Politecnico di Milano

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING
Master of Science – Mathematical Engineering

Cryptographic tokens: analysis and classification with a focus on a market index for cryptocurrencies

Supervisor
Prof. Daniele MARAZZINA

Co-Supervisor
Dott.ssa Nadia FABRIZIO

Candidate
Gian Marco PETTINE - Matr. 905690

Academic Year 2020 - 2021

Contents

List of Figures

List of Tables

Abstract

Cryptographic tokens are digital assets or authorisations created on a blockchain to exchange value, exchange verifiable data and achieve coordination between users. The development of these tools is happening very rapidly thanks to the blockchain technology that allows tokens to be created easily and at low cost. As a result, the implementation of token-based systems is becoming more and more frequent and this could lead to substantial changes in various economic and business models; moreover, cryptocurrencies, the most popular type of token at the moment, represent a new asset class that is gaining increasing interest in the financial sector. However, in order for tokens to become more widely used, it will be essential to develop the theory behind them in more detail and to solve some of the current problems, such as those related to regulatory aspects and network governance.

This thesis is structured in two parts. The first one provides a definition of tokens and collects, exposes and develops several research results in this field. In order to provide a clear and articulated vision of these instruments, their innovative features are outlined, the current issuing methods are exposed and the regulatory framework of reference is introduced. Particular attention is paid to the classification of tokens, elaborated on the basis of numerous previous studies. The second part of the thesis focuses on cryptocurrencies. In particular, a specific market index is studied, the CRIX (CRyptocurrencies IndeX), developed by the blockchain research unit of the Humboldt University of Berlin. The CRIX, which can be useful in various fields ranging from finance to further studies on tokens, is carefully analysed for its construction procedure, as it is very innovative since it is almost completely data-driven.

Sommario

I token crittografici sono asset digitali o autorizzazioni creati su una blockchain per scambiare valore, scambiare dati verificabili e conseguire il coordinamento tra gli utenti. Lo sviluppo di questi strumenti sta avvenendo molto rapidamente grazie alla tecnologia blockchain che permette di creare i token facilmente e a basso costo. Di conseguenza, l'implementazione di sistemi basati sui token è sempre più frequente e ciò potrebbe determinare cambiamenti sostanziali per vari modelli economici e di business; inoltre, le criptovalute, la tipologia di token più popolare al momento, rappresentano un nuovo genere di asset class che sta raccogliendo sempre crescente interesse nell'ambito finanziario. Tuttavia, perché si abbia una più ampia diffusione dei token, sarà essenziale sviluppare più dettagliatamente la teoria che ne costituisce la base e risolvere alcuni problemi attualmente presenti, come quelli legati agli aspetti normativi e alla governance della rete.

La tesi è strutturata in due parti. La prima fornisce una definizione di token e raccoglie, espone e sviluppa diversi risultati di ricerca in questo ambito. Al fine di fornire una visione chiara e articolata di questi strumenti, sono delineate le loro caratteristiche innovative, esposte le attuali modalità di emissione e introdotto il quadro normativo di riferimento. Particolare attenzione viene dedicata alla classificazione dei token, elaborata sulla base di numerosi studi precedenti.

La seconda parte della tesi è incentrata sulle criptovalute. In particolare, è stato studiato uno specifico indice di mercato, il CRIX (CRyptocurrencies IndeX), sviluppato dall'unità di ricerca in ambito blockchain dell'università Humboldt di Berlino. Del CRIX, che può essere utile in diversi ambiti che vanno dalla finanza ad ulteriori studi sui token, è stata analizzata accuratamente la procedura di costruzione, in quanto molto innovativa poiché quasi completamente data-driven.

Introduction

This thesis is the result of a research work in the blockchain sector carried out during the internship at Cefriel, a company based in Milan which operates in the field of innovation and research. The objective of the internship was to study tokens, as they are tools of increasing impact in the blockchain environment and may represent the real killer application of this new technology. Tokens are used on a blockchain to exchange value, exchange verifiable data and achieve coordination between users: this gives rise to a multitude of possibilities that are bringing increasing attention to these instruments by actors from different sectors. Today, more than three hundred thousands tokens were deployed just on the Ethereum blockchain[1]. One of them, Efforce, was released in November 2020 by Apple co-founder Steve Wozniak to transform and revolutionize the energy efficiency market [12]. Not only strong personalities in the IT industry, but also big actors in the financial sector are sending strong signals toward the blockchain sector. The October 21^{st}, 2020, Paypal decided to enter this market, allowing transactions of Bitcoins and other cryptocurrencies in its system. In the meanwhile, strong attempts by regulators to clarify the legal framework surrounding tokens appeared: in Europe, in the recently released EC digital finance package, the Markets in Crypto-Assets Regulation (MiCAR) [11] tries to find a legal framework for still unregulated crypto-assets.
Within the company reality, projects requiring the design of tokens are becoming more and more frequent and therefore it was deemed necessary to undertake a study on their characteristics, which can lay the foundations for this type of works.

The study of tokens is carried out on the basis of the current literature which, however, is still in-homogeneous and with gaps to be filled. The lack of organisation and uniformity in the theory of tokens can lead to difficulties when, for example, a token has to be designed, because each of its properties influences all the others and should be evaluated carefully.
For this reason, the main objective of this thesis is to provide a clear picture of cryptographic tokens, with a structured and comprehensive approach. A further focus is made on cryptocurrencies with a more financial analysis.
The thesis is divided in two chapters. The first one starts with a broad analysis of cryptographic tokens. It begins by providing a definition of them, as there is no agreed one. In addition, it covers several aspects, from the benefits and problems of developing tokens to legal issues, but focuses mostly on the tokens classification.
Subsequently, Bitcoin and Ethereum are also presented, mainly for two reasons. The first is that they represent the most important blockchains and cryptocurrencies cur-

[1] Source: https://etherscan.io

rently available, with the greatest market reach and capitalization. The second is that they are at the center of the studies of the second chapter, as the main components of the market index presented.
The second chapter describes the building procedure of a market index for cryptocurrencies called CRIX (CRyptocurrencies IndeX) [37] and introduces some of the reasons why it may be useful for future analysis. Indeed, this index could be of great interest for several financial applications, such as building portfolios made of cryptocurrencies, or capturing the differences between this market and others. Moreover, it could be used to understand further if there are characteristics of cryptocurrencies that influence their price.

In this thesis, especially in the first chapter, there is a frequent use of quite technical and specific terms regarding the blockchain technology. To clarify concepts for which a definition is not provided, it is recommended to rely on the reference glossary [18], developed by ISO, the International Organization for Standardization. It contains all the basic definitions concerning the blockchain technology.

Chapter 1

Tokens

The objective of this chapter is to provide the basics of cryptographic tokens, explaining what are these objects, what they are used for, how they are created and what is the legal framework surrounding them.

The main focus is on their classification, which is considered essential to fully understand these objects. Indeed, tokens can have many different facets and their function in applications is strongly influenced by the particular characterization given to them.

The last part of the chapter is dedicated to Bitcoin and Ethereum, the most famous blockchains currently in existence. Their native cryptocurrencies are the focus of the studies presented in the next chapter.

1.1 Cryptographic tokens: definition, benefits and future challenges

The concept of tokens has existed since before the emergence of blockchain networks and in several different fields, since they can represent any form of economic value or access right.

It is not difficult to find examples of tokens [38]: "casinos chips, vouchers, gift cards, bonus points in a loyalty program, coat check tokens, stock certificates, bonds, concert or club entry tokens represented by a stamp on your hand, dinner reservations, ID cards, club memberships, or train or airline tickets" are some possible examples. Even banknotes or coins are tokens. In computer science tokens may provide the right to perform some action or manage access rights. In psychology, tokens have been used as an incentive mechanism to stimulate patients to have a specific desired behavior.

In general, tokens always need a substrate to guarantee their validity and some built-in anti-counterfeiting measures to ensure that people do not cheat the system. Consequently, historically, there have always been entities that have created and managed tokens, acting as guarantors.

Now, tokens of any kind can be defined by everyone through a set of rules in software logic, expressed in a Turing-complete language on a blockchain: their validity and security is managed through an algorithm and the blockchain on which they are created. In this case they are called cryptographic tokens. There is no shared definition for these objects: the following one is developed on the basis of several researches on blockchain and tokens, such as [18], [28] and [38], which provide different points of view.

Definition 1 (Cryptographic tokens) *Cryptographic tokens represent digital assets*[1] *or authorizations issued on a blockchain to exchange value, exchange verifiable data and achieve coordination between users. They are managed by the blockchain protocol or by smart contracts.*

The aim behind the formulation of this definition has been to make it as comprehensive and generic as possible, since tokens have many facets, while keeping it simple and consistent with the rest of the work presented.

Before the advent of blockchain technology, developing tokens systems was already feasible, but not cost-effective to manage. The blockchain technology decreases the costs associated with developing tokens systems and it increases the likelihood that tokens will be adopted and spread.
A first fundamental characterization can be made to better understand the concept of cryptographic token and to introduce the central idea of asset "tokenization".
Some tokens could be native to a blockchain, i.e. built directly on-chain and living exclusively on the distributed ledger. In some cases they may be needed for the correct functioning of the blockchain itself. Examples of native tokens are Bitcoin and Ethereum.
Others are defined on the top of a blockchain through a smart contract. Examples are Ethereum tokens. Among these, some may be backed by an existing asset. Indeed, the development of the blockchain system has introduced a new technological support for the so-called "tokenization", defined in [17] as follows.

Definition 2 (Tokenization) *Tokenization (or tokenisation) is the process of representing in a digital way an existing real asset on a blockchain.*

Asset-backed tokens embed the rights and the value of the assets they represent. They will be further described later indeed, in the section 1.4 of this chapter, this first classification is recalled and further extended, introducing several other distinctive features.

The token economy [25], as described in [21], [29] and [38], offers a new perspective on some economic and business models. The main improvements relate to transparency, efficiency and new opportunities in the exchange of data, value and rights, as well as the possibility of creating completely new asset classes. However, there are also risks arising from the still premature development of this new frontier.
With blockchain and cryptographic tokens, data, assets, or authorizations of any kind can be shared among different markets and jurisdictions in a way that facilitates collaboration and improves trust and transparency between counter-parties. Once any kind of data, transaction, or token exchange is recorded on the blockchain, then it is immutable, verifiable by the entire network and securely stored because of the absence of a single point of failure.
An improvement in transparency and trust is also achieved by using smart contracts, which may automate some steps of several processes. In this way, it is sure that all

[1] From the glossary [18]: "an asset is anything that has value to a stakeholder", "a digital asset is an asset that exists only in digital form or which is the digital representation of another asset".

the conditions underlying a process are met before it is done.
Of course, the increased transparency may be desirable also for regulators, who thus have increased control over all operations carried out.
The automation provided by smart contracts affects heavily also the efficiency and the costs related to several operations: in many situations a third party can be excluded. A specific application of the increment of efficiency of a process is found for clearing and settlement services. Indeed, the market participants could benefit to a possible faster clearing and settlement, thanks to the blockchain technology. In the current system the settlement procedure is not as efficient as it could be, as explained in [32]. Both free delivery [2] and delivery vs payment (DVP) [3] present some problems related to settlement and counter-party risks, as well as possible inefficiencies due to the presence of intermediaries. Tokens on a blockchain could solve these issues because the technology guarantees that both transfers have to happen, in the sense that it is not technologically possible that only one transfer is executed if the other one is stopped. Moreover, the intermediary could be no longer necessary. Increased efficiency in clearing and settlement processes may have very strong implications for financial markets and may result in reduced counter-party and operational risks.
These potential benefits, certainly found in permissioned blockchains, may disappear in permissionless blockchains because of the uncertainty around settlement finality. However, through multi-chain atomic swaps, i.e. the exchange of two cryptographic tokens simultaneously in a single operation and without going through any centralised intermediary, the benefits are achieved also in a permissionless setting.

Another major benefit of these new technologies arises from the possibility of representing assets on a blockchain through tokens.
Indeed, tokenization could expand some markets such as the art, real estate and luxury ones, which are usually targeted at a limited audience. A larger number of investors, including retail, could be reach because the minimum investment required can be lowered, e.g. by offering the possibility of fractional ownership of an asset, which can be achieved through a token.
In general, the benefits for these products could be multiple and include increased liquidity and lower costs of price discovery. Indeed, issuers of tokenized assets will also be able to trade them on secondary markets of their choice, which can increase their liquidity by the access to a wider network of traders. This creates an advantage for both investors, by providing them more freedom in the sale, and sellers, because the tokens benefit from the reduced liquidity premium[4]. Moreover, in a scenario of widespread tokenized assets, the number and diversity of assets that would trade in public markets could increase.
Trading tokenized assets will be very different from traditional trading. Investors can be the broker-dealers for themselves, because the blockchain enables transactions in a

[2] A free delivery is a procedure whereby delivery of securities sold is made to the buying customer's bank without requiring immediate payment. Antithesis of delivery vs. payment.

[3] DVP is a securities industry settlement method that guarantees the transfer of securities only happens after the payment has been made.

[4] A liquidity premium is an additional value required by investors when any given asset cannot be easily and efficiently sold or otherwise converted into cash for its fair market value. An asset with an high liquidity premium is said to be "illiquid". In some sense illiquidity is an added risk for an investor which buys an asset, that therefore should be rewarded.

situation where there is partial or zero trust between the users, without the necessity of a central entity. This implies that an increased use of tokenization could lead to a disruption of the market-making model, and therefore of the structure and the proper functioning of the markets. The latter topics will not be explored further in this thesis, because a complete explanation would require details that are complex and off-topic. In [21] and [29] deeper analyses are presented.

While some of the benefits introduced are still being studied, the tokenization of some assets is already happening. For example, on the Ethereum blockchain there exist a token backed by gold bars, called Pax Gold[5]. Moreover, recently the tokenization of fine wine bottles[6] by a Swiss company happened.

All the implications previously described may also result in completely new use cases, business models, and asset types that before were not economically convenient to develop. In [9], the development of completely new value creation models is investigated: there is the possibility of a "democratization of entrepreneurship and innovation". Indeed, tokens:

- allow entrepreneurs to collect funds from investors all over the world, without intermediaries and without any special prerequisites. In this way, access to financial capital can be more democratic. The fundraising process may become way more efficient, more effective and less localized;

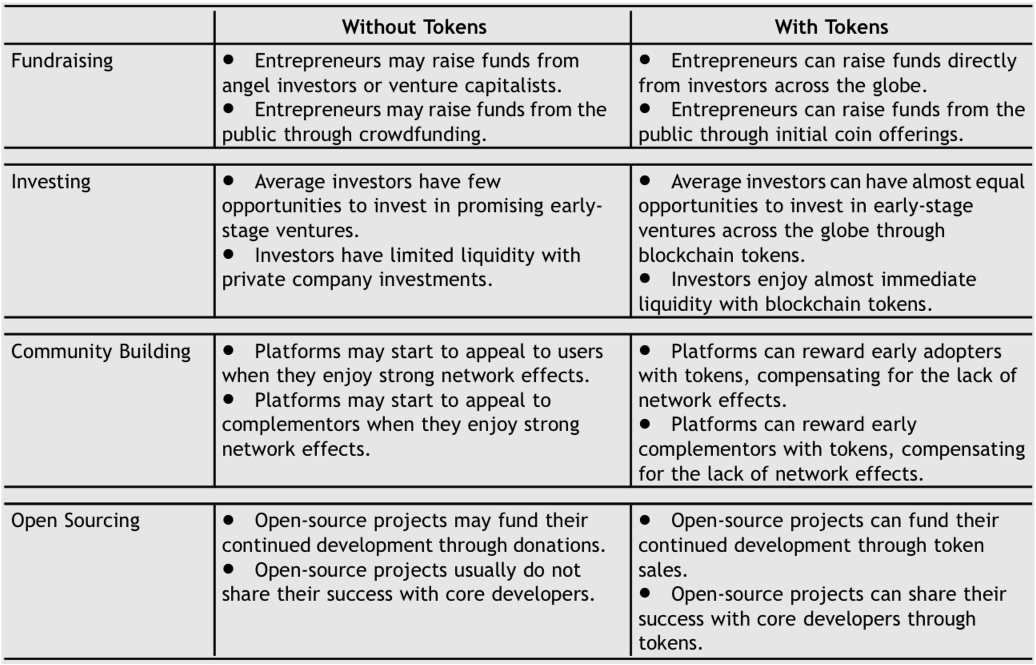
- provide investors with the possibility to invest in early-stage projects worldwide, making the access to investment opportunities open to everyone, easier and less risky.
  With tokens, the minimum investment period is affected because of their increased liquidity with respect to traditional assets. Indeed, investors could exchange them more easily, due to enhanced trading possibilities on secondary markets of their choice, thus making the investment less risky.
  In addition, the minimum investment amount can also be reduced in the case of tokens that can be divided into fractions. Thanks to the blockchain, splitting a token does not entail any additional costs.
  Some markets may become more inclusive for retail investors, who usually do not have large amounts of capital and are more risk-averse. Clearly, the increased possibilities that tokenized markets offer to retail investor must be followed by a public investment in financial education: in this way a mass adoption could be less difficult and problematic;

- facilitate the building of user and developer communities by reward mechanisms, making it easier, for instance, to encourage early adoption, usage of the network and contribution to a collective goal.
  The correct design of the token is essential to achieve this objective. In section 1.4 some possible choices are presented.

The latter concepts are schematically illustrated in Figure 1.1 from [9], which also compares the situations with and without tokens.

