Heartbeat of a Crypto-Economy:Transaction Information in a World withCentral Bank Digital Currencies

Heartbeat of a Crypto-Economy: Transaction Information in a World with Central Bank Digital Currencies

A THESIS PRESENTED
BY
TANCREDI CASTELLANO PUCCI DI BARSENTO
TO
THE DEPARTMENT OF COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF ARTS AND SCIENCES WITH HONORS
IN THE SUBJECT OF
COMPUTER SCIENCE

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MARCH 2020

Thesis advisor: Professor David C. Parkes
Thesis co-advisor: Daniel J. Moroz
Tancredi Castellano Pucci di Barsento

Heartbeat of a Crypto-Economy: Transaction Information in a World with Central Bank Digital Currencies

Abstract

Central Bank Digital Currencies (CBDCs) have captured the attention of world leaders. The current conversation is dominated by high-level motivations like the efficiency benefits of a cashless society. This discourse neglects the potential use of CBDC data for new analytics capabilities. This thesis makes three main contributions to that end. First, it identifies and explores the inherent challenges of analyzing blockchain data (as a proxy for future CDBC data), making future design recommendations where possible. Second, it develops a novel technique to extract useful sector-based macro-economic data from pseudonymous transaction data, using the Ethereum blockchain as a case study. This also enables a novel breakdown of the Ethereum ecosystem by actor type. Third, it unearths evidence new insights about the public blockchain ecosystem, for example that Ethereum users are becoming more sophisticated over time and that Initial Coin Offerings (ICOs) may have caused the 2018 cryptocurrency bubble.

Contents

Listing of figures

Dedicated to Filippo Brunelleschi

Acknowledgments

I would like to express my deepest appreciation to Professor David Parkes for his support as my teacher, mentor, and advisor. I am deeply indebted to him for critiquing my research, sharing his insights, reading over drafts of my thesis, and contributing to my growth as a young researcher. I absolutely could not have done this without him. Moreover I am also deeply indebted to Daniel Moroz for his research insights, encouragement, support and general excitement for the subject matter. I thank Professors Yiling Chen and Jim Waldo for generously offering to be my thesis readers. In addition I express my gratitude to Mark Chearavanont, Juan Pedro Crestanello, Catherine Kerner, Eric Green, Corey Gonzales, Emanuele Gualandri, Remi Pfister, Evan Mateen, and Liza Zheng for their support and encouragement throughout the process. Lastly, I express my warmest gratitude to my family, for their unconditional support.

1

Introduction

As of March 25 2020 the People's Bank of China (PBoC) has allegedly finished the basic development of a Central Bank Digital Currency (CBDC). [101] Given the numerous advantages it will provide to both the People's Republic of China and its citizens, it is tentatively planned to be implemented within the next 5 years. [63] A number of countries will most likely follow China's lead given current trends, first mover advantage and private money competition.

The implementation of CBDCs allows for the centralized collection of granular and timely trans-

action data that could be used for economic policy making, especially in regard to new analytics capabilities. The degree of usefulness of the data for both ends is highly dependant on choices that are made by the designers of said technology.

1.1 Overview

The essential function of CBDCs is to provide a single centralized digital currency. The data that these would produce would be capable of providing economic insight, of greater accuracy and of a more timely nature than is available today, using micro-level transactional data. After all an economy is simply the sum of its parts or, more accurately, its transactions. However, the design of these future CBDCs will instruct, limit and guide the type of economic data that can be elucidated from them. As such, the design of said technology will be crucial to its eventual analysis and associated policy decisions. This paper uses Ethereum as an illustrative case study to reflect the ways in which the design of a digital currency governs the particular difficulties that will be faced when seeking to engage with its raw data. Ethereum was chosen because future CBDCs could be Ethereum-like, and because it has the most diverse ecosystem of users and therefore the richest ledger data currently available. A technique for stream-lining this process was developed which was capable of extracting valuable economic signals according to, and, more importantly, despite the Ethereum design. The conclusions that economists and policy makers will draw from their CBDCs will therefore be directly influenced by the CBDC design. Yet, the analysis of tradeoffs of different designs have yet to find sufficient emphasis in the literature that has been produced thus far. Worryingly, the current discourse is dominated by a preoccupation with the economic consequences of CBDCs, without engaging with the technological underpinnings that will shape analyses.

1.2 Related Work

Given the relative youth of this field and the multidisciplinary nature of this work, multiple literature reviews are necessary. In brief, a literature regarding **CBDC Data** does not really exist, and has yet to break through into mainstream discourse. The isolated strands can nevertheless be grouped into four distinct areas. First, CBDC literature more generally is flourishing but mainly focuses on economic motivations [90] and its economic implications [16, 39, 47]; an exception is [7], which touches upon some technical nuances. Second, although a relatively new field, blockchain data, mainly cryptocurrency ledger data, extraction [3, 17] and the subsequent analysis of said data, have been studied. The latter mainly focused on understanding, and predicting price fluctuations; [8] and analyzing and detecting illicit activity including de-anonymization techniques [67, 64]. Third, economic data literature includes but is not limited to: data deficiencies [10], the importance of revisions [29, 28], and the use of big data [34] including the use of credit card transaction data. [5] Finally, there has been significant literature regarding transaction data privacy, including ways of protecting and attacking individuals' data [98]. Given the wide ranging of implications of CBDCs all the above are necessary to gain intuition for how they should be designed.

1.3 Contribution

The main contribution of this thesis is to underline the effect that CBDC design choices, in particular design choices related to data collection, have on the ability to analyze said data and extract useful signals. The following is a breakdown of the main contribution into its constituent elements:

- Three challenges are identified in regard to the analysis of blockchain data:
  - A general lack of computationally effective and efficient querying methods due to the inherent structure of ledgers. Ledger data is challenging to parse, and readily available techniques and libraries are still in their infancy.
  - The undefined nature of a transaction. The ability to run code on an blockchain and create 'smart' contracts and decentralized applications severely complicates one's ability to analyze the movement of said blockchain
  - Miscellaneous nature of identities. Not only is it challenging to deanonymize addresses but even connecting addresses owned by the same actor proves challenging.

  The identification of said challenges and the solutions used get around said challenges may prove useful to the initial process of designing a CBDC, future attempts at data analysis once the CBCD is in use or, at the very least, the data analysis of current cryptocurrencies.
- A technique is developed that leverages semi-supervised learning to extract sector-based signals from pseudo-anonymous transaction data. The approach involves training a classifier on a small labelled data set, created by scraping various websites, subsequently classifying a large number of addresses, and finally aggregating statistics of those addresses.
- New insights of the Ethereum ecosystem are provided such as:
  - Evidence that users are becoming more sophisticated.

- Evidence that ICOs may have partially caused the cryptocurrency bubble.
- A potential breakdown of the types of users on the Ethereum Blockchain.

Chapter 1 briefly underlines the importance of financial history followed by a short history of digital coins and CBDCs. Chapter 2 highlights why CBDCs are relevant and summarises the current literature on them. Chapter 3 highlights the weaknesses of current economic data and the potential benefits of analyzing transaction data. Chapter 4 introduces the reader to Ethereum. Chapter 5 discusses blockchain data including the associated analysis challenges. Chapter 6 provides a case study that develops a novel technique to extract sector based macro economic signals. Chapter 8 concludes.

The first panacea for a mismanaged nation is inflation of the currency; the second is war. Both bring a temporary prosperity; both bring a permanent ruin. But both are the refuge of political and economic opportunities.

Ernest Hemingway

2

Motivation and History

2.1 The Importance of Financial Innovation: A Look at History

When asked what some of the most important innovations in history are, one is likely to think of technological innovations like the wheel, Gutenberg's printing press, and the steam engine. Financial innovations are often overlooked. Yet history is rife with examples of financial innovations that

either fundamentally changed society or granted particular states prominence on the world stage. [49] Italy would not have been the cradle of the Renaissance were it not for its advances in double entry book keeping, letters of credit, and holding companies. [51] Similarly, Great Britain, Belgium and the Netherlands would not have become prominent imperial powers if not for their creation of joint stock companies. [30] More recently, the United States would not be the country it is today were it not for its role at the forefront of both financial and technological innovation.

Financial innovations do not lead to automatic progress; as we observed, as recently as 2008, unrestrained and unregulated innovations can lead to dire consequences. [12] For example, the Yuan and Ming dynasties in ancient China were among the first states to realize the potential of government backed paper money. [61] However, they lacked the foresight and the experience to realize that the continuous and reckless printing of money was not sustainable - a mistake that some countries still make today. Repeated independent events of hyperinflation not only caused havoc in sixteenth and seventeenth century China, but eventually led to the abandonment of government backed paper currencies - which were only reintroduced in China at the end of the nineteenth century. [82] These historical events demonstrate that financial innovations have the ability to push the world forward as much as they have the ability to propel it backwards.

Finally, financial innovations are surprisingly counterintuitive. They are maverick and creative endeavors that often deviate from current expectations. For example, the creation of paper money goes against the fundamental idea, historically held, that money must have intrinsic value - underlined by Marco Polo's reaction to "those pieces of paper" during his famous Oriental odyssey. [1] [86] Furthermore, interest bearing loans - so called usury - was prohibited for much of the middle ages, something unimaginable today.[2] Therefore, the scope and shape of future financial innovations

[1] Marco Polo was a Venetian Merchant circa 1245 who gained the trust of the Emperor of China, Kublai Khan, eventually serving as an advisor. He is known for his book *The Travels* that describe his extensive journey through Asia.

[2] Sharia-compliant finance still prohibits interest bearing loans to this day.

may be unknown and could take the form of innovations as absurd as hyper-bartering.[3] [19]

2.2 History of Digital Coins

The first digital transaction was sent in 1972 over the Arpanet between students at MIT and Stanford. [58] However, it was not until the creation of the World Wide Web in the 1990s and the emergence of online marketplaces such as Amazon and Ebay that digital transactions became mainstream. Since then, digital transactions have become relatively ubiquitous due to companies like Paypal and Square, and they will only become more omnipresent with the meteoric rise of mobile payment services like M-Pesa and Alipay.

The ability to exchange and store value digitally has opened a virtual Pandora's Box. It has facilitated and will continue to incentivize novel and often unexpected behavior. For example, the spread of multiplayer online role playing games (MMORPGS) allowed players to harvest and trade virtual goods for physical goods. [46] This presented a viable and much needed stable source of income for Venezuelan players, that was more immune to inflation and censorship than the Venezuelan Bolivar. [4] This example underlines the reasons why digital currencies were created though we already had a digital banking system. In particular, digital baking systems are not censorship resistant and prone to inflation.

Until 2008, it was not possible to exchange value over a network without a centralized third party due to the double spend problem, which refers to the potential flaw in digital cash brought upon by the ability to spend a single digital token more than once. Thanks to technological advancements, Chaum 1985 and Back 2002; Nakamoto 2008 was able to resolve the double spend problem by cre-

[3] Hyperbartering is the notion that once all items are tokenized there will be no need for money, instead artificially intelligent machine agents will be able to conduct exchanges with a matrix of liquid digital assets.

[4] Not completely immune to censorship because the games are controlled by centralized entities; moreover in game inflation can also exists but tends to be less than the insane amounts of inflation that have plagued Venezuela in the last few years. Last year inflation was over a million percent!

ating Bitcoin, the first version of blockchain technology. [5] The creation of blockchain technology led to the proliferation of digital coins commonly known as cryptocurrencies - which were mainly variations of Bitcoin with some added functionality. Most notably, Buterin 2014 outlined the design of a system that would be able to support self executing contracts, commonly known as smart contracts, allowing for the creation of decentralized applications (DApps). [6] This system eventually became Ethereum; which is currently full of DApps ranging from a whole ecosystem of decentralized finance (DeFi) to betting exchanges to games. [7]

The adoption of cryptocurrencies has been slow for a number of reasons. The main one is their inherent volatility. Most cryptocurrencies have an annualized volatility of over 100% while most S&P 500 companies have an annualized volatility of 10%.[8] Therefore, cryptocurrencies drastically fail in respecting one of the fundamental properties of money - a store of value. This has led to a second wave of digital coins commonly known as stable coins, most notably: Facebook's Libra, Maker's Dai, and Tether. [9] The creation and continued research of stable coins by large institutions such as Facebook and JPMorgan Chase has put nation-states on a back foot as the private sector tries to usurp one of the main functions of a modern government: controlling the money supply.

2.3 History of Central Bank Digital Currency

The idea of a Central Bank Digital Currency (CBDC) was first proposed by the Bank of England in its 2015 research agenda and has since gained momentum. [10] [14] In 2019, 80% of the 66 Central

[5] An append only data structure that is maintained by computers that are linked in a peer-to-peer network.

[6] A DApp is a computer program that executes on a distributed computing system. Can be arbitrarily complicated.

[7] Maker and Compound are example of DeFi DApps; Auger is an example of a betting exchange; and CryptoKitties is an example of a game.

[8] The annualized volatility (between October 2017 and February 2018) of ETH/USD was 120%; in the same timeframe, the annualized volatility of the S&P 500 was 13%.

[9] A stablecoin is a digital coin that is pegged to a fiat currency such as USD or a basket of fiat currencies.

[10] Definition to come in Section 3.1

Banks surveyed by the Bank of International Settlements (BIS) were working on digital currency related projects, geographic visualization of central banks involved can be seen in figure 2.1. Note that both Advanced and Emerging Market Economies are working on it. [16] Moreover, more than half of these central banks were already working on related experimental projects. For example, the Central Bank of Uruguay, as part of a wider financial inclusion program, launched a six month e-Peso pilot study. A bespoke platform was created to keep track of the records, decentralized ledger technology was not used. At the end of the six month trial period, it was phased out. The pilot study was deemed a success and is currently in an evaluation phase. [16]

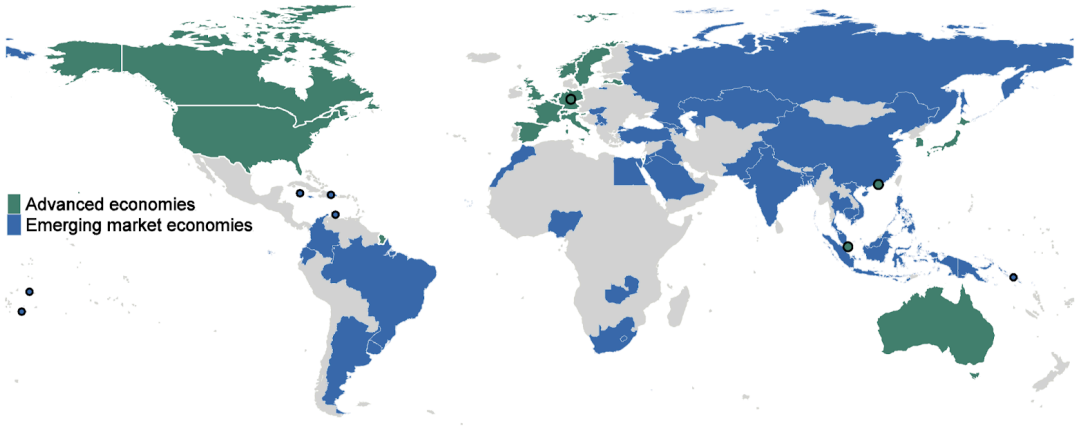

Figure 2.1: Respondents to the 2019 BIS Survey [11] [16]

The majority of Central Banks believe that CBDCs are valuable but that more research is required before a definite decision can be made. A few countries; however, are rushing to implement a CBDC. Notable examples are the Central Banks of Marshall Islands, the Bahamas and China. [12]

[11] The black circles represent the Cayman Islands, the Dominican Republic, the Dutch Caribbean, the euro area, Hong Kong SAR, Samoa, Singapore, the Solomon Islands and Tonga. "Advanced economies" and "Emerging market economies" as defined by the IMF World Economic Outlook country classification. The boundaries and names shown and the designations used in this map do not imply endorsement or acceptance by the BIS or by the writer of this work.

[12] The Republic of the Marshall Islands technically does not currently have a Central Bank since they use USD are their currency.

The governments of the Marshall Islands and the Bahamas have hired different third parties to help them implement CBDCs in attempts to gain first mover advantages; these implementations are similar to that of current cryptocurrency systems. [37, 96] On the other hand, the PBoC has taken a more hands on approach; underlined by the fact it has filed upwards of 80 CBDC related patents. [77] The system that it seems to be creating is distinctively different from other existing systems. [13] [63] Looking forward, it is unclear when CBDCs will be implemented on a wider scale; it is furthermore uncertain what specifications said technology will have.

