Cryptocurrency Price Evolution Analysis in relation with Complex Social Phenomenons and Network Effects

Author: Sebastian Krawczyk

Internship at Laboratoire d'Informatique en Images et Systèmes d'Information (LIRIS)
Data Mining and Machine Learning research group (DM2L) at LIRIS

Internship supervised by Professor Remy Cazabet (UCBL, LIRIS, IXXI)
Academic supervision by Professor Michel Grabisch (Univ. Paris I, PSE)

A raport presented for the degree of
M2 MMMEF Modélisation et Méthodes Mathématiques en Economie et Finance
Data Science and Optimization Theory Track

UFR 27 Mathématiques et informatique
Université Paris 1 Panthéon-Sorbonne
Paris, France
September 13, 2022

Contents

List of Figures

Acknowledgement

I want to express my deepest gratitude to Professor Remy Cazabet for his ongoing support, patience and very insightful discussions, suggestions, and explanations. Moreover, I would also like to thank him and show my most profound appreciation for the opportunity itself to work on this personally for me fascinating research in Laboratoire d'InfoRmatique en Image et Systèmes d'information. I would like to extend my sincere thanks to the entire teaching staff of the Université Paris I Panthéon-Sorbonne, especially the professors with whom I had classes during the last two years. I would like to offer my special thanks to Professor Jorgen Vitting Andersen, whose research on socio-finance has reshaped the way I perceive the financial markets and has significantly influenced the direction I took while working on this research. Lastly, I must express my very profound gratitude to my parents. Their belief in me has kept my spirits and motivation high, and I am sure I would not be here today without their support. I want to dedicate this work to Professor Didier Sornette, as his work, often reoccurring throughout my literature study, was most foundational for the conducted research.

Abstract

The cryptocurrency industry aspires to create a fully decentralized, alternative financial system using blockchain technology, with the global market capitalization of this market growing significantly in the last few years. Despite recurring bubbles, the adoption of cryptocurrency follows an exponential trend. Using the extended Log-Periodic Power Law Singularity Model, accounting for network effect based valuation, I provide an estimate of the cryptocurrency bubble crash. I also propose an improvement to the existing network effect based valuation framework, based on deanonymization and clustering of Bitcoin addresses. Moreover, using extensive blockchain data that provide transparency of cryptocurrency transactions, I model behavioural strategies of cryptocurrency users. Modelling using Extended Log-Periodic Power Law Singularity provided an estimated crash date consistent with the actual crash. Furthermore, preliminary research on behavioural strategies points to the existence of cryptocurrency-specific calendar effects related to the Halving of rewards of cryptocurrency miners. New feature ideas for behavioural analysis of Bitcoin users are also proposed, consistent with market impact literature.

Introduction

Cryptocurrencies originated in 2009 when financial markets still felt the repercussions of the Great Financial Crisis. Bitcoin (BTC) is a blockchain-based peer-to-peer fully decentralised payment system with embedded constraints to its total supply of money, design characteristic stated at the time of BTC orgin as, philosophically opposing Quantitative Easing policy that was being for the first time implemented in the US shortly before Bitcoin inception. Since then, this technology has been massively adopted, and outstanding efforts have been made to improve the underlying designs to introduce new use cases outside of simple transacting benefits. Narratives around cryptocurrencies throughout the years have been changing. It began as 'just an interesting tech idea', through money-loundering and terrorist financing narratives back in 2016 to currently discussed potential of a basis to reshape and rebuild financial markets with Central Banks discussing issuance of Central Bank Digital Currencies and crypto-backed ETFs issued on traditional stock exchanges.

Today's (September 2022) global cryptocurrency market capitalization is around one trillion USD, but not even a year ago, it was almost three trillion USD (November 2021). Anyone who looks at the chart of the global market cap for this market will see an obvious pattern of boom and bust cycles, indicative of repeating and cyclical bubble dynamics reoccurring in this asset class. Despite the abnormally high volatility, higher than any other traditional asset class, the number of crypto proponents and adopters is growing exponentially. What is particularly interesting about this asset class is that, for the first time, we can track and observe the behaviour of those who adopted crypto, as all of their transactions and wallets are stored in publicly broadcasted ledgers in the blockchain.

Those two dynamics: the boom and bust cycle, coupled with datasets enabling observation of agents' behaviour in this market, are the foundation for my research. Using an extended version of the Log-Periodic Power Law Singularity model that accounts network effects based-valuation methodology as a proxy for the fundamental price, estimation of the critical date for the cryptocurrency market crash is done. Moreover, using tools from complex network analysis and machine learning on the underlying blockchain data, I study how the composition of agents' behavioural strategies changes during the entire bubble cycle.

This report is divided into the following sections: Section 1 Describes what exactly bitcoin is and how it works. Most important aspects of its design in the context of this work are highlighted at the very beginning; afterwards, I describe the whole mechanism design concisely. Section 2 Introduce Log-Periodic Power Law Singularity Model together with confidence interval of critical date estimation procedure. Later, Metcalfe's Law valuation framework for bitcoin and how it can be improved using tools from complex networks is discussed. Finally, implementation of point estimation of crash critical date using Market-to-Metcalfe ratio is presented. Section 3 describes current research behind the study on the behavioural structure of users, preliminary results on a small sample are presented with interesting discovery of calendar effect specific to cryptocurrencies, and the final research design that will be conducted during the remaining part of my internship is also presented. Section 4 consists of a short literature review on the Ethereum Network and the data engineering problems behind the study of it. Some aspects of my research and my preliminary theory on the cyclicality of the price being caused by reflexivity in monetary policy, market microstructure, speculative herding, social phenomenons and tokenomics of cryptocurrencies is not included in this paper. However, I am happy to present it during the discussion briefly.

Laboratoire d'Informatique en Images et Systèmes d'Information

The LIRIS (Laboratoire d'Informatique en Images et Systèmes d'Information) is a joint research unit of CNRS, INSA de Lyon, Université Claude Bernard Lyon , Université Lumière Lyon 2 and Ecole Centrale de Lyon. It carries out upstream research in informatics along six lines of research, while developing know-how for the benefit of society in close collaboration with a wide range of fields of application: culture and heritage, environment and urban life, biology and health, ambient intelligence, human learning, digital leisure, scientific calculations.

The LIRIS is an informatics research unit and more generally studies information sciences and technologies officially recognised by the CNRS (UMR 5005). The scientific activities of these 12 research teams are divided into 6 internationally recognised areas of competence:

- Imaging and visual recognition (Imagine and M2DisCo teams): construction of algorithms to understand multimedia data (images, videos, digital documents, 3D scenes) and is broken down into acquisition/reconstruction, indexing, modelling, classification and automatic content recognition (objects, actions, concepts).
- Geometry and modelling (GeoMod and M2DisCo teams): 3D reconstruction (dynamic, controlled quality etc.), procedure-based modelling (complex or natural scenes, ecosystems etc.) and the geometric and topological analysis and processing of networking or discrete forms (descriptor calculation, indexing, compression, tattooing, segmentation, visualisation etc.).
- Simulation, virtuality and computational sciences (Beagle, R3AM and SAARA teams): development of computer formalism and software tools for modelling and simulating complex systems through synergies with the disciplines of biology/physics/medicine/engineering.
- Data science (BD, DM2L and GRAMA teams): design of new and robust representation, processing, access, exploitation and protection models for the masses of heterogeneous data, whether structured or not, and which may be continuously and quickly distributed and produced.
- Services, distributed systems and security (DRIM and SOC teams): development of new models, languages, protocols and tools for distributed services and systems ensuring service quality, service and data, and efficient information searches as part of big data and linked open data.
- Interactions and cognition (GRAMA and SILEX teams): study and design of dynamic systems in which several human or software agents are interacting with each other, based on individual and collective properties and the cognitive abilities of the agents; modelling of man-machine or agent-agent abilities to build up knowledge, to help the user or to analyse usages.

Internship Details

Internship has been carried out in Data Minning and Machine Learning team (DM2L) under supervission of Professor Remy Cazabet, associate professor of Université Claude Bernard Lyon 1, Deputy Head of DM2L team. The academic supervisor was Professor Michel Grabisch, full professor of Université Paris 1 Panthéon-Sorbonne and Paris School of Economics.

Internship has been financed by Agence Nationale de la Recherche under BITUNAM project (Bitcoin User Network Analysis and Mining) coordinated by Professor Cazabet.

Internship lasts for 6 months, from April until end of October 2022, with compulsory break for the month of August given university holidays period. Presented report is based on the work done until the end of July.

1 Bitcoin

In this section, I discuss Bitcoin, with the following subsection highlighting key design features in the context of my research. Next, Bitcoin design is more in-depth discussed and is less foundational for the research I work on, but the further section on deanonymization of Bitcoin addresses is more critical concerning the improvements I have made to Network Effects Based Valuation of Bitcoin.

1.1 Bitcoin - Key Features

The crucial feature of bitcoin network is that each user can theoretically create any amount of new addresses/wallets which Satoshi Nakammoto, anonymous creator of Bitcoin, even directly recommends as privacy preservation mechanism. From this, we may infer that bitcoin address does not equal bitcoin user. This makes tracking users of the bitcoin blockchain much more challenging. Moreover, at any given time, only a limited number of transactions can be performed. Therefore users who would like to make a transaction as fast as possible will have to pay a higher transaction fee to incentivise faster transaction validation. If, at any time number of users rushing to make a transaction would be abnormally high, transaction fees naturally would also be substantially higher. Transaction fees and so-called mining rewards are incentives paid to network validators, also known as miners, who maintain the network by solving complex cryptographic problems for a proof-of-work consensus mechanism. Every four years the mining reward is halved during so called 'Halving'. Halving has been introduced to account for growing computational resources and works as fixed intervention that maintains difficulty of cryptocurrency mining.

1.2 How Bitcoin Works?

At the beginning of my internship, I decided to understand the inner workings of bitcoin before defining the exact research plan. Reading the original white paper published by Satoshi Nakamoto has been a challenging task; the entire idea for an alternative virtual currency system is distilled into the 8-page long article. The paper was foundational for the development of the entire asset class, as all future cryptocurrencies replicate the core design idea of the original bitcoin (BTC). I have written below the explanation and core idea behind how bitcoin works as short as I could. (A rather rough description, I highly recommend reading the original bitcoin whitepaper). I omit the exact inner workings of some of the aspects, such as SHA256 cryptographic hash functions, as I did not study those in-depth.

Blockchains are timestamped distributed ledgers combined with cryptographic digital signatures that provide decentralized, trustless verification. As the first cryptocurrency, Bitcoin uses blockchain technology to establish a purely peer-to-peer version of electronic cash. All transactions between agents in the bitcoin system are stored in a shared timestamped public ledger. The spendable balance of any agent is stored in a wallet and is computed directly from previously verified transactions. Each user can create any number of new wallets at no cost.

Once a wallet is established, it generates its own public/private key pair. Public keys, responsible for data encryption, are used to send cryptocurrency between agents. In contrast, private keys are used for data decryption and transaction verification and are integral proof of ownership of a particular address. The actual electronic coin is a chain of digital signatures. Transfer of coins is done by digitally signing a hash of the previous transaction and the public key of the next owner. In order to prevent the double-spending of coins, the ledger is public, and there exists a well-defined consensus mechanism such that all the participants of the system are able to agree on a single history of the order of transactions. The term 'coin' is not really correct as there does not exist any coin in literal sense, more accurate intuition would be to think of bitcoin as liquid that flows through transactions from wallet to wallet, but I may use this term for simplicity of writing.

The consensus mechanism of bitcoin is called proof of work. At a high-level, proof of work assumes that the ledger to be trusted is the one that puts the highest computational work. In that case, the system can be made such that fraudulent transactions and conflicting ledgers would require infeasible computations to override a valid ledger. The protocol is based on the Hashcash system and involves the use of the cryptographic hash function SHA256.

In short, an input for the SHA256 can be any message, and the output, called hash, is a fixed length string of bits that is seemingly randomly generated. However, the generated hash is not random; for any given input, it will always generate the exact same hash. Even small changes to the input will unpredictably and substantially change the generated hash. Inverse computation of input knowing only generated hash is in practice almost impossible; the only way would be through brute force guessing. All transactions are stored within a ledger maintained within "blocks". Each block has its hash generated by SHA256, then the block is timestamped, and the hash is published. The timestamp proves data stored within a block must have existed at the time; moreover, each block also includes a hash of the previous block, forming a chain.

The proof-of-work involves scanning for value such that, when combined with a header of the previous block and hashed with SHA-256, it starts with predefined number of zero bits. Given how SHA-256 works, the only strategy to solve this problem is through brute force random guessing, which requires significant computational resources. Moreover, as the number of zero bits increases, the computational difficulty of finding such value increases exponentially. Once such value has been found, a new block has been created.

This mechanism makes it computationally impractical to reverse or change any existing validated blocks in the chain for malicious reasons. This is because once CPU effort to satisfy proof-of-work has been made, the block is validated and timestamped and cannot be changed without redoing the work for this particular block and, what's more important, all other blocks after the one considered for change. (Direct consequence of how SHA-256 works, any change inside any block will lead to significant and unpredictable change of the hash of the header for the block, leading to changes in all of the hashes of subsequent blocks). This mechanism further reinforces the validity of older blocks and makes them exponentially more difficult to change as new blocks are created. On the other hand, proof-of-work introduces majority decision-voting, as given how computationally difficult it is to solve the puzzle, one may think of the longest chain to be the majority vote. Until honest nodes control the majority of CPU power, the honest chain will grow the fastest and will outpace any competing chains (for example. chains with malicious actor who would like to transfer all the existing coins to his wallet). The proof-of-work uses difficulty adjustment mechanism in order to target an average number of blocks per hour; if blocks start to be generated too fast, the difficulty will increase. Because of this on average blocks are generated roughly every 10 minutes.