[5] https://www.paxos.com/paxgold/

[6] https://www.coindesk.com/crypto-bank-sygnum-tokenization-dlt



	Without Tokens	With Tokens
Fundraising	• Entrepreneurs may raise funds from angel investors or venture capitalists. • Entrepreneurs may raise funds from the public through crowdfunding.	• Entrepreneurs can raise funds directly from investors across the globe. • Entrepreneurs can raise funds from the public through initial coin offerings.
Investing	• Average investors have few opportunities to invest in promising early-stage ventures. • Investors have limited liquidity with private company investments.	• Average investors can have almost equal opportunities to invest in early-stage ventures across the globe through blockchain tokens. • Investors enjoy almost immediate liquidity with blockchain tokens.
Community Building	• Platforms may start to appeal to users when they enjoy strong network effects. • Platforms may start to appeal to complementors when they enjoy strong network effects.	• Platforms can reward early adopters with tokens, compensating for the lack of network effects. • Platforms can reward early complementors with tokens, compensating for the lack of network effects.
Open Sourcing	• Open-source projects may fund their continued development through donations. • Open-source projects usually do not share their success with core developers.	• Open-source projects can fund their continued development through token sales. • Open-source projects can share their success with core developers through tokens.

Figure 1.1. How blockchain tokens may change entrepreneurship and innovation [9].

Tokens could allow the creation of a new financial system, that is more democratic, more vast and more efficient than the current one. The following sentence from an article in Deloitte's Inside magazine issue 19 [21] is enlightening to understand all the power and enthusiasm surrounding this new technology: "*We foresee that tokenization could make the financial industry more accessible, cheaper, faster and easier, thereby possibly unlocking trillions of euros in currently illiquid assets, and vastly increasing the volume of trades.*".
However, there are still several challenges to face for a wider adoption of tokens [29].

1. Problems related to the underlying permissionless blockchain technology.

   (a) Scalability, settlement finality, and interoperability are common issues to many blockchains.

   (b) Operational risks, like for example network stability, cyber-risk, risk of hacking and 51% attacks. Moreover, the robustness of blockchain-based infrastructures may be undermined in the future due to rapid technological progress. In particular, the development of quantum technologies could threaten the security of cryptographic protocols based on symmetric and asymmetric cryptography.

   (c) Problems in the governance of the network, related to the absence of an owner or a single point responsible for the full network. Users can perform a 51% attack or can "fork", for instance to deviate to a different protocol.

2. It is well known that a big challenge for the development of these systems is making the legal/regulatory system clearer. This argument is addressed in the section 1.3 of this chapter.

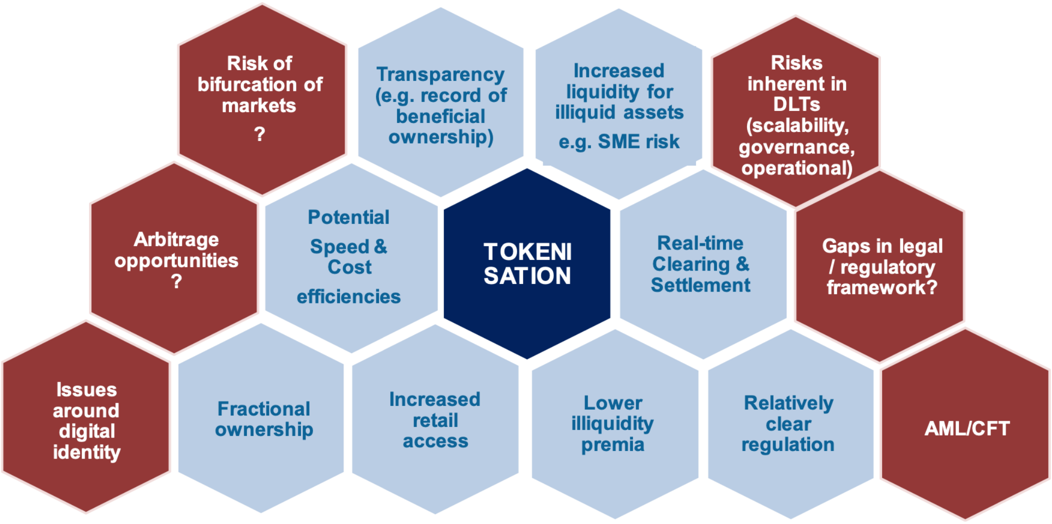

Risk of bifurcation of markets ?
Transparency (e.g. record of beneficial ownership)
Increased liquidity for illiquid assets e.g. SME risk
Risks inherent in DLTs (scalability, governance, operational)
Arbitrage opportunities ?
Potential Speed & Cost efficiencies
TOKENI SATION
Real-time Clearing & Settlement
Gaps in legal / regulatory framework?
Issues around digital identity
Fractional ownership
Increased retail access
Lower illiquidity premia
Relatively clear regulation
AML/CFT

Figure 1.2. Tokenization: benefits and future challenges [29].

3. The lack of systems for the development and management of a digital identity for users on blockchain could produce big issues. For instance, there are currently no systems to avoid many market manipulation methods, like wash trading[7]. These methods could be used to artificially affect the price of cryptographic tokens and could be performed in a framework where investors are not identified. Regulators may benefit from a better development and management of digital identity.

4. Generic risks concerning anti-money laundering (AML) and countering the financing of terrorism (CFT) have to be faced, in particular on blockchains where an high anonymity of users is guaranteed.

Through the Figure 1.2 from [29] it is possible to have a clear summary of the benefits and future challenges of tokenization.

1.2 Issuing tokens

Initial Coin Offerings (ICOs) (or Initial Token Offerings (ITOs)) are token sales mechanisms to exchange new cryptographic tokens with existing ones via a P2P trade on a blockchain, and became popular with the rise of the Ethereum network.
Usually, the main purpose of token sales is to fund new projects: in general, the released tokens are exchanged for a fiat currency or a very liquid cryptocurrency like Bitcoin or Ethereum [38].
ICOs, even for their name, can be reminiscent of traditional IPOs. However, there are huge differences between the two things, that make them almost impossible to associate, at least until now.

[7] Wash trading is a market manipulation technique in which an investor simultaneously sells and buys the same financial instruments to create misleading and artificial activity in the marketplace.

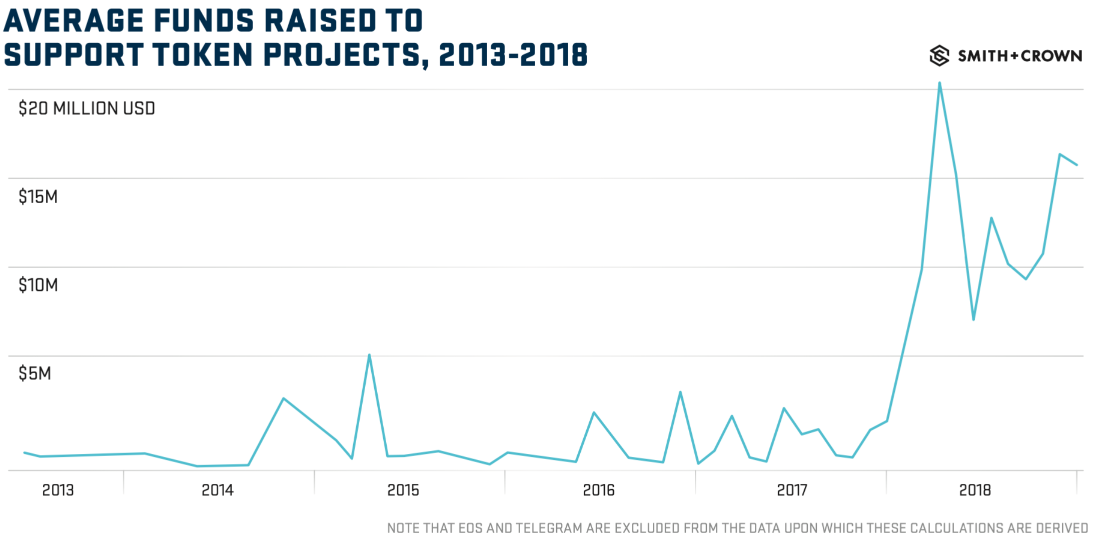


Figure 1.3. Funds raised in ICOs [8].

A first big distinction is from the regulatory point of view: ICOs regulations are not so stringent yet, a big difference with traditional IPOs. Moreover, legally they are treated like a crowdfunding operation or a pre-buy of a product, even if there are possibilities like the return on an investment that the latter categories usually do not provide.

The legal framework surrounding ICOs is still in a developing phase. The problem is that, in addition to some basic rules that can apply to ICOs of any kind, it is then necessary to reason in token-specific terms. In fact, tokens can be of completely different types that fall into opposite legal frameworks and this certainly affects the ICO phase. First, maybe, it is necessary to provide a clear legal classification of tokens and then to proceed with ICOs in a more structured way than today.

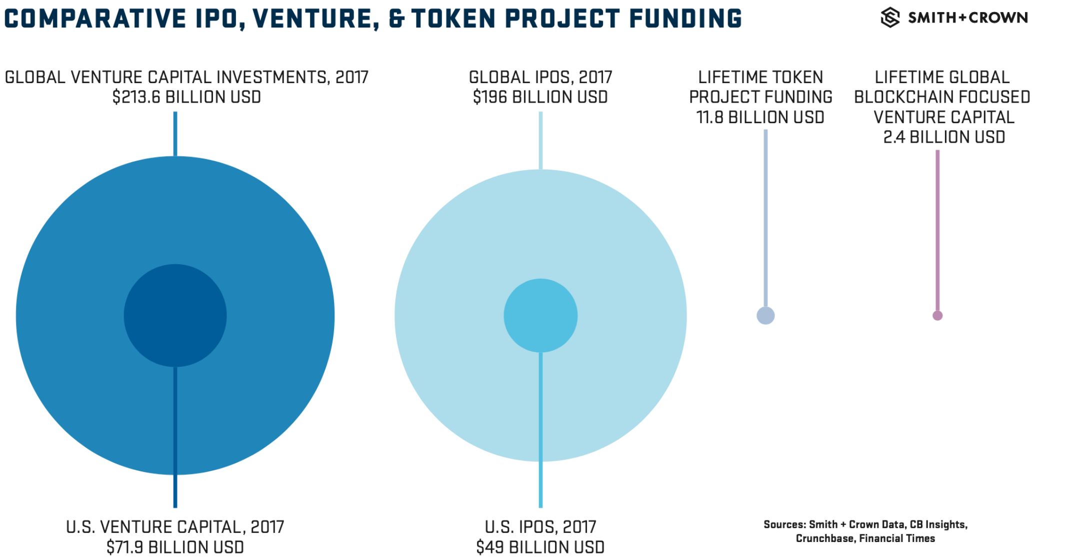
The lack of regulation has contributed to creating a negative aura around these funding methods over the years, especially among uninitiated, because there have been numerous scams related to sales of tokens.

Anyway, despite of that, funds issued through ICOs are growing year by year. In Figure 1.3 from [8] the growing trend of the average size of funds raised through ICOs is clearly visible.

However, there is still a long way to reach the funds raised with IPOs or global venture capital investments. Figure 1.4 from [8] clearly shows this difference.

Usually, in the pre-launch phase, token developers present a so-called whitepaper describing the technical specifications of the project and the business plan. There are not general rules about whitepapers contents, but practice has shown that there are some recommendations that may be useful to attract investors. These simple suggestions are presented in Figure 1.5 from [8].

The focus on ICOs and whitepapers will not go more in details since these arguments are not of central interest for this thesis. However, several more information can be found in the cited papers.



Sources: Smith + Crown Data, CB Insights, Crunchbase, Financial Times

Note: Token project fundraising includes all token projects that raised over $25,000 and did not return funds raised to participants. Amounts raised are valued according to average daily exchange rates on the date the fundraising closed. EOS's ongoing raise is valued according to the total raised during each auction period and grouped into monthly amounts. Rounds for the same project that are separated by more than 30 days are treated as separate fundraising events. Some data may be missing or subject to future revision.

Figure 1.4. Funds raised in 2017: venture capital investments vs IPOs vs ICOs [8].

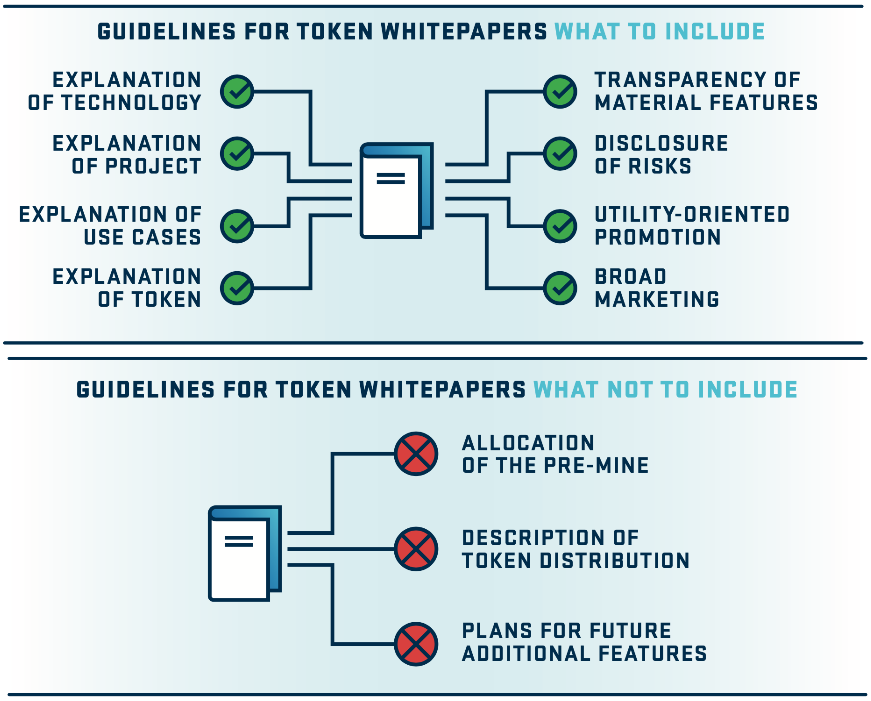

GUIDELINES FOR TOKEN WHITEPAPERS WHAT TO INCLUDE
EXPLANATION OF TECHNOLOGY
EXPLANATION OF PROJECT
EXPLANATION OF USE CASES
EXPLANATION OF TOKEN
TRANSPARENCY OF MATERIAL FEATURES
DISCLOSURE OF RISKS
UTILITY-ORIENTED PROMOTION
BROAD MARKETING
GUIDELINES FOR TOKEN WHITEPAPERS WHAT NOT TO INCLUDE
ALLOCATION OF THE PRE-MINE
DESCRIPTION OF TOKEN DISTRIBUTION
PLANS FOR FUTURE ADDITIONAL FEATURES

Figure 1.5. Guidelines for whitepapers: what to/not to include [8].

To complete this section, after describing how to launch a token and how a whitepaper should be structured, it should be specified when it is the case of issuing a token. This can be enclosed in simple rules.

1. It is suggested to issue a token if its role in the network is already clear before the ICO. There have been several cases of launches of tokens with an unclear purpose: in this case the lack of clearness could just reveal a weak idea from the creator. This could reflect on the investor which consequently may not give his trust or may lose his money.

2. It is not recommended to issue tokens just as a funding instrument. Indeed, issued tokens should always have a role in the network.

3. It is relevant to carefully evaluate cases where there might be conflict of interests between token holders. Indeed, a token may be seen:

    - as an utility asset. This is the case when it has a functionality in the network, for example to perform some kind of operations. These tokens are usually called *network tokens*;
    - as an investment asset. In this case the token is owned as an investment tool (*investment token*).

    In several cases some tokens may be seen, at the same time, in both ways: they are referred to as *entangled tokens* [23]. The latter are not desirable for a network as they may create a conflict of interest between token holders. This is because those who own an entangled token for its functionality in the network would prefer its price to be stable. Conversely, those who use the same token as a means of investment would like to see its price change. Eventually, the functioning of the network may be compromised.
    To avoid this problem one of the solutions is of disentangling the usage part of the token from its function as an investment vehicle. This could be accomplished, for instance, by issuing two different tokens, each with an unique purpose.

1.3 Overview of the regulatory framework

From a regulatory point of view there are many questions surrounding tokens, that heavily change from a legal system to another. Sometimes, a so-called “functional equivalence” approach has been used to provide them a legal classification: it consists in trying to relate tokens to already existing objects, without the need of a completely new and ad hoc regulatory framework. This is the case, for instance, of several tokens that are treated like securities or commodities in some legal contexts.
The question being investigated is if this is really sufficient to regulate such a disruptive technology.
Moreover, sometimes also for the simpler classification cases there are some issues to be solved. One of the most important, for instance, concerns the concept of ownership of a token.
Technically speaking, every token belongs to a blockchain address and is accessible

with a software, the wallet, that communicates with the blockchain and manages the public-private key pair. Using the private key one can, for instance, sign a transfer of the token or perform other actions depending on the token type. So, the person who has the private key for an address can be regarded as the owner of the respective tokens.

However, legally speaking, the ownership concept is not so clear. For instance, it is not always easy to connect a blockchain address with a physical entity, and sometimes the blockchain itself is designed to preserve this kind of privacy. The control of the physical owner of a token becomes even more difficult thinking about the exchanging possibilities that a blockchain offers. Indeed, there are opportunities for decentralized intercontinental trades, also between hostile countries with completely different laws, that can make the tracking of a token very difficult if not impossible.

These issues alone provide a clear view of the difficulty one may have in regulating products that offer so many possibilities previously unavailable. Several researches worldwide are being done in order to clarify regulatory concepts and speed up the development of clear rules.