[13] Technical documentation regarding China's CBDC was difficult to come across and the resources available are mostly in Chinese. A preliminary survey reveals a "two-tier" system. First layer involves the PBoC issuing said CBDC to commercial banks. The second layer involves the retail banks issuing the CBDC to retail market participants. The use of blockchain for either layer remains an option but not a requirement. [63]

Money is too important to be left to the private sector alone. Like the law, it is a foundational public good.

Lael Brainard, A governor of the US Federal Reserve

3

Central Bank Digital Currencies

3.1 What is a CBDC?

A Central Bank Digital Currency (CBDC) is a digital currency that has "Legal Tender" status. [1] [2] It differs from current digital transaction mediums by nature of being backed by the government one-to-one. [3] [4] If implemented correctly CBDCs may have the ability to enhance current financial systems. The specification of a CBDC is not clear and as a result neither are the subsequent effects on the banking system. This next section serves to both underline why CBDCs will likely be implemented, as well as a number of potential consequences of CBDCs.

3.2 Why are CBDCs relevant?

CBDCs have gained prominence in the last few years for three main reasons. First of all, the use of cash is on average declining. Secondly, central banks have to react to the myriad of private digital coins, in particular stable coins like Libra, as these coins may be threats to the Central Banks' sovereignty. Lastly, the creation of a CBDC may have far reaching geopolitical consequences since there may be a first mover advantage.

The use of cash in transactions is declining, especially in developed countries. [5] One of the more extreme cases is that of Sweden; where cash accounts for only 1% of GDP by value, and less than

[1] Coinage Act of 1965, specifically section 31 U.S.C 5103, entitled "Legal Tender" states "United States coins and currency (including Federal reserve notes and circulating notes of Federal reserve banks and national banks) are legal tender for all debts, public charges and dues". This would most likely have to be changed to include CBDCs. [81]

[2] Current literature separates CBDCs into two types: Retail and Wholesale. This work mainly preoccupies itself with the Retail type.

[3] Depending on where you are in the world when your bank defaults you are likely to lose a proportion of your money. In the U.S. the Federal Deposit Insurance Corporation (FDIC) generally insures up to $250,000 USD while in Europe the European Deposit Scheme (EDIS) insures up to EURO 100,000

[4] Accounts that are below these thresholds still participate in bank runs. [62]

[5] The situation is more nuanced; however, the consensus among academic is that the cash is decreasing. [90]

60% of Swedes use cash regularly. [6] While this may seem like an advance it is a potentially dangerous phenomena. The absence cash without a well thought out alternative not only risks alienating certain sections of society, in particular the elderly and rural dwellers, but also poses systematic risks to the financial system. Citizens in a cashless society without a CBDC would live in a complete fractional reserve banking system, meaning they would no longer have direct access to "safe" money. [7] This may be problematic as it puts a lot of trust in profit driven banks and financial institutions which interests may not be aligned with the public interest. Moving towards full reserve banking could alleviate these risks; however, this is highly improbably in the foreseeable future as it represents too radical a shift from current norms. [8] [9]

There exists a market for low-cost, high speed, frictionless and sophisticated transactions that the current financial system is unable to fill. This is one of the gaps that cryptocurrencies, in particular stablecoins like Libra, Maker's Dai, and Tether are trying to fill. [10] [2] This poses a threat to the current banking system for two reasons: the decrease in seigniorage revenue (approx $20B a year for the US) and the decrease in the effectiveness of monetary policy. [11] [50] The introduction of non-fiat money into the economy can decrease the effectiveness of monetary policy. [12] Competition between private money and public money in itself is not new, for example the US experienced a

[6] Compared to the rest of Europe and the US where coins and cash account for 10% and 8% respectively

[7] Fractional reserve baking involves banks only keeping a fraction of their reserves at any one time. So as to profit from the loaning out a proportion of the money collected. Regulations are country dependant by they generally mandate that banks keep approximately 10% of reserves. Fractional reserve banking can be an issue during times of crisis when bank runs tend to be commonplace.

[8] Full reserve banking requires banks to keep all of its customers funds in cash or liquid assets

[9] Literature regarding monetary reforms of this kind is extensive; the consequences of full reserve banking are unclear. Some economists like Laurence Kotlikoff and Murray Rothbard argue in favor of full reserve banking while others like Douglas Diamond and Philip H. Dybvig argue against.

[10] Innovation in the financial system is generally slow due to financial regulation and monopolies.

[11] Seignorage is the profit made by a government by issuing currency, especially the difference between the face value of coins and their production costs.

[12] For example the Libra foundation wanted its coin to be pegged to a basket of currencies. This would cause central bank monetary policies like changing interest rates to be less effective if a substantial proportion of transactions use Libra.

period of private banking from the end of the nineteenth century to beginning of the twentieth century. [52, 73, 89] Although, this could reappear in the future as private money is generally less liquid because it is not backed by the government [13] Generally speaking, fiat currencies have shown to be more resilient and given the increased societal dependence on transactions there are some who argue like Lael Brainard, a governor of the US Federal Reserve, that "Money is too important to be left to the private sector alone." The government is therefore left with two choices: innovate or regulate. However, regulating, and therefore de-facto stopping or at least slowing down innovation, may have geo-political risks - especially for the US.

We currently live in a world where the USD dominates, and has dominated since the Bretton Woods conference in 1944. This is highlighted by the fact that 50-80% of international trade is invoiced in USD, and 70% of the world's currencies are anchored to the dollar to varying degrees. [57] Being the dominant currency and the so called reserve currency of the world has a number of advantages that include but are not limited to liquid markets and lower borrowing costs. [14] [57] Therefore, a number of currencies, in particular the Euro, the Pound, and the RMB, are incentivized to try and become the global reserve currency or at least try and reduce the dominance of the USD.[15] [16] Without any paradigm shift it is unlikely that the USD will lose its throne any time soon. [26] Nevertheless as Mark Carney, the ex President of the Bank of England (BoE) has stated: "Technology has the potential to disrupt the network externalities that prevent the incumbent global reserve currency from being displaced." [17] [27]. As a result it may be in the best interest of particularly pow-

[13] Private institutions generally do not have the resources to act as a lender of last resort. There have been exceptions for example JP Morgan in 1907. [53]

[14] For an extensive survey of the advantages and disadvantages of being a reserve currency. [21, 99]

[15] One of the purposes of the EURO was to replace or at least decrease the dominance of the USD. It failed and is unlikely to succeed in its current form. [87]

[16] Non economic reasons for not wanting the USD as a reserve currency also exist. For example, not wanting the US government to be able to interfere and or surveil your activity. The SWIFT payment system which is responsible for a large proportion of international wires is controlled by the US. Moreover, there is evidence that suggests that the NSA monitors said system. [65]

[17] The technology he is referring to is not strictly confined to CBDCs. [55]

erful countries like China to try and implement a CBDC before the US does, if it wants to attempt to tackle dollar supremacy.

It is for these reasons that CBDC will most likely be implemented sooner rather than later.

3.3 Review of Current Literature

CBDC literature is still in its infancy, and consensus around some fundamental issues is only beginning to form. [15] Research in this particular area is sparse but flourishing. As previously discussed, upwards of forty Central Banks are currently active in this space. The most comprehensive literature can be compiled from the Bank of Canada, the People's Bank of China, the Bank of England, the European Central Bank, the Sveriges Riksbank (Sweden), the Bank of International Settlements (BIS) and MIT Media Lab's Digital Currency Initiative. Given most of the work is published by economists working for institutions like central banks, the literature tends to focus on the economic motivations for issuing a CBDC and its economic implications. As a result, technology questions are often overlooked, with an exception being Ali & Narula 2020. The existing scholarship tends to focus on three questions: One, the ramifications of a cashless economy. Two, the effect of CBDCs on the future of the financial system including financial stability. Three, efficiency gains caused by technological improvements as well as a slimmer financial institutional structure.

The consequences of a cashless economy fall under a few main categories. First, and currently the most topical given the economic circumstances of most developed countries, the potential implementation of unconventional monetary policies like negative interest rate and helicopter money. [18]

[18] A large proportion of developed countries, in particular Japan, western European countries and the US are facing record low inflation. Paired with record low interest rates central banks are struggling to stimulate the economy. Mainly, because of the effective lower bound (ELB) (Sometimes referred to as the zero lower bound (ZLB)). ELB is the notion that interest rates have a floor in a cash economy of approximately 0%. That is because if interest rates fell sufficiently below 0% individuals would be incentivized to take money out of banks. For a detailed review of how to potentially deal with the ELB and why it has become a problem Rogoff 2017 is recommended.

[19] [20, 47, 90] Second, a potentially drastic decrease in illicit activity including but not limited to informal economy activity. Cash allows for untraceable transactions and is therefore perfect for illegal transactions. [20] [47, 90] Third, an increase in financial inclusion since free bank accounts would have to be provided to all citizens. How else would they be able to interact in a cashless economy? [45, 47, 90] Fourth, the cost of storing and handling cash has been calculated to be equal to about 1% of a countries GDP. [21] [90]

Moreover, CBDCs have the ability to drastically change the financial system and alter its stability. [47, 39]. First, the widespread adoption of a CBDC would put into question the existence of commercial banks as we know them. Why would one put money in a bank if one could keep it in a secure digital personal account? [22] Second, the ability to swiftly and easily transfer money to other banks and/ or your personal wallets could create flash bank runs. [23] [22, 24] Third, because of the the aforementioned consequences it could narrow the banking system potentially leading to full reserve banking, [24] [18] completely changing the role current commercial banks fill. It is difficult to predict the consequences of a CBDC on a financial system. All the above are possible, but they are highly dependent on the design of and the regulations for a CBDC.

Finally, CBDCs may hold the key to a major update of the payment system, which not only include cheaper, faster, and smarter transactions but also the consequences said features would pro-

[19] Helicopter Money: a proposed form of unconventional monetary policy that involves central banks making direct payments to individuals. Can be thought of as a temporary and or partial universal basic income (UBI) used in downturns to stimulate the economy. The term was coined in 1969 by Milton Friedman.

[20] Rogoff 2016 advocates for the phasing out of large denomination bank notes due to their enormous prevalence in illegal transactions

[21] Costs include safely storing, transferring and counting physical cash

[22] A number of reasons remain raging from accruing capital through interest rates and added security. Nevertheless, the creation of a CBDC makes it feasible to store money safely and portably without the use of a bank.

[23] In the event an individual is unsure about the liquidity or solvency of its bank it could easily transfer money out of a bank.

[24] The narrower the banking system the less risky assets banks are allowed to borrow. The extreme case would be similar to the 1930s Chicago plan which was similar to full reserve banking.

vide. In particular, not only would they facilitate transactions that are currently foregone, but also allow for new types of transactions, which include but are not limited to micropayments and smart contracts. [25] [7, 59] The former would allow for the monetization of the web among other things. [26] [36, 100] Smart contracts are particularly useful in decreasing counterparty risk. [27] [28] Both behaviors would enhance the current system.

There is a general consensus that the development of CBDCs could result in a paradigmatic shift in the way financial systems currently operate . The uncertainty this would entail seems to have resulted in a reticence among existing scholars to advocate for their immediate implementation. This, in turn, has led to a potentially dangerous sidelining of practical considerations for the creation of CBDCs in academic circles. However, the clear opportunities they present have led to concentrated efforts to make CBDCs a reality, dragging the necessary theory kicking and screaming necessary alongside it.

[25] For example consumers who avoid online purchases due to security and privacy concerns or merchants who avoid selling online due to fees.

[26] The ability to send fractions of a cent allow individuals to easily pay to not see adds, to pay for content, and to sell their data

[27] A smart contract is a program that specifies (terms and conditions are laid out in the software) an agreement between two parties which normally include a transfer of value.

[28] Counterparty risk is the probability that actors involved in a transaction will default and not be able to complete their contractual obligation. A risk a German bank painfully discovered when Lehman Brothers filed for insolvency. [70]

Practical men, who believe themselves to be quite exempt from any intellectual influences, are actually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back.

John Maynard Keynes

4

Digital Currencies, Data Collection and Economic Policy

Regardless of their implementation, CBDCs collect real time transaction information that could be useful to economists and policy makers. This feature of digital currencies is often over-

looked by the current literature. [1] The data collection aspect of digital currencies is exciting for two intertwined reasons. First, it allows for the collection of more accurate economic information, which is currently notoriously unreliable. [10] Second, the data collection process itself has vast potential that is currently under-studied. The creators of digital currencies are making design choices that determine what type of data is collected and thus made available. This chapter is preoccupied with the former: the inadequacy of the current economic data, and the potential of using CBDC transaction data to make the allocation of scarce resources more efficient.[2]. Chapter 5 is preoccupied with the latter.

4.1 Notoriously Unreliable and Untimely

On the 1st of January 2001, Greece joined the European Union (EU), and three years later, it admitted to having misled European authorities by misrepresenting significant economic data. The Greek government claimed that their budget deficit was 2.0% while in reality it was 4.1%, which was over the 3% threshold that was needed to enter the EU. [25] Similarly, Chinese cities and provinces have admitted to tampering with economic growth data to impress the Central government. [32]

In both of these situations sub-optimal decisions were made. In one situation, a country was allowed to join a monetary union it probably should not have joined. [3] [48] In the other situation, much-needed subsidies were diverted because individuals and local governments wanted to impress their superiors. History is rife with examples of agents manipulating their data for the simple reason that they are incentivized to do so. A common example is that individuals and companies are incentivized to mislead tax authorities whenever possible all around the world. In fact, tax returns in Italy, Spain, and Greece do not mirror income patterns. [10] Moreover, companies are incentivized

[1] An exception being the PBoC.[63]
[2] After all economics is the study of scarce resources.
[3] The EU and the EURO are partially to blame for the severity of the Greek Debt crisis.

to hide information from competitors and to manipulate profits to pay fewer dividends. [10] Even governments, as previously discussed with Greece and China as examples, are incentivized to falsify data. [76]

Not only is economic data susceptible to manipulation but it is also notoriously challenging to measure. This difficulty is emphasized by the substantial revisions a number of macro-economic indicators undergo. [4] This is further underlined by the vast economic literature that exists on the subject; in particular, attempts to construct more accurate indicators as well as the effects of revisions on policy making. [5] [28, 29] More accurate indicators are self-evidently to be preferred yet the benefits are difficult to quantify. Nevertheless, work that quantifies the effects of revisions on economic policy-making indicates that initial data is not always accurate enough. Had revised data been used initially it would have led to different policies being implemented. [68, 84]

As a result there is a strong case to be made for the systematic collection of more timely and accurate (including less susceptible to manipulation) data. In the next section we will elaborate on why transaction data could aid in said endeavor.

4.2 Live Transaction Data

An economy is simply the aggregation, the sum, of all the individual micro-transactions made. Therefore, although computationally and algorithmically challenging, the most accurate way of understanding an economy is to parse through each and every transaction. Having the possibility to process every transaction in 'quasi-live' time has a number of benefits for economists and policy makers. The data is more accurate, more timely, and therefore less susceptible to revisions. Furthermore, misreporting and manipulation is less likely to occur because actors have to incur a cost

[4] GDP, Employment, etc indicators are revised multiple times.

[5] The creation of more accurate indicators currently relies on the use of big data e.g. using google trends [34]

if they want to manipulate the data. The benefits of having said data is clear, but is it a feasible endeavor to parse through this data and extract valuable insights?

Understanding the value and feasibility of analyzing transaction data is challenging as the literature is predominantly recent. Aladangady et al. 2019 "successfully filter, aggregate, and transform card transactions into economic statistics" making the case that transaction data is not only theoretically useful but that it can be used to extract useful macro-economic indicators. [5] This point is further underlined by the large number of hedge funds that buy credit and debit card data from AMEX, Mastercard and Visa. [35] [6] Transaction data can be used for macro-economic indicators as well as for the study of particular sub-sections of society: Alganday et al. 2016 used daily transaction aggregates to better understand how consumer spending is affected by unforeseen events such as hurricanes. [6] The ability to rigorously study a subset of society, including those commonly overlooked, will allow for better and more targeted economic policies. [7] To conclude, the transaction data market in the US is estimated to be currently worth between $1 and $5 Billion dollars, indicating the value of such data even in pre-CBDC societies.[8] [74]

If correctly executed, there is a strong case to be made that transaction data provided by CBDCs would be both quantitatively and qualitatively superior to the existing toolbox economists can play with.

[6] Some hedge funds supposedly analyze transaction data in real time. [35]

[7] In a cashless economy transaction data would be even more insightful. For example it would aid in our understanding of poverty and how individuals behave in relative and absolute poverty. As explored in the 2019 Economics Nobel Prize Winning work Banerjee & Duflo 2011 which illuminate how eliminating poverty requires observation and random control trials (Michael Kremer was also co-awarded the 2019 Economics Nobel Prize).