Bitcoin network technically is a collection of nodes that are independently storing and working on finding a difficult proof-of-work for the next block while collecting all the new transactions into that block. The first node to find a proof-of-work broadcasts it to all nodes, and if all the transactions stored in the new block are valid and nodes are accepting the block, a new block is added to the chain. The longest chain is always assumed to be the correct one; in the case of a tie, both chains are stored until one of them becomes longer.

In order to incentivize nodes to support the network, the first transaction of each block is a special transaction that transfers to the wallet of the creator of the block a reward. Moreover, for any transaction, if the output value is less than the input value, the difference is a transaction fee that is added to the incentive value for the block creator. Naturally, any node that wants to make a transaction, if it wants to be added to the nearest possible block and accepted as fast as possible, will have to propose a higher incentive by providing higher transactional fees. Not all transactions can be added instantenously as firstly, block size is limited to one megabite, translating it to at most 4000 transactions, moreover the average time for adding new blocks is equal to 10 minutes. Given constraints on the number of transactions that can be put into a block, if many nodes would like to perform transaction as fast as possible, network will get congested and transaction fees will increase. As for the size of the stored blockchain, in order to save disk space, saving of transactions is optimized. (all transactions are hashed several times in a tree like structure [Merkel Tree], with only the root included in the block's hash) Moreover, payments can be verified without the need to run the entire network node. It suffices for any user to use the block header of the longest chain as he can be sure that network nodes have accepted it before. It is also important to highlight that "coins" are divisible, smallest unit of Bitcoin is 1 satoshi with entire Bitcoin being equal to 100 million satoshis. Despite transactions being publicly broadcasted, privacy is maintained by keeping public keys pseudo-anonymous. It is nevertheless advised to use new key pair for each transaction to keep them from being linked to a common owner.

Summing up, the system is made of digital signatures and is secure from double-spending problem due to existing network based proof-of-work system with a record of public history of transactions. Assuming honest nodes controlling the majority vote, based on the CPU power, it is very difficult to break the honest consensus. Anyone can join or leave the network at anytime, assuming longest chain to be the ground truth at anytime. Incentive mechanism, based on transaction fees and mining rewards, maintains the network.

In terms of innovations after bitcoin, the first iteration was focused on transaction efficiency and transactional costs by changing parameters such as block size or average block time (case of direct copycats of Bitcoin such as Litecoin or Dogecoin). Next cryptocurrencies extended the pure accounting utility of the ledger and changed it into a "world state", which in fact, created a virtual machine able to run Turing-Complete programs such as in the case of Ethereum cryptocurrency. This creates a completely new application layer through the capacity to write smart contracts by which entire decentralized applications could be made. The latest iterations of cryptocurrencies, such as Cardano or Solana, try to overcome issues comming from the design of Proof of Work consensus mechanism. Nowadays most commonly discussed alternative is Proof of Stake, which is much more deeply enrooted in game theoretic considerations rather than economically/computationally expensive and ecologically questionable cryptographic computations. The consensus Mechanism design problem is quite fascinating, and there is much research at the intersection of mechanism design, game theory, decision theory and cryptography to tackle this problem [45] [22].

1.3 Deanonimization of Bitcoin Addresses

In order to understand the methodologies for address clustering, it is important to remind how bitcoin transactions are characterized. Firstly the intuitive idea behind bitcoin would be to think about it as a liquid that flows through transactions, from wallet to wallet. Formally each transaction consists of n inputs and m outputs. Each input, given chain structure, is in fact output of previous transaction which particular address has been recipient off. Each output is associated with the new destination stored in the blockchain as pseudo-anonymous public key (address). Below I provide graphical representation.

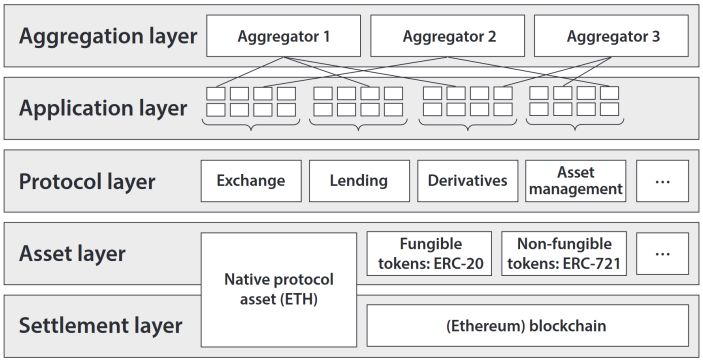


Figure 1: DeFi Stack (source: Schär, F. [2021])

The transaction fee equals the difference between the sum of transaction inputs and transaction outputs. Any user's balance at any time is then the sum of all previous transaction outputs that can be unlocked by private keys owned by the user that has not yet been spend. Transaction outputs, even if sent to the same unique address, are not directly summed together; they are recorded independently; for instance, assume transaction T_0 and transaction T_1 sending respectively amounts a_0 and a_1 to some address X. Then amounts sent in those transactions are identified respectively as output m_0 of transaction T_0 and output m_1 of transaction T_1, at no stage address X will receive a transaction with an amount equal to $a_0 + a_1$. Still, the spendable balance of address X will increase by $a_0 + a_1$, as it can be unlocked using the same private key [39].

Given the change mechanism, the user controlling output o_0 with y bitcoins; if he wants to use it as input to a transaction and send x bitcoins to another user, assuming $x < y$, he must send y bitcoins and will receive y - x of change to one of the addresses he controls, preferably to newly created address as recommended by Nakamoto [39].

Various methods exist for clustering addresses, but even simple heuristic-based approaches give surprisingly good results. I will describe the three most popular heuristics. In particular, the heuristic based on inputs has been used in the dataset I use in the network structure analysis.

Heuristic H1: All inputs used in the transaction are controlled by the same entity.

In order to cluster addresses using this heuristic, one must create a network representation of the bitcoin blockchain, where nodes are addresses. For each set of n addresses used conjointly as input of a transaction we add n-1 edges to create a path between them. Using such network we look for connected components, where each connected component can be interpreted as a user [39]. It is not ideal as there exist methods such as CoinJoin that help maintain privacy of user's. Nevertheless, multi-input heuristic could identify more than 69 per cent of addresses and assign them to specific entity.[23].

Another heuristic has been proposed by Androulaki [3] and is based on the change mechanism.

Heuristic H2: Assuming that transaction has only two-output addresses a_1 and a_2, and address a_1 has appeared before in blockchain, then address a_2 must be a newly created address and is a change address of the transaction.

Variation of the heuristic H2, generalizing it to a case with more than two output addresses has been propsed by Meiklejohn [34].

Heuristic H3: If address satisfies following properties it can be considered a one-time address:

1. the transaction is not a reward from minning
2. address is not reused, has not been among input addresses
3. it appears as output address for the first time.

Other methods exist, ranging from community detection algorithms to supervised machine learning, but for the sake of my research, I will concentrate on the methods described above. The primary reason is that despite their simplicity, they are pretty reliable, and there is sound underlying complex network reasoning why those heuristics will work, at least for the vast majority of users and critical agents in the system.

The importance of clustering addresses is visualized in Figure **??**, using Glassnode visualization tools for Bitcoin Blockchain.

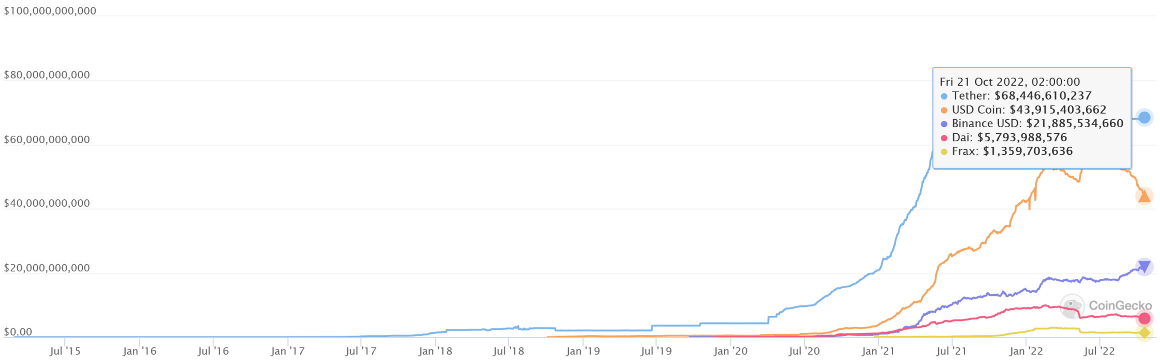


Figure 2: Top Stablecoins by Market Capitalization (source: coingecko.com)

By examining the chart below, we can observe that the cumulative sum of new entities is more stable than new addresses. Moreover, the sudden exponential growth of addresses and growth rates higher than new entities where observable during the times of a bubble (beginning of 2017 till beginning of 2018). The Bitcoin bubble in 2017 popped in the middle of December, with the market capitalization of bitcoin dropping by 50 per cent in less than a month. Following the price crash, we can also observe a sudden drop in the cumulative sum of addresses, and such a drop is not observed in the cumulative sum of new entities.

1.3.1 Bitcoin - Complex Networks View

Analysis of bitcoin transactional network from networks theory and analysis perspective has been conducted for quite some time. [32][5][31][44][29]. Those results have been shown over and over again. Therefore I decided not to pursue them and compute them myself. In particular, network analysis in this context, despite being theoretically easy to program, becomes quite challenging given the very significant size of the dataset (250-300GB+ that constitutes approximately 600 million of edges and approximately 800 million nodes in the graph). I will very quickly summarize the main results from the literature.

First of all, Bitcoin transactional graph is a scale-free network. Both in-degree (understood as the number of incoming transactions to an address) and out-degree (number of outgoing transactions from a particular address) obey power laws, with the α parameter being around 1.4 for both. This is one of the main reasons why the input-based heuristic is so effective. The vast majority of agents perform little number o transactions making them easily traceable in the network. The estimated clustering coefficient and the shortest-path lengths in 2021 are equal 0.0071 and 3.833, respectively, implying that there are many indirect transactions [44] .

Given estimates of clustering coefficient and shortest-path length, we may infer that small-world effect is present [44]. Largest strongly connected and weakly connected components are sizeable relative to the entire graph. Such result is not suprising, those connections are most probably going to important agent's within the bitcoin system such as cryptocurrency exchanges (institutions that play similar role to stock exchanges, where cryptocurrencies can be exchanged with other cryptocurrencies or "fiat money" such as US dollar or Euro) or minning pools (given exponentially growing difficulty of minning people started pooling computational resources together to mine bitcoin, individual minning is almost non-existant nowadays, at least minning of bitcoin. Side-comment, it is interesting how such dynamic leads to certain level of centralisation as computational power becomes controlled by few agents who agregates the computational resources of other small minners, to some extend it is against the core philosophical idea of Nakamoto of decentralisation of money. Some researchers are also pointing out this problem [1]).

2 Cryptocurrency Bubble Modelling

This section will introduce the modelling framework of bubbles in cryptocurrencies based on Log Periodic Power Law Singularity (LPPLS) Model, combined with Metcalfe's Law Valuation Framework. The first subsection, after a short word of introduction to financial bubbles, consists of a mathematical derivation of the LPPLS model and confidence intervals estimation procedure for bubble critical date. The second subsection briefly discusses problems with the valuation of cryptocurrencies and introduces Metcalfe's Law Valuation Framework. Next, Market-to-Metcalfe's ratio and improvement of it based on user clustering is presented. The final subsection describes implementation details and a briefly discuss problems that occurred while implementing the model.

2.1 Financial Bubbles

Financial bubbles intuitively are abnormal unsustainable price increases far above an asset's fundamental value. Once they achieve a peak and lack further buying pressure, they are followed by a substantial crash and rather brutal reversion towards reasonable valuation. Typically those events are explained by the "greater fool theory": asset prices are rising as buyers expect they will be able to sell the asset at a higher price to some other buyer in the future. Such events of theoretically unjustified price increases were present across financial history, from the Tulip bubble in the Netherlands in the XVII century and the South Sea Company bubble in the XVIII century in which physicist Issac Newton lost millions of dollars in today's terms, through railroads bubble in the US in XIX century to more recently Dot-Com era bubble in the early 2000s. Bitcoin also follows bubble-like dynamics [30][21]. Often observable to an outsider is increasing tension between euphoric proponents of further price growth in opposition to those who question the sustainability of exponential growth and highlight the lack of fundamental justifications for such. Those traits have been more formally described by Sornette [40] as transient super-exponential growth in price with accelerating periodic volatility fluctuations coming from the dissonance of opinions between bubble proponents and bubble adversaries, following phase transition in case of a bubble, crash or regime change [40].

2.2 Log-Periodic Power Law Singularity (LPPLS) Model

LPPLS model is based on [40]:

- the economic theory of rational expectations
- behavioral finance on imitation and herding of traders
- the mathematical and statistical physics of bifurcations and phase transitions

There are two types of agents in the model, rational agents with identical preferences and a group of noise, irrational agents whose herding behaviour driven by imitation leads to the development of the financial bubble. Agents live in a world of an ideal market with no dividends, and any constraints on market liquidity, risk aversion or interest rates are ignored. Since asset pays no dividend its fundamental value is p(t) = 0. Therefore any $p(t) > 0$ means the asset is in a speculative bubble. Given that the crash of the bubble is not a certainty in the model, rational agents will keep investing in an asset as a positive excess return above fundamental value compensates for the risk of the crash.