Other problems that are being studied regard, for example, concrete applications of the functional equivalence approach to understand if, legally, a given token is a security (in the US through the so-called Howey test) or a commodity (BTC and ETH are in the latter category), or, in general, how a token underlying purpose, generating mechanism and distribution impact the legal landscape. In Figure 1.6 from [8] some token types for legal classification are proposed. In the legal context, the *utility* class is usually reserved for tokens that are neither currencies, nor commodities, nor securities. In general, these types of tokens have a purpose and an utility confined to the platform where they were created. However, the most common type is that of *hybrid token*, i.e. a token which includes characteristics from several of the other types. Hybrid tokens are not included by many legal frameworks yet. Indeed, they are usually placed in the category, among the other four, that best represents their dominant characteristics.

The previous distinction is deeply explained and expanded in section 1.4 about classification. In particular, this is done for the utility type, which includes several subcategories.

A deeper classification could be necessary to expand the legal one, which is quite generic. Indeed, as of today, just on the Ethereum network there are hundred of thousands of issued tokens, all with different underlying purposes and developed from different innovative ideas. It may happen that some tokens may no fit well in the latter legal categories.

There are several benefits that a clear regulation may bring, perhaps leading to a mass adoption.

Indeed, the ability to predict the legal category of cryptographic tokens can avoid unnecessary problems that could obstruct the development and efficiency of new projects, and thus of the entire market. Also consumers could be advantaged as they can access more projects. Finally, legal predictability may help regulators themselves as it could keep away all those that are not interested in the development of a real project, but only in taking advantage of the poorness of regulatory guidance [8].

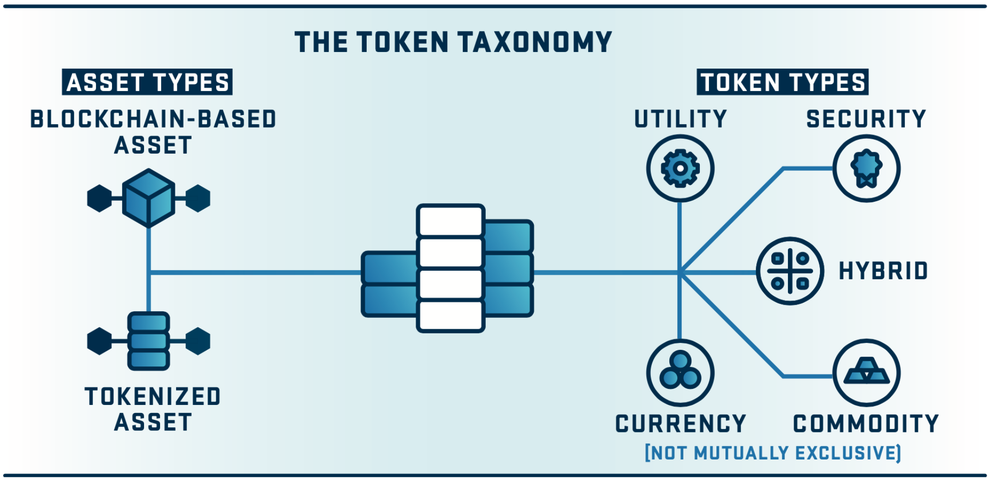


Figure 1.6. The token taxonomy: some categories for legal classification of tokens [8].

Lack of regulation increases confusion, so it's strictly necessary to intervene in an effective way.

Many jurisdictions are adapting to these new technologies and the following years may be increasingly characterized by the integration of tokens into the legislature. The European framework for blockchain and cryptographic tokens will change drastically in the next years. Right now, the legal situation is very fragmented and not harmonious between the various European states. Indeed, the development of regulations for blockchain and tokens is at very different stages from one country to another: some possibilities of "regulatory arbitrages" between jurisdictions may arise. In [14] the authors investigate on the development of the blockchain ecosystem in Europe, with a focus on 29 countries (27 EU member states, Switzerland and United Kingdom). The situation in each state is analyzed in two dimensions.

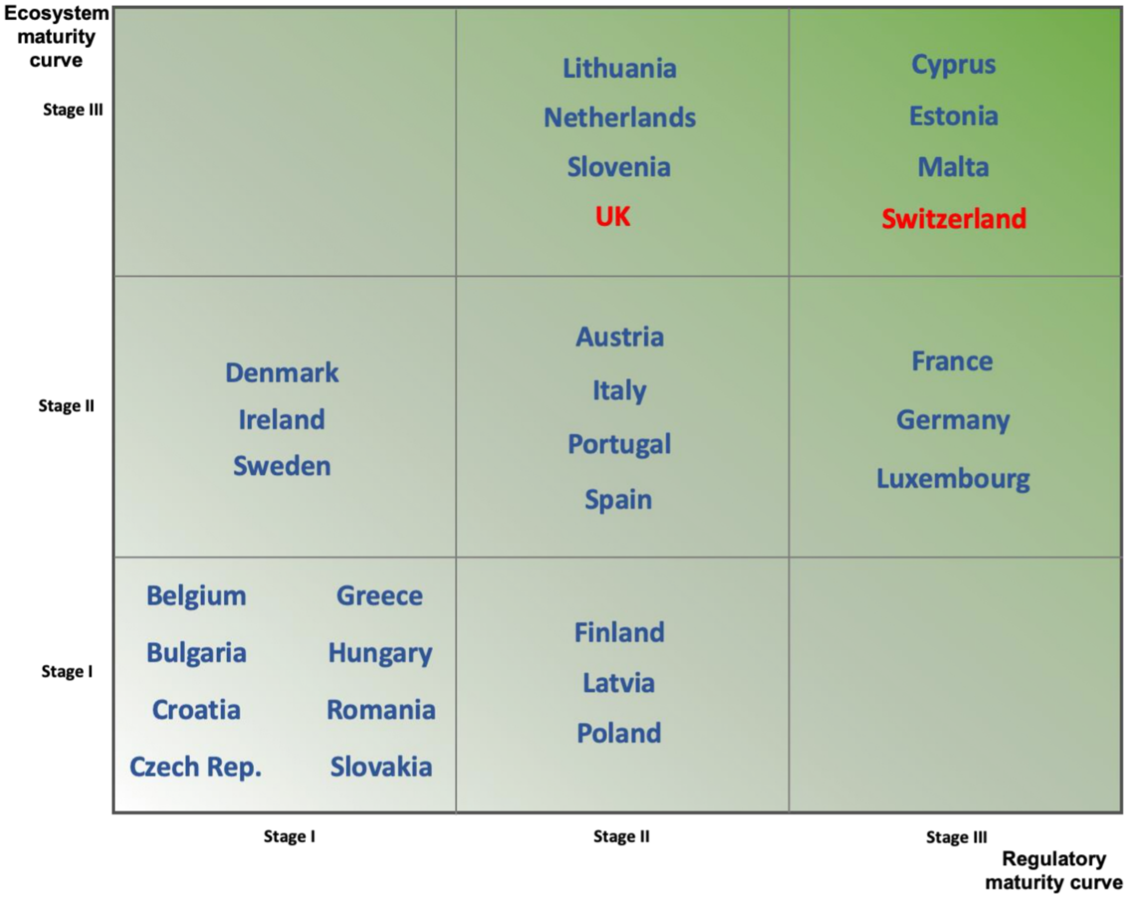
1. Regulatory maturity: it measures the status of legislation and the presence of government measures in support of the blockchain world. Each country is classified into three possible stages.

   - Stage 1: there is almost no legislation on blockchain and cryptographic tokens.
   - Stage 2: the government is involved in the field. Some regulations are object of discussion and there are government researches. There is an established framework for taxation and anti-money laundering.
   - Stage 3: there are laws for blockchain and tokens. The government is committed to strategies for the development of the technology.

2. Ecosystem maturity: it measures the level of development of the entire ecosystem in terms of three indicators:

   (a) presence of business ecosystems based on DLT technologies;

   (b) number of communities focused on blockchain and tokens;



Ecosystem maturity curve	Stage I	Stage II	Stage III
Stage III		Lithuania Netherlands Slovenia UK	Cyprus Estonia Malta Switzerland
Stage II	Denmark Ireland Sweden	Austria Italy Portugal Spain	France Germany Luxembourg
Stage I	Belgium Bulgaria Croatia Czech Rep. Greece Hungary Romania Slovakia	Finland Latvia Poland	

Figure 1.7. Classification of the EU countries in terms of regulatory and ecosystem maturity, as specified in [14].

(c) academic initiatives around the blockchain world.

For each of the indicators, a success threshold is established to determine whether a country has achieved sufficient development in the area described by the indicator. Also in this case, each country may fall in one of three stages.

- Stage 1: one or none of the indicators is above the threshold.
- Stage 2: two indicators are above the threshold.
- Stage 3: all the indicators are above the threshold.

The resulting classification is shown in Figure 1.7 from [14], which gives a clear and quick overview of the development of the framework of blockchain and cryptographic tokens in Europe.

In the Digital financial package of September 2020 produced by the European Commission, it has been introduced the Markets in Crypto-Assets Regulation (MiCAR) ([11] for the full text, [36] for an overview). It will, without a doubt, permanently alter the digital asset landscape in Europe.
The MiCAR focuses strongly on rules to regulate currently out-of-scope crypto-asset

types, such as stablecoins, as well as crypto-asset service providers. Moreover, it could make the European legal framework more harmonious between countries.
It has 4 broad objectives.

1. To provide legal certainty for crypto-assets not covered by existing EU legislation.
2. To establish uniform rules for crypto-asset service providers and issuers in EU.
3. To substitute existing national frameworks applicable to crypto-assets not covered by existing EU financial services legislation.
4. To establish specific rules for stablecoins.

The EU will put in place all the framework, enabling the uptake of distributed ledger technology and crypto-assets in the financial sector, by 2024.

1.4 Features and classification of cryptographic tokens

This section concerns the classification of cryptographic tokens which is, along with the topics described in section 1.1, the most crucial argument in the chapter.
Indeed, from the previous contents, it should be clear that the ability to provide a proper classification of a token could:

- help in its design in the pre-launch phase. Indeed, it is fundamental to design a token that fulfills at the best what the network is meant for. Having a detailed classification available can help to make the best choices. Moreover, it could be also a solid ramp for developing new concepts;
- help in finding investors, who have been shown to be more likely to invest in well-defined projects;
- increase the speed of development of a comprehensive regulatory framework. Legally, it is easier to develop rules for an object that is well known in all of his facets;
- generally speed up token deployment. A current big challenge for mass adoption is that tokens are easy to describe technically, but a general non-technical definition is still hard to give in a precise manner. This may alienate non-experts. A proper classification could help to overcome this problem and reach mass adoption.

As already mentioned, tokens are very versatile objects: technically, they can represent any asset of the existing economy and create new ones. The possibilities that arises with cryptographic tokens are huge and more have yet to be discovered. Thus, in the so called field of "Tokenomics"[8] [25], tokens classification is one of the main research

[8] "Tokenomics" is the study of tokens usage, utility, value and coordination with the broader ecosystem.

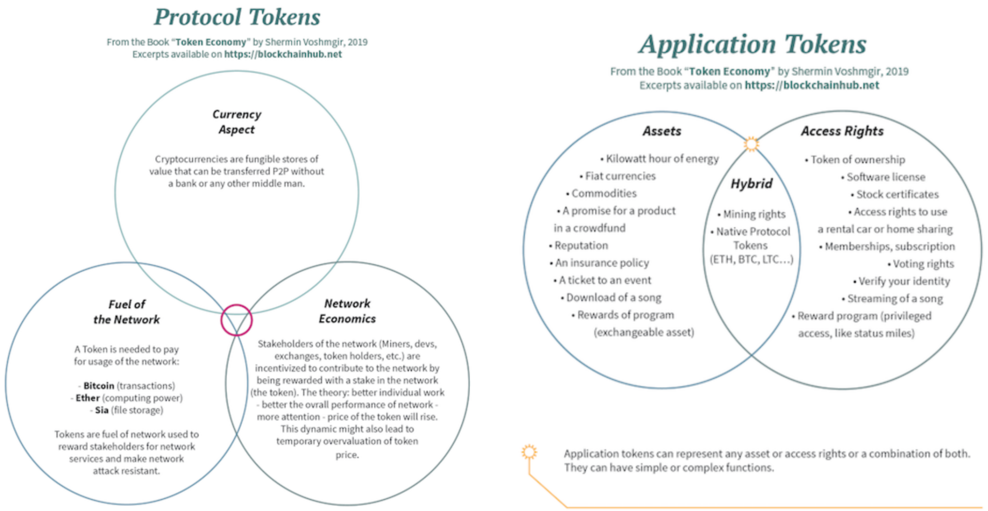

Protocol Tokens
From the Book "Token Economy" by Shermin Voshmgir, 2019
Excerpts available on https://blockchainhub.net
Currency Aspect
Cryptocurrencies are fungible stores of value that can be transferred P2P without a bank or any other middle man.
Fuel of the Network
A Token is needed to pay for usage of the network:
- Bitcoin (transactions)
- Ether (computing power)
- Sia (file storage)
Tokens are fuel of network used to reward stakeholders for network services and make network attack resistant.
Network Economics
Stakeholders of the network (Miners, devs, exchanges, token holders, etc.) are incentivized to contribute to the network by being rewarded with a stake in the network (the token). The theory: better individual work - better the ovrall performance of network - more attention - price of the token will rise. This dynamic might also lead to temporary overvaluation of token price.
Application Tokens
From the Book "Token Economy" by Shermin Voshmgir, 2019
Excerpts available on https://blockchainhub.net
Assets
• Kilowatt hour of energy
• Fiat currencies
• Commodities
• A promise for a product in a crowdfund
• Reputation
• An insurance policy
• A ticket to an event
• Download of a song
• Rewards of program (exchangeable asset)
Hybrid
• Mining rights
• Native Protocol Tokens (ETH, BTC, LTC...)
Access Rights
• Token of ownership
• Software license
• Stock certificates
• Access rights to use a rental car or home sharing
• Memberships, subscription
• Voting rights
• Verify your identity
• Streaming of a song
• Reward program (privileged access, like status miles)
Application tokens can represent any asset or access rights or a combination of both. They can have simple or complex functions.

Figure 1.8. Protocol tokens vs Application tokens [38].

topics.

Some of the most authoritative researches on tokens classification are based on empirical data studies, which are usually consulted to assess theoretical assumptions. This approach, which is the only possible one, will therefore make the topic of classification always central because the field is growing and for sure new ideas and concepts are coming.

The exploration of different types and roles of tokens is just at an early stage: everyday new studies on the topic are released. In this section some of these studies, quoted in the bibliography, are reported, linked together, and commented on. The focus is on the characteristics considered essential and to be verified first to properly describe a token.

The starting point is a very general classification, which immediately follows. Later, this first classification is expanded: new points of view are considered and finally new categories, that make it much more inclusive, are introduced.

A first possible characterization, already brought at the beginning of this chapter, is shown in Figure 1.8 from [38].

It is a quite technical classification, which depends strictly on the role of the token in the consensus protocol and on "where" it is defined.

Protocol tokens, also called *intrinsic*, *native* or *built-in* tokens, have a fundamental role in a public permissionless network, because they are directly involved in the functioning of the network itself. Indeed, protocol tokens keep it safe from attack, acting as block validation incentive, are involved in transaction spam prevention and may be needed to pay transaction fees in the network.

Conversely, *application tokens* can have any function or feature, indeed, they can represent anything from a physical or digital asset, or a right to execute an action in or outside the network. They are not involved in the functioning of the blockchain itself and for this reason they are said to be defined "on the top" of the blockchain. An example of both kind of tokens is easily found on the Ethereum network, where Ether (ETH) is the one and only protocol token, while there is a whole economy

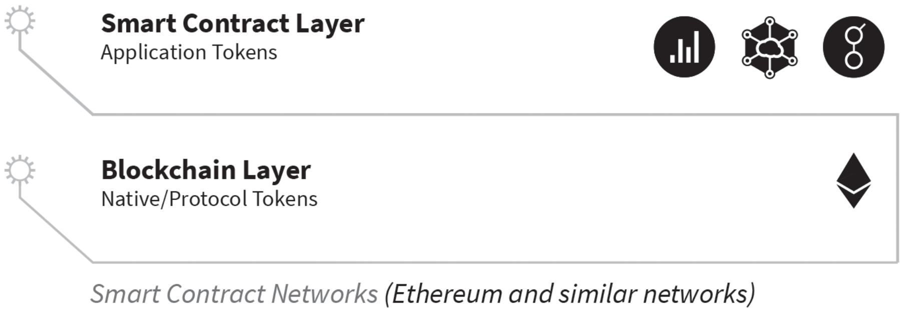


Figure 1.9. Layers of the blockchain [38].

