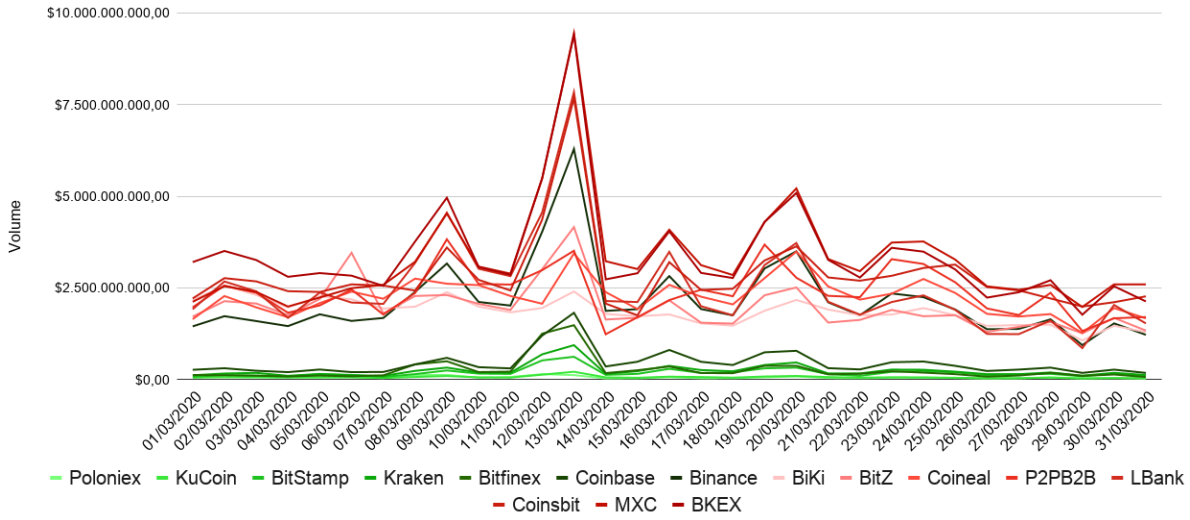


Figure 5.2: A possible scenario of inflated volumes, in green the "trustworthy" exchanges and in red exchanges suspected of boosting their volumes.

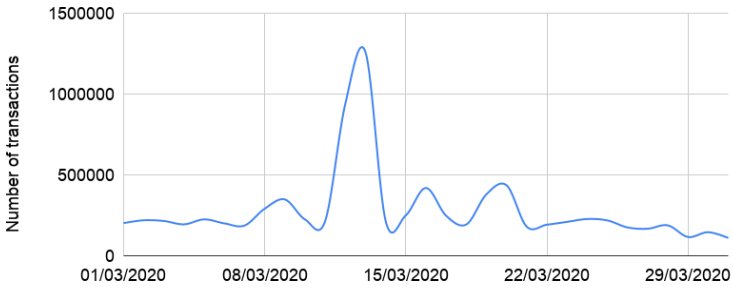
($r = 82,28\%$) and the highest for Kraken ($r = 100\%$). This could indicate that these exchanges may boost their true numbers or engage in wash trading in order to have major visibility and attract more customers, this potential "boost" could explain how despite the vast gap (the trend has five times less volume then the average "suspicious" exchange) they appear to behave in such a similar way.

5.2 Transaction data analysis

Another appreciable result comes from the analysis of the transactions of the four main exchanges taken in exam, in figure 5.3a the chart represents the reported volume for Kraken for the month of March 2020 whereas the figure 5.3b represents the number of transactions for Kraken for each day. In comparison the two charts are quite similar, in fact it is naturally expected to see a strong correlation between the trading volume and the number of transactions of a crypto exchange.

The same correlation was observed in the other main exchanges with the exception of LBank, moreover by analyzing its number of transactions we discovered that they remained approx-

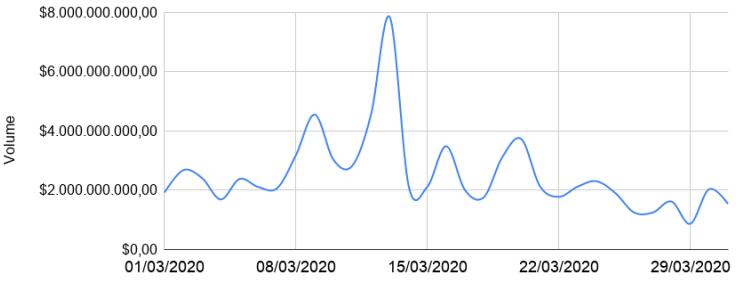
(a) Reported volume by Kraken.