In this model, a strong collective answer (as is the case for a crash) is not necessarily the consequence of one elaborated internal mechanism of global coordination. However, it can appear starting from imitative local micro-interactions, which are then transmitted by the market resulting in a macroscopic effect [20].

2.2.1 Macroscopic modelling

The hazard rate describing the imitative process of agents follows:

$$\frac{dh}{dt} = Ch^{\delta} \tag{1}$$

$C > 0$ is constant, and $\delta > 1$, represents the average number of interactions among traders, therefore increase in the number of interactions among traders leads to an increase in the hazard rate.

Integrating (1) we get [20]:

$$h(t) = (\frac{h_0}{t_c - t})^{\alpha}, \quad \alpha = \frac{1}{\delta - 1} \tag{2}$$

$\delta > 1$ and $\alpha > 0$ are required conditions for growth of h(t) as $t \to t_c$ and therefore a critical point in finite time. $\alpha < 1$ is required for price not to diverge at t_c. Combining those conditions, we get $2 < \delta < \infty$, which implies that any agent is at least connected with two other agents. Moreover, the crisis can result from a self-fulfilling prophecy, as investors collectively lose confidence in the market. This is modelled by the feedback process:

$$\frac{dh}{dt} = Dp^{\mu}, \quad \mu > 0 \tag{3}$$

The underlying idea is that the lack of confidence quantified by the hazard rate increases when the market price departs from its fundamental value. Therefore, the price has to increase to compensate for the increasing risk [20].

This can be interpreted as an increase in the probability of a crash at a particular time as the market price for an asset departs from its fundamental value.

2.2.2 Microscopic Modelling

Following [27] and [42], irrational agents are connected in a network, with each agent indexed with i = 1,...,I and N(i) being number of direct connections with other agents to agent i. Each agent has two possible states $s_i = 1$ or $s_i = -1$ for state "buy" and "sell" respectively. The Markov process determines the state of the agent:

$$s_i = sign(K \sum_{j \in N(i)} s_j + \sigma \epsilon_i) \tag{4}$$

where $sign(x)$ is sign function equal +1 or -1 for positive and negative values respectively. K is constant, and ϵ_i is a standard normal random variable.

K governs traders' imitation tendency, while their idiosyncratic behaviour is governed by σ. Order within the network increases as K increases or σ decreases. The higher the order within the network, the more agents imitate their neighbours' behaviour. If order starts to dominate sufficiently many agents start to imitate their neighbours and imitation spreads over the network.

Therefore there exists a critical K_c that determines the separation between order and disorder regimes. If $K < K_c$ disorder is governing network, and the bigger K, the more agents start collaborating, forming the same directional trading positions. When K is sufficiently big, a strong one-directional market with agents having the same position is present.

The system's susceptibility is sensitivity to widespread global influence or external perturbation. In case of the LPPLS model this would be probability that large group of agents would be in the same state given existing external influences in the network. [20] We add G as a measure of the global influence to (4):

$$s_i = sign(K \sum_{j \in N(i)} s_j + \sigma \epsilon_i + G) \tag{5}$$

Average state of the market would then be $M = (1/I) \sum_{i=1}^{I} s_i$. For $G = 0$, $E[M] = 0$ (by symmetry), for $G > 0$, $M > 0$ and for $G < 0$, $M < 0$ thus $E[M] \times G \geq 0$

Susceptibility is then defined as $\chi = \frac{dE[M]}{dG}\Big|_{G=0}$

We may interpret susceptibility as the sensitivity of M to a slight change in global influence [20]. Another interpretation is that, if one considers two agents, when one of them is forced to be a certain state, the impact of this intervention on the second agent will be proportional to susceptibility. Given those interpretations, suspectibility is believed to correctly measure the ability of the system of agents to agree on an opinion. From sufficient susceptibility global synchronisation from local imitation of agents can emerge, that is why Johansen et al [27] decided that hazard rate of crash will follow similar process.

2.2.3 Price Dynamics and Derivation of LPPLS Model

Rational agent is risk neutral and has rational expectations therefore the asset price p(t) follows a martingale process $E_t[p(t')] = p(t), \quad \forall t' > t$, where $E_t[.]$ represents the conditional expectation given all information available up to time t [20] [27]. Given non-zero probability of crash, jump process j is defined, which is equal to zero before the crash and one at the time of the crash t_c. t_c is stochastic variable with probability density function $q(t)$, cumulative distribution function $Q(t)$ and hazard rate given by $h(t) = q(t)/[1 - Q(t)]$. Hazard rate is then interpreted as the probability per unit of time of a crash happening, given that it has not yet occurred. Assuming price falls by a fixed percentage $k \in (0, 1)$ assumed asset price dynamics before the crash are given by:

$$dp = \mu(t)p(t)dt - kp(t)dj \tag{6}$$

$$\begin{aligned} E[dp] =& \mu(t)p(t)dt - kp(t)[P(dj = 0) \times (dj = 0) + P(dj = 1) \times (dj = 1)] = \\ & \mu(t)p(t)dt - kp(t)[0 + h(t)dt] = \mu(t)p(t)dt - kp(t)dt \end{aligned} \tag{7}$$

Given no arbitrage condition and rational expectations, $E[dp] = 0$ so $\mu(t)p(t)dt - kp(t)h(t)dt = 0$ therefore we get $\mu(t) = kh(t)$ Substituted to (6), we get differential equation which solution is:

$$\ln[\frac{p(t)}{p(t_0)}] = \kappa \int_{t_0}^{t} h(t')dt' \tag{8}$$

Reminding that there exists a critical point K_c that determines the separation between order and disorder in the system of agents, we now discuss the impact of susceptibility. Given $K < K_c$, only a small agreement between agents exists; therefore, susceptibility χ is finite. When K increases sufficiently, order starts to appear. The system can become highly sensitive to small global perturbation, with agents agreeing with each other and creating large clusters and imitation propagating over long distances. In Natural Science, such phenomena is called critical, formally meaning infinite susceptibility χ of the system (Reasoning is similar to the two-dimensional Ising Model of alignment of atomic spins to create magnetisation [27], [36]). This can be modelled according to the following power law [27]:

$$\chi \approx A(K_c - K)^{-\gamma} \tag{9}$$

with $A > 0$ constant and $\gamma > 0$ being the critical exponent of the susceptibility.

Assuming K evolves smoothly and $K(t_c) = K_c$ we use first-order Taylor expansion around the critical point with : $K_c - K \approx constant \times (t_c - t)$ Assuming hazard rate of crash follows similar dynamics we get following equation:

$$h(t) \approx B \times (t_c - t)^{-\alpha} \tag{10}$$

However it is important to highlight that t_c is not exact time of the crash but only a mode of the distribution of the time of the crash, otherwise the rational expectations hypothesis could not be assumed as agents would be able to anticipate exact time of the crash.

Plugging (10) to (8) we get:

$$log[p(t)] \approx log(p_c) - \frac{\kappa B}{\beta} \times (t_c - t)^{\beta} \tag{11}$$

with logarithm of the price before the crash following power law.

The standard approach of bi-dimensional Ising model is extended to incorporate financial market structure better. Actors in the market differ in trading size by many orders of magnitude, moreover there exists higher level influences on assets (for instance currencies exchange rates influencing stock prices). Therefore one should assume that financial market has characteristic of hierarchical system. Therefore the financial market structure is now given by a hierarchical diamond lattice. Structure works in the following manner, first, let us consider two agents linked together. Substituting the link between those agents with four new links diamond is formed. Two starting agents have now connected with two new agents. In each iteration, we repeat the process of substituting each link with four new connections, and after n iterations, we get $N = (2/3) * (2 + 4^2)$ agents and $L = 4^n$ links among them (see figure 3 for visualisation of hierarchical diamond lattice structure).

The general solution for susceptibility given hierarchical diamond structure has been solved by Derrida et al. [14] and takes following form :

$$\begin{aligned} \chi &\approx Re[A_0(K_c - K)^{-\gamma} + A_1(K_c - K)^{-\gamma + iw} + ...] \\ &\approx A_0'(K_c - K)^{\gamma} + A_1'(K_c - K)^{\gamma} cos[\omega \ln(K_c - K) - \phi] + ... \end{aligned} \tag{12}$$

The power law incorporates now log-periodic oscillations that are accelerating, with frequency exploding as they reach critical time. Assuming hazard rate behave in similar fashion, and assuming financial structure being hierarchical diamond lattice the final hazard rate is approximated by following equation:

$$h(t) \approx B_0(t_c - t)^{-\alpha} + B_1(t_c - t)^{-\alpha} cos(\omega log(t_c - t) - \phi') \tag{13}$$

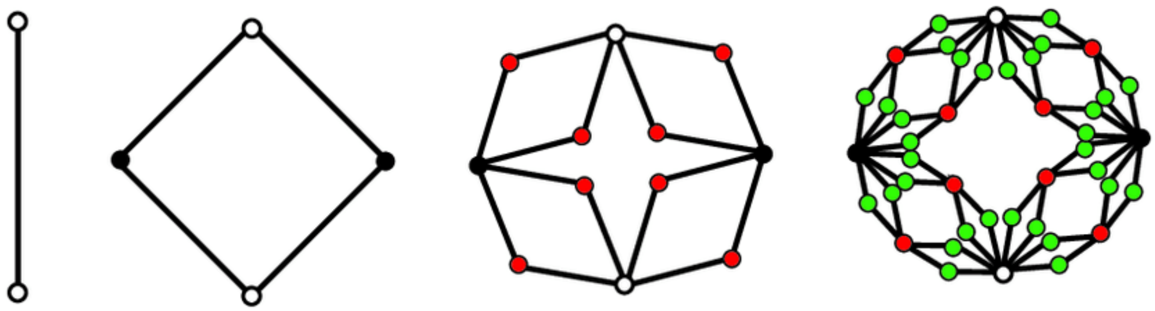


Figure 3: Hierachical Diamon Lattice Structure Creation (source: Sirca et al. [2016])

.

Assuming such hazard rate, risk per unit of time of a crash, given it did not occur, increases drastically as interactions between investors become sufficiently strong. This acceleration is also influenced by sequential decreases of risks that are modeled by log-periodic oscillations.

Finally we get:

$$ln[p(t)] \approx ln[p_c] - \frac{\kappa}{\beta}[B_0(t_c-t)^\beta] + B_1(t_c-t)^\beta cos(\omega ln(t_c-t)-\phi) \tag{14}$$

which can be rewritten to:

$$ln[p(t)] \approx A + B(t_c-t)^\beta\{1 + Ccos[\omega ln(t_c-t)-\phi]\} \tag{15}$$

with A $>$ 0 being value of $ln[p(p_c)]$ at actual critical time, B $<$ 0 increasing $ln[p(t)]$ over time if C is sufficiently close to zero. C controls the magnitude of oscillations around the exponential growth. ω controls oscillation frequency while parameter ϕ is phase parameter with $0 > \phi > -2\pi$. $(t_c-t)^\beta$ is power law singularity that represents positive feedback loop of herding behavior of the noise traders, it leads to the formation of the bubble. Parameter β is between zero and one in order to ensure finite price at critical time t_c [40]. The log periodic function $\cos(\omega\ln(t_c-t)-\phi)$ takes into account the existence of a possible hierarchical cascade of panic acceleration punctuating the growth of the bubble resulting either from a preexisting hierarchy in noise trader sizes and/or from the interplay between market price impact inertia and nonlinear fundamental value investing [40].

We can also write (14) as

$$ln(p(t)) \approx A + B(t_c-t)^\beta + C(t_c-t)^\beta cos(\omega ln(t_c-t)-\phi) \tag{16}$$

The term $C\cos(\omega\ln(t_c-t)-\phi)$ can be extended into two linear parameters $C_1 = C\cos\phi$ and $C_2 = C\sin\phi$ which allows to rewrite the LPPLS into:

$$\ln p(t) = A + B(t_c-t)^\beta + C_1(t_c-t)^\beta cos(\omega\ln(t_c-t)) + C_2(t_c-t)^\beta sin(\omega\ln(t_c-t)) \tag{17}$$

This reformulation decreases the number of nonlinear parameters to three and get rid at the same time of interdependence between the phase ϕ and the angular log-frequancy [18].

From now on, equation (17) defines $LPPLS(t_i, t_c, \beta, \omega, A, B, C_1, C_2)$.