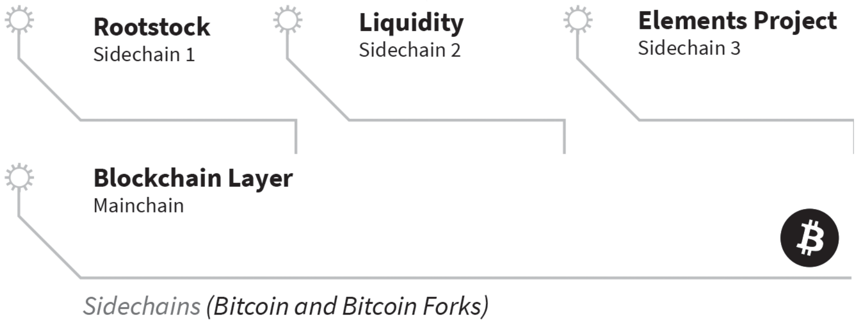


Figure 1.10. Some sidechains on the Bitcoin blockchain [38].

of application tokens running on the top of the network (ERC-20, ERC-721 and other standards). Bitcoin itself is a protocol token. Indeed, without it, the Bitcoin blockchain could not work. Bitcoin and Ethereum blockchains, and the role of their native tokens, are further explored in sections 1.5 and 1.6.
Protocol tokens and application tokens exist, in general, in different so-called layers of the blockchain, as shown in Figure 1.9 from [38]. The Layer-1 represents the underlying blockchain architecture, where the consensus protocol takes action. Instead, Layer-2 is the overlying network located at the top of the underlying blockchain, where application tokens may be defined through smart contracts. In [39] some pros and cons of the two blockchain layers are presented together with some examples .

The previous distinction between protocol and application tokens based on the different layers on which they are defined is not always valid, because there are some blockchains where the concept of layer is drastically changed. New solutions based on a different network design have been released.
Now, some examples are reported. A complete explanation of these concepts is not in the scopes of this thesis but, anyway, these new solutions have to be introduced, to reinforce the idea that the difference between protocol and application tokens is not always applicable in the way explained before. For those interested in learning more about the topics, in-depth papers are cited and related websites are linked.
The first example are *sidechains tokens*, that cannot be included in the distinction made earlier. Sidechains, [2] and [35], are separate blockchains, connected to the

primary one, the mainchain, with a two-way-peg [9]. Sidechains could greatly increase the possibilities offered to users by the mainchain, because they have their own protocol and implementation, which can be completely different from the one of the primary blockchain. A whole new token system can be created on sidechains (sidechain tokens) for example with smart contracts, even if they were not originally conceived on the mainchain. The two-way-peg mechanism allows users to obtain tokens created on the sidechain starting from tokens they already own on the main blockchain.

Clearly, the distinction between protocol and application tokens given before hardly fits in this context, because a sidechain is a different solution from a Layer-2.

Figure 1.10 from [38] shows a schematic view of some real sidechains created on the Bitcoin blockchain.

The second and last example concerns solutions where multiple tokens can be created in the native level, i.e. in the Layer-1. The blockchains offering these possibilities are called multi-asset ledgers. In this case not all tokens are involved in the functioning of the blockchain, being directly involved in the protocol, but anyway are defined on the Layer-1. The definition of protocol and application token given before does not cover this case. Ripple[10] and Stellar[11] are some real applications of this solution.

But then, if the definition of protocol and application tokens does not cover several examples, why was it made?

- Because it is a good starting point to understand tokens, their basic functionalities and their potential.
- Because it helps in understanding the basic functioning of the blockchain.
- Because it is still applicable to several blockchains, including Bitcoin and Ethereum.
- Because, anyway, it is a very general definition which could be easily modified to cover particular cases like the ones mentioned.

The previous classification, mostly technical, can be enriched adding another point of view useful to distinguish different tokens: the rights a token could represent.

Indeed, as pointed out also when tokenization was defined, a token can, for instance, represent a right to some underlying economic value. In general, the rights provided by possession of a token can be varied:

1. right to an asset. A token can be pegged to an asset (asset tokenization), thereby acquiring its value and all the rights attached to its possession. A token pegged to an off-chain asset is called *asset-backed*.
   This kind of right can be further extended to the case of tokens which, like digital shares, provide profits or dividends (for example by representing the possession of a real share). A token falling in the latter case is called *security/equity token*. In some cases, for instance in [30], these kind of tokens are called *tokenized securities*;

[9] A two-way peg is a device used to perform bidirectional transfer of tokens between the mainchain and the sidechain at a fixed or pre-deterministic exchange rate.

[10] https://ripple.com

[11] https://www.stellar.org

2. right to do a work. A *work token* is a token, usually native, which gives the holder the right to do work, and, in return, receive a portion of the fees associated with the work done. Work tokens are considered a possible breakthrough for the role of workers in a network [22].
   Sometimes, the term work token is used also for tokens which are just rewards for users who complete a certain action or show a certain behaviour;

3. right to use a good or service that others possess. It refers to the legal right to use something in someone else ownership. These tokens are sometimes called *usage tokens*;

4. voting right. Possessing a token may give a right to vote for something. The vote may regard the blockchain itself or something external. These tokens are called *voting tokens*;

5. multiple rights at once. The previously listed rights do no exist in isolation: rather, a single token may grant many of them. In this case it is called *hybrid token.*

Just considering the first two points of view introduced, the technical one and the one related to the rights derived from the possession of a token, the possibilities are many. However, there are still some fundamental aspects to consider. One of them, for example, is the possibility of interaction between the users participating in a blockchain and owning tokens. In fact, in a network involving connected users, such as a blockchain, one must consider the possibility of token exchanges. Most of the examples previously introduced, such as those related to fractional ownership of assets, suggest that this possibility is always guaranteed. In reality this is not always true. In fact, let's imagine a token that is tied to the identity of an individual. This, representing an identity, will certainly be unique and non-exchangeable. This concept can be extended to many cases and creates the need for another essential characteristic to distinguish tokens: fungibility.
Fungible tokens have two main properties [38]: 1) only quantity matters, meaning that units of a fungible assets of the same kind are indistinguishable; 2) any amount can be merged or split into a larger or smaller amount of it and still be indistinguishable from the others. Being identical, fungible tokens are interchangeable with other tokens of the same kind. This means, for example, that if 10 fungible tokens are exchanged for 10 other tokens of the same type, the whole operation is equivalent to keeping the initial tokens.
The exchange may also involve only a fraction because of the divisibility property.
Fungible tokens can be compared to commodity money and are therefore referred to as crypto-commodities or digital coins (fungibility is an important property of any currency or commodity). They have characteristics that open doors to new kind of markets. Indeed, they can represent new or existing assets or even derivatives.
An example of a new (digital) asset which is a fungible token is Bitcoin, that satisfies all the above properties. In general, all the *cryptocurrencies* belong to this category. Indeed, the term cryptocurrency refers to fungible tokens which act as a store of value and unit of account.
In the physical world several assets like art, real estate or luxury goods can not be

split up and traded at global scale. The lack of these properties makes them less easily exchangeable. Now, at low cost, potentially they can be tokenized in a fungible token, allowing fractional ownership (with, obviously, some limitations to prevent possible problems, for instance dusting attacks [12]) and creating completely new, large-scale business models based on these products.
Regarding derivatives, put and call options can be defined as tokens that grant rights to a collateral deposited in smart contract vaults. This mechanism ensures, without the need of third parties, that the option can only be exercised if specific parameters, dictated by the smart contract, are met [28].
Non-fungible tokens (NFTs), instead, are unique, in the sense that each one differs from the others of the same type for its exclusive attributes. For this reason they are not interchangeable, representing unique value or access rights. Moreover, they are non-divisible. Sometimes they are referred to as crypto-goods. An example of non-fungible token could be a so-called identity token: this is bind to anything that uniquely represents a person, like the ID. Such a token should inevitably be non-fungible. In recent months other applications of non-fungible tokens have been gaining in popularity. These are related to the creation of unique digital items, usually referred to as crypto-art and crypto-collectibles. Artists can create unique digital artworks by tokenizing them, i.e. linking them to a non-fungible token on a blockchain. While everybody can look at the art work, the blockchain guarantees both that it is owned by the one who purchased it and its originality through the artists signature. A new frontier of digital art based on non-fungible tokens is emerging. There are numerous platforms that allow artists to first tokenize their digital works and then to sell them. Some examples are SuperRare[13] and OpenSea[14], marketplaces to buy, sell and collect digital artworks. Other platforms, like Async Art[15], allow the creation of works that are changeable and customisable over time. This new emerging world can represent a new frontier for artists creativity, for mainstream access to art and for the art market.
In Table 1.1 the main characteristics of fungible and non-fungible tokens are summarized.
On the Ethereum blockchain fungible and non-fungible tokens are mainly represented by two different standards, that are respectively the ERC-20 and the ERC-721. In section 1.6 these two standards are further described and several examples are added.

Given the potentially infinite use cases, it is possible to imagine many other viewpoints from which to look at tokens. Some of them, in addition to those already listed, are discussed in some previous works, such as [28] and [38]. Now, these points of view are reported and commented on, and new ones are added.

- **Purpose** viewpoint: it refers to the main purpose behind the creation of a token on a blockchain. This perspective is one of the first things to be generally

[12] A dusting attack is an attack in which a tiny amount of cryptocurrency, called dust, is sent to a large number of wallet addresses to discover the identity of the owners of the addresses, to spread messages and to spam a network.

[13] https://superrare.co

[14] https://opensea.io

[15] https://async.art

Fungible tokens	**Non-fungible tokens**
Identical	Unique
Interchangeable	Non-interchangeable
Divisible	Non-divisible
Examples: Bitcoin, Ethereum	Examples: crypto-art, identity tokens

Table 1.1. Fungible vs Non-fungible tokens.

explored when designing a new token or studying an existing one. Blockchain technology uses tokens mainly to exchange value, exchange verifiable data, and achieve coordination among the users by encouraging a new form of collective value creation, in a situation where there is partial or zero trust between the users. The possible uses could also include exploiting the tokens to encourage users in interacting with the platform in a determined way, for instance by entering it, using it for a certain period or leaving it. Thus, in some cases, the purpose of the token can be related to the governance of the network.

- **Privacy** viewpoint: the privacy of token holders is usually evaluated carefully because it could compromise the functioning of the whole system.
  In general, users of a blockchain are identified through their addresses, which are alphanumeric strings generated using asymmetric cryptography. An address does not give any information on the physical entity which created it. Moreover, each physical entity could potentially create any number of addresses: the blockchain results in a partial/zero trust environment where users do not know each other, but they can still trustfully send and receive tokens.
  In the common financial system, professionals that provide any kind of financial service are subject to the KYC[16] requirements: such policies are not mandatory when a blockchain address is created.
  For all these reasons, in common belief, anonymity is an ever-present feature of any blockchain. This is not true. Indeed, the high transparency of blockchain networks could make users non-anonymous. In several cases, for instance in Bitcoin, it is more correct to talk of pseudo-anonymity. Indeed, the public transaction data recorded on the blockchain can be studied with some techniques that, going backward into transactions and connecting them, could find the real owner of the token. Tokens for which the privacy of holders is compromised may incur undermined fungibility.
  New blockchains, such as the Monero[17] one, are developing techniques for a better privacy.
  From the regulators point of view the problem of privacy is not easy to deal with and generates many challenges that could slow down the development of a clear regulatory framework.

[16] Know your costumer (KYC) refers to a set of policies that affect the relationship between a service provider and its customers. The provider must verify the identity of the costumers to identify possible risks or illegal intentions. KYC fits in the broader anti-money laundering policies.

[17] https://www.getmonero.org

- **Regulatory/legal** viewpoint: regulation, as mentioned in the dedicated content (1.3), is a very complex topic. Belonging to a different legal category could completely change the possibilities that a token can offer, so this perspective should be deeply studied in the token design phase.

- **Supply** viewpoint: it concerns the supply of a token, in terms of its maximum available quantity and method of distribution. The token supply management model is implemented in the protocol (for protocol tokens) or in the smart contract that manages the token itself.
  Depending on the token type, this model can be [28]:

  - developed as a new independent system. This is, for instance, the case of native tokens involved in the protocol of the network, or of the so-called *freestanding tokens*. The latter represent unique assets, which have no prior meaning, are not linked to anything off-chain, and, after being issued, exist independently of the issuer;

  - based on existing data collected on-chain, or off-chain via oracles[18].
    For instance, a token which gives access to a service, i.e. an usage token, should have a supply depending on the maximum users capacity that the service provider could handle. Or, the supply of asset-backed tokens is naturally linked to that of the assets they represent.
    In both previous examples, there are data that need to be collected off-chain and then transferred to the blockchain in order to manage the supply of tokens.

  In general, initially the supply is empty or pre-mined, i.e. tokens are distributed to certain accounts at launch. In both cases, if the supply is not fixed at the initial value, there is a subsequent mint operation[19]. It can be just scheduled in some simple way, for instance by time, or one-off, i.e. new tokens are created and distributed to accounts according to rules established in the blockchain protocol or in the token smart contract.
  The supply greatly influences the value of a token. For example, Bitcoin supply is managed by the protocol and is limited to 21 million units. This generates a so-called scarcity[20], which greatly affects Bitcoin price.
  In general, the supply management of a token is strictly linked to its “life cycle”, presented in the next point.

- **Life cycle** viewpoint: it refers to the set of stages a token goes through after it is issued, up to its eventual removal.
  The life cycle of a token can generally be of three types:

  - straight: the token is created, assigned to a beneficiary, consumed and finally destroyed. In this case, it is common to say that the token is “burnt”

[18] A blockchain oracle is a device/service that provides the blockchain with off-chain data. In [10] an overview of this topic is presented.

[19] “A mint operation creates and distributes new token units or objects” [28].

[20] The term scarcity is used in economics to describe the situation of a good which, because it exists in limited quantity, may have a supply that does not cover demand.

  after use. There is also the possibility that it may expire after a certain period of time. An example is a token representing a voucher which, after being assigned, can only be used once before its expiry date;

  - circular: the token is used in circuits, as is the case with currencies;
  - chip: it is the case of tokens provided with an exchange service with other tokens or fiat currencies, and vice versa.

  However, in some cases the life cycle of a token includes features from several of the above categories.
  The correct design of the life cycle could be used in order to include a desired inflationary/deflationary behaviour.
  Another possible viewpoint introduced in some works is the so-called *spendability*. A token can be spendable in the network to obtain something. A common example of spendable tokens can be found in several video games, where it is possible to convert fiat money into an in-game currency to buy, for example, power ups. However, this perspective is not presented on its own because it can be derived by studying the life cycle of the token.

- **Transferability/tradability** viewpoint: it refers to the possibility to transfer the ownership of a token to one user to another, for instance by selling it. In most cases fungible tokens tend to be transferable, while for non-fungible tokens this is not always possible.

- **Durability** viewpoint: it refers to the ability of a token to perform its function correctly for its entire design lifetime. If the token is native, its durability is closely connected to the resilience of the network. Indeed, as already mentioned, permissionless blockchains work properly as long as they use the native token as a reward for users who complete specific actions and show a desired behaviour. Protocols are implemented in such a way as to make correct behaviour more profitable than incorrect behaviour through an incentive based mechanism.
  Two strictly related key factors for a native token durability are its value and its supply. Indeed, the strength of the incentive is of course proportional to the value that the token has for users and to the ease with which it is obtained.

- **Price stability** viewpoint: short-term price stability is a property that should be carefully evaluated in the design phase of specific tokens (for instance for those intended as a medium of exchange). Indeed, in some cases their price is a fundamental aspect for their durability. Moreover, as mentioned at the end of section 1.2, keeping a stable price could avoid entangled tokens. Of course, there may be also cases in which the token is intended for speculative purposes and the price stability is not desirable.
  This viewpoint is inserted last because price stability is influenced by all the characteristics of a token, and in turn strongly influences them.
  Older tokens, like Bitcoin, do not present characteristics that aim at price stability. The protocol alone, as it is designed in Bitcoin, is not sufficient to guarantee it, especially at times when more and more investors are entering the market.

However, there are more sophisticated protocol mechanisms or smart contracts that dynamically change the token supply according to some rules and that may achieve the stability. A possible way is by using bonding curves. A *bonding curve* is a function, usually increasing, which relates the token price with its supply. In this way the price is deterministic and can be managed: to increase it it is sufficient to mint new tokens, while to decrease it some tokens must be burned. However, the implementation of a proper function is very complex because it should take into count also all the other parameters that control the token. A deep discussion on the design parameters of bonding curves is not in the scopes of this thesis. For those interested, in [20] the author makes further considerations on this topic.

Finally, there exist fungible tokens whose price is stable, being backed by an underlying asset (for example a fiat currency). These tokens are called *stablecoins*. Examples of stablecoins are Tether[21] and Paxos Standard[22].

The design of this kind of token is open to several possibilities, but the most common and general cases are the two defined below. The following definitions are given in [28]. However, for a deeper analysis, [24] is recommended.

1. "*Redeemable* or *convertible* stablecoins are backed by a collateral or reserve and grant redemption value for the underlying assets".
2. "*Synthetic* stablecoins are designed to track the price of one or more underlying assets but do not grant any redemption value for these assets".

At this point, the main aspects that characterize a token have been introduced. It is interesting to observe how these strongly influence each other: a single wrong choice in the design phase can compromise the whole use of the token. This is why insisting on classification and trying to make it clearer and more complete is a task that is fundamental to token development.

There exist several works which try to provide a comprehensive and useful classification tool for tokens. Now two of these, [13] and [30], which are considered among the most comprehensive around, are described, with a particular focus on the second one.

In [13] the author identifies "multiple angles from which you can look at tokens", namely five dimensions to characterize a token. The proposed scheme encapsulates many of the features described before. This classification tool is represented and briefly explained in Figure 1.11 from [13].