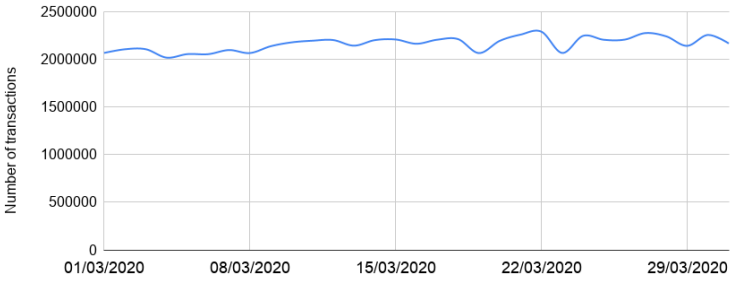
(b) Number of transactions reported by Kraken.

Figure 5.3: The trading volume and number of transactions have a strong correlation.

imately the same throughout the time, with no great variation. As shown in figure 5.4a the trading volume of LBank has a similar pattern to the reported volume by Kraken but that is not the case for the number of transactions, which, as shows in figure 5.4b, follow an almost linear trend (the values are quite close to the mean).



(a) Reported volume by LBank.



(b) Number of transactions reported by LBank.

Figure 5.4: The trading volume and number of transactions follow a different trend.

From figures 5.3 and 5.4 it is noticeable that in the exchange Kraken (previously defined as "trustworthy") there is a natural correlation between trading volumes and number of transactions, whereas the exchange LBank (previously defined as "suspicious") shows an artificial pattern in its number of transactions, which appear to remain the same despite the movements in the volumes. This could indicate fraudulent printing, meaning that the exchange has algorithms that systematically generate fake transactions in order to inflate the reported volumes, it is also likely that the exchange may be engaging in wash trading, in this case buying and sell-

ing at the same time so as to achieve a higher volume [14]. Finally we determined the volume per transaction of each exchange by simply diving the volume of the day for the number of transactions that were reported for the same day. As expected, in figure 5.5 we can see that the "trustworthy" exchanges Kraken and Binance have a constant trend, their range goes from a minimum of $496 USD to a maximum of $1296 USD for transactions, while HitBTC and LBank (considered "suspicious") present a more irregular trend with great climbs and downhills, in fact their range goes from a minimum of $401 USD to a maximum of $3657 USD volume for transaction.

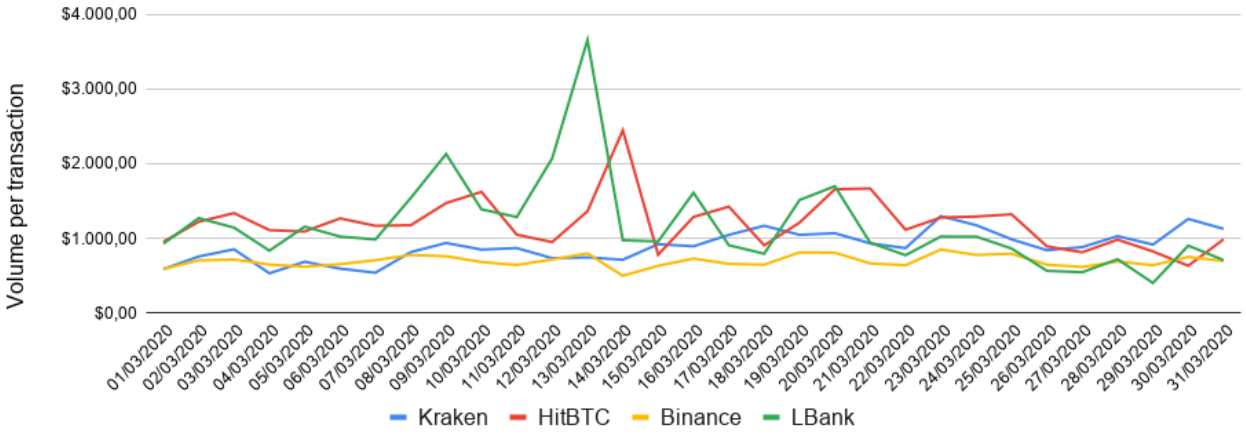
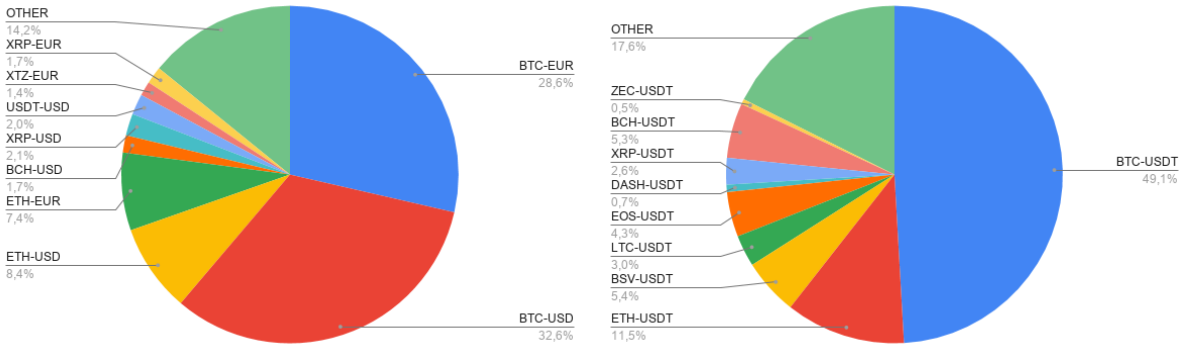


Figure 5.5: Volume per transaction for the exchanges Kraken and Binance (which have a constant trend) and the exchanges HitBTC and LBank (which have a irregular trend).