[8] This estimate includes using transaction data for marketing purposes.

There were 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days.

Eric Schmidt, Executive Chairman of Google

5

Ethereum

Ethereum is the second most popular cryptocurrency, accounting for approximately 10% of the $250 billion cryptocurrency market cap. [1] It is fundamentally different from other cryptocurrencies because not only is it censorship resistant and non-inflationary like Bitcoin but it also allows individuals to create smart contract and DApps. This has resulted in the development of a diverse

[1] For more information regarding Ethereuum visit https://github.com/ethereum/wiki/wiki/What-is-Ethereum, https://ethereum.org/what-is-ethereum/, https://blog.ethereum.org/2014/08/18/building-decentralized-web

ecosystem of users, use cases and applications which are not present in other cryptocurrencies.

5.1 Motivations and Limitations of Using Ethereum?

In the absence of a CBDC to analyze, Ethereum was used as a proxy. [2] Ethereum was chosen instead of other cryptocurrencies for two reasons. Although one can argue, Ali & Narula 2020 and do that simpler systems should be implemented as they have a smaller surface of attack, the added functionality of Ethereum-like cryptocurrencies outweighs the security risks. In particular, we do not believe the security risks are that high given the possibility of a hard fork in the case of an emergency, a solution that the Ethereum community used after the DAO hack in 2016. [3] [95] The idea that future CBDC will be similar to Ethereum is further underlined by current developments in the cryptocurrecy universe. Libra, a cryptocurrency recently launched by Facebook with CBDC aspirations, is similar to Ethereum. It draws significantly from Ethereum but it has a number of significantly different and unique features - most notably it is a permissioned blockchain. The analysis of data collected by Libra would be an interesting endeavour. However, only the test-net is live as it has faced severe push back from governments around the world. [69, 71] Second, the Ethereum blockchain is used by a variety of different actors. This has led to particularly rich and diverse data that is not available on any other platform. The proliferation of DApps ranging from complex collectible games like CryptoKitties to a whole ecosystem of Decentralized Finance (DeFi) creates the most realistic snapshot of what a CBDC data may look like.

[2] Some people believe that CBDCs may be built on top of Ethereum [1]

[3] The DAO hack occurred in 2016 when an attacker was able to drain 3.6million ether (approximately $50million) from a Decentralized Autonomous Organization (DAO) thanks to flaws in the contracts that were used to build the organization. After the hack and after much discussion the Ethereum community decided to hardfork the Ethereum ecosystem. A hardfork happens when 51% or more of the nodes decided to collude and either go back to a previous block nullifying all transactions that happened after that block including the hack or adopt new software. This process can be repeated if future vulnerabilities are discovered. Although not ideal, this mechanism to turn back time makes blockchain systems ironically more secure in our eyes.

5.2 Technical Specifications

Ethereum consists of a decentralized virtual machine that can execute programs commonly called smart contracts. These are written in a Turing complete byte code language called Ethereum Virtual Machine (EVM) Bytecode. [4] [5] Each program consists of a permanent store in which to save data and a set of functions that can be invoked by either individuals or other programs. Programs and individuals can own ether which they can send to other individuals or programs. Ethereum allows for two types of transactions: plain value transfers and contract executions. Plain value transfers are simply the movement of ether from one account to the other. Contract executions, however, are transactions with associated EVM (Ethereum Virtual Machine) code. In this work plain value transactions will be referred to as transactions and contract executions will be referred to as traces. The further subtleties of transactions and traces are covered in Section 6.2.

Executing a program has a cost that is proportional to the computational complexity of the program. This ensures that programs will always terminate. This feature of Ethereum allows for its existence as it basically eliminates the risk of malicious or incorrect programs running forever and bringing the Ethereum platform to a halt. Individuals use an internal currency made of ether called gas to pay for the execution of a program or programs. An individual has to specify *a priori* the maximum amount of gas it is willing to spend on said execution. If too little gas is specified the execution will fail and the gas is not redeemable.

As transactions transfer value it is paramount that their execution is performed correctly. Therefore, each transaction is processed by a decentralized network according to a consensus protocol - which is currently a Proof of Work (PoW) Protocol. [6]

[4] EVM is a quasi Turing complete state machine because all executions are finite.

[5] Smart contracts are generally programmed in a higher level language like Solidity which compile to EVM byte code.

[6] The Ethereum Foundation has announced plans to shift to a Proof of Stake Protocol (PoS).

5.3 General Ethereum Metrics

A basic understanding of macroeconomic Ethereum indicators is necessary to engage with this paper effectively. This allows for the development of some intuition vis-a-vis trends the Ethereum network is seeing, it is also related to the macroeconomic signals that economic policy makers, in particular central banks, frequently consider. It is important to note that most cryptocurrencies including Ethereum are U.S. centric, the majority of Ethereum nodes are in the US as displayed by Figure 5.1. [7] [8] [9]Therefore using ETH/USD as an exchange rate and the S&P 500 as an economic indicator make the most sense. [10]

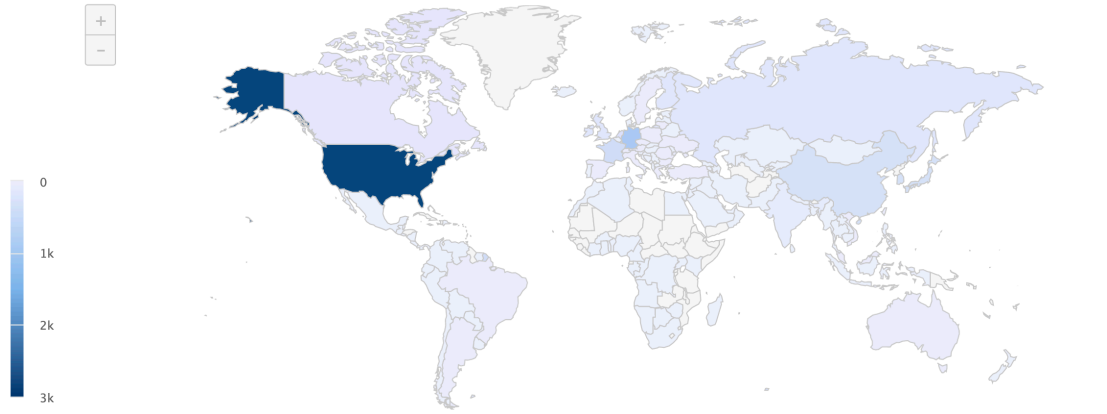

Figure 5.1: Geographic Distribution of Ethereum Nodes[11]

[7] As discussed in Section 3.2 most of the financial world is heavily centered around the US and the USD

[8] There are a number of different types of nodes for the purpose of this paper a node can be defined as an entity that acts as a communication point and may perform different functions

[9] With sufficient resources it is possible to use de-anonymization techniques to figure out the IP address of large proportions of individual wallets [43, 97]

[10] Although the US is the most prominent region one could argue that the most accurate indicators to use would be Ethereum activity weighed indicators. Instead of using ETH/USD or the S&P 500 one would use synthetic values weighed by how prominent each region is. Given the relative dominance of the U.S. and the purpose of this case study such accuracy was not necessary.

[11] Image from https://etherscan.io/nodetracker

The summary statistics graphs displayed in Figure 5.2 represent a selection of signals chosen and created by the author of this work that he believes are particularly relevant to understanding the Ethereum network. They reflect four distinct trends. Ethereum on-chain statistics like activity were extracted from the Ethereum Google Big Query Dataset [41]; Ethereum Price and S&P data was taken from Yahoo Finance. [102]

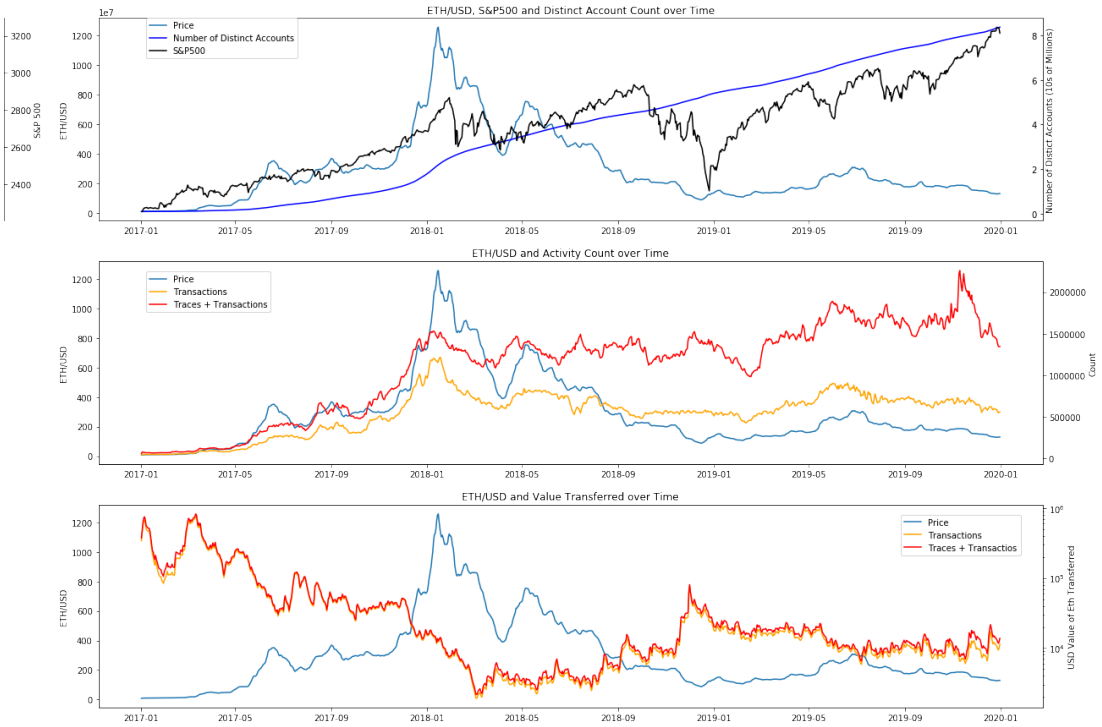

Figure 5.2: Summary Statistics of Ethereum Trends [12]. The top graph displays Ethereum Price, S&P 500 and the number of distinct accounts over time; The middle graph displays Ethereum Price, Transaction Activity and Trace plus Transaction activity by count; The bottom graph displays Ethereum Price, Transaction Activity and Trace plus Transaction activity weighed by value

As a first observation that should inform the understanding of this data, note that there was a speculative price bubble towards the end of 2017 and beginning of 2018. The speculative bubble affected all cryptocurrencies. At its peak, the cryptocurrency market cap reached $800Bn. A few

[12] Number of Distinct Account Data came from Etherscan and S&P data came from Yahoo finance.

months later it lost three quarters of its value. [40] In line with most speculative bubbles it had long lasting consequences. In particular, it attracted significant numbers of new users, some of which remained on after the bubble collapsed. It also significantly changed the type of activity on Ethereum.

As a second observations, we hypothesize that users have become increasingly sophisticated. Here we define more sophisticated as users who are more knowledgeable of the Ethereum blockchain which can be measured by how intricate their transactions and traces are. For example, a user who simply sends transactions versus a user that uses smart contracts and potentially writes smart contracts. In this particular case the increase in sophistication is highlighted by the increase in trace activity relative to the increase in transaction activity. As shown in figures 5.3 the percentage of transactions as a percentage of overall movements of ether have fallen both by number and by value. Indicating an increase in the overall use of decentralized applications and smart contracts. [13] The difference in transaction count percentage over transaction value percentage underlines that the increase in trace activity is most likely caused by an increase in the use of decentralized applications instead of individual smart contracts. That is because DApps tend to increase the number of traces by count more than by value more as they tend to involve a large number of small interactions. [14] On the other hand smart contracts tend to produce a smaller amount of traces but move more value across the ecosystem.

A third observation, with the exception of the speculative bubble, the number of addresses has been increasing fairly linearly. Therefore, either the number of users has been increasing, the number of addresses per user has been increasing, or both. It is unlikely that any one actor created a significant amount of addresses without using them because it costs approximately $0.10 USD to create an address. The cost of creating a new address that shows up on the blockchain is equivalent

[13] Decentralized applications use smart contracts but smart contracts do not necessarily have to be part of decentralized applications.

[14] For example, interacting with the game CryptoKitties involves buying and trading collectibles all which create multiple traces. Just using a smart contract that executes after a certain time or given certain conditions tends to produce a smaller number of traces.

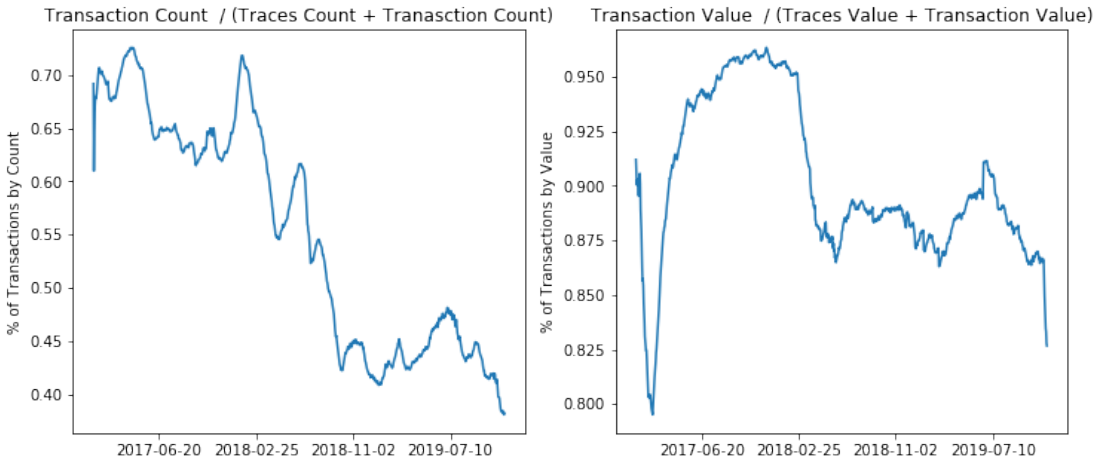

Figure 5.3: Trace vs Transaction Trends: The left graph displays the proportion of Ethereum transaction over total movements. The right graph displays the proportion of Ethereum transaction over total movements weighted by the value.

to the transaction cost since in order for an address to show up on the blockchain it has to take part in a transaction. Transaction costs are approximately $0.10 subject to some daily volatility based on how busy the network is at a particular time. Therefore, if an actor is creating and using more addresses they are self-evidently becoming more sophisticated. There are a number of reasons an actor would want to use more addresses that include but are not limited to wanting more privacy.

Fourth, the relationship between Ethereum activity, Ethereum prices, and the S&P 500 indicates that recent lack of correlation between Ethereum price and the S&P 500 may not be sustainable. [15] Displayed by a decoupling between Ethereum Activity, S&P500 and Ethereum price. In more detail Ethereum Activity and the S&P500 have remained correlated throughout the time frame analyzed. On the other hand both the correlation between Ethereum Price and S&P 500 and Ethereum Price and Ethereum activity have decreased over time. Intuitively the price of Ethereum should be correlated to its activity. Moreover since the Ethereum activity is positively correlated with the S&P

[15] This insight was proven to be correct as this paper was being written. Ethereum and other cryptocurrencies faced severe losses in line with the S%P 500 losses during the March 2020 market turmoil.

500 this should imply that the price is also positively correlated. The latter is interesting because it signifies that actors on Ethereum are affected by the US economic status.

Summarizing the insights that were gained from the above graphs. First, the cryptocurrency bubble had long lasting effects on the ecosystem. Second, are becoming more sophisticated. Third, the lack of correlation between Ethereum Prices and S&P500 is suspicious. Graphing general trends although insightful lacks the granularity that is needed to better understand why things happened. For example, what caused the bubble? or which users are becoming more sophisticated? In Chapter 7 we attempt to answer these questions by breaking these signals into their constituent parts.

There were 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days.

Eric Schmidt, Executive Chairman of Google

6

Blockchain Data

Blockchain data is a side effect of blockchain technology. When Nakamoto, Buterin and others were creating their respective cryptocurrencies their main goal was to create a decentralized trustless immutable ledger; not a centralized transaction information database. Blockchains by design distribute the data to any and all nodes, as opposed to a standard centralized system. Data extraction and the subsequent data analysis was not prioritized, this need not be the case in the future as designers will have complete agency, within technical limitations, of how to build a CBDC. There-

fore, understanding the current difficulties in analyzing blockchain data could allow for technical recommendations on how a CBDC should be built.