In [30] the latter classification is expanded, making it more comprehensive. The proposed classification tool is represented in Figure 1.12. It is obtained starting from the study of several real projects, all very different. The red-marked parameters represent the four main dimensions to classify a token:

- purpose parameters refer, as previously mentioned, to the aim of the token in the network. In the proposed tool, the purpose of the token is influenced by its class, which includes the three categories mostly used in the legal classification

[21] https://tether.to

[22] https://www.paxos.com/pax/

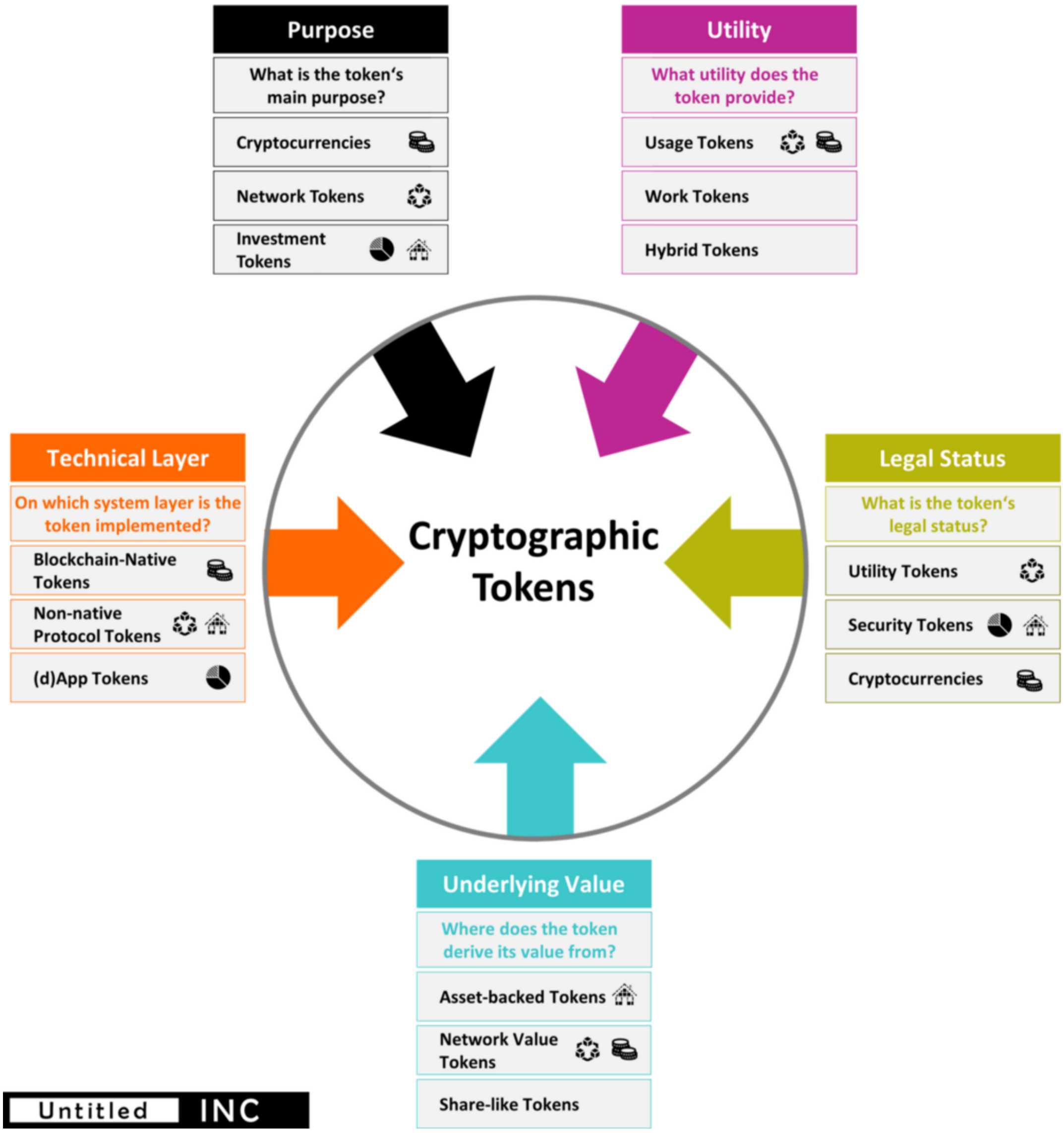
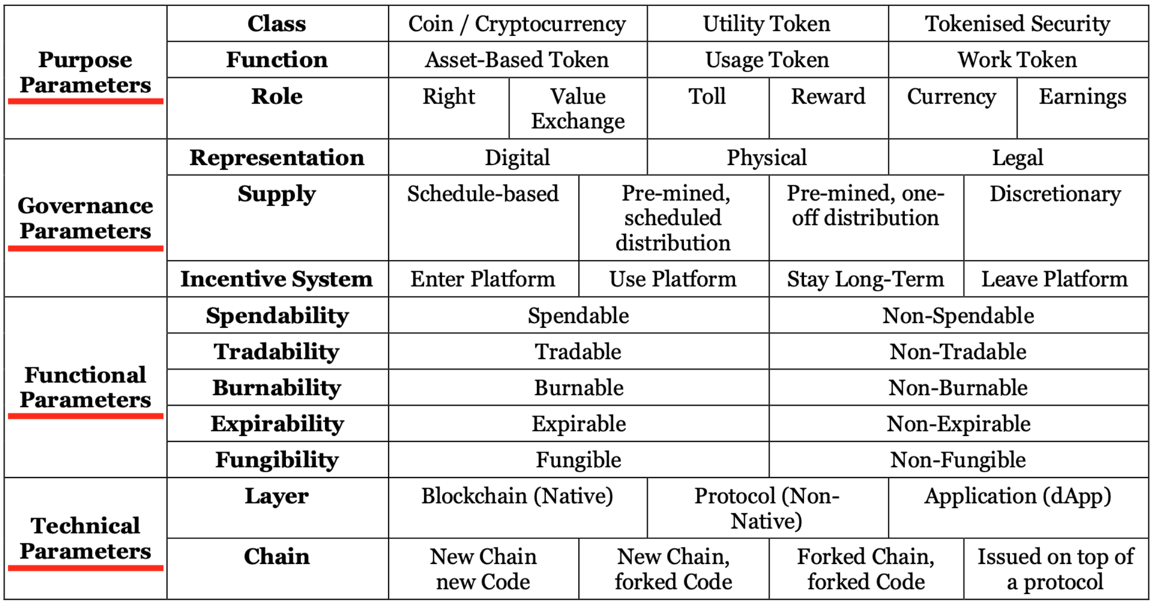


Figure 1.11. Classification scheme for tokens based on five dimensions [13].



<table>
<tr><td rowspan="3"><b>Purpose Parameters</b></td><td><b>Class</b></td><td colspan="3">Coin / Cryptocurrency</td><td colspan="2">Utility Token</td><td colspan="3">Tokenised Security</td></tr>
<tr><td><b>Function</b></td><td colspan="3">Asset-Based Token</td><td colspan="2">Usage Token</td><td colspan="3">Work Token</td></tr>
<tr><td><b>Role</b></td><td>Right</td><td colspan="2">Value Exchange</td><td>Toll</td><td>Reward</td><td colspan="2">Currency</td><td>Earnings</td></tr>
<tr><td rowspan="3"><b>Governance Parameters</b></td><td><b>Representation</b></td><td colspan="3">Digital</td><td colspan="2">Physical</td><td colspan="3">Legal</td></tr>
<tr><td><b>Supply</b></td><td colspan="2">Schedule-based</td><td colspan="2">Pre-mined, scheduled distribution</td><td colspan="2">Pre-mined, one-off distribution</td><td colspan="2">Discretionary</td></tr>
<tr><td><b>Incentive System</b></td><td colspan="2">Enter Platform</td><td colspan="2">Use Platform</td><td colspan="2">Stay Long-Term</td><td colspan="2">Leave Platform</td></tr>
<tr><td rowspan="5"><b>Functional Parameters</b></td><td><b>Spendability</b></td><td colspan="4">Spendable</td><td colspan="4">Non-Spendable</td></tr>
<tr><td><b>Tradability</b></td><td colspan="4">Tradable</td><td colspan="4">Non-Tradable</td></tr>
<tr><td><b>Burnability</b></td><td colspan="4">Burnable</td><td colspan="4">Non-Burnable</td></tr>
<tr><td><b>Expirability</b></td><td colspan="4">Expirable</td><td colspan="4">Non-Expirable</td></tr>
<tr><td><b>Fungibility</b></td><td colspan="4">Fungible</td><td colspan="4">Non-Fungible</td></tr>
<tr><td rowspan="2"><b>Technical Parameters</b></td><td><b>Layer</b></td><td colspan="3">Blockchain (Native)</td><td colspan="2">Protocol (Non-Native)</td><td colspan="3">Application (dApp)</td></tr>
<tr><td><b>Chain</b></td><td colspan="2">New Chain new Code</td><td colspan="2">New Chain, forked Code</td><td colspan="2">Forked Chain, forked Code</td><td colspan="2">Issued on top of a protocol</td></tr>
</table>

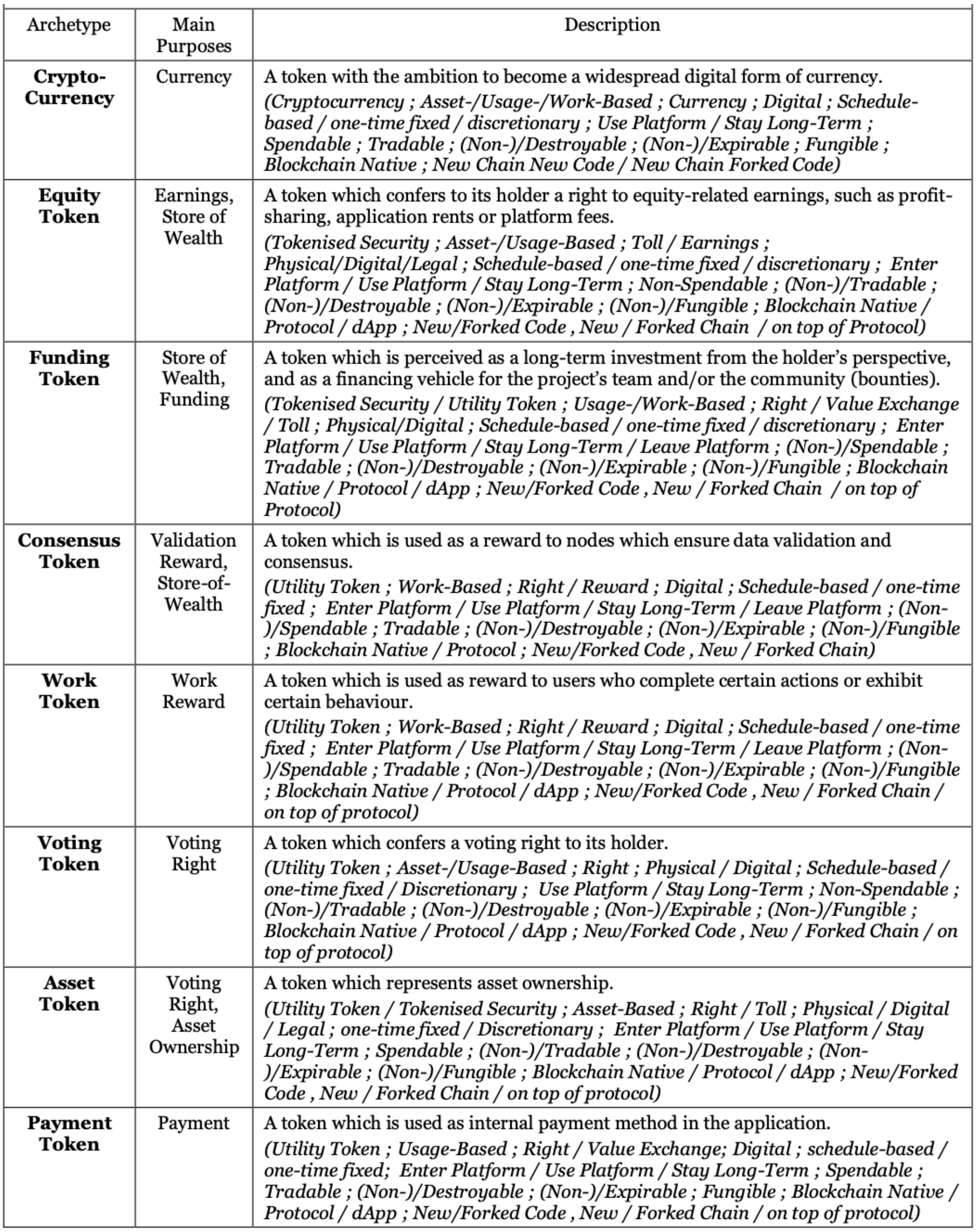
Figure 1.12. Extended tokens classification [30].

framework, by its function, i.e. the reason why the token exists and finally by its precise role, which makes the characterization even more specific;

- governance parameters refer to characteristics of the token that are related to the way the platform is governed and managed;
- functional parameters refer to the token characteristics that may alter its ownership or its existence;
- technical parameters refer to the the technical attributes of the token, in terms of layer (i.e. of its role in the blockchain functioning) and chain on which it is based on. It should be pointed out that the notation of this classification makes a distinction between native token (what was previously called native or protocol token) and protocol token, which could be, for instance, a token defined on the Layer-1, but not critical to the basic functioning of the blockchain.

Each of these classes is further divided in related subcategories, which cover most of the characteristics of a token previously introduced. Being based on data of real projects, the authors of [30] show that this classification tool permits to capture most of the patterns of existing tokens. Moreover, the parameters that characterize a token are linked together and schematised in such a way that Figure 1.12 can be an useful tool also to design new tokens.
Some archetypes, i.e. tokens with specific characteristics that are frequently observed in existing real life applications, are reported and defined in Figure 1.13 from [30].

To conclude this section on token classification, it is crucial to repeat that cryptographic tokens do not exist in an independent manner, but are part of the wider context of the whole blockchain system. Indeed, as explained in [13], the whole system is made of three layers: that of tokens, that of governance and that of technology. In some sense,



Archetype	Main Purposes	Description
Crypto-Currency	Currency	A token with the ambition to become a widespread digital form of currency. *(Cryptocurrency ; Asset-/Usage-/Work-Based ; Currency ; Digital ; Schedule-based / one-time fixed / discretionary ; Use Platform / Stay Long-Term ; Spendable ; Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; Fungible ; Blockchain Native ; New Chain New Code / New Chain Forked Code)*
Equity Token	Earnings, Store of Wealth	A token which confers to its holder a right to equity-related earnings, such as profit-sharing, application rents or platform fees. *(Tokenised Security ; Asset-/Usage-Based ; Toll / Earnings ; Physical/Digital/Legal ; Schedule-based / one-time fixed / discretionary ; Enter Platform / Use Platform / Stay Long-Term ; Non-Spendable ; (Non-)/Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of Protocol)*
Funding Token	Store of Wealth, Funding	A token which is perceived as a long-term investment from the holder's perspective, and as a financing vehicle for the project's team and/or the community (bounties). *(Tokenised Security / Utility Token ; Usage-/Work-Based ; Right / Value Exchange / Toll ; Physical/Digital ; Schedule-based / one-time fixed / discretionary ; Enter Platform / Use Platform / Stay Long-Term / Leave Platform ; (Non-)/Spendable ; Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of Protocol)*
Consensus Token	Validation Reward, Store-of-Wealth	A token which is used as a reward to nodes which ensure data validation and consensus. *(Utility Token ; Work-Based ; Right / Reward ; Digital ; Schedule-based / one-time fixed ; Enter Platform / Use Platform / Stay Long-Term / Leave Platform ; (Non-)/Spendable ; Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol ; New/Forked Code , New / Forked Chain)*
Work Token	Work Reward	A token which is used as reward to users who complete certain actions or exhibit certain behaviour. *(Utility Token ; Work-Based ; Right / Reward ; Digital ; Schedule-based / one-time fixed ; Enter Platform / Use Platform / Stay Long-Term / Leave Platform ; (Non-)/Spendable ; Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of protocol)*
Voting Token	Voting Right	A token which confers a voting right to its holder. *(Utility Token ; Asset-/Usage-Based ; Right ; Physical / Digital ; Schedule-based / one-time fixed / Discretionary ; Use Platform / Stay Long-Term ; Non-Spendable ; (Non-)/Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of protocol)*
Asset Token	Voting Right, Asset Ownership	A token which represents asset ownership. *(Utility Token / Tokenised Security ; Asset-Based ; Right / Toll ; Physical / Digital / Legal ; one-time fixed / Discretionary ; Enter Platform / Use Platform / Stay Long-Term ; Spendable ; (Non-)/Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; (Non-)/Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of protocol)*
Payment Token	Payment	A token which is used as internal payment method in the application. *(Utility Token ; Usage-Based ; Right / Value Exchange; Digital ; schedule-based / one-time fixed; Enter Platform / Use Platform / Stay Long-Term ; Spendable ; Tradable ; (Non-)/Destroyable ; (Non-)/Expirable ; Fungible ; Blockchain Native / Protocol / dApp ; New/Forked Code , New / Forked Chain / on top of protocol)*

Figure 1.13. Tokens archetypes [30].

token layer is in between these other two and connects them, as shown in Figure 1.14 from [13]. This aspect must be taken into account in the design of a token.