Next we focused on how reported volume was composed to understand what are the main cryptocurrencies exchanged and the most profitable markets for each crypto exchange. We analyzed Kraken and HitBTC which have a similar size, the first is more focused on fiat to crypto exchange whereas the latter accepts only crypto to crypto trades, (acquiring cryptocurrencies is supported by an external service). As shown in figure 5.6a for the exchange Kraken pairs BTC-EUR and BTC-USD impose a clear dominance with an average of 61% of the total volume, also the pairs ETH-USD and ETH-EUR have a significant relevance with an average of 15% of the total volume.

The same pattern is present in the exchange HitBTC, as shown in figure 5.6b Bitcoin on av-

erage takes up half of the total volume, whereas Ethereum on average take up the 11% of the total volume. These two percentages are close to the 61% of Bitcoin and 15% of Ethereum of the collected data from Kraken. Also we can notice that for HitBTC in every one of the main markets USDT (Tether) is present, this is due to the fact that Tether is a stable-coin that preserves the same value of 1 USDT = 1 USD. In this way whoever trades Tether avoids the volatility typical of crypto currencies but maintains the operability of a crypto.



(a) Partition of the Kraken volume.

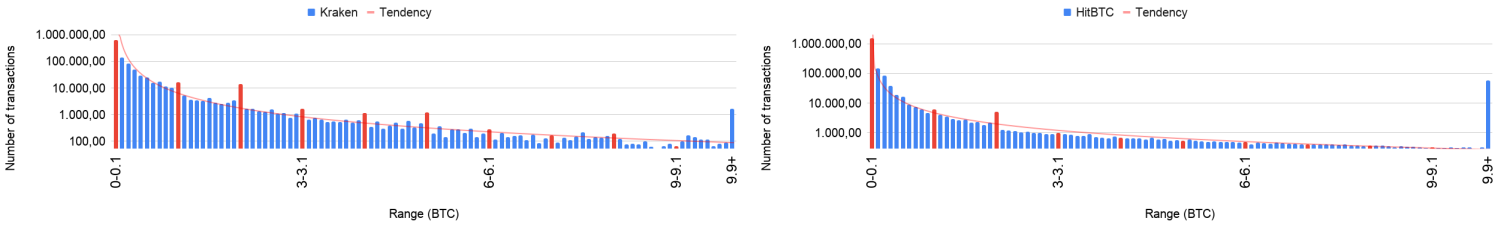
(b) Partition of the HitBTC volume.

Figure 5.6: Comparison between Kraken and HitBTC on their volume partitions.

In the next analysis we examined in detail transactions that involved Bitcoin[5], the pair BTC-USD was selected as benchmark since it is the largest market (see 5.6) and the most common. We investigated on the distribution of the number of transactions over the quantity exchanged, further we tried to discover possible patterns in association with the size of the exchange and its level of trust. First we counted the number of trades by using the "quantity" field of the transactions, this data was grouped in a total of hundred intervals of 0.1 BTC that go from 0 to 10 or more BTC[6]. Figures 5.7a and 5.7b represent the distribution for the exchanges Kraken and HitBTC, the horizontal axis represent the intervals of 0.1 BTC and the vertical axis the number of transactions that fall in that range (note that the a logarithmic scale was applied to the vertical axis to better visualize the distribution). Figure 5.7 shows the comparison between the distributions of the medium sized exchanges Kraken and HitBTC, as expected the majority of trades are concentrated on the first intervals and, as the intervals progress, the

[5]the period examined is from March 1 to March 31.

[6]the first interval is 0-0.1 BTC, the second 0.1-0.2 BTC and so on, the last interval is 9.9 BTC or more.



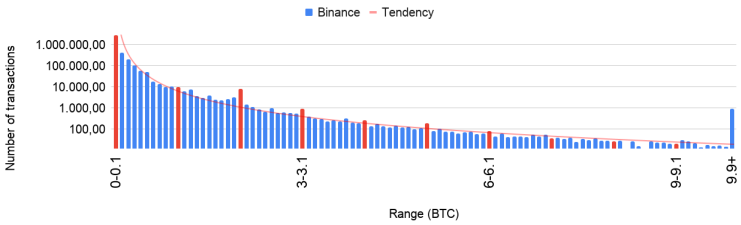
(a) Distribution of Kraken.

(b) Distribution of HitBTC.