After reformulation the LPPLS model has 3 nonlinear parameters t_c, β, ω and 4 linear parameters A, B, C_1, C_2, and the phase ϕ is contained by C_1 and C_2. By using L^2 norm, the sum of squared residuals is then described as:

$$\begin{aligned} F(t_c,\beta,\omega,A,B,C_1,C_2) = \sum_{i=1}^{N}[&\ln p(t_i) - A - B(t_c-t_i)^\beta - \\ &C_1(t_c-t_i)^\beta\cos(\omega\ln(t_c-t_i)- \\ &C_2(t_c-t_i)^\beta\sin(\omega\ln(t_c-t_i))]^2 \end{aligned} \tag{18}$$

Slaving the 4 linear parameters A, B, C_1, C_2 to the remaining 3 nonlinear parameters t_c, m, ω yields the following cost function

$$F_1(t_c, \beta, \omega) = \min_{A,B,C_1,C_2} F(t_c, \beta, \omega, A, B, C_1, C_2) \tag{19}$$

where solution for will be unique coming from matrix equation, this problem can be solved for instance using LU decomposition.

$$\begin{pmatrix} N & \sum f_i & \sum g_i & \sum h_i \\ \sum f_i & \sum {f_i}^2 & \sum g_i f_i & \sum f_i h_i \\ \sum g_i & \sum f_i g_i & \sum {g_i}^2 & \sum h_i g_i \\ \sum h_i & \sum f_i h_i & \sum g_i h_i & \sum {h_i}^2 \end{pmatrix} \begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{C_1} \\ \hat{C_2} \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i f_i \\ \sum y_i g_i \\ \sum y_i h_i \end{pmatrix} \tag{20}$$

where $y_i = \ln p(t_i)$, $f_i = (t_c - t_i)^\beta$, $g_i = (t_c - t_i)^\beta \cos \omega \ln t_c - t_i))$
and $h_i = (t_c - t_i)^\beta \sin \omega \ln (t_c - t_i)$

Final problem, after parsing and solving linear part as a function of nonlinear, is reduced to nonlinear optimization problem in 3-dimensional space.

$$\min_{t_c,\beta,\omega} F_1(t_c, \beta, \omega) \tag{21}$$

Filimonov [18] proposes to further reformulate the optimization problem, given how crucial the estimation of t_c is.

$$F_2(t_c) = \min_{\omega,\beta} F_1(t_c, \beta, \omega) \tag{22}$$

$$\beta(\hat{t_c}), \omega(\hat{t_c}) = \arg\min_{\beta,\omega} F_1(t_c, \beta, \omega) \tag{23}$$

$$\hat{t_c} = \arg\min_{t_c} \widetilde{F_2}(t_c) \tag{24}$$

There are several methods proposed to solve this problem, with robust and stable calibration proposed by one of the authors of the original model being metaheursitic local search methods like Tabu search or simplex inspired Nelder-Mead Method. [18] Due to computational complexity and difficulty for the implementation Dean Fantazzini proposed simpler approach using Broyden-Fletcher-Goldfarb-Shanno Algortihm (BFGS) which is Quasi-Newton method [15]. There has been also propositions to use evolutionary algorithms like genetic algorithm [25] or covariance matrix adaptation evolution strategy (CMA-ES) which was very successful in solving the problem at very low granularity of data for real time prediction of bubbles (30min time intervals) [40].

2.2.4 Confidence Intervals Estimation

Following Confidence Interval Estimation is based on procedure developed by Filimonov et al. [17] specificly for the needs of Log Periodic Power Law Singularity Models.

Given that we are particularly interested in the estimation of the critical time that is parameter t_c we treat all other parameters of the model as nuisance parameters $\psi = \{\beta, \omega, A, B, C_1, C_2\}$.

Equation (18) can be viewed as minimizing sum of squared residuals ($\epsilon(t_i; t_c, \psi) = lnp(t_i) - LPPLS(t_i; t_c, \psi)$) as in typical Ordinary Least Squares method.

$$SSE(t_c, \psi) = \sum_{i=1}^{n} ((\epsilon(t_i; t_c, \psi))^2 = \sum_{i=1}^{n} (ln(p(t_i)) - LPPLS(t_i; t_c, \psi))^2 \tag{25}$$

Assuming residuals are normally distributed Ordinary Least Squares is equivalent to Maximum Likelihood Estimation.

The likelihood function takes form:

$$L(t_c,\psi,s) = (2\pi s)^{-n/2}exp\Big(-\frac{SSE(t_c,\psi)}{2s}\Big)\rightarrow max_{t_c,\psi,s} \tag{26}$$

with $s = \sigma^2$ being variance of the residuals $\epsilon(t_i;t_c,\psi)$ and n being the number of data points in the fit. Likelihood is meaningful only up to arbitrary positive constant therefore we are able to omit constant pre-factors.

The estimate for the σ^2 is

$$\hat{\sigma^2} \equiv \hat{s} = \frac{1}{n}SSE(\hat{t_c},\hat{\psi}) \tag{27}$$

Given primary interest in estimation of the critical time, other parameters $\eta = \{\psi, s\} = \{\beta, \omega, A, B, C_1, C_2, s\}$ are treated as nuisanse parameters. In order to eliminate nuisanse parameters we construct profile likelihood and replace them by their MLE at each fixed value of the parameter of interest.

Profile Likelihood $L_p(t_c)$ is defined as:

$$L_p(t_c) = max_\eta L(t_c,\eta) \equiv L(t_c,\hat{\eta_{t_c}}) \tag{28}$$

where $\hat{\eta_{t_c}} = argmax_\eta L(t_c,\eta)$ is MLE for η for a fixed value of t_c.

Important observation is fact that profile likelihood approach is technically identical to optimization of the function $F_2(t_c)$ discussed in previous subsubsection (22) [17]. MLE estimates of parameters $\hat{\psi}_{t_c}$ is given by the solution of the OLS: $\hat{\beta_{t_c}}$ and $\hat{\omega_{t_c}}0$ are derived from (22) while estimates of $\hat{A_{t_c}}, \hat{B_{t_c}}, \hat{C_{1,t_c}}, \hat{C_{2,t_c}}$ are from (20). Moreover $\hat{s_{t_c}}$ if we replace $\hat{t_c}$ by t_c it takes similar form as (27).

$$\hat{s_{t_c}} = \frac{1}{n}SSE(t_c,\hat{\psi}) \equiv \frac{1}{n}F_2(t_c) \tag{29}$$

Plugging (29) to (26) we get:

$$L_p(t_c) \propto (\hat{s_{t_c}})^{-n/2} \propto (F_2(t_c))^{-n/2} \tag{30}$$

Likelihood (similarly profile likelihood) is meaningful only up to a constant therefore we consider relative likelihood (similarly relative profile likelihood), which is normalized to one by its maximum and takes values in [0,1]:

$$R(t_c) = \frac{L_{t_c}}{max_{t_c}L(t_c)} \tag{31}$$

Using relative profile likelihood we can create likelihood intervals that will be used in place of standard Wald Confidence Intervals:

$$CI(t_c) = \Big\{t_c : R_p(t_c) = \frac{L_p(t_c)}{L_p(\hat{t_c})} > 0.05\Big\} \tag{32}$$

2.3 Network Effects and Metcalfe's Law Valuation

In this section, I briefly describe current ideas on the valuation of cryptocurrencies in the literature. Next explanation of network effects and network effects-based theories used in finance are mentioned, and finally, the exact Metcalfe's Law valuation framework is described.

One of the main problems with working on Bitcoin, or any other cryptocurrency for that matter, is that there does not exist a commonly agreed upon valuation framework for cryptocurrencies. Some of the recently developed frameworks, such as Equilibrium Bitcoin Pricing [6] focus on transactional benefits and costs and trade-offs between the use of traditional central bank issued vs crypto-based currencies. Model has been derived in rational expectations framework similarly to OLG model known in Macroeconomics. Julien Pratt et al. [38] have introduced one of the first micro-founded token pricing models based on token-in-advance constraint with an endogenized velocity of circulation of tokens. Still, the model is quite challenging in calibration, given significant data requirements and only a limited number of tokens that can be analyzed using this framework. The key finding of the model is that each token initially must provide excess financial returns to speculators to drive adoption to actual users. As adoption progresses, volatility decreases and the price of the underlying token increases. A similar result has also been derived by Lin Cong et al. [11] as he showed that most tokens prices should follow S-curve-like dynamics assuming users adopt them. Moreover, he showed inter-temporal feedback exists between user adoption and token price, and such feedback accelerates the adoption. Furthermore, a standard cross-section analysis of returns indicates a network effect premium as one of the factors [11].

Assuming bitcoin is a currency, we enter into a philosophical discussion of the currency's value and where the currency's value comes from. Citing Varian [46], he sees the value of the US dollar coming from the direct power of the government and the general people's acceptance to use it as a payment method. Still, he concludes that the primary driver of the value of the US dollar comes from network effects.

Network effects is the phenomenon by which the value or utility a user derives from a good or service depends on the number of users of compatible products. They occur in network effect industries such as telecommunication, transportation, electricity, banking and health care. Positive network effects in any good is additional utility consumers gain as the total number of customers in this market increases. Therefore network effects are increasing with increasing adoption by users [28].

In a similar spirit, Cauwels and Sornette developed an original valuation method for social network firms based on the economic value of users' demographics, focusing on the diffusion of the technological idea and its adoption. They were able to predict ex-ante the performance of companies such as Facebook, Zynga and Groupon after their IPOs [10] [19].

Metcalfe's Law is one framework that tries to account for network effects while valuing networks. It is more a rule of thumb than existing Law, but it states that the value of the network is proportional to the square of the number of users. Despite being rather a simplistic assumption, it has been empirically validated in evaluating social media companies [49].

The professional investing community have tried to create a simple valuation framework for bitcoin based solely on Metcalfe's Law. Metcalfe's Law is also overwhelmingly often cited in financial or crypto-focused media as one of the major themes driving the price [37]. Simple regression on the logarithm of the market capitalization of bitcoin against the active bitcoin addresses gives the R^2 significantly above 80 per cent. The valuation equation for bitcoin based on Metcalfe's Law would be given by the following equation:

$$p = e^{\alpha_0} u^{\beta_0}, \quad where \quad \beta_0 = 2 \tag{33}$$

Where p is the market capitalization of bitcoin, u is the number of active addresses in the bitcoin network, and the value of beta zero comes from the assumption that the value is proportional to the square of the number of users.

Following this idea, Wheatley and Sornette [48] discuss the idea of Metcalfe's Law in the valuation of Bitcoin, pointing out that, despite the regression of market capitalization against active users severely violating the assumption of the independent and identically distributed errors, Metcalfe's Law still can be a pretty helpful analytical tool. Moreover, the significant residuals are, in fact, the bubble and crashes that occur in cryptocurrency markets.

Generalized Metcalfe's Law empirically estimates relevant parameters with log linear regression:

$$\ln p = \alpha + \beta \ln u + \epsilon \tag{34}$$

The achieved coefficient of determination is around 95 percent when fitted on data from 17 October 2010 till 26 February 2018.

2.4 Combining Metcalfe's Law Valuation and LPPLS modelling

Building upon the LPPLS models, where the bubble is defined as unsustainable growth in the price above fundamental value, and assuming the fundamental value of bitcoin can be estimated with generalized Metcalfe's Law, Wheatley, Spencer and Sornette [48], created an analytical tool that gives a warning of bubble's existence and reasonable confidence interval for its bursting time.

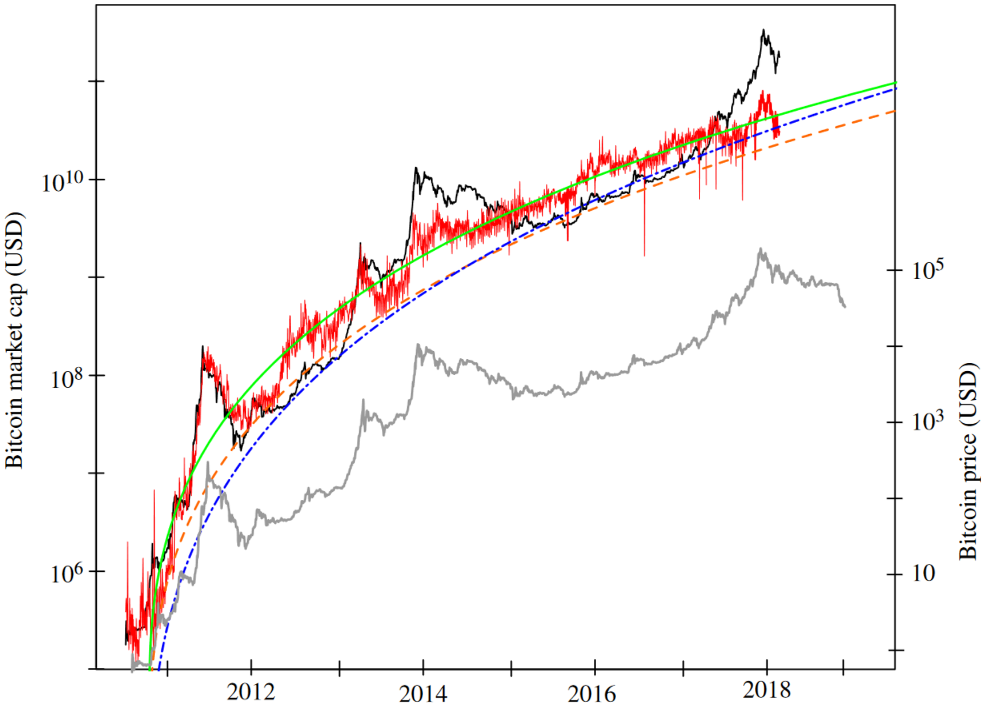


Figure 4: Bitcoin Market Cap vs Generalized Metcalfe's Law (source: Wheatley et al. [2019])

Assuming there exists a feedback loop, correct with the rationale of the LPPLS model, that price increase will lead to an influx of buyers, there will be an endogenous relationship in the regression of price against a number of active wallets. The reasoning is straightforward, if a new user joins the network, this will lead to a price increase, leading to more interest, and again new users joining the network again increasing the price and so on. In order to solve the problem of endogeneity, active addresses have been smoothed out using Locally Estimated Scatter-plot Smoothing, which is a generalization of the moving average using local regression.