6.1 Literature Review of Blockchain Analysis

The analysis of blockchain data is relatively novel due to the simple fact that blockchain technology has only recently celebrated its 12th birthday. Nevertheless, it is a burgeoning field as highlighted by the tens of millions of dollars that are being invested in blockchain data analytics startups. [1] [66]. Most of the data analysis which is done by profit seeking companies is centered around two questions: first, understanding and predicting price fluctuations [4, 8] and second, analysing and detecting illicit activity. [56, 60, 64, 67, 92] The existent small body of academic work centers around the same broad questions with little practical focus. A significant portion of published works concentrate on Bitcoin since it has been around for longer. However, given the substantial difference in the purposes of Ethereum and Bitcoin, the structure of the transaction network and the behavior of the agents may be significantly different. [2] [3] Finally, it is important to note that the analysis of blockchain data is not limited to transactions but can include smart contract data as well. [85]

6.2 Blockchain Transaction Data Challenges

Four main challenges arise when analyzing blockchain transaction data.[4] It should be noted that these insights are derived from the analysis of Ethereum. [5] Although, blockchain systems can be

[1] About $100 million USD was raised in the blockchain analytics space in 2019

[2] Transaction network not be confused with P2P network. Which is also an active area of research. [11]

[3] Although, outside the scope of this work, underlining the differences between different types of cryptocurrency transaction networks would be an interesting research question.

[4] Blockchain data is not limited to transaction data. For example, some blockchain systems like Ethereum store information about the smart contracts that are executed on it

[5] Bitcoin was also analyzed; however, the results are not displayed in this paper.

very diverse it is argued that the following four challenges will likely exist in most blockchain systems, especially cryptocurrencies.

6.2.1 Computational Tractability

Blockchain data is challenging and expensive to query given inherent structure of the ledger and the lack of readily available effective and efficient specific querying techniques.

As was alluded to at the outset of this chapter, blockchain data is a side effect. The primary purpose of storing transaction data on a ledger is to ensure the desired properties of a blockchain e.g. resolve the double spend problem without a centralized bookkeeper. Therefore, transaction data tends to be stored on the ledger in ways that are unfriendly to data scientists. Unfriendliness of data, to data scientists is loosely measured by the complexity and computational efficiency of queries.

It is neither cost effective nor simple to directly query the ledger. The more advanced ledger querying tools that exist focus on DApp-to-ledger interactions rather than the extraction of historical data. [6] [83] Therefore, ledger data is generally transferred into centralized databases before it is queried by data scientists. Best practices are still being developed with regards to how the data should be extracted, modeled and stored. [3, 17, 75] Current existing techniques and libraries are based on specifically designed and *ad-hoc* engineered approaches, not suited for general purpose analysis. [17].

In addition, to extract ledger data one has to either run a full Ethereum Node or access an API like Infura, both of which are expensive. [7] Moreover, the storing of the data is non-trivial given its relative size: 1.2Tb and 0.25Tb for Ethereum and Bitcoin respectively. Even if one has access to the data in a queryable database e.g. the Google Big Query cryptocurrency datasets, querying it is

[6] For example Web3j is extremely useful when one is building a DApp on top of a blockchain like Ethereum; however, it lacks some of the desired functionality a data scientist would want.

[7] It costs approximately 100$ per month to run a full node on AWS, and 250$ a month to access Infura Historical Data (100,000 requests a day)

expensive and cumbersome. [8] [41, 42] Use cases, especially nuanced ones, are still being discovered and and as result established ways of accessing them do not yet exist.

Although, this is currently a problem it is likely (with the exception of the cost) that some of these difficulties will be abstracted away as new use cases are tested and techniques developed. [17, 75] Private blockchain data analysis companies like Chainalysis and Cmorq have probably already done so. Moreover, the efficiency of a blockchain at large should be prioritized over the ease at which data scientists can extract data, as long as extracting said data is eventually possible.

6.2.2 Undefined Nature of Transactions

The ability to run code on an blockchain and create 'smart' contracts and decentralized applications severely complicates one's ability to analyze the movement of said blockchain

This section will briefly elaborate on said challenge by outlining the describing the challenge in the Ethereum environment. This challenge is significantly reduced in Bitcoin like blockchains which do not allow for traces.

Section 5.2 briefly explored the different types of transactions that are permitted on Ethereum: plain value transfers referred to as transactions; and contract executions referred to as traces. Defining a transaction is simple: Bob transfers 2 ETH to Alice today. Defining the movements of ETH related to a trace is more complex. Bob decides to create a contract today that transfers money to Alice in ten days only if it rains and only upon his receipt of a confirmation by Alice that she recognises that it has indeed rained. Should one count the transaction today? in ten days? what happens if it doesn't rain and Bob gets the money back? This trivial example serves to illustrate the complexity of defining movements of ether. We turn to the technical specifications to better understand how the Ethereum developers decided to define these nuances. Looking at the Ethereum specifications

[8] Google Big Query charges 5$ per Tb of data analyzed. Expenses are highly dependant on the type of data a researcher is trying to aggregate.

we see that all movements of ETH are recorded. In other words someone analyzing this interaction would notice 10 ETH leaving Bobs account today and in ten days he would notice 10 ETH either entering Bobs account or Alice's account. Without an analysis of the smart contract one would not be able to understand the nature of the transaction.

In order to better understand the nuances of analyzing traces we will briefly elaborate on the state of affairs in Ethereum. Transactions can be directly extracted from the Ethereum ledger. Traces need to be extracted from a full record of the EVM upon the execution of every transaction. Traces are always initialized by a transaction (or by a trace whose ancestor was a transaction). To get a comprehensive and accurate trace data set one needs the state of the EVM at any moment in time and the series of EVM bytecodes that were executed at the subsequent moments. This is abstracted by the actual nodes into a construct aptly termed a trace. [9] Traces are subdivided into the following four categories [10]:

- **Create**: Trace executed to deploy a smart contract. Can be initiated by an individual or by another smart contract.
- **Call**: Trace executed to transfer money or messages through different Ethereum accounts.
- **Suicide**: Trace that can be executed by a smart contract at the end of its existence. Causes the smart contract to delete its code and refund the value to a specific account.
- **Reward**: Trace that miner receives when they mine a block.

The trace abstraction allows us to see the movements of ETH but cannot distinguish traces from each other at a more granular level than the 4 types just defined. Traces and their respective

[9] The abstraction of EVM bytcodes to traces is far from perfect and is dependant on the implementation client being run on the node. Underlined by the following bug report which indicates that 1 million traces were not accurately captured by one of these implementation clients. The main node protocols are Geth and Parity.[33]

[10] https://geth.ethereum.org/docs/dapp/tracing

programs emit. logs which can be analyzed to better understand the purpose of a particular trace. Software exists to translate raw hex log data into higher level code; however, general software that categorizes either the hex log data or the higher level code into types of traces does not exist. Instead it generally requires a human touch making the general analysis of traces challenging. Developers can choose to emit events. Events are constructs that log when certain events or series of events have taken place. They are often used by DApp developers to change front end interfaces.

The deciphering of traces is of upmost importance to gain a better understanding of how ether is moving. Machine learning techniques are being developed to classify the log data into types of traces; however, we posit that the developers could have enforced an event like piece of data describing the trace. Although, it is unclear how it would be enforced, creating a standard for trace type would enable the more granular analysis of traces.

6.2.3 Miscellaneous Nature of Accounts

Miscellaneous nature of identities. Not only is it challenging to deanonymize addresses but even connecting addresses owned by the same actor proves challenging.

Most cryptocurrencies that currently exist today are pseudo-anonymous and allow individual users to create multiple accounts. This makes extracting signal from the data particularly challenging since the large number of arbitrarily used accounts causes a lot of noise. It should be noted that on some blockchains like Bitcoin it is advised never to use an account more than once. [88] The following two types of grouping techniques exist.

- Off-Chain Deanonimization Attacks:

  These involve using metadata leaks to connect accounts to physical people or locations. The most common form of metadata that is used to try and deanonomize accounts is the senders IP address. [67] Other forms of metadata have also been used for example online cookie data

that is collected when individuals use cryptocurrencies to buy products online. [56] The effectiveness of said techniques is unclear and dependant on both the specific blockchain and the security precautions that individuals take. Once you have the physical identities of accounts it is relatively easy to group addresses.

- On-Chain Grouping Attacks:

  These involve using on-chain data and heuristics to estimate which accounts are connected. A number of creative heuristics have been developed ranging from stylometry on smart contracts [72] to specific transaction behavior. For example: if an account sends all of its value down to the last fraction of a cent to another account it is reasonable to assume both accounts belong to the same individual. [56, 64] The effectiveness of said techniques are also unclear and severely dependant on the type of blockchain they are used on. The success of these tactics can be particularly vulnerable to coin joins.[11] [80]

Most of the successful deanonimization tactics described above have been tested on Bitcoin. It is unclear whether Ethereum is more or less vulnerable to deanonimization attacks; however, transposing attacks that worked on Bitcoin to Ethereum is generally not feasible. [43]

From a data science perspective it would be a lot simpler if identities were known and individuals were only allowed one account. For privacy reasons that are discussed in the conclusion this is most likely not feasible. The degree of anonymity that may be allowed for CBDCs is very unclear. Future CBDC will most likely have to abide by Know-Your-Customer (KYC) regulations but what that entails from a technical perspective and the effects that has on the ability to analyze the transac-

[11] A tactic whereby multiple users send coins to a single account who then forewards it to a different set of account. Coinjoin contracts like Tornado Cash and the utilization of said contracts are becoming more common

tion data is uncertain. [12] [13] [96] What is clear though is that a discussion needs to be had about the trade-offs between data and privacy.

[12] The PBoC claims that its CBDC will have controllable anonymity, it is not clear what that entails and the technical specifications are not currently available.

[13] SOV, the third party contracted to build the Marshallese CBDC, is planning on building a pseudo-anonymous blockchain. KYC regulations are respected by only allowing individuals to access the system with a third party. As a result the third parties store the maps between accounts and individuals.

It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories instead
of theories to fit facts.

Sherlock Holmes

7

Ethereum: A Case Study

Although these issues are fascinating from a theoretical perspective, an empirical analysis often sheds light on details that would have otherwise gone unnoticed. To that end, it would be beneficial to study a CBDC. Unfortunately, these does not yet exist. In the absence of a better alternative, Ethereum was used as a proxy. Motivations for the use of Ethereum and the subsequent limitations were covered in Chapter 5. The primary goals of this case study are twofold, to gain a

practical understanding of the challenges of extracting signals from pseudo-anonymous transaction data, and to develop a technique to extract sector-based macroeconomic signals.

One way, among many, to better understand an economic indicator is to break it down into its constituent parts. For instance, in order to better understand the origins and consequences of the financial crash of 2008, one must look beyond the S&P 500 and examine each sector of the economy. By looking at each sector, one notices discrepancies and irregularities such as the housing market bubble. In a similar vein, the overall signals of Ethereum can be broken down into sector-based signals.

In the absence of known entities and sectors the following methodology was used: A number of websites were scraped for known Ethereum accounts which were then manually classified into 6 sectors. Transaction data summary statistics of those accounts was extracted from the Ethereum ledger and used to train a classifier. The classifier was then used to classify unknown accounts. Sector aggregate information about the activity of those accounts over time was then extracted, resulting in signals akin to sector-based macroeconomic signals.

The final results are similar to the summary statistics displayed in chapter 5 but broken up by actor type. Which allow us to better understand the development of the Ethereum Ecosystem.

7.1 Data Used

Due to the challenges of dealing directly with ledger data described in 6.2.1 the Ethereum datasets hosted on Google Big Query were used. [1] When the Ethereum blockchain had to be accessed directly, it was done through Infura, an API that allows quick access to the Ethereum blockchain. We mainly used Infura to make sure the statistics we were extracting from the Google Big Query were

[1] The datasets were stored in a relational way and therefore easy to access with SQL. Google Big Query charges users by size of data analyzed by query and therefore some queries in particular can be expensive. For more information visit: https://cloud.google.com/blog/products/data-analytics/ethereum-bigquery-public-dataset-smart-contract-analytics

correct.

[2] Generally, this paper only analyzed data between 01/01/2017 - 01/01/2020.

7.2 Constructing A Small Labelled Data Set

As was discussed in section 6.2.3 most cryptocurrencies including Ethereum are pseudo-anonymous. Therefore, Etherscan and Twitter were scraped for known accounts which were then classified into one of the following six sectors: decentralized exchange (DEX), exchange, game, initial coin offering (ICO), individual, and miner. [3] Etherscan is one of the better known Ethereum block explorers. A block explorer is a search engine that allows individuals to easily extract basic information from a particular blockchain. For example, it identifies whether a transaction has been validated. Etherscan also contains a variety of useful metrics and information. In particular, it contains a rolodex of known accounts and their owners [4] Although extremely useful, Etherscan's rolodex lacks individual accounts which make up an important portion of the Ethereum ecosystem. Therefore, data extracted from Twitter was used to supplement the dataset. [5] Owing to the technique used to scrape Twitter, the accounts collected mostly belong to Ethereum developers and investors. As a result, they may not be an accurate reflection of the average individual that uses Ethereum. [6] Moreover, it is very likely that these experts have multiple accounts and therefore only parts of their overall activity is observed. In total, scraping twitter and etherscan produced 733 accounts.

As mentioned in 6.2.3, grouping addresses significantly reduces the amount of noise in a dataset. Due to the lack of known successful deanonimization heuristics for the Ethereum blockchain, general grouping algorithms were not used in this work. [43] However, of the accounts collected

[2] https://infura.io

[3] Broad categories were chosen due to the relative small size of the final data set and because a number of actors are not well defined at a more granular level.

[4] For information regarding the Etherscan rolodex, see Appendix A.1.1

[5] For detailed information regarding how Twitter was scraped. Appendix A.1.2

[6] A BFS of depth 3 was launched on the friends of Vitalik Buterin, see Appendix A.1.2

some belonged to the same actors and were therefore grouped. [7] [8]

The process of grouping together accounts that belong to the same actor compressed 733 accounts into 509 actors. Of these 509 actors those that had negligible activity were omitted. For this project, an account has negligible activity if it satisfies one of the following two conditions: either (I) having been part of less than 50 transactions in the given time frame, or (II) having transferred less than 1ETH (100USD) and being part of less than 10 transactions in the given time frame. This reduced the dataset from 509 actors to 437. These actors are omitted because their activities are negligible and omitting them helped to increase the signal-to-noise ratio of the data-set. Appendix A.3 includes more information regarding classes and the class breakdown.

7.3 Training a classifier

The goal of the classifier is to accurately predict the type of an account based off of transaction summary statistics.

We extracted summary statistics for the 437 addresses from Google Big Query. [9]. The summary statistics included features like: Average Size of Incoming Transaction, Average Number of Transactions Received in a Day, etc. For a detailed overview of feature engineering and classifiers tested visit A.3 and A.4.

When choosing a classifier we were extremely conscious of over-fitting. We therefore opted for a classifier that was less susceptible to over-fitting, interpretable and easy to modify. As a result, this project employs the One-vs-All Multinomial Logit Classifier because it performed similarly to other models we tested on the training and testing set but was more interpretable.[10]. Figure 7.1 displays

[7] Grouping accounts creates some changes in how information regarding a group is collected. For information regarding how grouping effects transactions summary statistics, data used by classifier section 7.3, see Appendix A.3.3

[8] For information regarding how grouping affected the accuracy of classifiers Appendix A.4

[9] For detailed overview of feature extraction and engineering see Appendix A.3

[10] For more information regarding the classifiers and how the classifiers were chosen see Appendix A.4

the magnitude of the statistically relevant features, p-values ≤ 0.05. [11] Figure 7.1 is comforting because the results are in line with the intuitive predictions. Moreover there are a few interesting observations. For example, an individual generally has less transactions a day and an exchange generally has more receivers.

[11] For this project, the significance level is set at 0.05.