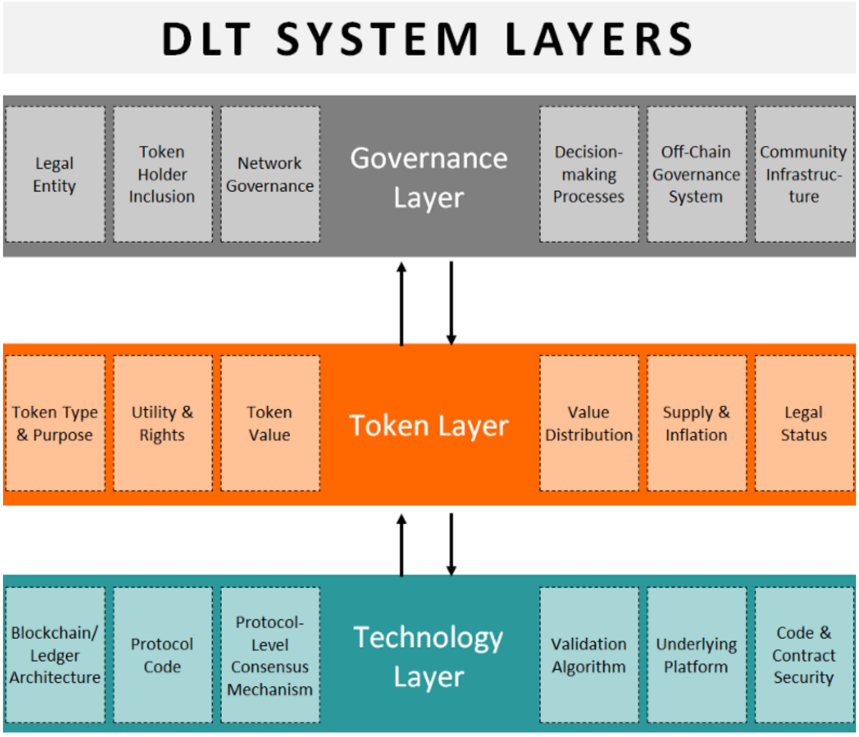


Figure 1.14. The whole DLT system as the connection of three layers [13].

1.5 Bitcoin

Bitcoin (BTC) is both a protocol and a cryptocurrency created in 2008 by an unknown person or group of people using the name Satoshi Nakamoto and started in 2009, when its implementation was released as open-source software [27].
The Bitcoin blockchain has a native token, Bitcoin, which is a decentralized digital currency that can be sent between users through P2P exchanges, without the need for intermediaries. There are no central banks or administrators involved, indeed transactions are verified by network nodes through cryptography and recorded in the blockchain. Bitcoins are created with a process called mining, which is connected to the reward mechanism [40].
Moreover, it is also possible to create tokens on the top of the Bitcoin blockchain, through the Bitcoin scripting language. This possibility offers some use cases, for instance to manage real world assets on the blockchain with simple rules. An example of these tokens are Colored Coins[23].
However, the possibilities offered by the Bitcoin blockchain for derived tokens are quite limited. First, its script is simple and not Turing-complete. Moreover, it does not have a state machine functioning which provides the possibility of executing smart contracts. Thus, the infinite possibilities of token creation through smart contracts,

[23] https://en.bitcoin.it/wiki/Colored_Coins

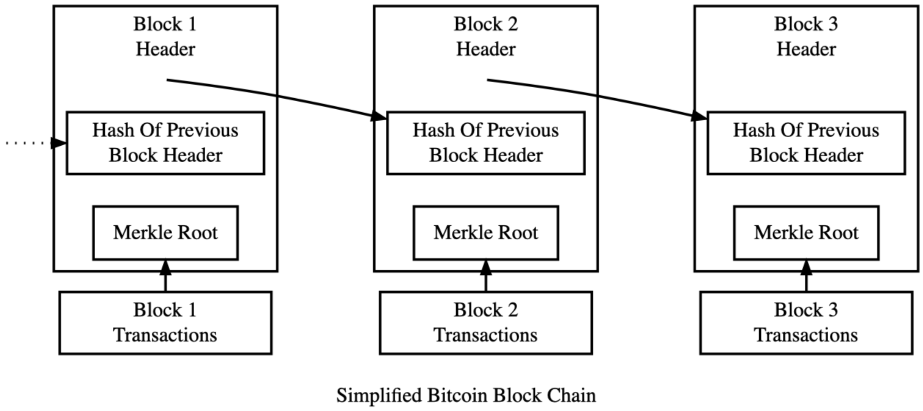


Figure 1.15. Bitcoin blockchain: a simplified view [4].

for instance provided by the Ethereum blockchain, cannot be replicated on the Bitcoin blockchain. In section 1.6 this topic is further expanded and various citations for further studies are provided.
In the following subsections the main features of BTC are described. The information presented is mostly derived from [4] and [27], where further details can be found.

1.5.1 Blockchain

The Bitcoin blockchain is a public[24], permissionless[25] and open-source ledger which records ordered and timestamped transactions. It is used to prevent double spending and the alteration of previous transaction records, without a third party acting as an intermediary in transactions. Through a consensus mechanism, (called Proof of Work and described later) a coordination between the nodes who have downloaded the ledger is reached. Several nodes are *in consensus* when they have the same blocks in their blockchains.
In Figure 1.15 from [4], a simplified version of a blockchain is shown. Each block is made of several components.
The Merkle root, is responsible of "recording" all the transactions included in that particular block. The Merkle root is obtained thorough a simple procedure. Given a list of all the transactions included in the block, each of them is hashed, the hashes are then paired and hashed, then these hashes are paired again and hashed again and so on until a single hash remains. The resulting single hash is the Merkle root stored in the block header. This process for four transactions is schematized in Figure 1.16. By this procedure, the Merkle root has a predefined length, given by the hash function itself, and its value depends on all the transactions inserted in the block. If only one of these transaction is changed then the Merkle root will change accordingly for the no-collision property of hash functions.

[24] Public means that anyone can use the blockchain just as a payment system, or download the protocol and the whole ledger and verify transactions (full nodes). There are no requirements to enter the system.

[25] Permissionless means that anyone can become a "miner", i.e. can download the protocol and the ledger and write transactions to it.

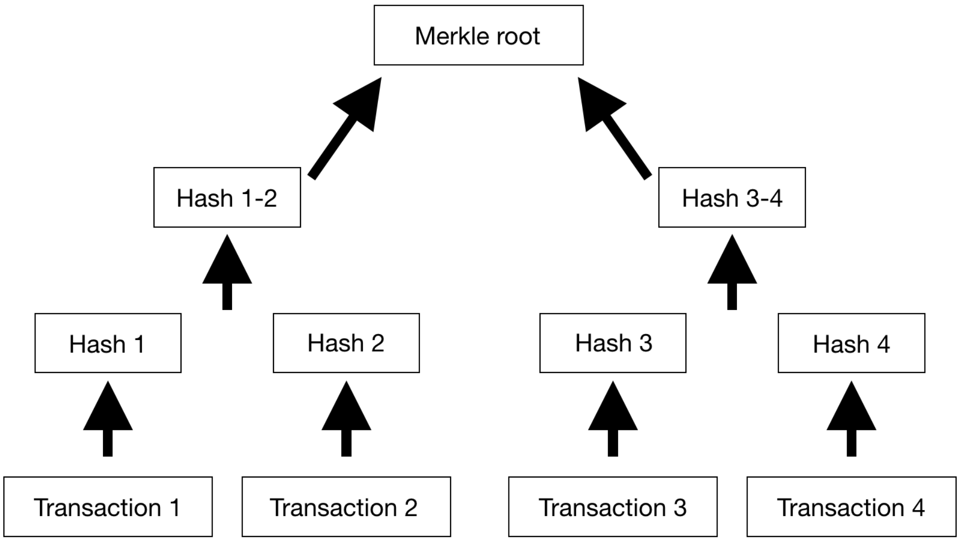


Figure 1.16. Schematized procedure for Merkle root computation.

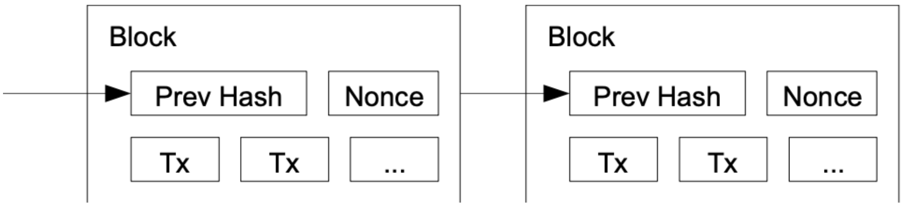


Figure 1.17. How blocks are connected in the Bitcoin blockchain [27].

Another fundamental quantity stored by a block is the hash of the previous block's header. This is the chaining mechanism between subsequent blocks, which ensures that even a single transaction in whatever block cannot be modified without modifying the block itself where it is recorded and all the following blocks. A schematic view of the way the blocks are connected is shown in Figure 1.17 from [27]. In [4] it is possible to find more details on the blockchain and what transaction exactly means in the Bitcoin network.
Other components are a timestamp, the difficulty target and the nonce. These are introduced subsequently with proof of work.

1.5.2 Proof of work

The Proof of Work (PoW) system is at the core of the Bitcoin blockchain functioning, because it solves the problem of selecting one node to create a new block in each round of the consensus, by a computational power competition. In Figure 1.18 from [43] the flow of PoW is shown. In general, the Proof of Work involves searching for a value whose hash begins with a specific number of zero bits. In Bitcoin, the hash is computed with the hash function SHA-256[26]. This mechanism allows to establish the difficulty target of the problem, indeed the average work needed is exponential in

[26] SHA-256 (Secure Hash Algorithm-256) is a cryptographic hash function.

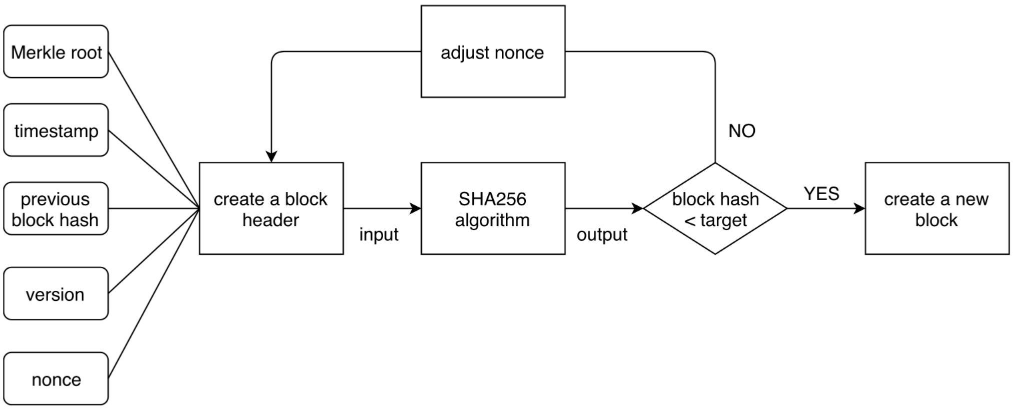


Figure 1.18. Schematic view of Proof of Work [43].

the number of zero bits required. After a proper value is found its correctness can be verified just by executing a single hash. For the characteristics of a proper hash function it is very rare that there are two different values that have the same hash. Moreover, the only possible way to find the required value is by trial and error.

The nonce contained in a block is used in the Bitcoin blockchain to implement the PoW. Indeed, miners increment its value until a nonce is found that gives the sought number of zero bits to the block hash.

Proof of work was intended as a "one-CPU-one-vote" method in a majority decision system for determining a new block: the first miner who solves the hash problem delivers the new block (the precise mechanism will be further described later on). A possible issue is when two or more (with decreasing probability) miners produce a block at roughly the same time. This produces a fork in the blockchain because some nodes may update the chain with the block from one miner and some others with the block from the other miner. In these cases the longest chain, where the greatest computational effort has been invested, represents the majority decision. The whole PoW mechanism keeps the blockchain tamper-proof because in order to change a block is necessary to redo all the work done to satisfy the proof of work. Moreover, as later blocks are chained after it, the work to change the block would include redoing all the blocks after it: to modify a block, an attacker would have to redo the proof-of-work of the block itself and of all blocks after until he surpasses the honest chain. If a majority of CPU power is controlled by honest nodes, the honest chain will grow the fastest and outpace any competing chains: after being added to the blockchain, a block is immutable. It has been shown that a scenario in which an attacker outpaces the honest chain is possible only if the attacker has more computational power than the rest of the network (51% attack). It is proved that the probability of an attacker with less computational power than the honest nodes of catching up diminishes exponentially as subsequent blocks are added [27]. The Proof of work system of Bitcoin is designed to target a certain deterministic number of blocks created over time (≈ 1 block each 10 minutes). The mechanism modifies the difficulty target if blocks are generated too fast. In this way hardware developments over the years will be managed.

However, despite being a disruptive method for a decentralized consensus mechanism, PoW protocol in the Bitcoin blockchain presents several issues.

- Decentralization. While decentralization was one of the main goals in creating Bitcoin, the develop of huge mining pools [27], which are very few and collect almost all the CPU power of the BTC network, frustrate this objective. This form of collaboration could create some points of centralization because the new blocks are always created by few entities which manage the pool.

- Public and permissionless blockchain. With the development of mining pools and the resulting increase in computational power required to create blocks, the idea of a public and permissionless blockchain could fail. Indeed, with a normal commercial hardware, it is almost impossible to mine Bitcoins since the computational effort required is too high. Being a miner is not open to anybody: a strong and expensive CPU is needed.

- Scalability. A blockchain system is scalable if it could handle an increasing amount of users, while adapting its performance. This property is crucial for a network to reach mass adoption. The Bitcoin system, if thought as a mean of payment, has several problems in this sense. Indeed, the blocks creation rate and the time for blocks finality are not fast enough to obtain a number of transactions per second comparable with other means of payment.
  However, to help improve scalability, new solutions, mostly off-chain, have been adopted. Among various sidechain solutions, a famous example is the lightning network [31].

- Anonymity and privacy. The problem of privacy was already addressed when discussing tokens in generality and the Bitcoin blockchain is a good example of it.
  The Bitcoin blockchain system is highly transparent: anybody can have a clear view of the transactions recorded on it. This quality is not always valuable, indeed, there are some drawbacks related to anonymity and privacy of users. Data analysis techniques could be used to study transactions and go back to the Bitcoins holder. The compromised privacy of the holder makes the token less fungible. To tackle this problem, new privacy preserving protocols have been developed.

- Sustainability. PoW system is not environmentally sustainable because the great amount of electricity that powers the CPUs is often derived from unsustainable sources, such as coal.

These problems have created the need to develop various alternative systems that are present in more modern blockchains than Bitcoin and that seek to remedy them.

1.5.3 Network

Using the previously introduced mechanisms, the network, schematically, operates in the following way.

[27] The term mining pool refers to the situation where individual miners collaborate by sharing their CPU to solve the hash problem in generating new blocks. A mining pool operator manages a full node and creates the blocks. The reward from the creation of a block is then divided according to the contribution.

1. New transactions to be included in the blockchain are spread to all nodes.
2. Each node selects some transactions and collects them into a block.
3. Each node works on the PoW for its block.
4. When the PoW is solved by a node, it sends the block to all the others.
5. Nodes check the block and accept it only if it contains all valid and not already spent transactions.
6. Nodes communicate that they have accepted the block by starting working on creating a next block, linked to the accepted one.

The longest chain, i.e. the one on which more computing power has been invested, is always considered by nodes to be the correct one, so they will be working on it. In the case of two different blocks sent simultaneously by two nodes, then some nodes may get one or the other first. In that case, they should work on the first one they receive, so two different chains start to grow. The draw is ended when one chain becomes longer than the other, in the sense that more computational effort has been put in it. At this point all nodes switch to the longer chain.

1.5.4 Validation Incentive

In each block added to the chain, by convention, the first transaction is a special one that produces a new coin for the creator of the block.
This mechanism is used as an incentive for miners to support the network. Moreover it is a way to distribute new coins into circulation without the need of a central entity issuing them.
The block validation incentive was introduced to reward the cost that is incurred by miners for the hardware and for the energy consumption and to encourage nodes to be honest. Indeed, for an attacker, even with more computational power than all the other nodes together, it is not convenient to defraud people, steal their coins and in this way destroy the network. For him is more profitable to play by the rules and collect more coins than everyone else combined than threatening the system and the value of its wealth.
Another method to fund the incentive for miners is the presence of transaction fees that are added to the block creation reward. The introduction of transaction fees solves another problem that the network will face. Indeed, the total number of Bitcoins that could be mined is of 21 Million. This makes the asset scarce, a strong analogy, among with others like mining, to gold. Once this number of coins have entered circulation, the incentive will transition entirely to transaction fees, the system will be completely inflation free and still working safely.

1.6 Ethereum

Ethereum is a decentralized open source blockchain conceptualized in 2013 by Vitalik Buterin [7] and Gavin Wood, featuring smart contract functionality.

In the origin of this system, there was the intent to deploy an alternative to Bitcoin for using in internet use cases, with the idea of having a system capable to exchange value over the Internet, for instance for art or for authorship. Ethereum went live in 2015, after a crowd sale for development funding which happened in 2014. In the crowd-sale 72 million coins were pre-mined. During the years, Ethereum has established itself as one of the most important blockchains, counting million of users and thousands of tokens created on it. Today, to unlock all the potential of the technology and to reach mass adoption, some key updates are being done. These are referred to as Eth2[28]. The main issues addressed by the Eth2 upgrades regard:

1. scalability;
2. security;
3. sustainability.

The Ethereum blockchain was originally designed with a Proof of Work system, similar to the BTC one, but Eth2 includes moving to Proof of Stake [29, 30]. This could help to improve the security, providing a stronger mechanism to encourage a correct behaviour of users, and the sustainability. Indeed, the PoW requires a huge quantity of energy for the computing power, often making it the subject of criticism for its strong environmental impact. In [43] the authors provide a deeper comparison of PoW, PoS and other consensus protocols.
At the same time, to solve the scalability problem, the updates will enable around 2000 transactions per second, while right now there are only about 15.
The contents that follow regard the main features of Ethereum and are fully derived from the Ethereum whitepaper [7] and the Ethereum development documentation on the official website[31], with the exception of the part regarding tokens. The description provided in this thesis is brief, but still sufficient to understand some key concepts behind this blockchain, that is far more complex than the Bitcoin one. In the cited documents more details can be found.