Metcalfe's Law valuation will be used as a proxy for Bitcoin's fundamental value. Persistent price derivation above the level consistent with a valuation based on Metcalfe's Law can then be perceived as a bubble. Therefore one may define the Market-to-Metcalfe ratio, which can then be used in the fit of the LPPLS model. Moreover, valuation is not performed directly using the data on daily active addresses but rather based on smoothed support line computed below the log-graph of active addresses to further account for sudden potential growth in addresses during the end of the bubble (black line indicates actual Bitcoin market cap, the rough red line is the precise prediction using unsmoothed Metcalfe's law, and possible support lines based on Metcalfe's Law are visualised as blue, orange and green lines see Figure 4).

The slight improvement over the methodology proposed by Wheatley and Sornette, coming directly from how the bitcoin transnational graph is created, is to use the number of entities instead of the number of addresses in Metcalfe's Valuation. The bitcoin address is not equal to the bitcoin user; moreover, given the fact that each user can theoretically create any number of new addresses in the network at no cost and such behaviour is even directly encouraged by Nakamoto as a privacy preservation mechanism, a number of active addresses will significantly and non-linearly overstate the number of entities. The cumulative number of entities is a more stable metric and does not suffer from high exponential explosion and crash during bubbles. Moreover, after change to entities, less significant smoothing is required. Lastly, using estimate of the number of users is also more aligned with valuation frameworks developed by other researchers [38][12].

Generalized Market-to-Metcalfe's value ratio is:

$$MMV_i = \frac{p_i}{e^{-\alpha} u_i^{\beta}} \tag{35}$$

p_i is the actual market cap at time t_i, and the denominator in the ratio is a valuation based on Metcalfe's Law. α and β are parameters from the generalized Metcalfe's law regression.

The standard LPPLS model provides only a point estimate of the crash date. Extending this approach, we can obtain confidence intervals of the predicted crash date, from which we can get a stable advanced warning of the market crash.

The procedure goes as follows: data for each fit constitutes Market-to-Metcalfe's ratios in an interval defined by T_1 and T_2, where T_1 is an assumed starting time of the bubble and T_2 is the assumed turning point. All T_2's are dates in a set from one year before the actual turning point, being, in the case of the bubble in 2017, a December 17th to two weeks beyond. For each T_2, a range of bubble starting times, T_1, are considered, that is, all the dates between 360 to 250 days before the assumed turning point T_2. Then for each T_2, for all assumed bubble starting dates T_1, crashing dates using LPPLS are estimated. For each estimated crashing date (one per each T_1), the confidence interval is computed using profile likelihood (see Section 2.2.4). Then all confidence intervals are aggregated as a simple mean giving the final confidence interval. The aggregated curve gives the final estimate of the confidence interval at particular T_2.

Given the convergence of the estimated confidence interval, the closer we get to the actual bubble bursting date, we can get an advanced warning signal.

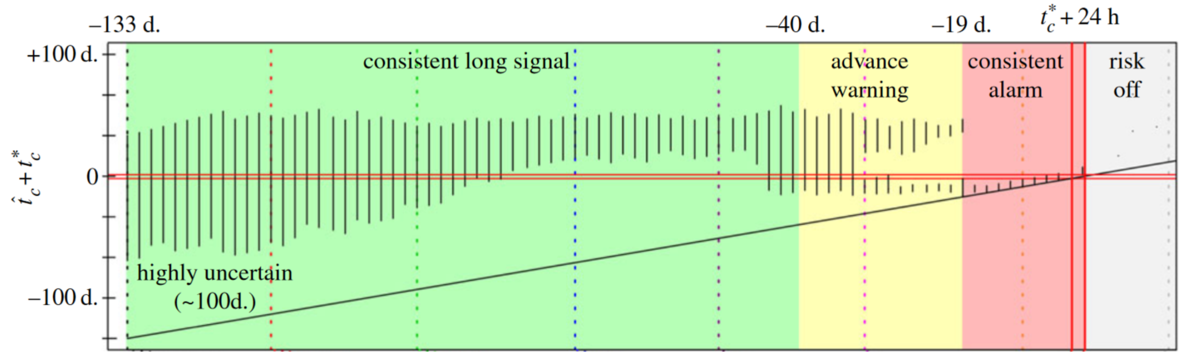


Figure 5: Confidence Intervals for the estimate of the critical time (source: Wheatley et al. [2019])

2.4.1 Implementation

The model has been implemented using Python accordingly to the robust and stable calibration proposed by Filimonov, and Sornette [18], that is, slaving the 4 linear parameters to 3 non-linear parameters, solving with LU decomposition and 3 non-linear parameters has been estimated using Nelder-Mead Method, where exact implementation details in terms of code have been similar to Takagi [43].

Smoothing of addresses has been done using R programming language and LOESS function with five degrees of freedom.

The market capitalization of Bitcoin data was downloaded from Quandl, while blockchain data has been downloaded from Coinmetrics.

Assumed bubble starting and ending dates are same as Wheatley et al. [48].

Fitting the LPPLS model to generalized Market-to-Metcalfe's ratio by using 95 per cent of bubble length data provides astonishingly good point estimation for the crash date. Testing this methodology on different bubbles gave similar results.

2.4.2 Difficulties of implementation

Implementation of this model took me several weeks, mainly due to the problem of slaving the linear parameters into non-linear ones. Unfortunately, neither myself nor Professor Cazabet knew the immediate answer to how such a problem could be solved.

```
datetime.date(2017, 12, 18)
```

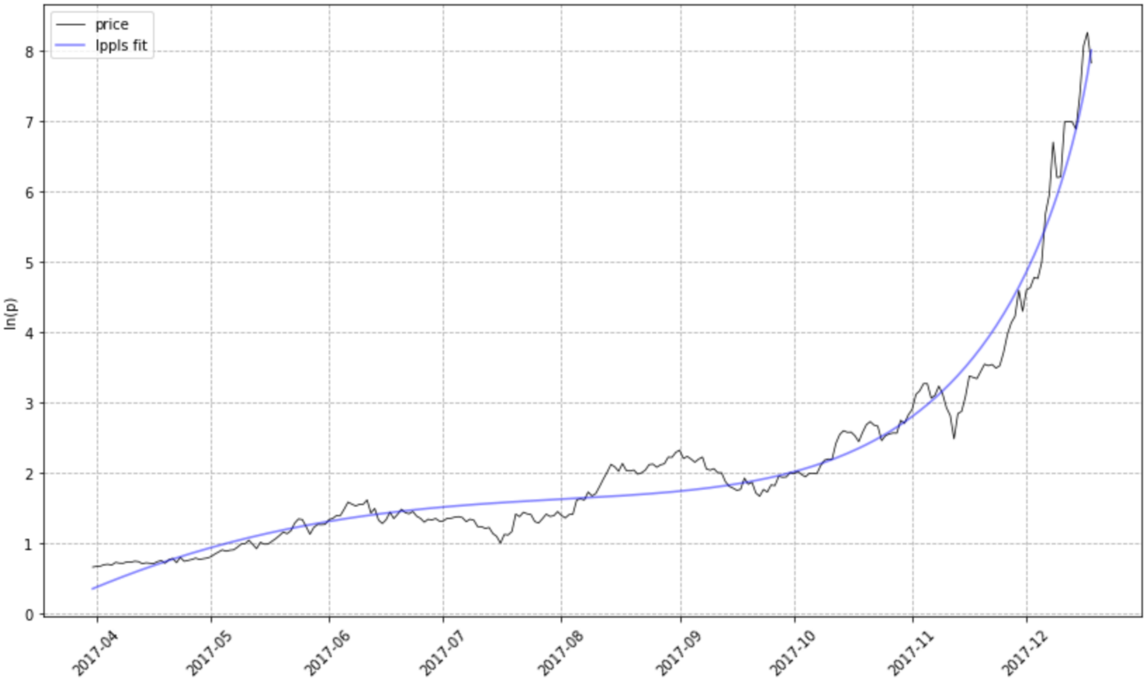


Figure 6: Point estimation of the bubble bursting time exactly matching actual bursting time date.

One of the proposed solutions by another member of the DM2L team was to use the package SymPy for symbolic computations in Python. I have programmed the entire problem with SymPy. The first problem arose with slaving linear parameters to nonlinear, as I wanted to use standard matrix inverse and then multiply the respective vectors to get estimates for A, B, C_1, C_2. Unfortunately, despite the significant computational time, it was still impossible for the program to converge. Afterwards, I implemented a well-optimized function for LU Decomposition in the SymPy package, and I finally got the solution. Nevertheless, it was still impossible to find the final solution for the sum of squared residuals as one of the objects, by the design of the package, changed it is the class type to immutable class and computation of MSE because of it was impossible.

I kept looking for other solutions. Firstly I have found the work of Jeremy [26], where he mentioned the results of Professor Fantazzini [15], [20], [16]. Using a package created by Professor Fantazzini in R programming language and his approach based on BFGS, I got the first results for the estimation of the t_c. Nevertheless, I kept testing this approach and found odd linearity for the predictability of the t_c. When I moved linearly by a month or two, both the starting date and ending date of my dataset, I found that the predicted t_c moved almost perfectly by the exact amount as the starting and ending dates of the new dataset. Given how the model is supposed to be based on accelerating variation, which clearly must have been different in the new dataset, it intuitively did not make sense that such linearity would occur. In the paper on how to compute modified profile log-likelihood for the LPPLS model, a similar claim for sensitivity to perturbation has also been stated [17]. Moreover, given BFGS method is quasi-Newton, it may be prone to get stuck at local solutions. Reading several papers on the LPPLS model, depending on the formulation of the problem, I knew that it is not a smooth convex function that would guarantee equivalence of a local solution with a global solution, which was another drawback of this proposed methodology. Trigonometric reformulation (see (17)) seems to help when tested in simulations [18], but there is still no formal proof that this is always the case and that there is a single local minimum.

Confidence interval computation will be finished during the second part of the internship but is significantly less challenging than the initial solving of the LPPLS as, given very specific to LPPLS construction of profile likelihood, significant part of already existing code can be reused.

3 Socio-Finance and Cryptocurrencies

WAŻNE DO DORZUCENIA: LINK POMIĘDZY OLD NODES SELLING INTO BUBBLES, BEHAVIOR OF THOSE WHO ARE RICH IS DIFFERENT THAN THOSE WHO ARE NOT, RICH GETS RICHER ALE NA DANYCH Z GIEŁD DODATKOWO POKAZANY ŻE TO BUBBLE. COMMENT:! There are short lived benefits of transacting with bitcoin particularily related to transaction costs, if transaction costs is higher given even adjustments for short lived price volatility then it is beneficial for agent to make a transaction with btc. Otherwise BTC should be viewed not as a currency but as risk asset. TU SIE KONCZY COMMENT

This section will firstly describe the current literature on the relationship between the price formation of cryptocurrencies and the structure of the underlying graph of transactions. Next, the study done by Apsembitove et al. is described, where the framework for mining the behavioural structure of user composition is presented. The following subsection presents the first results on a small sample for behavioural mining and subsequent discovery of the Halving Effect - crypto-specific calendar effect reoccurring in this market. Finally, the extended study design is discussed in the last section.

3.1 Bitcoin Transactional Graph and Price - state of the literature

Following the work of Jørgen Vitting Andersen, socio-finance describes price formation in financial markets as the sociological phenomenon that relates individual decision-making to the emergent social level. Price results from the social dynamics of interacting individuals or groups of individuals. Interaction occurs either indirectly through the price or by direct communication [2].

The natural link to the work of Professor Andersen and Cryptocurrencies is the design of how Cryptocurrencies operate - through the publicly available blockchain of transactions. This gives researchers an entirely new source of information as, for the first time, agents' decisions and their relation to price formation, at least partially, can be directly observed and studied.

The number of studies on the price of cryptocurrencies and the state of the underlying blockchain of transactions is nevertheless quite limited. Bovet et al. have established a Granger causal relationship between transaction network properties and bitcoin price. In particular, all higher moments of the out-degree distribution provide information about future price movements. For instance, the higher out-degree standard deviation, meaning a more heterogeneous distribution of payments, Granger causes a price decrease. Opposite causality also exists; price impacts the moments of the degree distributions. Moreover, he also found that the causality structure based on the user's network (as defined by previously defined heuristics) has changed when considering the daily time scale, while it remained relatively stable on the weekly scale when comparing the period pre Mt-Gox and post Mt-Gox hack. This indicates that major regime shifts/structural changes may occur as time passes. (Note: MtGox has been the largest cryptocurrency exchange, responsible for over 70 per cent of all bitcoin transactions in early 2014. Following the security breach, thousands of bitcoins, worth hundreds of millions of dollars - at 2014 prices - have been stolen.) [9].

Interestingly, herding behaviour [7], a common trait of bubble dynamics, can be observed in the bitcoin transactional network. During the bubble's peak, many users behave similarly, moving their coins mostly between the cryptocurrency exchanges and their wallets. Once the market starts to crash and bubble bursts, the connectivity becomes more heterogeneous, and degree distribution widens. [9].

Bovet et al. [8] have found a link between the global structural properties of the user network and price behaviour. In particular, after constructing a temporal network representation of a transactional graph on weekly scales, he showed a strong correlation between the number of users and the price, with a strengthening correlation during bubble periods.