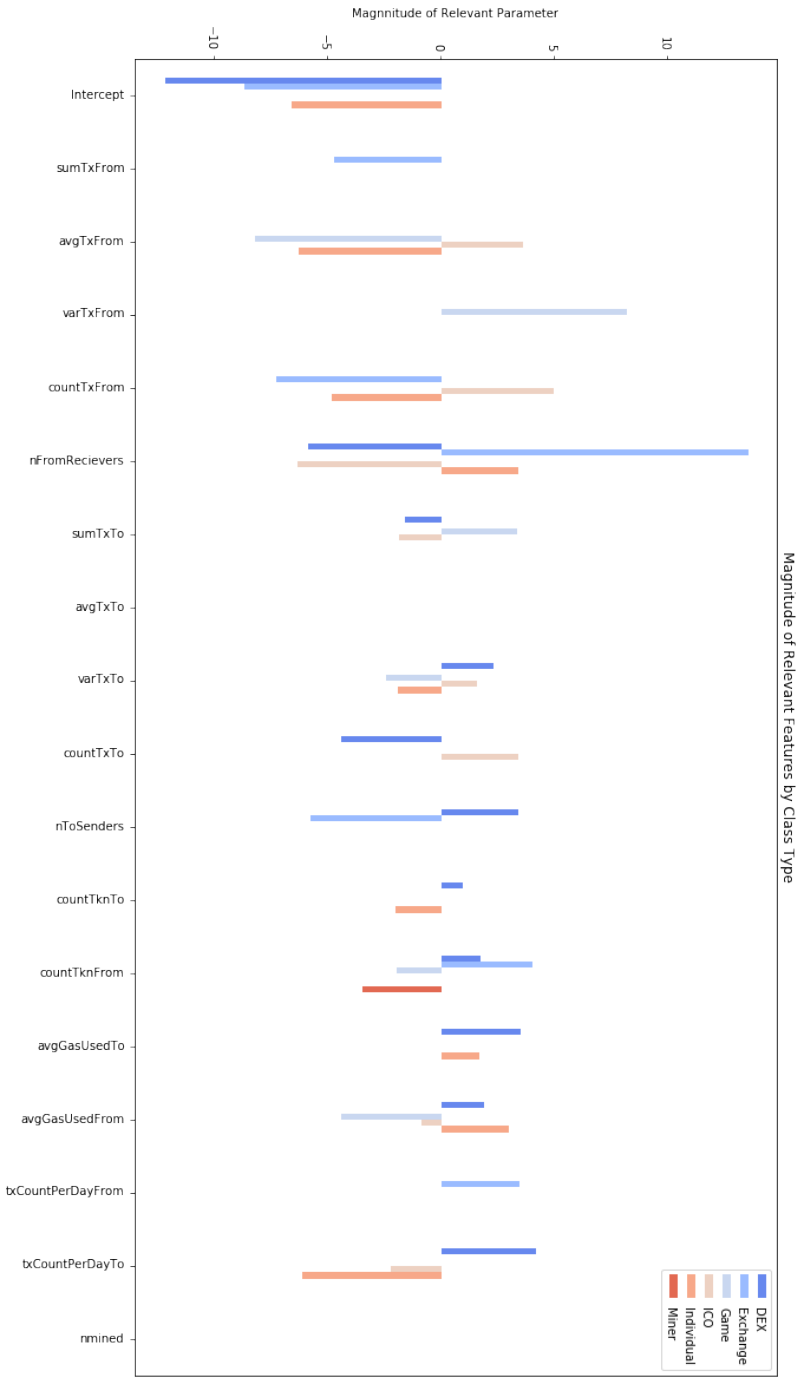

Figure 7.1: Magnitude of Statistically Relevant Features - One-vs-All Multinomial Logit

The concern for over fitting led to a calibration of the model such that although its overall accuracy fell slightly its precision increased. For information about this calibration turn your attention to Appendix A.5.[12] The final accuracy of the model on the test set was 0.68 (instead of 0.81 without calibration).

7.4 Unknown Accounts

Once a classifier is trained and appropriately calibrated it can then be used to classify the unlabelled data set. [13] Using the classifier on all sixty million Ethereum accounts would be expensive and futile since 1% of the wallets make up for 62% percent of value transfer and 69% percent of transfers by count. [14]. Therefore, we only used our classifier on the top 1 million accounts (1.7%) sorted by ether received and sent. Table 7.1 displays the results of the classifier in terms of category proportions and provides a significant addition to current Ethereum literature. Note not all accounts were classified because of how we calibrated the classifier, see chapter 7.3 and Appendix A.5

[12] The calibration involved a successful scoring functions, unsuccessful reliability curves

[13] The classifier was trained on all of the training data. Contrarily to the previous sections where a test set was withheld.

[14] These statistics were calculated using data from the Google Big Query Dataset

Table 7.1: Class Proportion in Ethereum

Type	Proportion
DEX	6.72%
Exchange	1.28%
Game	0.04%
ICO	15.69%
Individual	41.03%
Miner	1.05%
All	58.98%

It is important to note that the percentages reported belong to the top 5% of Ethereum accounts. If one were to take into account all 60 million accounts it is likely that these values would be diluted - the exception being individuals. The results are in line with expectations given equation: **??**.[15]

$$\text{Total Accounts} = \text{Average number of Accounts Per Actor in Specific Category} \times \text{Number of Actors in Category} \tag{7.1}$$

For example, we expected there to be a large number of individuals albeit unclear whether we would be able to accurately classify them given the nature of individuals in the training data. In addition, taking the average exchange as having around 10 addresses the above results would indicate that there are 1280 exchanges.[16] Due to the dilution mentioned earlier it is unlikely that a large proportion of the remainder of the addresses belong and are used by exchanges and thus we do not need extrapolate the proportion across all 60million addresses. An estimate made in 2018 put the number of cryptocurrency exchanges at about 500 which, including growth, is in line with the 1280 we supposedly classified. [94] Using similar calculations Games is inline with intuition but there

[15] Using this equation assumes the classifier can pick up on the variety of different accounts an actor has, which is most likely not true.

[16] Based off of our labelled dataset this is on the generous side. However, it is very likely that the exchanges we had information on use accounts that were not in our dataset

may be a bit too many Miners, ICOs and DEXs.

A possible extension could have been to retrain the classifier on the newly labelled data set and use it either to further classify points or to interpret the differences between newly classified points and existing labelled points by producing the equivalent of Figure 7.1.

7.5 PCA

Dealing with multidimensional data is challenging because it is hard to visualize. In an attempt to better understand whether the classifier was accurate we decided to use Principle Component Analysis.[17] Our results are displayed in Figures 7.2, 7.3, 7.4, 7.5. Before we go over the results we will briefly explain the process as there is some nuance. PCA identifies the axis that account for the largest, second largest, etc amount of variance in a dataset.[54] Therefore, projecting and visualizing the data onto the first few components of PCA can often allow us to visualize clusters. Note that the axis created are dependent on the data set used. In our case we extracted axes for both the labelled and unlabelled dataset. [18] Once we have the axis we can project other datasets on to them, as long as they have the same features.

Figure 7.2 and Figure 7.5 underline the difficulties of trying to separate the data. In particular, the lack of shape and separation of the data points in Figure 7.5 outline the potential low signal to noise ratio of the data. Figure 7.4 underlines that with the exception of the loose cluster of data points in the bottom right corner the majority of the data points in the unlabelled data set are not radically different from the ones with the labelled data set. Moving to Figure 7.3 we notice that the labels that are predicted tend to group together. This is an expected behavior but may be a weakness

[17] This is generally a good idea because of the Manifold hypothesis which is explained in Appendix **??**

[18] It is important to standardize the data set before one performs PCA if not the features with the largest magnitude will be weighted more favourably. In addition balancing does affect PCA analysis for more information see Appendix **??**

[19] Data points that were not classified were omitted from this projection.

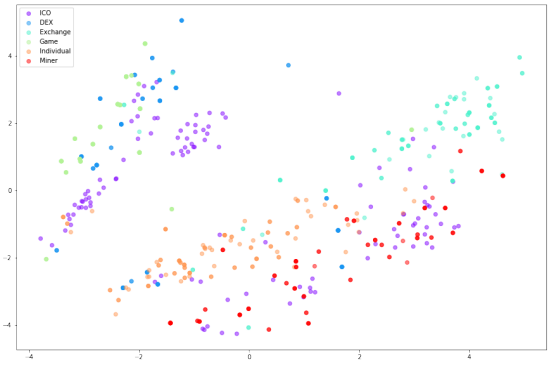

Figure 7.2: Labeled Dataset on Labeled Dataset PCA Axis

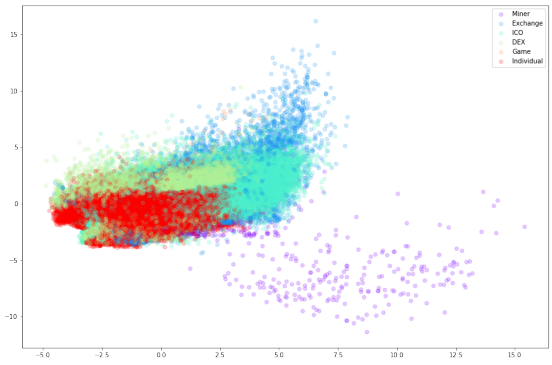

Figure 7.3: Unlabeled Dataset with predicted Labels on Labeled Dataset PCA Axis

19

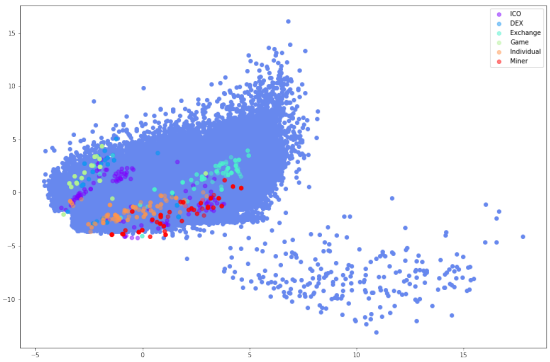

Figure 7.4: Labelled Dataset and Unlabeled Dataset on Labeled Dataset PCA Axis

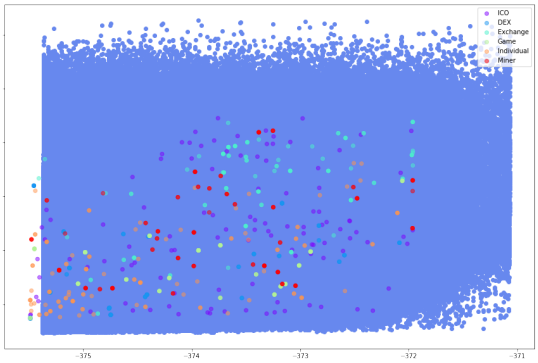

Figure 7.5: Labelled Dataset and Unlabeled Dataset on Unlabeled Dataset PCA Axis

of the classifier given the data does not seem to be easily separable. Finally, still from 7.3 we can see that the classifier is most likely to miss classifying miners as it is unlikely that the loose cluster in the bottom right are miners given how distant they are from the known miners in image 7.4

7.6 Group Level Indicators

Next, we extracted sector-based signals by aggregating account transaction information. This was done for both the labelled and unlabelled data sets so that the resultant contrast may provide possible insights. Deciding which signal/signals to extract was challenging, since some metrics like activity are very noisy and susceptible to being dominated by a small number of accounts, we settled for the daily balance, Equation 7.2, of the accounts given it is the most concrete metric.

$$b_{i+1} = b_i + I_i - O_i \tag{7.2}$$

Where b_i is the balance at day i, I_i are the trace inflows at day i and O_i are the trace outflows at day i. [20]

Figure 7.6 and 7.7 display the daily balance for the known data set; and 7.12 displays the daily balance for the unknown dataset. Figures 7.6 and 7.7 have the following interesting characteristics - note that the following insights are limited as they are only derived from around 700 accounts. First, the increase in the ICO signal seems to precede the price bubble indicating that the ICOs may have been one of causes the bubble happened. The fact that multiple ICOs cause this peak rather than a small number as can be seen in 7.7 further emphasize this point. ICOs generate a lot of advertisement and buzz which could have lead to an influx of actors. More research would have to be done with regard to the cryptocurrency universe since the bubble was not an Ethereum specific phenomena.

[20] The query that was used to calculate the balance is not perfect since Transactions were used instead of traces.

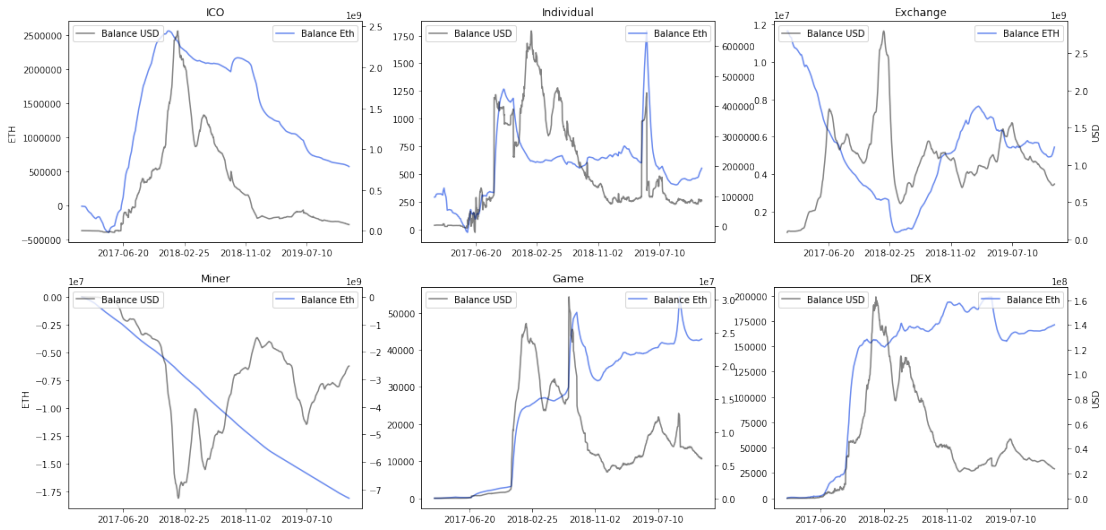

Figure 7.6: Ethereum Aggregate Trends Known Dataset: each graph shows the total amount of Eth and USD held in Eth over time by a subsection of accounts who's type is listed at the top of each graph.

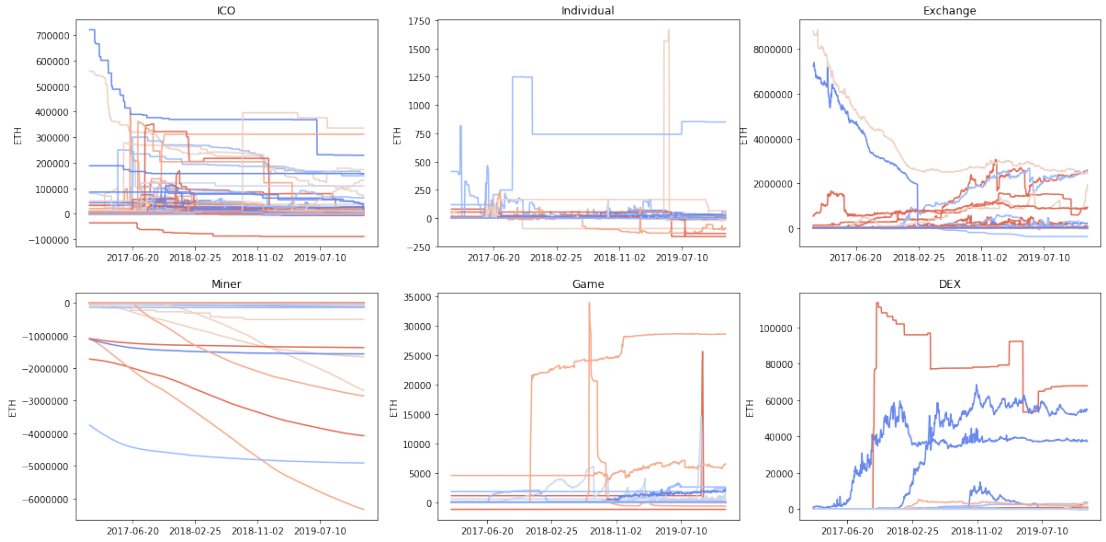

Figure 7.7: Ethereum Individual Trends Known Dataset: each graph shows the amount of Eth and USD held in Eth over time by a specific accounts subsection of addresses listed who's type is listed at the top of each graph.

Second, the Individuals sector is mainly made of up of small not very active accounts. Which underlines that the addresses posted on Twitter are most likely not the main account of these individuals. The spikes in balance in Figures 7.6 are caused by either Vitalik Buterin or Josh Johnson moving ETH around (an arbitrary Ethereum developer) see Fig 7.10. [21] It is unclear why Josh Johnson decided to move a quarter of a million dollars worth of ETH in the summer of 2019. Based off of his twitter activity around that time one could hypothesise that those funds were used to fund some open source projects. This invasive example was shown to underline the potential lack of privacy that can exist.

Third, the decrease in balance of exchanges before the bubble and the subsequent rise is counter intuitive - see Figure 7.6. One would expect more advanced Ethereum users to not store their Ethereum on an exchange and the opposite to be true for less advanced users. One exception being actors that previously used Bitcoin as they would have already had experience with cryptocurrencies and the associated technological challenges of storing them. Another explanation can be derived by looking at the equivalent DEX graph. In other words perhaps the creation of DEXs cannibalized exchange activity. This hypothesis is strengthened by Etherdelta, see Figure 7.8. On the same figure the spike in Airswap was caused by its ICO. Revealing some of the challenges of classifying data into sectors.