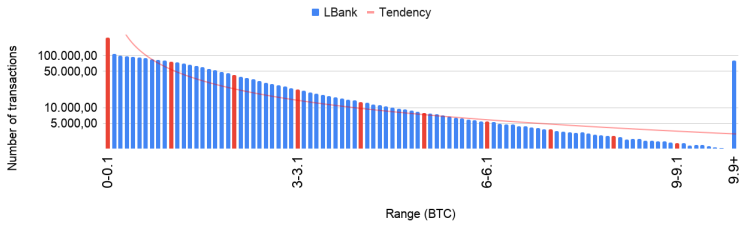
Figure 5.7: Comparison between the distributions of the number of BTC-USD transactions for ranges of 0.1 BTC (Kraken and HitBTC).

number of trades decreases. Both distributions follow a trend similar to the positive branch of the hyperbola of the function $f(x) = \frac{1}{x}$. Additionally it is noticeable that there are various spikes (colored in red) in proximity of whole numbers, for instance 1-1.1 and 2-2.1 BTC show a clear increase compared to the numbers in their proximity, this behaviour seems natural since people tend to deal with whole numbers instead of decimal numbers.

Next we analyzed the large sized exchanges Binance and LBank by applying the same procedure, in figure 5.8a we discovered that also Binance behaved in a very similar way compared to the medium sized exchanges, presenting the same characteristics. For LBank the situation is quite different, despite some similarities such as the majority of its trades being concentrated on the first intervals and the pattern of decreasing over the quantity of BTC, its distribution is unusual because it appears to be more linear and without spikes for whole numbers. Further, we compared the distributions of Kraken, HitBTC and Binance with the distribution of LBank by calculating their **coefficient of variation** (c_v), this value is a standardized measure of dispersion of a frequency distribution. It is defined as the ratio of the standard deviation $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ to the mean $\mu = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$, $c_v = \frac{\sigma}{\mu}$ [34], where $n = 100$ in our case represents the number of intervals. We discovered that LBank had a coefficient of variation of 137,88% whereas Kraken, Binance and HitBTC had a coefficient of variation of 585,77%, 748,61% and 768,03% respectively. A higher c_v means that data is more sparse, thus it has more variability compared to the arithmetic mean, whereas a lower value means that data is more concentrated around the mean, the latter appears to be the case for LBank, which scored the lowest coefficient value out of the four exchanges. This could indicate that the



(a) Distribution of Binance.



(b) Distribution of LBank.

Figure 5.8: Comparison between the distributions of the number of BTC-USD transactions for ranges of 0.1 BTC (Binance and LBank).

number of transactions may be fabricated in order to reach a higher trading volume and as a result explain the supposedly more linear distribution of LBank.

Finally we analyzed the transactions of the four exchanges in order to discover some insights on when and where the trades are usually taken place. Every transaction has a "timestamp" field which indicates the time when the trade took place (this value is expressed in epoch time), in our analysis we used the timezone GMT+1 (Greenwich Mean Time) in order to stay consistent with the time of the volume data gathered. Four periods of time, composed of six hours, were chosen to represent a single day: **night**, which goes from midnight to six, **morning** (six to midday), **afternoon** (from midday to eighteen) and **evening** (from eighteen to midnight). We counted the number of transactions that fell in each one of these four intervals, then we made a daily average for each interval and calculated the percentage of transactions compared to the total average for the day. We observed, as figure 5.9 shows, that for the exchanges Binance, LBank and HitBTC the intervals have similar percentages, this could indicate that they have a worldwide market since there is no discrepancy, whereas Kraken has a greater number of transactions during the afternoon and the night, this indicates that its customers are more concentrated in Europe and America.

In order to corroborate this assumption and understand which countries have more activity for every exchange, we used website traffic data provided by SimilarWeb, this data shows where the visitors of the exchange's website come from (the April 2020 analysis was used). We discovered (see figure 5.10) that in fact Kraken website is more accessed among European states