3.2 Minning of Behavioral Structure of Users of Bitcoin

The ability to track user behaviour gives us the rare opportunity to observe how the behavioural strategies of those users change over time, particularly during different market regimes. By implementing machine learning methods, Aspembitova et al. [4] have studied the underlying transactional network of Bitcoin and Ethereum cryptocurrency. She has found that different clusters of users with different persistent behavioural patterns are observable in the blockchain data. Moreover, she found that those patterns change as the price regime changes. However, Ethereum networks show a much more stable behavioural composition than bitcoin.

3.2.1 Minning of Behavioral Structure - Research Design

Firstly network representations of users' transactional graphs are created in different periods using monthly data, particularly during different pricing regimes that is a bubble, price increase, stable price, price decrease and crypto winter (Note: period of prolonged slow price decline following the 2017 crash with significantly negative sentiment around cryptocurrencies). Created graphs are directed, temporal, and weighted for each defined period, where each link (i, j, w, t) is a transaction between two nodes i and j at time t with w coins.

For each user, following features are extracted:

- Total degree of a node i at time interval t - this reflects the frequency of transactions of a user i
- In-degree of a node i at time interval t - reflects the number of receiving transactions of a user i
- Out-degree of node i at time interval t - reflects the number of sending transactions of user i.
- Out-going value - sum of all weights w of node i directed outwards of i at time interval t- reflects the total sum of send bitcoin
- Incoming value - sum of all weights w of node i directed towards i at time interval t - represents the total sum of bitcoin received
- Total balance - represents the net number of coins in the account balance of node i at time interval t.
- Total transacted value - sum of all outgoing and incoming values of node i node i in time interval t.

Some of the users are discluded from the dataset; the "noisy" users are those whose total degree for whole transaction history is less than two and traded value is less than 1 per cent of all transacted values at that time step. Those 'noise' users may be important if we consider the modelling of bubbles, herding and influx of new bitcoin adopters. Still, for the sake of modelling persistent behavioural structures, they must be excluded for a better fit of machine learning algorithms. Otherwise, the fit of algorithms would be too much overfitted to a noise generated by nodes not actively participating in the network.

The authors decided to use k-means clustering, with the elbow method that determines the optimal number of clusters, to find different behavioural strategies of bitcoin users. Once data points are clustered, they are qualitatively examined to asign them interpretation. Next fit of each datapoint to a particular cluster is statistically valideted by computing Silhouette Score. By removing points that do not follow rules defined by a cluster, new adjusted feature and label sets are created. Next, the Support Vector Machine is trained on adjusted feature sets and adjusted label sets. Finally trained SVM algorithm performs classification for all datasets.

The algorithm can be summarized as follows:

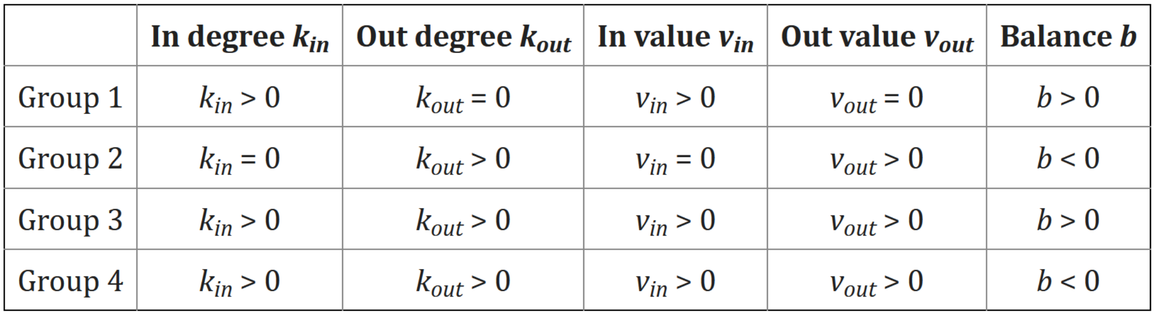
1. By elbow method on all features define optimal number of clusters
2. Use k-means clustering on the unlabeled feature matrices A_c and obtain vector of labels V_c

$$A_c = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1j} \\ a_{21} & a_{22} & ... & a_{2j} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nj} \end{pmatrix}, \quad V_c = \begin{pmatrix} v_1 \\ v_2 \\ ... \\ v_n \end{pmatrix} \tag{36}$$

3. Qualitatively examine clusters to find their specific properties in each group.
4. Statistically validate clusters with Silhouette Score.

5. Remove data points that do not follow the rules defined by clusters and create adjusted A_c and adjusted V_c.
6. Train SVM model on adjusted A_c and adjusted V_c.
7. Using a trained SVM model, set labels to a set of features from all periods.

The defined distinct properties of the majority nodes in each cluster are summarized in the following table.



	In degree k_{in}	Out degree k_{out}	In value v_{in}	Out value v_{out}	Balance b
Group 1	$k_{in} > 0$	$k_{out} = 0$	$v_{in} > 0$	$v_{out} = 0$	$b > 0$
Group 2	$k_{in} = 0$	$k_{out} > 0$	$v_{in} = 0$	$v_{out} > 0$	$b < 0$
Group 3	$k_{in} > 0$	$k_{out} > 0$	$v_{in} > 0$	$v_{out} > 0$	$b > 0$
Group 4	$k_{in} > 0$	$k_{out} > 0$	$v_{in} > 0$	$v_{out} > 0$	$b < 0$

Figure 7: Properties of node's in each cluster (source: Aspembitova et al. [2021])

Each group has been assigned behavioural interpretation in the following manner:

- Optimists (Group 1) - Users actively invest in the currency as they only accumulate coins.
- Pessimists (Group 2) - User's who sell the currency, only sell coins in the analyzed period.
- Positive Traders (Group 3) - User's who both buy and sell coins in the analyzed period but with a preference for accumulation of coins.
- Negative Traders (Group 4) - User's who both buy and sell coins in the analyzed period but with a preference for distribution of coins.

Using those rules, we may examine the evolution of behavioural types during different periods and price regimes. Moreover, we may also investigate if there are significant differences in behavioural strategies of users in two separate blockchain systems: Ethereum and Bitcoin.

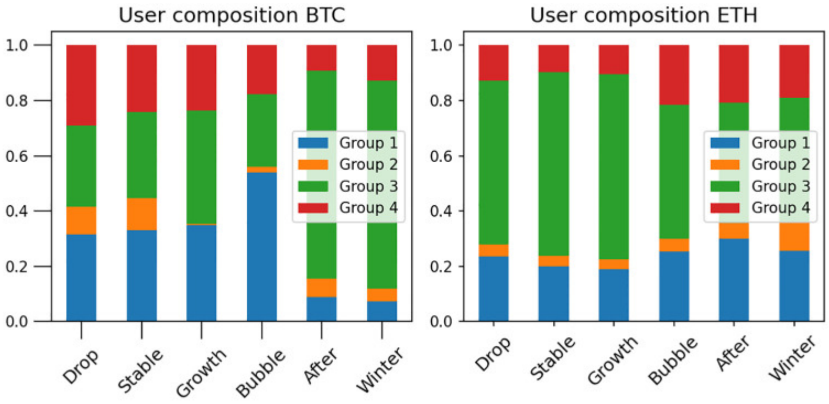


Figure 8: User composition during different price regimes

Looking at BTC's user composition (see Figure 8), we can see that people's behaviour is more volatile depending on the price movement compared to Ethereum. During price drops and stable slow price growth composition of users remain similar, but once the market starts to grow, the number of pessimists - those actors that are only selling coins - diminishes to almost zero. Once we enter the bubble, most agents (around 70 per cent) have a positive bias, with an evident dominance of pure optimists - agents that are only accumulating coins.

These results directly link to the previous result established by [8] - There is explicit herding behaviour of the users to buy more coins during the bubble. Moreover, the dominance of homogenous connectivity within the network can also be seen in the user's composition as groups of users involved in trading (buying and selling coins) become much smaller during bubbles. I hypothesize the difference in the stability of the user's composition in Ethereum is a direct consequence of the existence of the multiple-layer usage.

3.2.2 Behavioral Structure of Users and Halving

This subsection describes results on a small sample of bitcoin transactional graph using behavioural mining, but to understand this result, it is crucial to remember what halving means. Following the bitcoin protocol's design, at a fixed time of around four years, the miner's rewards are halved to account for the growing computational capacity of computer systems. This leads to a decrease in the supply rate of new bitcoins following the halving.

During the hackathon organized for the conference "Crypto-monnaies et autres monnaies alternatives numériques" organized by Professor Cazabet I tested with Victor Favre, another LIRIS intern, the clustering rules defined by Apsembitova et al. around the halving time and got very interesting results.

The data provided during the hackathon included all the transactional information a day before, on the day and the day after the halving that occurred at block 420 000, on the 9th of July, 2016.

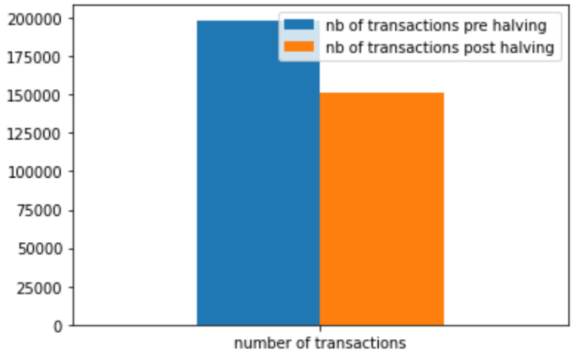


Figure 9: Number of transaction's pre and post halving.

Looking at the transactions data, one of the first discoveries was that the number of transactions pre-halving was significantly higher than the number of post-halving transactions.

Looking at the average transactions fee, important and counterintuitive, discovery is that the average transaction fee has been higher pre-halving than post-halving. Given how the miner's reward has been halved due to the halving, the opposite was expected. This could be seen as the first indication of network congestion pre-halving, as mentioned in the part describing how bitcoin works, during periods of congested network transactions, fees can be artificially higher than during normal times.

Actor classification following rules previously defined has also been analysed (see Figure 7). The visible trend is that before the halving, the group of positive traders has been significantly more prominent than post-halving, indicating the existence of positive, accumulative bias. There is a deeper reason why traders may have been more motivated to accumulate right before halving.

First, the previous halving has led to parabolic movement in the bitcoin price a few weeks after the halving occurred. The traders may have been motivated to accumulate more coins because of this particular reason. Given the time constraint and fear of increasing transactional costs, they may have rushed to buy, creating congestion in the network. This would explain above average and

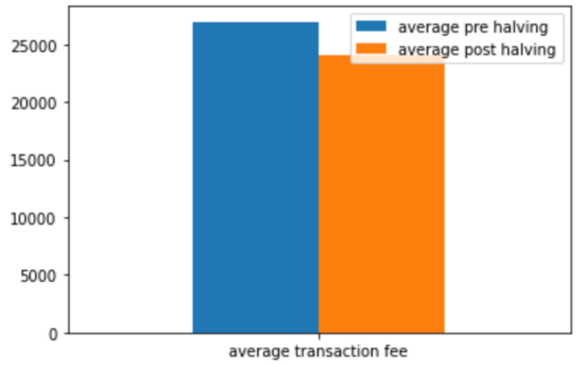


Figure 10: Average Transaction Fee comparison

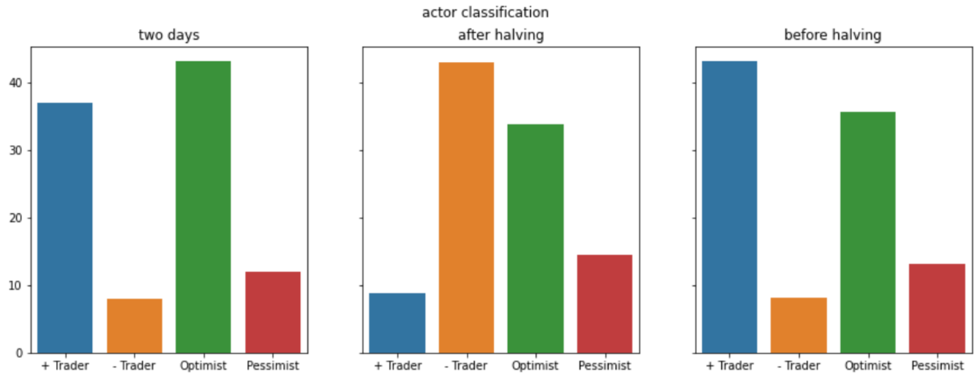


Figure 11: Actor Classification pre-halving and post-halving

counterintuitive transactional costs pre-halving. Interestingly those traders who acquired coins were right, as it seems there exists a so-called "halving effect".

The rationale behind the halving effect is quite simple, assuming the demand for bitcoin stays the same, but due to halving, the amount of new bitcoin issued is halved, limiting the supply of new coins; naturally, the market value for the cryptocurrency should increase. The halving effect has been discussed in the literature [35], but even a simple chart analysis would indicate that parabolic price movement after halving has been a reoccurring event.

First halving may have caused a more significant disturbance in the actual supply of bitcoin. However, given that new coin issuance follows the logarithmic trend, a subsequent marginal new coin issued per block decreases due to halving were significantly lower than the one caused by the first halving and most probably were not a cause for significant supply disturbance. Changes in miners' behaviour could indicate herding and accumulation (see Figure 12). For instance, during the last halving in 2020, we see a significant accumulation of coins right after halving. Similar behaviour occurred in the past. Similarly to what has happened in the past, the new bull market started in 2020 following the halving.