Fourth, The query that was run did not capture reward traces which is why some of the miners have negative balances. The slight kink in the otherwise relatively constant downwards sloping line reflects the reduction of mining rewards with the Ethereum Constantinople upgrade. On the 28th of February 2019, mining rewards were reduced from 3ETH to 2ETH. [38] Note that most of the mining actors are mining pools.

Fifth, actors of type Games tend to not be very successful for very long with the exception of CryptoKitties. Activity metrics may have been a better metric for this particular actor type.

[21] Some analysis of whether the individual sector had an advantage in predicting the price of ether was conducted on this data set. From the movements in this very restricted data set the notion that these experts may be better at buying and selling cryptocurrencies was not validated.

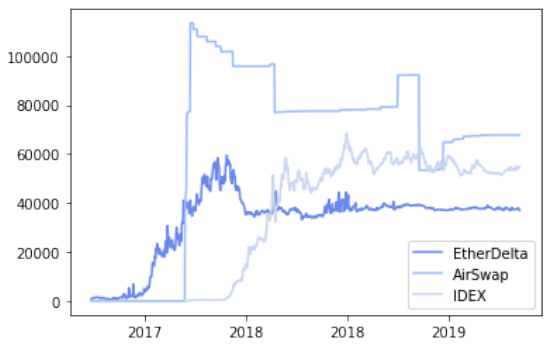

Figure 7.8: DEX Specific Trends

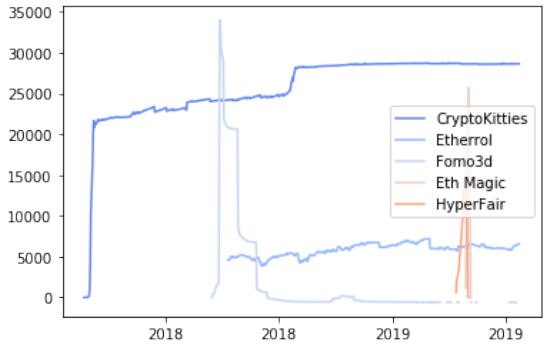

Figure 7.9: Game Specific Trends

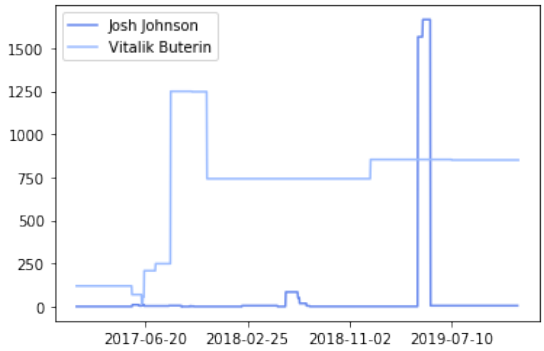

Figure 7.10: Individual Specific Trends

Figure 7.11: Sector Specific Trends: each graphs shows the total amount of Eth and USD held in Eth over time by a individual accounts who's type owner is listed in the legend and type is listed below the graph.

Sixth, Decentralized Exchanges are dominated by EtherDelta and IDEX.

All of the insights are subject to the weakness of the relatively small size of the data set which is why we trained a classifier in the hopes that a larger dataset may provide more robust and accurate signals. The signals created by the larger data set are however susceptible to miss classified accounts.

Looking at the Fig 7.12, it was not possible to graph the equivalent of 7.7 because that would entail plotting tens of thousands of lines in some cases and because querying the individual daily activity of an Ethereum account is not cost effective. The main insights and difference from 7.12:

First, there is a pretty large difference between the ICO signals in Figures 7.6 and 7.12. The negative value is likely due to the extended use of traces by ICO actors. Although volatility is to be expected with ICOs the size of the steps are surprising. Furthermore, we are not sure we can explain the general decrease. More information vis-a-vis ICO is necessary.

Second, the Individuals signals are strikingly different. The positive slope of the line could indicate a larger number of expert individuals using Eth. Which is intuitively correct as we expect more people to have joined the platform and the ones who were already on it have potentially become more sophisticated as we hypothesized earlier in this paper.

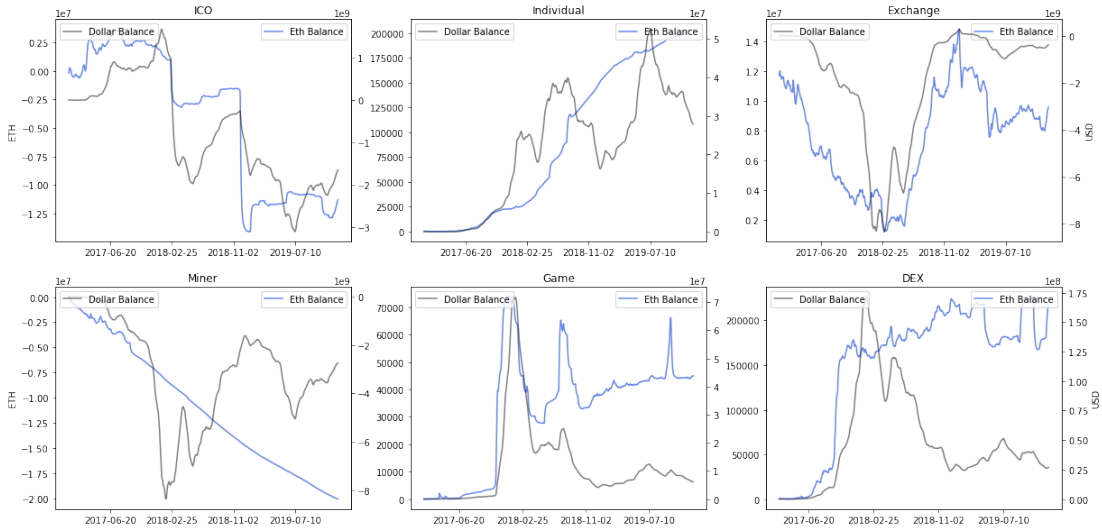

Figure 7.12: Ethereum Aggregate Trends Classifier Labels: each graphs shows the total amount of Eth and USD held in Eth over time by a subsection of accounts who's type is listed at the top of each graph.

The noticeable differences between using only the smaller data set and the the augmented data set underline the necessity for techniques that extract useful information from the unknown data points.

7.7 Potential Improvements

There are a number of things that could be done to extract more accurate sector-based statistics.

- **Richer features:**

  The features the classifier was trained on were extremely noisy, skewed and correlated with each. This was further underlined by lack of clusters in the PCA results. It is not clear how one would go about extracting more features albeit there are some possibilities e.g. smart contract data, trace type and embedding of neighbour graphs.

- **Larger Initial Data Set:**

  The initial data set was created by scraping Etherscan and Twitter. A more extensive search for addresses could be attempted.

- **Grouping Heuristics and De-anonymization attacks:**

  Using grouping heuristics on both the labelled and unlabelled data sets would most likely increase the classifiers accuracy.

- **Classifier Improvements:**

  The accuracy of the final classifier was reduced severely due to over-fitting fears. A more extensive research and calibration of classifiers may be useful. Although, we believe the other areas are more likely to lead to greater improvements.

- **Signal Extraction:**

  A number of metrics could have been further extracted that might have lead to interesting insights. For example, number of daily active users etc.

7.8 Conclusion of Case Study

Extracting more granular data from pseudo-anonymous transaction information is challenging but feasible. With more research and the implementation of some of the recommendations of Section 7.7 this method could reveal key insights about the development of the Ethereum, and potentially other cryptocurrencies. There are some privacy concerns which will be addressed in the conclusion.

Civilization is the progress towards a society of privacy. The savage's whole existence is public, ruled by the laws of his tribe. Civilization is the process of setting man free from men.

Ayn Rand

8

Conclusion

Our understanding of privacy has significantly changed in the last few decades and especially in the last twenty-four months. Collecting and analyzing data has never been easier and with that infringing on traditional understandings of privacy. This was brutally underlined by the 2018 Cambridge Analytica Data Scandal. Data about individuals is no longer only being used for targeted commercial marketing but also for social engineering. The extent to which Cambridge Analytica succeeded in social engineering is unclear; however, this attempt in itself highlights the very real threats of so-

cial engineering in the near future. [1]

Transaction information is particularly sensible and vulnerable to social engineering for a number of reasons. Firstly, unlike social media data or other forms of digital trails we leave behind, it is costly to obfuscate and manipulate. In other words, transaction data represents, to a greater extent than other data, a ground truth of individuals' preferences and behaviors. Unless one is willing to incur a cost, one cannot simply cover up his or her true preferences and behavior. The accuracy of transaction data is frightening: Target, the American retailer, was able to predict that a teenage woman was pregnant before her parents knew merely by analyzing her acquisitions. [44] Secondly, transaction information can be used to actively discriminate and incriminate people. You most likely don't want your health insurer to know that you are a chain smoker. Similarly, one may not want the government to know one is underpaying taxes. [2] Moreover, digital transaction data will only become more accurate as societies transition to cashless economies, since the data of a greater percentage of transactions will become easily accessible. Finally, transaction information is extremely vulnerable to linkage attacks [79]. For example, one could link anonymized transaction data with anonymized subway usage data using quasi-identifiers to not only potentially deanonymize users but also to gain a scarily accurate representation of individuals. [98] As a result, society as a whole should carefully evaluate who should have access to transaction data. The ramifications of this data falling into the wrong hands could be (and has been) dire. One important consideration to bear in mind is that governments expect to have access to said data or parts of said data to identify and prevent illicit activity. How does society balance an individual's right to privacy with society's need to enforce its laws and regulations? This dilemma is, unfortunately, beyond the scope of this essay. However, it is an important question that needs to be revisited as new technologies including, but not limited to,

[1] Cambridge Analytica or what was left of it claims that they were very successful. Given it is in their best interest to laud their work, I am skeptical. Worrylingly, Facebook has not released the data necessary to evaluate these claims.

[2] Various estimates put the tax cheat rate at 80-95% for people who employ baby-sitters , housekeepers or health aides.

CBDCs are drastically undermining our ability to control privacy. [3]

We hope to have convinced the reader that Central Bank Digital Currencies may become a reality and that if they do they have the potential to reshape the fabric of society. Similar to Roth's call on Economist to be Engineers we call on Engineers to be Economists and Philosophers as the systems they may create will dramatically change how we live our lives. [93] This work has humbly attempted to make additions to the existing literature. It attempts to partially re-frame the discussion surrounding CBDCs to incorporate the technology underlining them and the ramifications of the data said technology will collect. In doing so it also proposes a methodology to collect sector-based economic signals from pseudo-anonymous transaction data; and unearths some insights regarding the development of the Ethereum platform.

[3] There is a notion that increases in surveillance increases pressure on an individual to conform. Said notion is commonly referred to as social cooling and for the better or for the worse it is changing society

A

Methodology

A.1 Scraping Labels For Ethereum Accounts

Since Ethereum is pseudo-anonymous labels of known accounts had to be collected.

A.1.1 On Etherscan

The Etherscan Label Word Cloud (https://etherscan.io/labelcloud) and the Etherscan Directory (https://etherscan.io/directory) can be used to find known accounts. The classes that are given by Ethereum are confusing and not always accurate so they have to be manually checked.

A.1.2 On Twitter

Scraping Twitter is a complicated endeavor the API (https://developer.twitter.com/en/docs) was used. Since the API only allows a certain number of requests per 15 minutes the techniques used were throttled (Purposefully slowed down - in an ideal world one would parallelize). Account labels were collected by crawling through the names and bios of twitter accounts in search of human readable Ethereum addresses which are provided by the Ethereum Name Service (ENS) (https://ens.domains/). ENS names can be compared to website naming: instead of using '0xd8da6bf26964af9d7eed9e03e53415d37aa96045' one uses 'vitalik.eth'. A breadth first search (BFS) algorithm of depth 3 on the friends of Vitalik Buterin, the founder of Ethereum (@Vitalik.Eth) was used. [1]. Because the BFS was launched from Vitalik Buterin the accounts collected belong to experts and therefore their activities may not reflect the activities of the majority of individuals on Ethereum. Moreover, it is very likely that these experts have multiple accounts and therefore only parts of their overall activities are observed. [2] A second methodology was attempted to try and collect accounts of individuals that were not experts. Accounts that were posted in response to fraudulent posts e.g. "ETH TOKEN GIVEAWAY". [3] were collected. However, not only

[1] Due to the Twitter API request restrictions the BFS was interrupted a few times and therefore some individuals may have been missed. It took three days to run the BFS even only up to depth 3

[2] Interesting observation: 15% of the ENS accounts that were collected from twitter did not map to an Ethereum accounts. Which would stipulate that some individuals are just adding .eth after twitter their name without actually using ENS.

[3] https://twitter.com/DappCentre/status/1240390239438409728

was this technically challenging due to Twitter's API restrictions but also unproductive since a large proportion of the accounts being posted had negligible activity. Therefore, this methodology was not used.

A.2 Class Descriptions and their Respective Breakdown

Table A.1: Class Descriptions

Type	Example Actor	Details
Decentralized Exchange (DEX)	EtherDelta	Exchanges that do not require a third party bookkeeper. Instead smart contracts are used to connect buyers and sellers. No couterparty risk is preset.
Exchange	Poloniex	Regular Exchanges.
Game	CryptoKitties	Games, mostly collectible games.
ICOs	Augur	Probably the most diverse class. Includes decentralized Applications like Maker and Auger that have released tokens as part of their application. The purpose.
Individual	Vitalik Buterin	Ethereum Developers or Investors.
Miner	AntPool	Mining Pool addresses.

Table A.2: Breakdown of Classes

Type	Count	Percentage%
DEX	30	6.87%
Exchange	86	19.68%
Game	23	5.25%
ICO	132	30.21%
Individual	117	26.77%
Miner	49	11.21%
All	437	100%

A.3 Transaction Summary Data and its respective Features and Feature Engineering

This section outlines the nature of the features collected including a description of the features, how grouping affected feature extraction, and the use of transformations on the features.

A.3.1 Transaction Summary Data Feature Definitions

Table A.3: Feature Description

Features	Description of Features
sumTxFrom	The sum of all outgoing transactions
avgTxFrom	The average size of an outgoing transaction.
varTxFrom	The variance of outgoing transactions.
countTxFrom	The number of outgoing transactions.
nFromRecievers	Number of different addresses Ethereum was sent to from this particular address.
sumTxTo	The sum of all incoming transactions.
avgTxTo	The average size of an incoming transaction.
varTxTo	The variance of incoming transactions.
countTxTo	The number of incoming transactions
nToSenders	Number of different addresses Ethereum was sent from to this particular address
countTknTo	The number of incoming Tokens.
countTknFrom	The number of outgoing Tokens.
LifeSpan	Number of Days between first transaction and last transaction within timeframe.
nmined	Number of blocks mined
txCountPerDayFrom	Average number of transactions sent per day.
txCountPerDayTo	Average number of transactions received per day.
avgGasUsedTo	Average amount of gas used by senders
avgGasUsedFrom	Average amount of gas used when sending a transaction

A.3.2 The Effect of Grouping on Feature Extraction

Grouping accounts changes how some features are extracted. In particular the features affected were:

1. sumTxFrom, avgTxFrom, varTxFrom, countTxFrom ,sumTxTo, avgTxTo, varTxTo, countTxTo: Movement of ether within group is omitted.

2. countTknTo, countTknFrom : Movement of tokens within group is omitted.

3. nToSender, nFromRecievers : Accounts in group are omitted, and accounts are not double counted (If account A sends ether to two or more different accounts in a group A should only be counted once).

4. LifeSpan: Number of days between first activity of any account in the group to last activity of any account in a group.

The remaining features that were directly extracted from the blockchain can be calculated normally. The features that were engineered from the features extracted e.g. txCountPerDayTo can be recalculated normally as long as the upstream features have been extracted correctly.

Note: Querying grouped statistics is particularly expensive in Google Big Query due to the details of how one is charged.

A.3.3 Feature Transformation

The features extracted from the data were particularly right skewed and therefore transformations had to be applied to the data so that a classifier could be appropriately trained. Figure A.1 displays the distribution of features before they are transformed. Figure A.2 displays the distribution of the features after they were subject to $log(log(x + 1) + 1)$ transformation and a standardization (mean removal and variance scaling), the lifeSpan feature was only standardized as the data in its raw form was not right skewed. Figure A.2 represents the data that was ultimately used to train the classifiers, subject to rebalancing.