1.6.1 General Characteristics

Ethereum is designed with the purpose of creating an alternative protocol for building decentralized applications.
It relies on a blockchain which has, except for the basic idea behind it, several differences with the previously described BTC blockchain.
Indeed, Ethereum provides a built-in Turing-complete programming language which allows anyone to write smart contracts and decentralized applications with arbitrary rules for ownership, transactions and state transition functions.

[28] https://ethereum.org/en/eth2/

[29] Proof of Stake (PoS) is a consensus mechanism used in a blockchain to reach distributed consensus. The basic idea of PoS is that the creator of the next block is selected randomly with a probability which depends on the amount of native tokens he possesses.

[30] The web pages *https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/* and *https://consensys.net/blog/blockchain-explained/what-is-proof-of-stake/* provide numerous details on PoS and how it will be implemented in Ethereum.

[31] https://ethereum.org/en/

To permit these complex features, Ethereum state is conceptualized as a data structure which holds in addition to all accounts and balances, also a machine state, which can change from block to block according to a predefined set of rules, and which can execute arbitrary machine code. The rules behind the change of state are specified by the Ethereum virtual machine.
These features unlock several possibilities for smart contracts that the Bitcoin scripting could not offer. All the applications developed on Ethereum need to register their state in the distributed ledger through a transaction, therefore require fee to run.

1.6.2 Accounts

Ethereum blockchain, unlike Bitcoin, does not rely upon unspent transaction outputs (UTXOs) for balances, but on a state made of accounts. There are two main types of accounts: externally owned accounts, handled by key pairs (cryptocurrency wallet), and contract accounts, managed by their contract code.
To understand better the differences between this two kinds of accounts it is sufficient to introduce four fields contained in every kind of account:

1. nonce, a counter used to prevent double spending;
2. the Ether balance;
3. the account's contract code (not always present). The code is written in a low level Turing-complete language that enables to implement also very complex smart contracts;
4. the storage (empty by default).

Ether is the native token of Ethereum, used to pay transaction fees. An externally owned account has no code and it can send transactions. Instead, a contract account has a code and activates it every time it receives a message: this kind of account can, through the script, read and write to internal storage, send other messages or create other contracts. Contract accounts should be seen like "autonomous agents" in the Ethereum execution environment, which always execute a specific piece of code with a deterministic result and have direct control over their Ether balance and their own key/value store to record some variables.
Another difference between externally owned accounts and contract accounts is in the way they exchange value/information between each other: the first ones send *transactions*, while the second ones *messages*. In general, as in Bitcoin, in order to send a transaction or a message it is necessary to pay a fee to the miner who includes it in a block. In Ethereum, this fee is calculated with an unit of account called *gas*.
Now, transactions and messages can be defined.

- a transaction is a "signed data package that stores a message to be sent from an externally owned account" [7]. It contains:
  - some information to identify the receiver and the sender and the amount of Ether to be sent;

  - a *STARTGAS* value, which imposes the maximum number of computational steps the transaction execution can have in terms of gas. The more are the steps required, the larger is the amount of gas spent.
  - a *GASPRICE* value, which is the amount paid by the sender for each computational step;
  - a non-mandatory data field, which could contain additional data needed for a contract execution.

  The STARTGAS and GASPRICE fields are used to prevent a problem that derives from the Turing completeness of the scripting language. Indeed, there could be scripts with infinite loops or other computational wastage that eventually could slow down the entire network (like a denial of service). For this reason, each transaction is required to set a limit to how many computational steps of code execution it can use. In addition to the computational expense, the amount of gas required by an operation depends also on the amount of data that must be stored as part of the state: an attacker has to pay proportionately for every resource used. The fee mechanism based on gas is a way to discourage bad behaviour.

- A message is essentially like a transaction, produced by a contract when is called. With messages, contracts can have relationships with other contracts in exactly the same way that external actors can.

1.6.3 Ethereum state transition function

To better explain the Ethereum state transition function, it is easier to start moving back to Bitcoin. Technically, the ledger of Bitcoin can be thought of as a state transition system: the “state” is made of the ownership status of all Bitcoins and there is a “state transition function” that, given the current state and a transaction, gives a new state. A parallel can be drawn between this system and a standard banking system. In the latter, the state is a balance sheet and a transaction is a demand to move a value X from A to B. The state transition first checks everything is coherent with the transaction and then simply updates the balance sheets of A and B removing X from A’s account and adding the same amount to B’s account. Going back to Bitcoin, more precisely, the “state” is the collection of all coins (technically UTXO) that have been mined and not spent associated with each address. Each transaction contains:

- some inputs, each containing a reference to an UTXO and a cryptographic signature made with the private key of the address of the sender.
- some outputs, each containing a new UTXO, i.e the amount to be sent to a recipient which will generate for him an UTXO associated with its address.

The Ethereum state transition function is an extension of the one of Bitcoin because it includes several more possibilities mainly related to contract code execution. Before the update of the state, it should check not only that the quantity sent has not already been spent and is actually available, but also what concerns the gas. Indeed, the

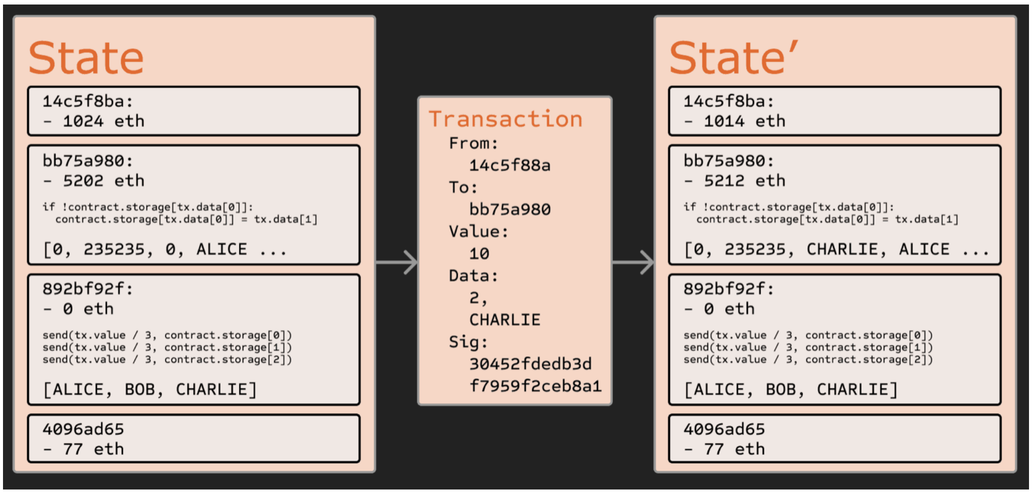

State
14c5f8ba:
- 1024 eth
bb75a980:
- 5202 eth
if !contract.storage[tx.data[0]]:
  contract.storage[tx.data[0]] = tx.data[1]
[0, 235235, 0, ALICE ...
892bf92f:
- 0 eth
send(tx.value / 3, contract.storage[0])
send(tx.value / 3, contract.storage[1])
send(tx.value / 3, contract.storage[2])
[ALICE, BOB, CHARLIE]
4096ad65
- 77 eth
Transaction
From:
14c5f88a
To:
bb75a980
Value:
10
Data:
2,
CHARLIE
Sig:
30452fdedb3d
f7959f2ceb8a1
State'
14c5f8ba:
- 1014 eth
bb75a980:
- 5212 eth
if !contract.storage[tx.data[0]]:
  contract.storage[tx.data[0]] = tx.data[1]
[0, 235235, CHARLIE, ALICE ...
892bf92f:
- 0 eth
send(tx.value / 3, contract.storage[0])
send(tx.value / 3, contract.storage[1])
send(tx.value / 3, contract.storage[2])
[ALICE, BOB, CHARLIE]
4096ad65
- 77 eth

Figure 1.19. Ethereum state transition function: update of the state after a transaction [7].

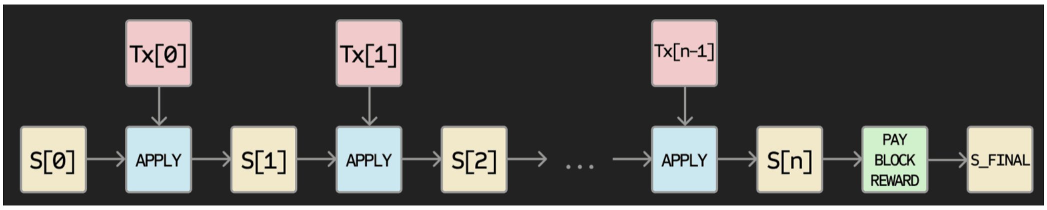


Figure 1.20. Update of the state according to n transactions in a new block [7].

costs regarding the gas should be checked and included: if the receiving account is a contract, its code is executed until completion or until execution runs out of gas. In the first case the fees for the remaining gas are sent back to the sender and the others are payed to the miner who included the transaction in the block. If the receiving account does not exist then it is created.

1.6.4 Ethereum Blockchain

To understand how to join the previous introduced characteristics it is necessary to describe briefly the blockchain and mining protocols.
Ethereum blockchain shares several characteristics with the one of Bitcoin but each block contains a copy of both the transactions list and the most recent state.
When a block with n transactions is added to the chain, the state must be modified accordingly: the state transition function takes transactions one at a time and changes the state. So, for n transactions there are as many changes of state. Figure 1.19 from [7] shows how the state is updated after a single transaction of 10 Ether from one account to another, while Figure 1.20 from [7] shows the succession of state updates following the addition of n transactions in a block.
Contract code execution is part of the validation algorithm, so if a transaction is added into a certain block, the code execution spawned by that transaction is executed by all nodes that download and validate the block.

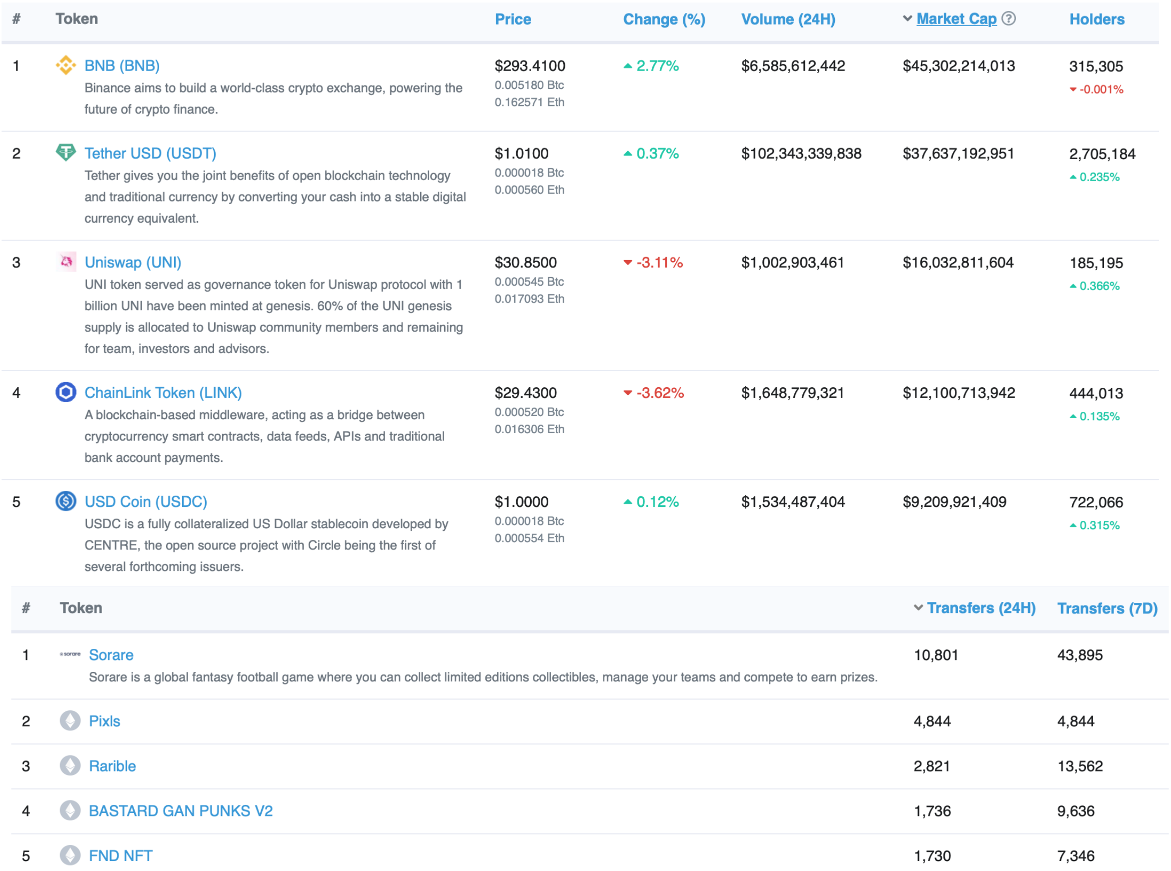
ERC-20 tokens	**ERC-721 tokens**
All tokens are identical	Each token is unique
Interchangeable	Non-interchangeable
Divisible	Non-divisible

Table 1.2. ERC-20 vs ERC-721: main differences.

1.6.5 Tokens in Ethereum

The characteristics of the Ethereum blockchain and the Turing completeness of its script make it possible to create applications for the management of Ether or other tokens. The latter can be deployed on the blockchain through start contracts.
Ethereum tokens are extremely popular, hundred of thousands, because they are easy to create and manage and because the possibilities are almost endless. All the tokens characteristics presented in the previous sections can be replicated in Ethereum.
The two main standards for tokens are the ERC-21 and the ERC-721. The first one was developed in 2015 and sets a common list of rules that an Ethereum token has to implement. It is a standard used to represent fungible tokens developed on the Ethereum network. Binance coin (BNB)[32] is currently the first ERC-20 token for volume (24h) and market capitalization. It can be used as a payment device, but also as an utility token in the Binance ecosystem, for instance to pay commissions or for participation in token sales.
Non-fungible tokens are represented by the ERC-721 standard, which was introduced in late 2017 by the popular game CryptoKitties[33]. Sorare[34] is the top ERC-721 token for number of daily transfers. It is used to represent crypto-collectibles in the soccer based game "Sorare". The main differences between ERC-20 and ERC-721 tokens are reported in Table 1.2. Moreover, the two standards differ also for the available methods and events.
The website Etherscan[35] provides a complete list of the most popular Ethereum tokens, both ERC-20 and ERC-721. In Figure 1.21 both of these lists, with their top five components, are shown.
For additional details on the characteristics of the two standards refer to [5] and [19]. There are also other standards that try to fix some issues related to the previous two, but whose description is not in the interest of this thesis.

[32] https://www.binance.com/en
[33] https://www.cryptokitties.co
[34] https://sorare.com
[35] https://etherscan.io



#	Token	Price	Change (%)	Volume (24H)	Market Cap	Holders
1	BNB (BNB) Binance aims to build a world-class crypto exchange, powering the future of crypto finance.	$293.4100 0.005180 Btc 0.162571 Eth	▲ 2.77%	$6,585,612,442	$45,302,214,013	315,305 ▼ -0.001%
2	Tether USD (USDT) Tether gives you the joint benefits of open blockchain technology and traditional currency by converting your cash into a stable digital currency equivalent.	$1.0100 0.000018 Btc 0.000560 Eth	▲ 0.37%	$102,343,339,838	$37,637,192,951	2,705,184 ▲ 0.235%
3	Uniswap (UNI) UNI token served as governance token for Uniswap protocol with 1 billion UNI have been minted at genesis. 60% of the UNI genesis supply is allocated to Uniswap community members and remaining for team, investors and advisors.	$30.8500 0.000545 Btc 0.017093 Eth	▼ -3.11%	$1,002,903,461	$16,032,811,604	185,195 ▲ 0.366%
4	ChainLink Token (LINK) A blockchain-based middleware, acting as a bridge between cryptocurrency smart contracts, data feeds, APIs and traditional bank account payments.	$29.4300 0.000520 Btc 0.016306 Eth	▼ -3.62%	$1,648,779,321	$12,100,713,942	444,013 ▲ 0.135%
5	USD Coin (USDC) USDC is a fully collateralized US Dollar stablecoin developed by CENTRE, the open source project with Circle being the first of several forthcoming issuers.	$1.0000 0.000018 Btc 0.000554 Eth	▲ 0.12%	$1,534,487,404	$9,209,921,409	722,066 ▲ 0.315%

#	Token	Transfers (24H)	Transfers (7D)
1	Sorare Sorare is a global fantasy football game where you can collect limited editions collectibles, manage your teams and compete to earn prizes.	10,801	43,895
2	Pixls	4,844	4,844
3	Rarible	2,821	13,562
4	BASTARD GAN PUNKS V2	1,736	9,636
5	FND NFT	1,730	7,346

Figure 1.21. Lists of the top five ERC-20 and ERC-721 tokens by Etherscan (https://etherscan.io).

Chapter 2

CRIX – CRyptocurrencies IndeX

The purpose of this chapter is to describe the CRIX[1], a market index [2] for cryptocurrencies developed by the blockchain research unit in the Humboldt University of Berlin [37].

Among the various types of tokens described in the previous chapter, cryptocurrencies are currently the most popular: their market value has recently exceeded $1.5 trillion, indicating an increasingly widespread collective interest.