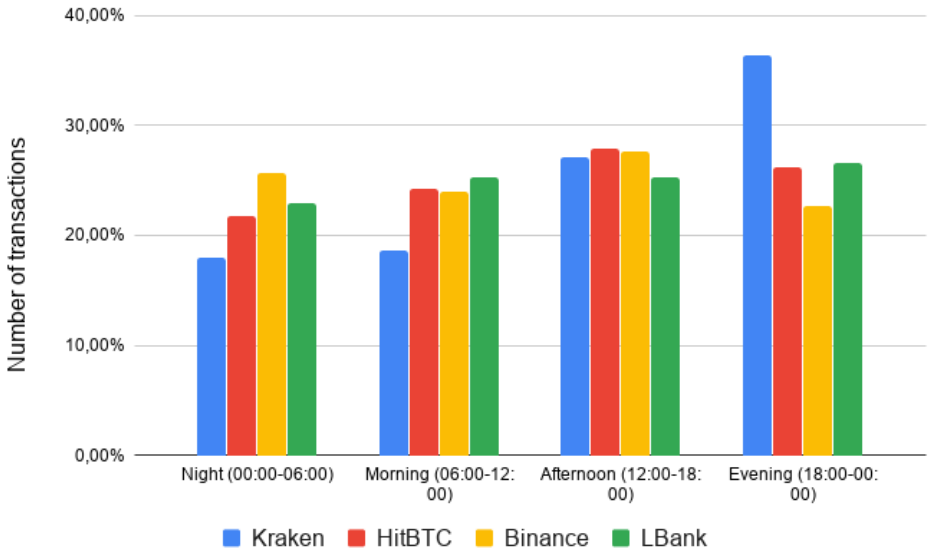
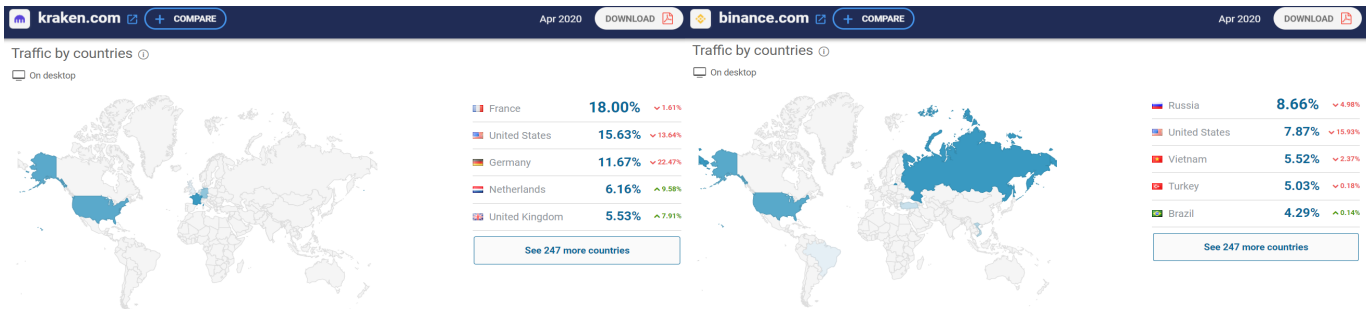


Figure 5.9: Distributions of the number of transactions over the four intervals of time.

such as France (18% of total visits), Germany (11,67%), Netherlands (6,16%) and the United Kingdom (5,53%) and in the U.S. (15,63%). Whereas Binance in addition to the U.S. (7,87%) has also a strong Asian market with Russia (8,66%) and Vietnam (5,52%).



(a) Website traffic for Kraken.

(b) Website traffic for Binance.

Figure 5.10: Comparison between the website traffic of Kraken, which is more accessed in European countries and in the U.S, and Binance, which also have a strong Asian market.

5.3 Wash trading

In our final analysis we focused more on the general state of the cryptocurrency spot market, uncovering dishonest the dishonest practice of wash trading. This activity could impact the public's view on digital currencies' and falsely portray immense volumes exchanged in the crypto market, where in reality its magnitude is significantly smaller. Wash trade is a form of market manipulation where an exchange platform simultaneously sell and buy large amounts of a commodity (crypto in our case) [35] by creating artificial transactions. The goal is to significantly increase its trading volume in order to show a higher liquidity and activity. Higher volumes in exchanges imply a more authoritative perception of the public view and the likelihood to attract more customers. In the 2019 report of Bitwise [14], it was discovered that that roughly 95% of reported trading volume in Bitcoin is fake or non-economic in nature, also Bitwise revealed that, contrary to the common belief, fake volume does not influence price discovery[7] in the real Bitcoin spot market, due to arbitrage[8] systems that keep accurate prices on real global Bitcoin spot exchanges, eliminating sufficient pricing discrepancies in short time.

We tried to estimate real volumes for the month of February 2020, i.e. volumes that are not affected by the distortion of wash trading and fake volumes, this was done by using web traffic data provided by SimilarWeb, which should give an approximate estimate of the real number of users and transactions. This method does not account for users trading on API and mobile app, which we believe are a minority and can be ignored in order to give an estimation. We are aware that this is a simplification but a required step to better present exchange volume data. The analysis followed this process: first we scraped monthly volume data from the listing of the top hundred exchanges (by reported volume) provided by CMC. Then we obtained the monthly web visits for the websites of the exchanges by using SimilarWeb (or WebsiteOutlook[9] when SimilarWeb did not provided data for the website's exchange). We determined a ratio of reported volume per website visit defined as $r = \frac{vol}{visits}$ for each one of the hun-

[7] the price discovery process is the process of determining the price of an asset in the marketplace.

[8] the simultaneous buying and selling assets in different markets in order to profit from differing prices.

[9] https://www.websiteoutlook.com/

dred exchanges. Next we created a baseline by averaging the volume per visit ratio of the trustworthy exchanges (Binance, Kraken, Coinbase, Poloniex, Bitstamp, Bitfinex) as follows, $\mu = \frac{1}{n}\sum_{i=1}^{n} r_i$ obtaining a mean of $1641 USD volume per visit. We used this value to deduce an estimation of the normalized volume by multiplying the baseline for the number of visits as $vol_{norm} = \mu \times visits$. Finally we compared the normalized volume with the reported volume to obtain a percentage ratio as $\alpha = \frac{vol_{norm}}{vol_{reported}} \times 100$ that defines how much of the reported volume is real.