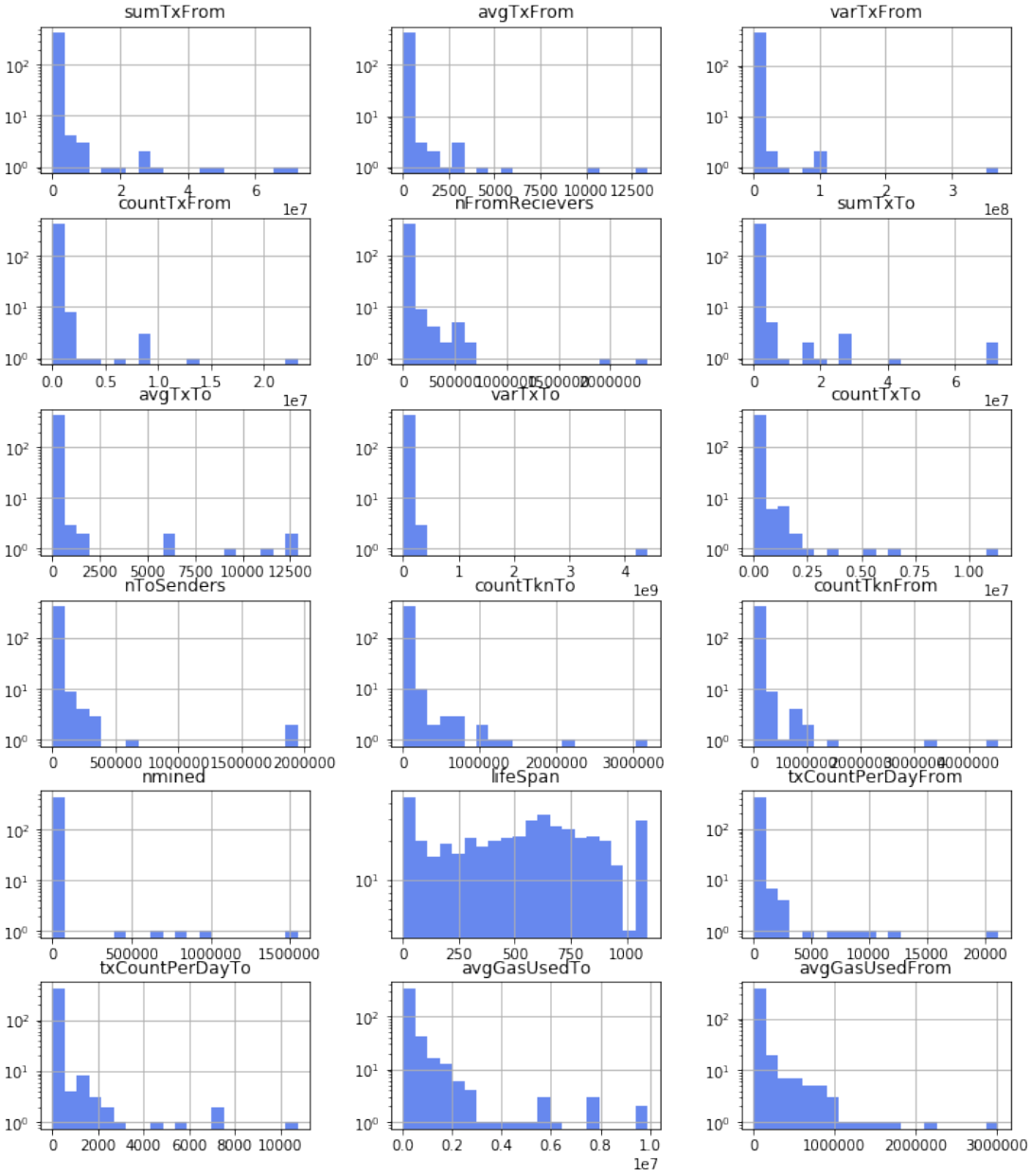

Figure A.1: Histograms of Transaction Summary Features before Transformations

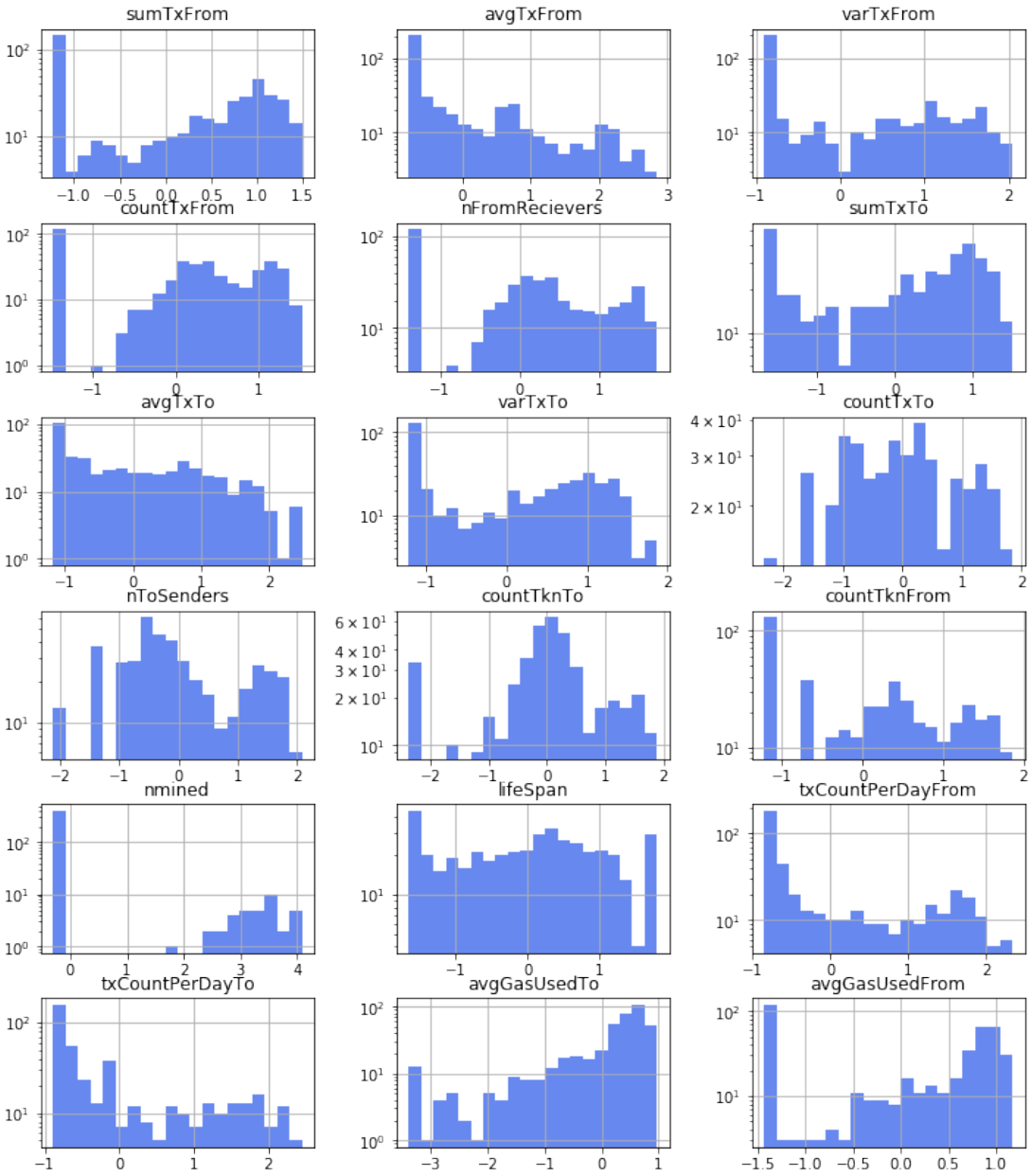

Figure A.2: Histograms of Transaction Summary Features after Transformations

A.3.4 Balancing

As you can see from A.2 the data set was extremely unbalanced therefore we used random up-sampling with replacement. Moreover, balancing was very effective on increasing the accuracy of

the predictors as displayed in A.5. It also made all-vs-one Multinomial Logit classifier more interpretable. Figure A.3 is the equivalent of 7.1 on the unbalanced data set; if you compared the two images you will notice that balancing the data set brings out some significant p-values that are intuitively reasonable. For example, a smaller average transaction size for games.

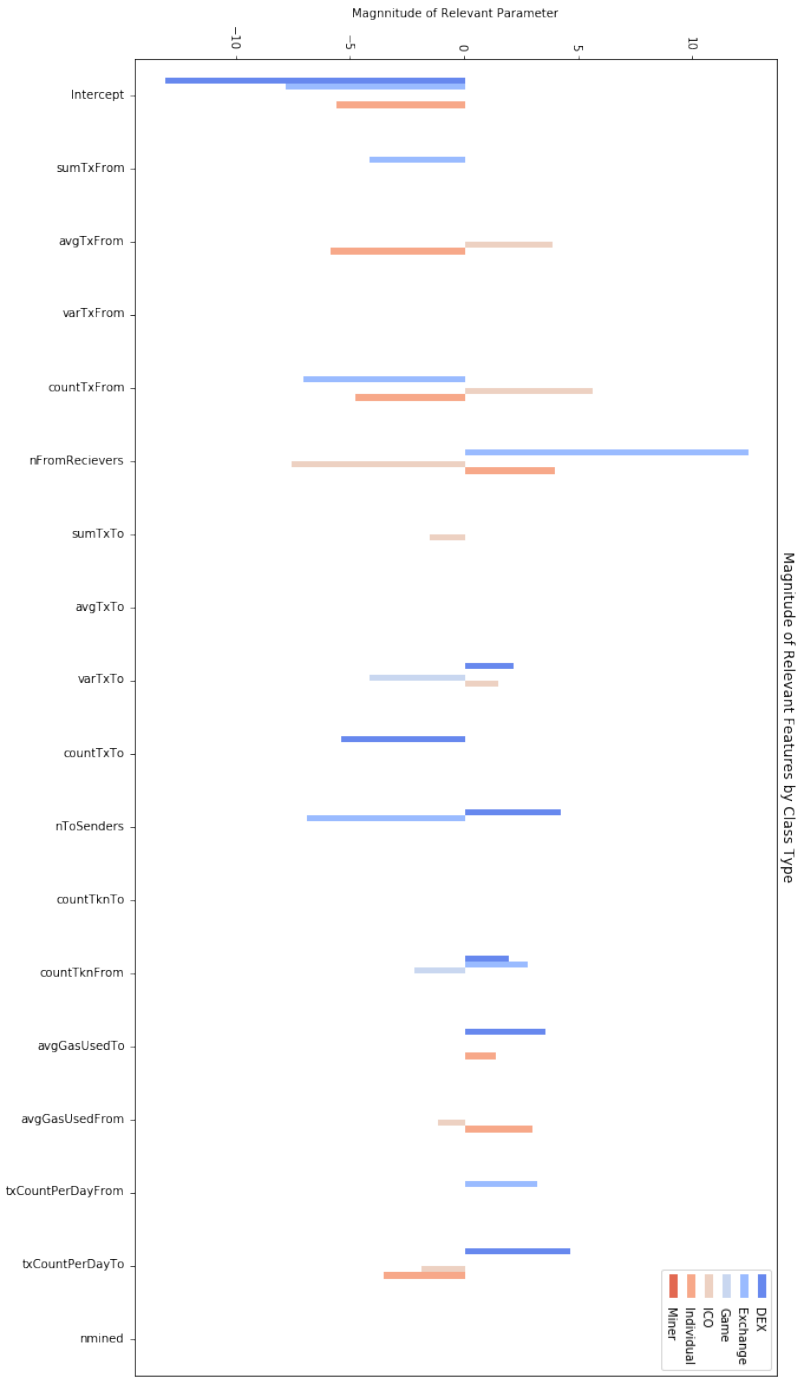

Figure A.3: Magnitude of Statistically Relevant Features- One-vs-All Multinomial Logit Unbalanced

A.4 Classifier Selection

Table A.5 illustrates the effects of transformations on various different classifiers. Table A.5 describes those transformations. The transformations and classifiers shown represent an indicative sample of techniques that were tested. Table A.5 illustrates that by transforming, grouping and balancing the data set classifiers got better underlining the importance of these manipulations.

[4]

Table A.4: Data Transformations Descriptions

Transformation	Details
Standardization (S)	Mean removal and variance scaling.
Transformation (T)	Performs $\log(\log(x+1)+1)$ on all features except LifeSpan.
Grouping (G)	Groups together accounts that belong to same actor.
Balancing (B)	Resample smaller classes until dataset is perfectly balanced.

Table A.5: Classifier Test Set Accuracy subject to Data Manipulation and Trasformations

Classifier	S	TS	TSG	TSGB
Random Forest Classifier	0.81	0.84	0.87	0.94
DecisionTreeClassifier	0.73	0.73	0.81	0.96
Logistic Regression Multinomial	0.37	0.67	0.70	0.75
Logistic Regression One-vs-All	0.47	0.60	0.75	0.81
Logistic Regression One-vs-All Calibrated [5]	NA	NA	NA	0.61
K Nearest Neighbour Classifier	0.59	0.59	0.65	0.82

[4] https://www.statsmodels.org/stable/generated/statsmodels.discrete.discrete_model.MNLogit.html
[5] For more information regarding the calibration see Appendix A.5

Scikit-Learn Library was used for Random Forest Classifier, Decision Tree Cliassifier, and K Nearest Neighbour.[6] While stas model was used for the Logistic Regression. [7]

[6] https://scikit-learn.org/stable/

[7] https://www.statsmodels.org/stable/index.html

A.5 Multinomial Classifier Calibration

As we mentioned in the main section of this paper we were particularly worried about over fitting therefore not only did we chose a classifier that was more rigid and therefore less prone to overfitting but we also attempted to calibrate the model. Although, not all of the methods attempted were successful they were all telling. Three main calibration methods were attempted: bespoke scoring function, reliability curves and physical calibration using PCA analysis. We only used the bespoke scoring function.

A.5.1 Scoring Function

Since we are using a one-vs-all predictor we ended up with 6 classifiers one can modify the scoring function that selects which class to choose. The most common scoring function is displayed by equation A.1

$$= argmax(p_{Dex}, p_{Exchange}, p_{Ico}, p_{Individual}, p_{Miner}, p_{G}ame) \tag{A.1}$$

Although, a reasonable way of selecting which class to choose we decided to use the following scoring rule as it is even less susceptible to overfitting.

$$= \begin{cases} argmax(P) & L(P, t, 2) < 2 \text{ AND } max(P) < t \\ -1 \end{cases} \tag{A.2}$$

Where $P = [p_{Dex}, p_{Exc}, p_{Ico}p_{Ind}, p_{M}, p_{G}ame]$ and $L(P, t, n)$ if the count of arguments above t is below n than return TRUE else return FALSE. Following Figure A.4 shows the accuracy and the precision of the two scoring rules. Where precision defined as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositives} \tag{A.3}$$

The scoring rule developed had a higher precision and therefore was used. A threshold of 0.5 was set.

[8] Scoring rule 1 refers to A.1 and Scoring rule 2 refers to Equation A.2

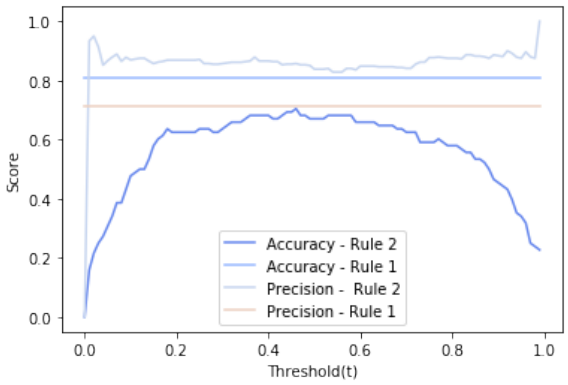

Figure A.4: Scoring Rule Comparison [8]

A.5.2 Calibration Curves

We attempted to use calibration curves but based of A.5 it appears that we were not using enough data for the calibration curve to be useful. In particular we believe that the underlying inbalance in the data is what caused the reliability curves not to be useful underlined by the spikes in the graph especially for the smaller classes.