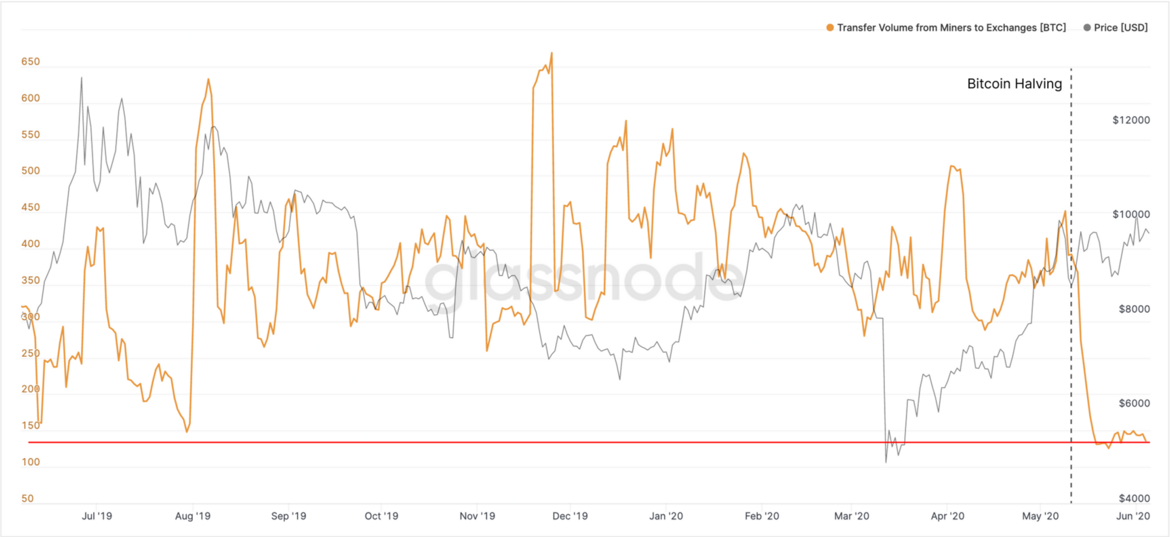


Figure 12: Miners Accumulation during Halving in 2020. (source: Glassnode)

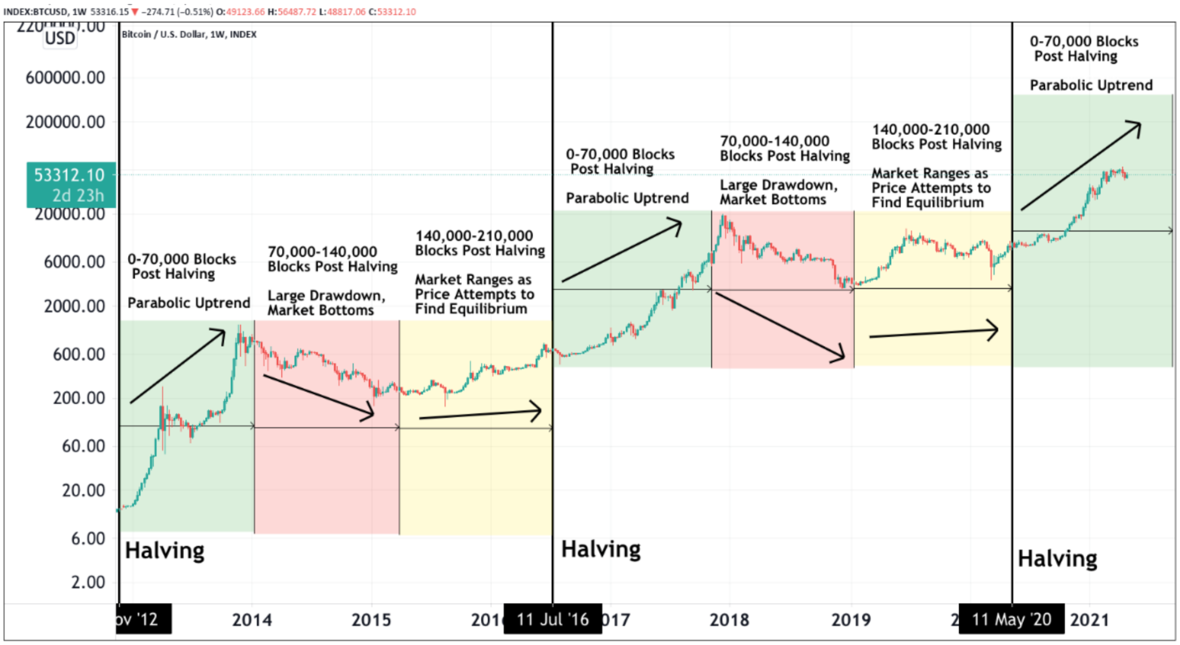


Figure 13: Bitcoin Halving and Parabolic Moves (source: TradingView)

3.2.3 Behavioral Structure of User's: extended study

Analysis of the transactional graph and data stored in the blockchain can be, given significant effort, extended with additional information. Cryptocurrency users throughout the years have found, through often cumbersome own transactional efforts, links between addresses and actual real-life entities. One of the sources that aggregate this information is Wallet Explorer. Using this information, we can assign names to users and tags related to their activities, such as exchanges, wallets, mining, gambling and others.

Labelling could also account for the node's size, particularly given the market impact literature [13]; this would be interesting. The so-called whale coin movement is likely more indicative of future price levels. Given the timestamp at each block, we can also assign an estimated monetary value in the US dollar for each transaction, estimating each node's total monetary loss/gain. If the vast majority of nodes are in profit, the risk of a crash is much higher; at the same time, if this coincides with the slowing growth of new nodes, then the buying pressure may wear off, further increasing the risk of a crash. Moreover, certain features of particular users can be also assigned if user's are analysied since the inception of Bitcoin.

Furthermore, because of timestamping, We can assign labels of long-term/short-term holders to each node. Coin movement is characterized by an exponential decay in the probability that coin will be moved. If a coin has not been moved soon after being acquired by an address, the probability it will be moved again decreases at an exponential rate. This exponential decay can be visualized by analyzing probability curves that a coin is moved within 7, 14, 30, 60, 90 and 120 days. Glassnode team proposed 155 days as a cutoff for assigning the label of the long-term holder.

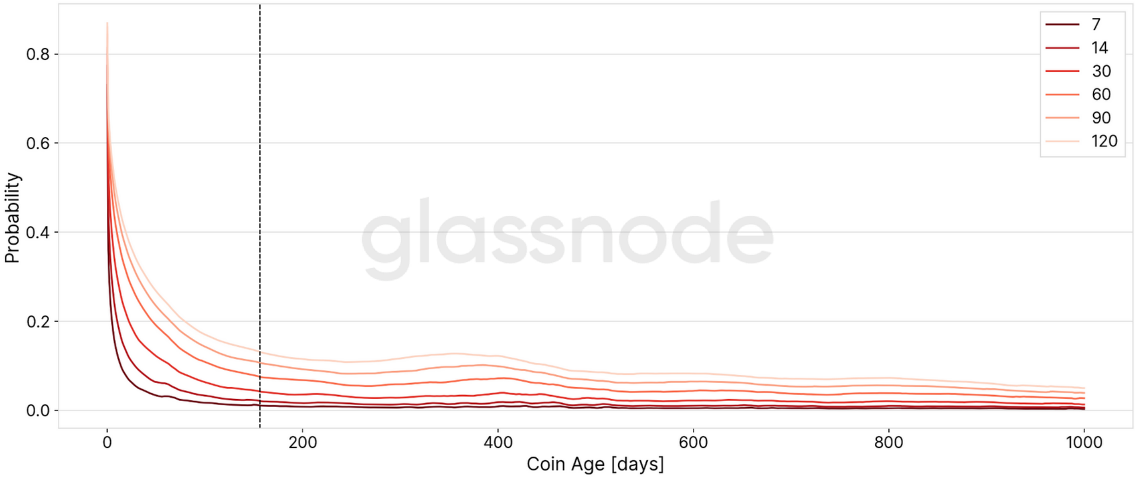


Figure 14: Exponential Decay of Probability UTXO being spend (source: Glassnode)

Using an extended information set, I plan to repeat the study of Aspembitova et al. on an extended feature set [4].

In particular the set of new features would include following:

- total degree of a node i at time interval t
- in-degree of node i at time interval t
- out-degree of node i at time interval t
- outgoing / incoming values in satoshi
- total balance in satoshi
- total transacted value
- total balance in usd
- ocean labels - depending on the size of the balance, label is assigned to a node, for instance whales being nodes with above 1000BTC stored while shrimps are nodes with less than 1 BTC
- long-term/short-term holder
- average number of transactions per block
- average number of transactions per day, per week
- standard deviation of number of transactions per week
- relative range of transaction intervals

- median transaction time intervals
- active time ratio
- transaction days to lifetime ratio
- type of agent label (exchange, minning etc.)

I plan to mine for new ways for clustering users. Excluding influential nodes such as mining, exchanges etc. would give a much clearer picture of the behaviour of speculators and investors, or vice versa analysis of the action of only particular types of agents might have crucial signalling information about future price behaviour. Accounting for size may shed some light if the 'whale' movement creates market impact problems similar to those described in market micro-structure literature. Particularly interesting would be movements directed from and into the cryptocurrency exchanges (the motive for cold storage of coins - storage on the external, self-custodial wallet to diminish risk related to exchange being hacked - or motive for selling coins). Moreover, analysis of the profit/loss of traders can be indicative of buying/selling pressure exhaustion. Some of the features are inspired by still unpublished research from a team of researchers from the University of Cardiff that was presented during a webinar organised by the Royal Statistical Society [33]. Presented research indicates that a very small subset of nodes (300 vs 30 million analysed) seem to behave like informed traders, while most nodes transact randomly.

This project is going to be my main focus for the remaining part of the internship.

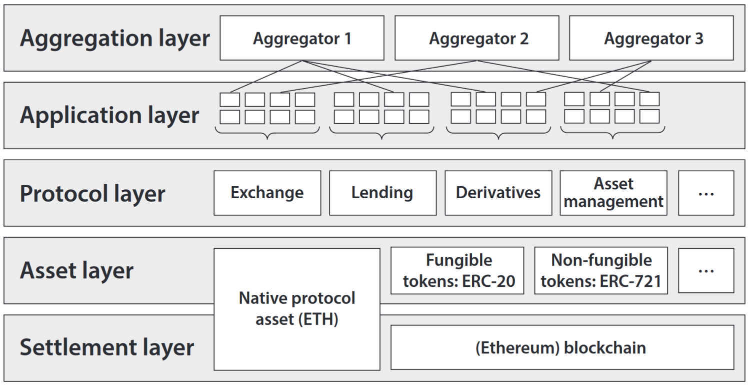


Figure 15: Schär[33] [2022])

4 Ethereum Network

First subsection consists of short literature review on ethereum cryptocurrency, particularily focused on complex networks. Following section describes data engineering problems behind working on ethereum network.

Ethereum is the second most popular cryptocurrency, characterized as a permission-less blockchain system that enables Turing Complete-smart contracts deployment. Users can not only transact with Ethereum coins but also can build and publish distributed applications via smart contracts that will have their respective tokens/coins. Ethereum itself creates an interesting ecosystem of human users and autonomous agents (contracts).

4.1 Ethereum Blockchain Analysis - short review of literature

There are four types of interaction between Ethereum accounts:

1. User-To-User (transaction or token transfer)
2. User-to-Contract (call or kill)
3. Contract-to-User (transaction or token transfer)
4. Contract-to-Contract (create, call, kill or hard-fork)

Interactions from the user or contract to the null address 0x0 denote the creation of smart contracts, while transactions to accounts without a from address denote the generation of ether as a mining reward. Using those interactions, four separate interaction networks are created: Trace Network with all user and contract information, Contract Network with only transactions between contracts, Transaction Network with only transactions between users, and Token Network where only transactions with particular tokens are stored [50].

One of the first results is that the number of new vertices and arcs added to graphs is almost of the same order as the total number of vertices and arcs in the graph at that time, which implies that Ethereum interaction networks are growing at a fast speed and account information is updated at a fast pace as well [50].

All networks get sparser over time when looking at the evolution of densities. Interestingly there is a high correlation coefficient between the number of new vertices and the number of vertices in the previous year that, is 0.99481, 0.99911, 0.99965, 0.99995 for Transaction Network, Trace Network, Token Network and Contract Network, respectively as for 2021. This indicates the preferential attachment phenomenon, the more particular node is connected, the more likely it grows new arcs [50].

This has an intuitive behavioural explanation as well; users that were active in the network will probably remain active in the future, with similarily contracts and tokens. The percentage of new vertices connected to old vertices is also high. For 2019 it was around 74 per cent, indicating that as time progresses, the Ethereum network matures with more accounts remaining active and more than half of the new vertices interacting with old vertices [50].

Preferential attachment and maturity lead to an interesting research hypothesis regarding the price behaviour of tokens. Consistently used tokens will have a higher probability of adoption and, given often artificial constraints in the supply, may also start following bubble like dynamics similar to the one observable in Bitcoin. This is also coupled with capital flows, as the more adopted token is, the more cryptocurrency exchanges list it, fueling the bubble.

Not suprisingly the most active accounts in term's of transactions remain active throughout the years as they represent important entities such as major exchanges, minning pools or very popular tokens like Tether [50].

Somin et al. [41] found that entire ERC20 tokens transaction network (explanation: ERC20 is a well specified commonly accepted standard for smart contract creation) follows strong power law dynamics displaying similar connectedness structure to other real-world networks, presenting non-negligible number of extremely connected nodes with majority of nodes having small number of connections.