In figure 5.11 are represented the reported volumes (in blue) and normalized volumes (in red) for the hundred exchanges examined. It is clear that most of the exchanges are inflating their volumes, especially when the reported volume appeared excessive (generally amounts greater $10B USD showed low normalized volumes), this result perfectly reflects the situation revealed by Bitwise, which stated that 95% of reported trading volume in bitcoin is fake.

Furthermore the analysis showed that 71 exchanges had a ratio $\alpha \leq 10\%$, which strongly suggests a possible situation of dishonesty, 12 of them had a ratio $10\% < \alpha \leq 50\%$, which indicates a moderate suspect of wash trading or false reports, whereas a ratio $50\% < \alpha \leq 100\%$ indicates a really low or non existing situation of fake volumes. Finally, only 6 exchanges had a ratio $\alpha > 100\%$ that is evidently an overestimation of the volume reported. In figure 5.12 are depicted the top hundred exchanges examined, in this case each one of the bars represent its exchange's ratio which are colored in red if $\alpha \leq 50\%$ or in green otherwise.

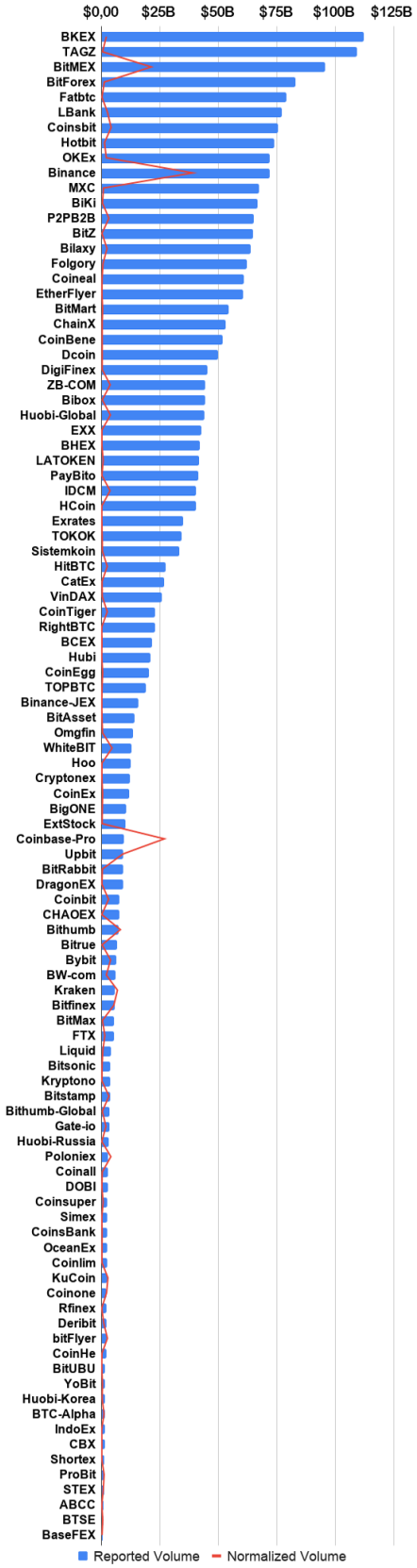


Figure 5.11: The 100 exchanges with reported (in blue) and normalized volume (in red).