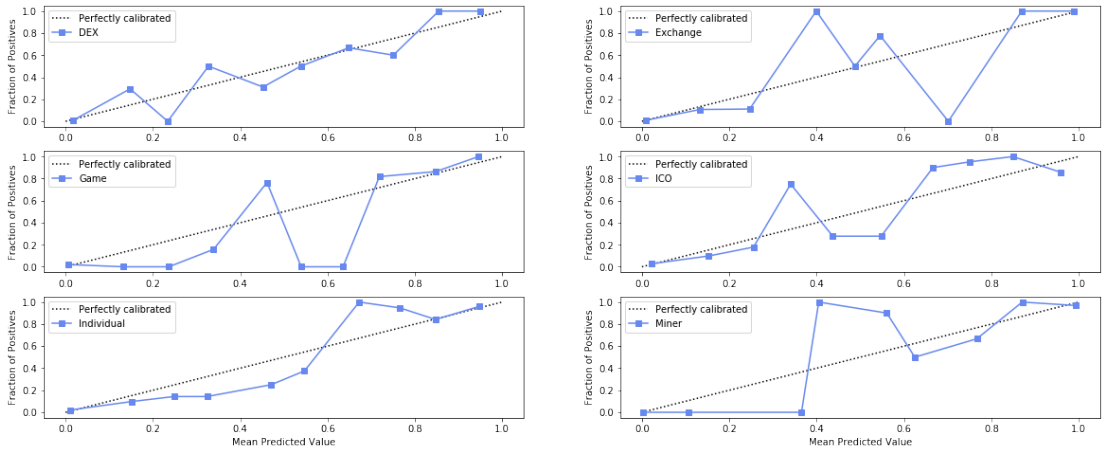

Figure A.5: Reliability Curves

References

[1] Adams, R. S. (2020). Fedcoin on ethereum (market monday - lite). https://bankless.substack.com/p/fedcoin-on-ethereum-market-monday-bc2,Accessed[04/03/2020].

[2] Adrian, T. & Griffoli-Mancini, T. (2019). The rise of digital money.

[3] Akcora, C. G., Dixon, M. F., Gel, Y. R., & Kantarcioglu, M. (2019). Blockchain data analytics. *Journal of IEEE Inteligent Information*, 20(1).

[4] Alabi, K. (2017). Digital blockchain networks appear to be following metcalfe's law. *Electronic Commerce Research and Applications*, 24, 23–29.

[5] Aladangady, A., Shifrah Aron-Dine, W. D., Feiveson, L., Lengermann, P., & Sahm, C. (2019). From transactions data to economic statistics: Constructing real-time, high-frequency, geographic measures of consumer spending. *Finance and Economics Discussion Series 2019-057. Washington: Board of Governors of the Federal Reserve System*.

[6] Alganday, A., Dine, S. A., Dunn, W., Feveison, L., Lengerman, P., & Sahm, C. (2016). The effect of hurricane matthew on consumer spending. *FEDS Notes. Washington: Board of Governors of the Federal Reserve System*.

[7] Ali, R. & Narula, N. (2020). Redesigning digital money: What can we learn from a decade of cryptocurrencies? *Digital Currency Iniative (DCI). MIT Media Lab*.

[8] Athey, S., Parashkevov, I., Sarukkai, V., & Xia, J. (2016). Bitcoin pricing, adoption, and usage: Theory and evidence. *Stanford Institute for Economic Policy Research (SIEPR) Working Paper*, (17-033).

[9] Back, A. (2002). Hashcash - a denial of service counter-measure.

[10] Bagus, P. (2011). Morgenstern's forgotten contribution: A stab to the heart of modern economics. *American Journal of Economics and Sociology*, 70(2), 540–562.

[11] Bailey, M., Kim, S. K., Ma, Z., Mason, J. J., Miller, A., & Mural, S. (2018). Measuring ethereum network peer.

[12] Baily, M. N., Litan, R. E., & Johnson, M. S. (2008). The origins of the financial crisis. fixing finance series. *Brookings*.

[13] Banerjee, A. V. & Duflo, E. (2011). Poor economics: A radical rethinking of the way to fight global poverty.

[14] BankofEngland (2015). One bank research agenda.

[15] Barker, J., Clayton, E., Dyson, B., & Meaning, J. (2018). Broadening narrow money: monetary policy with a central bank digital currency. *Bank of England Staff Working Paper*, (724).

[16] Barontini, C. & Holden, H. (2019). Proceeding with caution a survey on central bank digital currency. monetary and economic department. bank for international settlements paper no. 101. *Bank for International Settlements*.

[17] Bartoletti, M., Bracciali, A., s Lande, & Pompianu, L. (2017). A general framework for blockchain analytics. *Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*.

[18] Benes, J. & Kumhoff, M. (2012). The chicago plan revisited. *IMF Working Paper*.

[19] Berg, C., Davidson, S., & Potts, J. (2018). Beyond money: Cryptocurrencies, machine mediated transactions and high frequency bartering.

[20] Bordo, Michael D; Levin, A. T. (2017a). Central bank digital currency and the future of monetary policy. *NBER Working Paper Series*, (23711).

[21] Bordo, Michael D; McCauly, R. N. (2017b). Triffin: Dilemma or myth>. *Bank of International Settlements*, (684).

[22] Broadbent, B. (2016). Central banks and digital currencies. *Speech to the London School of Economics*.

[23] Buterin, V. (2014). Ethereum whitepaper: A next generation smart contract & decentralized application platform.

[24] Callesen, P. (2017). Can banking be sustainable in the future? a perspective from danmarks nationalbank. *Speech at the Copenhagen Business School 100 years celebration event, Copenhagen*.

[25] Carassava, A. (2004). Greece admits faking data to join europe. https://www.nytimes.com/2004/09/23/world/europe/greece-admits-faking-data-to-join-europe.html[Accessed: 02/02/2020].

[26] Carbaugh, R. & Hedrick, D. (2008). Losing faith in the dollar : Can it remain the world's dominant reserve currency? *Challenge*, 51(3), 93–114.

[27] Carney, M. (2019). The growing challenges for monetary policy in the current international monetary and financial system.

[28] Chang, A. C. (2018). The fed's asymmetric forecast errors. *FEDS Working Paper*, (026).

[29] Chang, Andrew C; Li, P. (2015). Measurement error in macroeconomic data and economics research: Data revisions, gross domestic product, and gross domestic income. *Finance and Economics Discussion Series*.

[30] Chaudhuri, K. N. (1999). *The English East India Company: The Study of an Early Joint-Stock Company 1600-1640*. Routledge.

[31] Chaum, D. (1985). Security without identification: Transaction systems to make big brother obsolete. *Commun. ACM*, 28(10), 1030–1044.

[32] Chen, Y. & Woo, R. (2018). Another chinese city admits 'fake' economic data. https://www.reuters.com/article/us-china-economy-data/another-chinese-city-admits-fake-economic-data-idUSKBN1F60I1, Accessed[02/02/2020].

[33] Chiplunkar, A. (2019). Traces missing for oog calls to builtin contracts #10156. https://github.com/openethereum/openethereum/issues/10156, [Accessed: 04/01/2020].

[34] Choi, H. & Varian, H. (2008). Predicting the present with google trends.

[35] Cohan, P. (2018). Mastercard, amex and envestnet profit from $400m business of selling transaction data. https://www.forbes.com/sites/petercohan/2018/07/22/mastercard-amex-and-envestnet-profit-from-400m-business-of-selling-transaction-data/#6fed29467722,[Accessed: 02/24/2020].

[36] coil (2019). https://coil.com/about,[Accessed: 02/24/2020].

[37] Comben, C. (2020). Bahamas races ahead with its 'sand dollar' digital currency. https://bitcoinist.com/bahamas-races-ahead-with-digital-currency/ [Accessed: 02/02/2020].

[38] consensys (2019). The thirdening: What you need to know. https://media.consensys.net/the-thirdening-what-you-need-to-know-df96599ad857, Accessed[04/03/2020].

[39] Constancio, V. (2017). The future of finance and the outlook for regulation. *Remarks at the Financial Regulatory Outlook Conference organised by the Centre for International Governance Innovation and Oliver Wyman*.

[40] Cornish, C. & Murphy, H. (2018). What next after cryptocurrency bubble burst? https://www.ft.com/content/7ed0c3b8-a1f3-11e8-85da-eeb7a9ce36e4,[Accessed: 02/02/2020].

[41] Day, A. & Medvedev, E. (2018). Ethereum in bigquery: how we built this dataset. https://cloud.google.com/blog/products/data-analytics/ethereum-bigquery-how-we-built-dataset, [Accessed: 03/04/2020].

[42] Day, A., Medvedev, E., AK, N., & Price, W. (2019). Introducing six new cryptocurrencies in bigquery public datasets—and how to analyze them. https://cloud.google.com/blog/products/data-analytics/introducing-six-new-cryptocurrencies-in-bigquery-public-datasets-and-how-to-analyze-them, [Accessed: 03/02/2020].

[43] Dijkhuizen, T. & Klusman, R. (2018). Deanonymisation in ethereum using existing methods for bitcoin.

[44] Duhigg, C. (2012). How companies learn your secrets. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html[Accessed: 02/02/2020].

[45] Dyson, Ben; Hodgson, G. (2016). Why central banks should start issuing electronic money. *Positivemoney2016*.

[46] Economist, T. (2019). Venezuela's paper currency is worthless, so its people seek virtual gold. the law of supply and demand is ignored in venezuela but not online. https://www.economist.com/the-americas/2019/11/21/venezuelas-paper-currency-is-worthless-so-its-people-seek-virtual-gold, [Accessed: 02/02/2020].

[47] Engert, Walter; Fung, B. S. (2017). Central bank digital currency: Motivations and implications. *Canada Staff Discussion Paper*.

[48] Feldstein, M. S. (2011). The euro and european economic conditions. *NBER*, (17617).

[49] Ferguson, N. (2008). *The Ascent of Money: A Financial History of the World*. Penguin Group.

[50] Frankel, A. S. (1998). Monopoly and competition in the supply and exchange of money. *Antitrust Law Journal*, 66(2), 313–361.

[51] Fredona, Robert; Reinart, S. A. (2017). Merchants and the origins of capitalism. *Harvard Business School*.

[52] Friedrich, H. (1976). Denationalisation of money: The argument refined.

[53] Frydman, C. (2012). The panic of 1907: Jp morgan, trust companies, and the impact of the financial crisis.

[54] Geron, A. (2017). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*.

[55] Giles, C. (2019). Mark carney calls for global monetary system to replace the dollar. https://www.ft.com/content/a775b55a-c5c2-11e9-a8e9-296ca66511c9,[Accessed: 02/02/2020].

[56] Goldfeder, Steven; Kalodner, H. R. D. A. (2017). When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies.

[57] Gourinchas, P.-O. (2019). The dollar hegemon? evidence and implications for policy makers.

[58] Grabowski, M. (2019). *Cryptocurrencies: A Primer on Digital Money*. Routledge.

[59] Halaburda, Hannah; Fung, B. S. (2016). Central bank digital currencies: A framework for assessing why and how. *Canada Staff Discussion Paper*.

[60] Hamacher, Kay; Katzenbeisser, S. O. M. (2013). Structure and anonymity of the bitcoin transaction graph. *Future internet*, 5(2), 237–250.

[61] Horesh, N. (2014). *Chinese Money in Global Context: Historic Junctures Between 600 BCE and 2012*. Stanford University Press.

[62] Iyer, Rajkamal; Puri, M. (2008). Understanding bank runs: The importance of depositor-bank relationships and networks.

[63] Jinze, E. (2019). First look: China's central bank digital currency. https://research.binance.com/analysis/china-cbdc[Accessed: 02/02/2020].

[64] Jordan, G., Levchenko, K., McCoy, D., Meiklejohn, S., Pomarole, M., Savage, S., & Voelker, G. M. (2013). A fistful of bitcoins: Characterizing payments among men with no names.

[65] Kessler, R. (2008). *The Terrorist Watch: Inside the Desperate Race to Stop the Next Attack*.

[66] Khatri, Y. (2019). Blockchain data analytics firm elementus raises $3.5m in new funding. *The Block Crypto*.

[67] Koshy, P., Koshy, D., & McDaniel, P. (2013). An analysis of anonymity in bitcoin using p2p network traffic. *Financial Cryptography and Data Security, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg*.

[68] Kozicki, S. (2001). How do data revisions affect the evaluation and conduct of monetary policy?

[69] Kuhn, D. (2019). Facebooks new crypto faces scrutiny from european authorities. https://www.coindesk.com/facebooks-new-crypto-faces-scrutiny-from-european-authorities,[Accessed: 04/01/2020].

[70] Kulish, N. (2008). German bank is dubbed 'dumbest' for transfer to bankrupt lehman brothers. https://www.nytimes.com/2008/09/18/business/worldbusiness/18iht-kfw.4.16285369.html, [Accessed: 02/24/2020].

[71] Libra-Association-Members (2019). An introduction to libra.

[72] Linoy, S., Stakhanova, N., & Matyukhina, A. (2019). Exploring ethereum's blockchain anonymity using smart contract code attribution.

[73] Luther, W. J. (2011). Friedman versus hayek on private outside monies: New evidence for the debate.

[74] Mckinsey&Company (2019). Global payments report 2019: Amid sustained growth accelerating challenges demand bold actions. https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/tracking%20the%20sources%20of%20robust%20payments%20growth%20mckinsey%20global%20payments%20map/global-payments-report-2019-amid-sustained-growth-vf.ashx [Accessed: 03/03/2020].

[75] Medvedev, E. (2018). Blockchain etl. https://github.com/blockchain-etl,[Accessed: 03/05/2020].

[76] Michalski, T. & Stoltz, G. (2013). Do countries falsify economic data stategically? some evidence that they might. *The Review of Economics and Statistics*, 95, 591–616.

[77] Murphy, H. & Yang, Y. (2020). Patents reveal the extent of china's digital currency. https://www.ft.com/content/f10e94cc-4d74-11ea-95a0-43d18ec715f5 [Accessed: 02/02/2020].

[78] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[79] Naraynan, A. & Shmatikov, V. (2008). Robust de-anonymization of large datasets.

[80] Nick, J. D. (2015). Data-driven de-anonymization in bitcoin.

[81] of-the Treasury, U.-D. (2011). Legal tender status. https://www.treasury.gov/resource-center/faqs/currency/pages/legal-tender.aspx[Accessed[02/02/2020.

[82] Onge, P. (2017). How paper money led to the mongol conquest: Money and the collapse of song china. *The Independent Review*, (2).

[83] Opensource (2016). Web3j sdk. https://www.web3labs.com/ Accessed[03/25/2020].

[84] Orphanides, A. (1997). Monetary policy rules based on real-time data.

[85] Pinna, A., Ibra, S., Baralla, G., Tonelli, R., & Marchesi, M. (2019). A massive analysis of ethereum smart contracts, empirical study and code metrics.

[86] Polo, M. (2015). *Il Milione*. Rusconi Libri.

[87] Posen, A. S. (2008). Why the euro will not rival the dollar. *International Finance*, 11(1), 75–100.

[88] Project, B. (2009). Protect your privacy. https://bitcoin.org/en/protect-your-privacy,[Accessed: 02/24/2020].

[89] Radford, R. A. (1945). The economic organisation of a p.o.w. camp. *Economica, New Series*, 12(48), 189–201.

[90] Rogoff, K. (2016). *The Curse of Cash*. Princeton: University Press.

[91] Rogoff, K. (2017). Dealing with monetary paralysis at the zero bound. *Journal of Economic Perspectives*, 31(3), 47–66.

[92] Ron, D. & Shamir, A. (2012). Quantitative analysis of the full bitcoin transaction graph. *Department of Computer Science and Applied Mathematics. The Weizmann Institute of Science*.

[93] Roth, A. (2002). The economist as engineer: Game theory, experimentation, and computation tools for design economics.

[94] Sedgwick, K. (2018). The number of cryptocurrency exchanges has exploded. https://news.bitcoin.com/the-number-of-cryptocurrency-exchanges-has-exploded/[Accessed: 04/02/2020].

[95] Siegel, D. (2016). Understanding the dao attack. https://www.coindesk.com/understanding-dao-hack-journalists,[Accessed: 04/01/2020].

[96] SOV-Foundation (2019). The marshallese sovereign (sov): Fair, sustainable money.

[97] Spagnuolo, M., Maggi, F., & Zanero, S. (2014). Bitiodine: Extracting intelligence from the bitcoin network. *International Conference on Financial Cryptography and Data Security*, (pp. 457–468).

[98] Sui, P. & Li, X. (2017). A privacy-preserving approach for multimodal transaction data integrated analysis. *Neurocomputing*, 253, 56–64.

[99] Triggs, A. (2019). Reducing the dominance of the us dollar. *Financial Review*.

[100] WebMonetization (2019). Web monetization explainer. https://webmonetization.org/docs/explainer, Accessed[03/05/2020].

[101] Xuanmin, L. (2020). China's central bank moves closer to issuing digital currency: insiders. https://www.globaltimes.cn/content/1183579.shtml, Accessed[03/27/2020].

[102] Yahoo (1997). Yahoo finance. https://finance.yahoo.com/, Accessed[04/01/2020.