Similar power law dynamics are also present in the tokens popularity among sellers and buyers of tokens.

Power-law dynamics disappear, however, if one study only a particular token network, in which each edge represents the transfer of a specified amount of the respective token between two addresses. Victor et al. [47] have analysed Ethereum transactional graph at specified token networks, with the entire set of ERC20 tokens comprising 64.393 networks capturing around 45 per cent of all unique Ethereum addresses.The smallest 500 networks consisted of only one node, while the two most extensive networks captured 1.5 million nodes. More than 80 per cent of all token networks have less than ten nodes.Studying degree distribution at each token network level, it was possible to fit the power law model for all of them, but for only 10 per cent of them, the power law hypothesis was statistically significant.

Moreover, the authors hypothesise that even when the power law hypothesis cannot be rejected, if we suppose the network contains addresses to cryptocurrency exchanges, multiple high indegree addresses will likely be present. Estimated exponents of power laws are very high, meaning the degree distributions are decaying quickly. While in social networks, the popularity of most connected nodes can grow organically, in cryptocurrency networks, the most popular nodes are limited institutionally as they represent exchanges, mining pools etc.

This relates to an issue of trust. Users are trading their tokens primarily with well-established and credible nodes. Moreover, given that the initial token distribution often follows the preferential attachment phenomenon, the entire ERC token graph is likelier to follow the power law. Preferential attachment is caused by the existence of so called 'Airdrops', which are free token allocations for particularly active users. Those airdrops will then become connectors between individual token networks.

The small-world phenomenon is present in the entire Ethereum token network, with measures of global clustering coefficient and average local clustering coefficient being even higher than for the Bitcoin addresses network. This indicates the potential for the existence of communities.

Labelling nodes that are well-known exchanges also paints an interesting picture of the type of activity users are primarily interested in. The authors have collected data from Etherscan and various internet forums on exact exchange addresses. Together with exchange labelling and assuming that two nodes with the highest outdegree on 10 per cent of initial transfers are distributors of particular tokens, they have analysed the behaviour of initial receipients of coins.

Scatterplot analysis of the fraction of active recipients to all initial recipients (active recipients are defined as nodes that have performed at least one transfer) to fraction with active recipients with an existing path to an exchange node showed there exists a clear relationship that the more active recipients where in the network, the more they tend to move token allocation to an exchange.

This can be interpreted that the primary use case for tokens is trading of them on cryptocurrency exchanges. Moreover, in 50 per cent of token networks, the mean shortest path to exchange is smaller or equal to two, further indicating that the primary use case is trading. Removal of distributing and exchange addresses from the graph showed that the median fraction of edges remaining in the networks is 42 per cent which is another indication that large parts of token networks exist only for speculative purpuses.

4.2 Ethereum Blockchain Analysis - Implementaion Issues

Due to the complexity of possible contracts to be deployed, combined with relative ease in creating new contracts, Ethereum has been widely used, which has led to the astonishing growth of the size of its underlying blockchain. Currently, the fully synced Ethereum blockchain data size is around 880GB and consists of vast information on user-to-user, user-to-contract, contract-to-user and contract-to-contract interactions. Those interactions can be modelled with complex networks. Unfortunately, mining all the data and creation of network representations for each type of interaction is complex software and data engineering task.

In order to mine the data, I have decided to deploy a tool developed by a team from Nanyang

Technological University: EtherNet [24]. The authors of EtherNet, wanted to provide other researchers with an automated end-to-end tool that would manage the Extract, Transform, Load (ETL) pipeline for massive and complex blockchain data and would easily create graph equivalent representations for the Ethereum network. Input data to this software comes directly from Google Cloud Platform, on which blockchain analytics firm Nansen provides them free of charge, updated in real-time.

Given my interest in analyzing the Ethereum network, I have decided to deploy the above-mention tool on the Google Cloud Platform. After setting up a virtual machine running on Linux on the GPC platform and following the instructions of the authors of the paper, I, unfortunately, got stuck just at the end of the process.

The IP and port addresses mentioned in the paper were no longer working. I have contacted Professor Voon Hou Su but unfortunately did not receive any reply. Other alternative methods, like dune analytics or blockdeamon services, were too costly or too simplistic for full network analysis. Running my own node and setting up the entire data management process, given my lack of experience in data engineering with working on data sets of such significant size, felt like a difficult task, mainly also due to the computational and data storage limits of my personal computer.

Because of problems with acquiring entire Ethereum blockchain data, I have abandoned the Ethereum analysis project.

References

[1] Humoud Alsabah and Agostino Capponi. "Pitfalls of Bitcoin's Proof-of-Work: R&D arms race and mining centralization". In: *Available at SSRN 3273982* (2020).

[2] Jørgen Vitting Andersen, Andrzej Nowak, et al. *An introduction to socio-finance*. Springer, 2013.

[3] Elli Androulaki et al. "Evaluating user privacy in bitcoin". In: *International conference on financial cryptography and data security*. Springer. 2013, pp. 34–51.

[4] Ayana T Aspembitova, Ling Feng, and Lock Yue Chew. "Behavioral structure of users in cryptocurrency market". In: *Plos one* 16.1 (2021), e0242600.

[5] Annika Baumann, Benjamin Fabian, and Matthias Lischke. "Exploring the Bitcoin Network." In: *WEBIST (1)* 2014 (2014), pp. 369–374.

[6] Bruno Biais et al. "Equilibrium bitcoin pricing". In: *Available at SSRN 3261063* (2020).

[7] E Bouri, R Gupta, and D Roubaud. *Herding behaviour in cryptocurrencies. Finance Res. Lett.* 2018.

[8] Alexandre Bovet et al. "Network-based indicators of Bitcoin bubbles". In: *arXiv preprint arXiv:1805.04460* (2018).

[9] Alexandre Bovet et al. "The evolving liaisons between the transaction networks of Bitcoin and its price dynamics". In: *arXiv preprint arXiv:1907.03577* (2019).

[10] Peter Cauwels and Didier Sornette. "Quis pendit ipsa pretia: Facebook valuation and diagnostic of a bubble based on nonlinear demographic dynamics". In: *The Journal of Portfolio Management* 38.2 (2012), pp. 56–66.

[11] Lin William Cong, Ye Li, and Neng Wang. "Tokenomics: Dynamic adoption and valuation". In: *The Review of Financial Studies* 34.3 (2021), pp. 1105–1155.

[12] Lin William Cong et al. "Value premium, network adoption, and factor pricing of crypto assets". In: *Network Adoption, and Factor Pricing of Crypto Assets (December 2021)* (2021).

[13] Rama Cont and Lakshithe Wagalath. "Risk management for whales". In: *Available at SSRN 2739227* (2015).

[14] Bernard Derrida, L De Seze, and Claude Itzykson. "Fractal structure of zeros in hierarchical models". In: *Journal of Statistical Physics* 33.3 (1983), pp. 559–569.

[15] Dean Fantazzini. "Modeling bubbles and anti-bubbles in bear markets". In: *TRADING* (2010), p. 365.

[16] Dean Fantazzini. "Quantitative Finance with R and Cryptocurrencies". In: *Amazon KDP, ISBN-13: 978-1090685315* (2019).

[17] Vladimir Filimonov, Guilherme Demos, and Didier Sornette. "Modified profile likelihood inference and interval forecast of the burst of financial bubbles". In: *Quantitative finance* 17.8 (2017), pp. 1167–1186.

[18] Vladimir Filimonov and Didier Sornette. "A stable and robust calibration scheme of the log-periodic power law model". In: *Physica A: Statistical Mechanics and its Applications* 392.17 (2013), pp. 3698–3707.

[19] Zalán Forró, Peter Cauwels, and Didier Sornette. "When games meet reality: is Zynga overvalued?" In: *arXiv preprint arXiv:1204.0350* (2012).

[20] Petr Geraskin and Dean Fantazzini. "Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask". In: *The European Journal of Finance* 19.5 (2013), pp. 366–391.

[21] Jan-Christian Gerlach, Guilherme Demos, and Didier Sornette. "Dissection of Bitcoin's multiscale bubble history from January 2012 to February 2018". In: *Royal Society open science* 6.7 (2019), p. 180643.

[22] Yossi Gilad et al. "Algorand: Scaling byzantine agreements for cryptocurrencies". In: *Proceedings of the 26th symposium on operating systems principles*. 2017, pp. 51–68.

[23] Martin Harrigan and Christoph Fretter. “The unreasonable effectiveness of address clustering”. In: *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*. IEEE. 2016, pp. 368–373.

[24] Voon Hou Su, Sourav Sen Gupta, and Arijit Khan. “Automating ETL and Mining of Ethereum Blockchain Network”. In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022, pp. 1581–1584.

[25] Emilie Jacobsson. “How to predict crashes in financial markets with the Log-Periodic Power Law”. In: *Master diss., Department of Mathematical Statistics, Stockholm University* (2009).

[26] Marc Jeremy. “Prediction of financial bubbles and backtesting of a trading strategy”. In: *Master diss., Department of Mathematics, Imperial College London* (2020).

[27] Anders Johansen, Olivier Ledoit, and Didier Sornette. “Crashes as critical points”. In: *International Journal of Theoretical and Applied Finance* 3.02 (2000), pp. 219–255.

[28] Andreas Kemper. *Valuation of network effects in software markets: A complex networks approach*. Springer Science & Business Media, 2009.

[29] Dániel Kondor et al. “Do the rich get richer? An empirical analysis of the Bitcoin transaction network”. In: *PloS one* 9.2 (2014), e86197.

[30] Nikolaos Kyriazis, Stephanos Papadamou, and Shaen Corbet. “A systematic review of the bubble dynamics of cryptocurrency prices”. In: *Research in International Business and Finance* 54 (2020), p. 101254.

[31] Jiaqi Liang, Linjing Li, and Daniel Zeng. “Evolutionary dynamics of cryptocurrency transaction networks: An empirical study”. In: *PloS one* 13.8 (2018), e0202202.

[32] Matthias Lischke and Benjamin Fabian. “Analyzing the bitcoin network: The first four years”. In: *Future Internet* 8.1 (2016), p. 7.

[33] Angi Liu et al. “Behavioural Fingerprint of Bitcoin-traders”. In: *https://www.youtube.com/watch?v=Wu9xWnx-xjg* (2022).

[34] Sarah Meiklejohn et al. “A fistful of bitcoins: characterizing payments among men with no names”. In: *Proceedings of the 2013 conference on Internet measurement conference*. 2013, pp. 127–140.

[35] Artur Meynkhard. “Fair market value of bitcoin: Halving effect”. In: *Investment Management and Financial Innovations* 16.4 (2019), pp. 72–85.

[36] Lars Onsager. “Crystal statistics. I. A two-dimensional model with an order-disorder transition”. In: *Physical Review* 65.3-4 (1944), p. 117.

[37] Timothy Peterson. “Metcalfe’s Law as a Model for Bitcoin’s Value”. In: *Alternative Investment Analyst Review Q* 2 (2018).

[38] Julien Prat, Vincent Danos, and Stefania Marcassa. “Fundamental pricing of utility tokens”. In: (2019).

[39] Cazabet Remy, Baccour Rym, and Latapy Matthieu. “Tracking bitcoin users activity using community detection on a network of weak signals”. In: *International conference on complex networks and their applications*. Springer. 2017, pp. 166–177.

[40] Min Shu and Wei Zhu. “Real-time prediction of Bitcoin bubble crashes”. In: *Physica A: Statistical Mechanics and its Applications* 548 (2020), p. 124477.

[41] Shahar Somin, Goren Gordon, and Yaniv Altshuler. “Social signals in the ethereum trading network”. In: *arXiv preprint arXiv:1805.12097* (2018).

[42] Didier Sornette and Frank Cuypers. “Why stock markets crash: Critical events in complex financial systems”. In: *Physics Today* 57.3 (2004), pp. 78–79.

[43] H Takagi. “Exploring the Endogenous Nature of Meme Stocks Using the Log-Periodic Power Law Model and Confidence Indicator”. In: *Journal of Matehmatical Finance* 12 (), pp. 263–274.

[44] Bishenghui Tao et al. “Complex network analysis of the bitcoin transaction network”. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 69.3 (2021), pp. 1009–1013.

[45] Nitin K Tyagi, Mukta Goyal, and Adarsh Kumar. “Game Theory-Based Proof of Stake Mining in Blockchain for Sustainable Energy Efficiency”. In: *International Conference on Artificial Intelligence and Sustainable Engineering*. Springer. 2022, pp. 121–132.

[46] Hal Varian. “Economic Scene; Paper Currency can have Value without Government Backing, but such Backing adds Substantially to its Value”. In: *The New York Times* (2014).

[47] Friedhelm Victor and Bianca Katharina Lüders. “Measuring ethereum-based erc20 token networks”. In: *International Conference on Financial Cryptography and Data Security*. Springer. 2019, pp. 113–129.

[48] Spencer Wheatley et al. “Are Bitcoin bubbles predictable? Combining a generalized Metcalfe’s law and the log-periodic power law singularity model”. In: *Royal Society open science* 6.6 (2019), p. 180538.

[49] Xing-Zhou Zhang, Jing-Jie Liu, and Zhi-Wei Xu. “Tencent and Facebook data validate Metcalfe’s law”. In: *Journal of Computer Science and Technology* 30.2 (2015), pp. 246–251.

[50] Lin Zhao et al. “Temporal analysis of the entire ethereum blockchain network”. In: *Proceedings of the Web Conference 2021*. 2021, pp. 2258–2269.