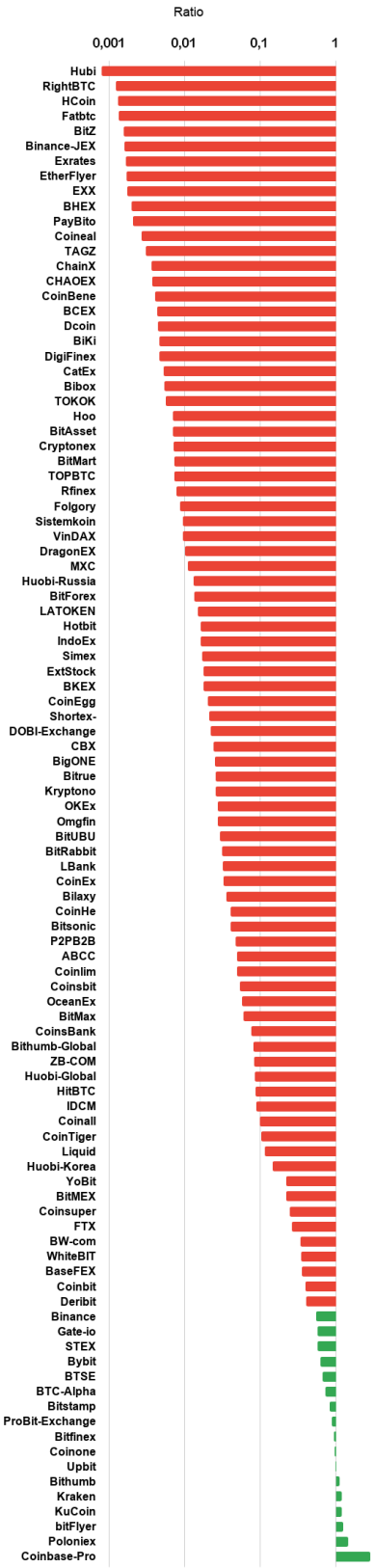


Figure 5.12: Ratio of normalized volume over reported volume

Chapter 6

Conclusion

In this thesis we have shown the current state of cryptocurrency exchanges, the ecosystems surrounding Bitcoin and also presented the state of the art about crypto exchange analysis. Moreover, we classified exchanges as trustworthy or suspicious based on a trust score and a transparency rating, this helped us evaluate our sample of exchanges analyzed. Additionally, we presented some possible cases of fake volumes, demonstrated the existing strong correlation between volumes and transactions and analyzed the general situation about wash trading. We took a closer look at what kind of data the exchange provided using REST API and compared volume data with data provided by third party data aggregators to verify their accuracy. Thus, we used web scraping to gather and collect data about the top 100 crypto exchanges by reported volume in order to have a better understanding of the most relevant crypto exchanges. We used this data to extract a general trend of the trusted exchanges, that we used as benchmark for testing suspicious exchanges. For every exchange we determined the variations in volume for each day examined and then compared them with the variations of the general trend. This work revealed that some exchanges might not be honest about their volumes, in fact discrepancies between trusted exchanges and suspicious exchanges were discovered. We also tried to estimate the real volumes of 100 exchanges by using web traffic data provided by a third party website. As before we used trusted exchanges as benchmark to verify the validity of reported volumes, we calculated a percentage ratio of normalized volume (our estimation

of real volume) over reported volumes. This analysis revealed that 83 exchanges out of the 100 examined reported less then 50% of the real estimated volume. Finally, we believe that a better regulation by the authorities is needed, this could prevent a false depiction of the actual cryptocurrency exchange market and increase awareness among crypto exchange users.

6.1 Future works

In the future a longer longitudinal analysis regarding crypto exchanges could be done in order to reveal new patterns or behaviours previously unknown. As seen in the work of Yue et al. (2018) [12] the use of software similar to BitExtract could help gather and extract new Bitcoin exchange intelligence. Another interesting opportunity of research could be the analysis of the order book of different exchanges, in this way we could determine if some kind of manipulation is present and how it affects the market. Further, since a strong interest in crypto exchange failures and hacks has risen over the years, a new study could be conducted to understand the risks associated with the use of crypto exchanges, thus discovering which are the effective solutions to prevent mishandle and theft of fund and possible countermeasures to adopt.

Bibliography

[1] Satoshi Nakamoto and A Bitcoin. A peer-to-peer electronic cash system. *Bitcoin. "https://bitcoin.org/bitcoin.pdf"*, 2008.

[2] Frank Chaparro. *Bitcoin went bonkers in 2017 — here's what happened as the cryptocurrency surged more than 1000%.* https://www.businessinsider.com/bitcoin-price-in-2017-review-2017-12?IR=T, 2017. [Online; accessed 04/06/2020].

[3] Luqin Wang and Yong Liu. Exploring miner evolution in bitcoin network. In *International Conference on Passive and Active Network Measurement*, pages 290–302. Springer, 2015.

[4] Nicolas Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224, 2013.

[5] Steven Goldfeder, Joseph Bonneau, Rosario Gennaro, and Arvind Narayanan. Escrow protocols for cryptocurrencies: How to buy physical goods using bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 321–339. Springer, 2017.

[6] UW Chohan. The cryptocurrency tumblers: Risks, legality and oversight. ssrn, 2017.

[7] Morten Brandvold, Peter Molnár, Kristian Vagstad, and Ole Christian Andreas Valstad. Price discovery on bitcoin exchanges. *Journal of International Financial Markets, Institutions and Money*, 36:18–35, 2015.

[8] Pengcheng Xia, Bowen Zhang, Ru Ji, Bingyu Gao, Lei Wu, Xiapu Luo, Haoyu Wang, and Guoai Xu. Characterizing cryptocurrency exchange scams. *arXiv preprint arXiv:2003.07314*, 2020.

[9] Tyler Moore, Nicolas Christin, and Janos Szurdi. Revisiting the risks of bitcoin currency exchange closure. *ACM Transactions on Internet Technology (TOIT)*, 18(4):1–18, 2018.

[10] Sinan Krückeberg and Peter Scholz. Decentralized efficiency? arbitrage in bitcoin markets. *Financial Analysts Journal*, page 1, 2020.

[11] Christian Decker, James Guthrie, Jochen Seidel, and Roger Wattenhofer. Making bitcoin exchanges transparent. In *European Symposium on Research in Computer Security*, pages 561–576. Springer, 2015.

[12] Xuanwu Yue, Xinhuan Shu, Xinyu Zhu, Xinnan Du, Zheqing Yu, Dimitrios Papadopoulos, and Siyuan Liu. Bitextract: Interactive visualization for extracting bitcoin exchange intelligence. *IEEE transactions on visualization and computer graphics*, 25(1):162–171, 2018.

[13] Lin William Cong, Xi Li, Ke Tang, and Yang Yang. Crypto wash trading. *Available at SSRN 3530220*, 2019.

[14] Matthew Hougan Hong Kim and Micah Lerner. *Economic and Non-Economic Trading In Bitcoin: Exploring the Real Spot Market For The World's First Digital Commodity*. https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5574233-185408.pdf, 2019. [Online; accessed 04/06/2020].

[15] B. Brindley, J. Law, and J. Smullen. *A Dictionary of Finance and Banking*. Oxford Paperback Reference. OUP Oxford, 2008.

[16] C.F. Lee. *Encyclopedia of Finance*. Encyclopedia of Finance. Springer US, 2006.

[17] Binance Academy. *What Are Makers and Takers?* https://www.binance.vision/economics/what-are-makers-and-takers/, 2020. [Online; accessed 04/06/2020].

[18] Binance Academy. *Crypto Wallet Types Explained.* https://www.binance.vision/blockchain/crypto-wallet-types-explained/, 2020. [Online; accessed 04/06/2020].

[19] M.N. Ogun. *Terrorist Use of Cyberspace and Cyber Terrorism: New Challenges and Responses.* NATO Science for Peace and Security Series - D: Information and Communication Security. IOS Press, 2015.

[20] Usman W Chohan. *The Problems of Cryptocurrency Thefts and Exchange Shutdowns. Available at SSRN 3131702*, 2018.

[21] Bitpoint. *Notice of suspension of all services.* https://www.bitpoint.co.jp/news/info/info-2019071202/, 2019. [Online; accessed 04/06/2020].

[22] Shane Hickey. *$32m stolen from tokyo cryptocurrency exchange in latest hack.* https://www.theguardian.com/technology/2019/jul/12/tokyo-cryptocurrency-exchange-hack-bitpoint-bitcoin, 2019. [Online; accessed 04/06/2020].

[23] Global Legal Research Directorate Staff. *Regulation of Cryptocurrency Around the World.* https://www.loc.gov/law/help/cryptocurrency/world-survey.php, 2018. [Online; accessed 04/06/2020].

[24] IEEE. *Impact of Successful DDoS Attacks on a Major Crypto-Currency Exchange.* https://ieeexplore.ieee.org/document/8671642, 2019. [Online; accessed 04/06/2020].

[25] Danny Nelson John Biggs. *Upbit Is the Seventh Major Crypto Exchange Hack of 2019.* https://www.coindesk.com/upbit-is-the-sixth-major-crypto-exchange-hack-of-2019, 2019. [Online; accessed 04/06/2020].

[26] Lindsay X Lin. Deconstructing decentralized exchanges. *Stanford Journal of Blockchain Law & Policy*, 2(1), 2019.

[27] Christopher P Buttigieg and Christos Efthymiopoulos. The regulation of crypto assets in malta: The virtual financial assets act and beyond. *Law and Financial Markets Review*, 13(1):30–40, 2019.

[28] Binance Academy. *A Beginner's Guide to Candlestick Charts*. https://www.binance.vision/economics/a-beginners-guide-to-candlestick-charts, 2020. [Online; accessed 04/06/2020].

[29] Ali Martinez. *HitBTC An Insolvent Scam Operation Claims Cybercrime Investigator*. https://cryptobriefing.com/hitbtc-insolvent-scams-users-cybercrime/, 2019. [Online; accessed 04/06/2020].

[30] Adrian Zmudzinski. *Two Exchanges Overtake Binance On CMC Rankings But Research Suggets Volume Is Fake*. https://cointelegraph.com/news/two-exchanges-overtake-binance-on-cmc-rankings-but-research-suggests-volume-is-fake/, 2019. [Online; accessed 04/06/2020].

[31] R. Mitchell. *Web Scraping with Python: Collecting More Data from the Modern Web*. O'Reilly Media, 2018.

[32] Sze Jin. *Trust Score 2.0: CoinGecko Updates Trust Score to Improve Exchange Transparency*. https://blog.coingecko.com/trust-score-2/, 2019. [Online; accessed 04/06/2020].

[33] Clay Collins. *Nomics' Exchange Transparency Ratings (And The Correlation Between Transparency & Fake Volume, and Toxic Activity)*. https://blog.nomics.com/essays/transparency-ratings/, 2019. [Online; accessed 04/06/2020].

[34] B.S. Everitt. *Cambridge Dictionary of Statistics*. Cambridge University Press, 1998.

[35] Tom CW Lin. *The New Market Manipulation. Emory LJ*, 66:1253, 2016.