Cryptocurrencies now seem increasingly destined to become part of investors portfolios. This makes it necessary to conduct a more in-depth financial study of them, which may be useful to understand whether some of their characteristics, identifiable in the "viewpoints" described in the previous chapter, may influence their value. The CRIX index could be fundamental in this kind of future researches indeed, for instance, it can be used to compare a particular subset of cryptocurrencies with the whole market, to understand which of their unique characteristics influences more their prices. Moreover, it could also be used to build portfolios based on it.

The main focus of this chapter is on the building procedure of CRIX, developed by the authors of [37]: all the applied techniques are deeply explained and commented. Moreover, in the last section of this chapter, the analysis of the performance of CRIX is briefly described.

2.1 Introduction and Motivation

Cryptocurrencies are gaining more and more importance and credibility. An increasing number of institutions or companies are starting to include this new kind of assets in their portfolios, believing in their value. Recently, for instance, Tesla bought $1.5 billion worth of Bitcoin. Such a move is a strong signal that the main cryptocurrency is starting to be seen as a valid investment, unlike a short time ago. Another example is Paypal, who is accepting Bitcoin as a payment method on its platform.

Of course, also other cryptos are growing rapidly in terms of capitalization and trust: the total market capitalization of cryptocurrencies is seeing a fast explosion of its

[1] https://thecrix.de

[2] A market index is a portfolio of holdings that represents a segment of interest in the financial market and whose value is derived from that of its constituents.

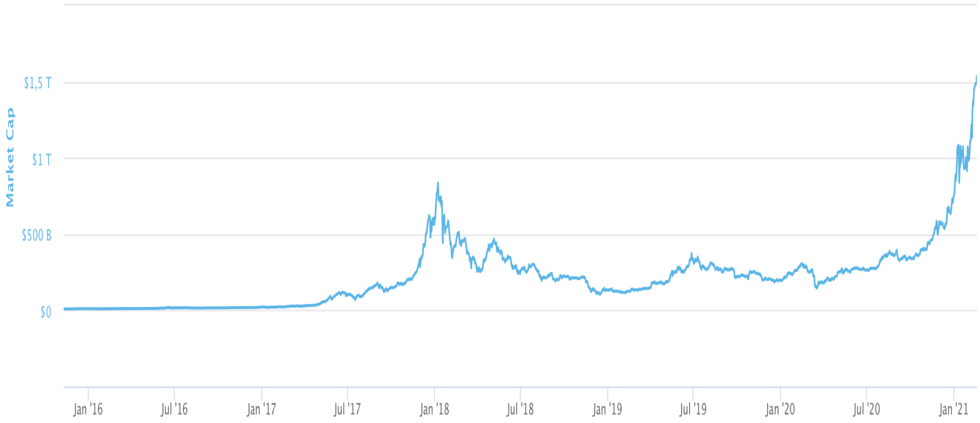


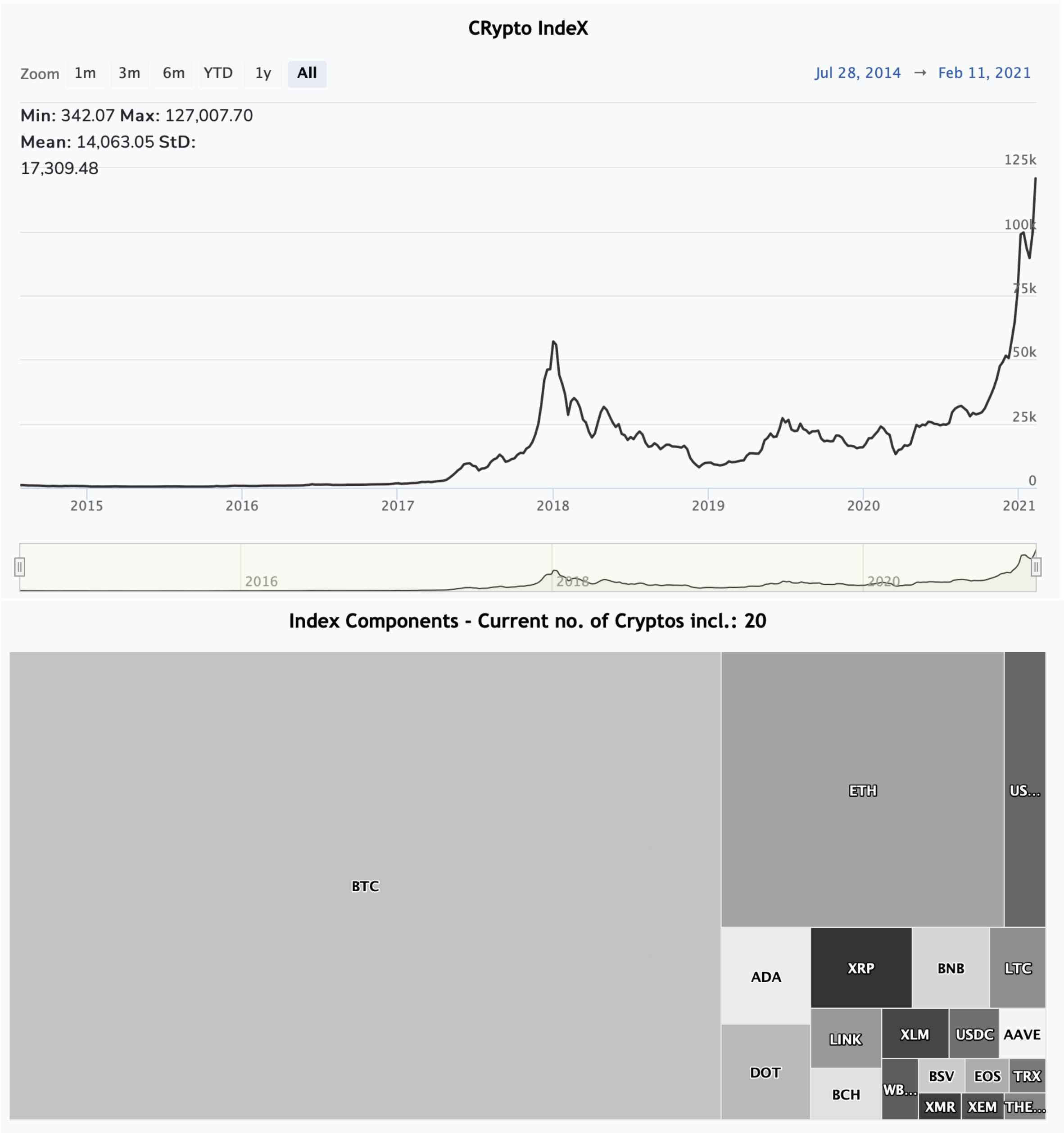
Figure 2.1. Evolution of the total market capitalization of cryptocurrencies (coinmarketcap.com).

value. This rapid growth is shown in Figure 2.1 from coinmarketcap.com.

Now that the market is so large and is attracting more and more investors there is a real need for a market index, which should aid in cryptocurrencies picking when building a portfolio and representing investors sentiment. Moreover, it could be used as an investment tool for those who want to invest in the whole market.
For researchers in the field of "Tokenomics", it could be used to get new insights on the unique characteristics that drive the price of cryptocurrencies, or to find new ones. To build an index with this characteristics it is fundamental to choose properly the constituents, which should:

1. be sufficient in number to represent accurately the market;
2. be enough liquid and not too high in number as not to compromise the purpose of create an investible index.

To do this an almost purely statistical approach is selected: just the bare minimum of rules are chosen a priori.
These rules are imposed specifically to describe better the cryptocurrencies market. One of them, for instance, concerns the frequency of reallocation of the index constituents. This choice is crucial since the CCs (cryptocurrencies) market changes very fast and the index should be able to reproduce such dynamism. A frequent reallocation, while being optimal for the tracking purpose, could be a problem in a traditional market compromising the investability purpose. Indeed, continuously changing index constituents means that also investment portfolios based on it should be changed, with potentially high transaction costs. However, in this framework this should not be a problem because the transaction costs are very low, so a more frequent reallocation is economically feasible.
Figure 2.2 from thecrix.de shows the trend of CRIX and its current components.



Min: 342.07 **Max:** 127,007.70
Mean: 14,063.05 **StD:** 17,309.48

Figure 2.2. CRIX evolution and components as of February 2021 (thecrix.de).

2.2 Index typology

The basic principle behind of any index is to study the evolution of a certain basket of assets from a starting period. This idea is formalized in the following definition.

Definition 3 (Laspeyres index) *Given a basket of k assets, the Laspeyres index P^L compares it against a base period.*

$$P^L_{0t}(k) = \frac{\sum_{i=1}^{k} P_{it} Q_{i0}}{\sum_{i=1}^{k} P_{i0} Q_{i0}},$$

where P_{it} is the price of asset i at time t and Q_{i0} the quantity of asset i at time 0.

For instance, for stock indices, Q_{i0} is a number of shares and P_{it} the correspondent price at time t. Their product results in the market capitalization.
The following definition formalizes the idea of an index which is representative for the market.

Definition 4 (Benchmark) *A benchmark is a measure which is made up of a selection of cryptocurrencies which represent the market.*

The CRIX should be build in order to be a benchmark for the CCs market and the Laspeyres definition is clearly not sufficient for this purpose. Indeed, the CCs market is frequently changing and very volatile, so the desired index should be able to represent such dynamism by changing its structure adequately.
The main problem of the Laspeyres index is that a change of constituents results in a change of the index value not due to price changes. Discontinuities may arise.
For this reason, in the construction of the CRIX the following modification of the formula is applied.

Definition 5 (Adjusted formula of Laspeyres) *The Crix index at time t, with k constituents weighted by β, is defined by the adjusted formula of Laspeyres as:*

$$CRIX_t(k,\beta) = \frac{\sum_{i=1}^{k} \beta_{i,t_l^-} P_{it} Q_{i,t_l^-}}{Divisor(k)_{t_l^-}}.$$

- *Q and P are defined as before,*
- *β_{i,t_l^-} is the adjustment factor of asset i at time point t_l^-, where l indicates that this is the l^{th} adjustment factor,*
- *t_l^- is the last time point when the quantities in the index were updated,*
- $Divisor(k,\beta)_0 = \dfrac{\sum_{i=1}^{k} \beta_{i0} P_{i0} Q_{i0}}{starting\_value}.$

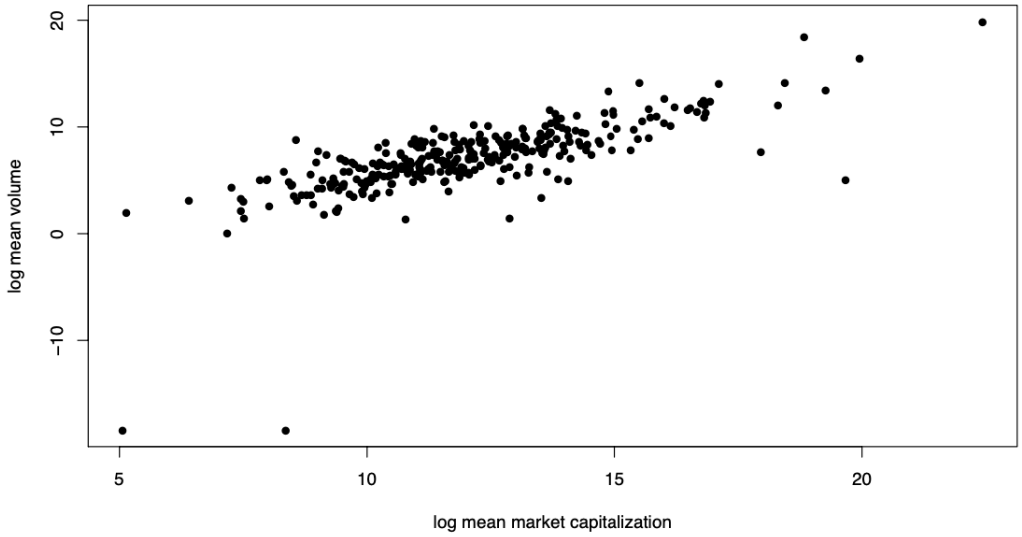


Figure 2.3. Log mean market cap. vs log mean volume for some existing cryptocurrencies [37].

The *Divisor* is a fundamental quantity that solves the problems presented before. First, by its definition, it ensures that the index has a predefined starting value, which usually is 100, 1000 or 10000. It also ensures that any development from the base period is revealed.

Moreover, if the structure of the index is modified, the *Divisor* is adjusted in such a way that only price changes are effectively reflected, preventing discontinuities of the curve. Suppose, for instance, that at time t_l^- there is an update, which changes the number of the constituents from k_1 to k_2 and the weights. Then, the *Divisor* is also adjusted so that the value of the index does not change. Its new form is obtained by solving the following equation:

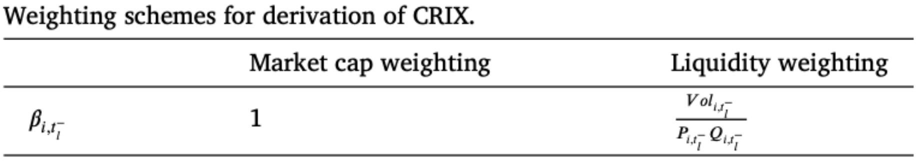
$$\frac{\sum_{i=1}^{k_1} \beta_{i,t_{l-1}^-} P_{i,t-1} Q_{i,t_{l-1}^-}}{Divisor(k_1,\beta)_{t_{l-1}^-}} = CRIX_{t-1}(k_1,\beta) = CRIX_t(k_2,\beta) = \frac{\sum_{j=1}^{k_2} \beta_{j,t_l^-} P_{j,t} Q_{j,t_l^-}}{Divisor(k_2,\beta)_{t_l^-}},$$

with $Divisor(k_2,\beta)_{t_l^-}$ as unknown.

This procedure is similar to the one used for the S&P Dow Jones indices [15], but with the fundamental difference that in the case of CRIX the number of constituents is not fixed. The problem of determining the constituents of the index is tackled in the section 2.4.

Before beginning the explanation of the statistical procedure used to construct the index, it is important to discuss another property of the CCs market. In Figure 2.3 from [37] it can be seen that for some cryptocurrencies an high market capitalization is not always followed by an high trading volume.

The trading volume could maybe be considered important for the cryptocurrencies that are selected to compose the index due to the investability purpose. For this reason, another index, the LCRIX (Liquidity CRIX), is constructed in [37]. The assets



Weighting schemes for derivation of CRIX.

	Market cap weighting	Liquidity weighting
β_{i,t_l^-}	1	$\frac{Vol_{i,t_l^-}}{P_{i,t_l^-} Q_{i,t_l^-}}$

Figure 2.4. Weighting schemes for CRIX and LCRIX [37].

that constitute it are weighted proportionally to their trading volume.
The weighting schemes of both CRIX and LCRIX are shown in Figure 2.4 from [37]. The LCRIX has the potential to diminish the strength of cryptocurrencies such as Bitcoin, which is dominant in terms of market capitalization but, as shown in other studies and in the next chapter, is not able to describe the whole market by itself. Liquidity CRIX is built following the same procedure used for CRIX, as described in the next section. The weighting scheme is the only thing that changes.

2.3 Introduction to components selection

The choice of the number of constituents of the index is a central task both in importance and difficulty.
It could be worthwhile to include a large number of assets: this can improve the way the index represents the market, which is crucial because CRIX should be a benchmark. In reality, there are already a number of stock indices whose sole role is to best track a given market called *Total Market Indices (TMI)*. They include the stocks of all the companies present on the target market. An example is the S&P Total Market Index[3]. Of course, CRIX could be built as a TMI, but this is not optimal. Indeed, there are several reasons to prefer a sparse index, i.e. an index with a low number of constituents:

1. financial practice has shown that smaller indices receive more attention;

2. CRIX should be an investable index. In the future ETFs or contingent claims based on it could be created. For this reason it should include only a low enough number of liquid assets: adding illiquid and non-inveastable asset could make impossible the management of a tracking portfolio.

The reasons introduced lead to the following hypothesis.
Hp: between investment portfolios with equal performance, the one with least assets is chosen.

As already pointed out, the whole procedure for the choice of the constituents is statistical.
Now, some quantities necessary for the construction procedure are defined.

Definition 6 (Total market) *The total market (TM) consists of all the existing CCs and its value is the combined market value of all the CCs.*

[3] https://www.spglobal.com/spdji/en/indices/equity/sp-total-market-index-tmi/

$$TM(K)_t = \frac{\sum_{i=1}^{K} P_{it} Q_{i,t_l^-}}{Divisor(K)_{t_l^-}}.$$

- *The Divisor is used as a normalization coefficient for the TM to make it comparable to a benchmark candidate.*
- *K is the total number of cryptocurrencies in the CC universe.*

Definition 7 (Log returns) *Given a positive time series $\{x_t\}_{t \in T}$, where T is a discrete time interval, its log returns are defined as:*

$$\epsilon_t^x = log(x_t) - log(x_{t-1}),$$

for every $t, t-1 \in T$.
For the TM and the $CRIX$, the respective log returns are:

$$\epsilon(K)_t^{TM} = log(TM(K)_t) - log(TM(K)_{t-1}),$$

$$\epsilon(k, \beta)_t^{CRIX} = log(CRIX(k, \beta)_t) - log(CRIX(k, \beta)_{t-1}),$$

where $CRIX(k, \beta)_t$ is the value of CRIX with k constituents at time t.

Now, the problem of finding the constituents of the index and their weights can be stated.
Problem: find k and β which achieve

$$\min_{k,\beta} ||\epsilon(k,\beta)||^2 = \min_{k,\beta} ||\epsilon(K)_t^{TM} - \epsilon(k,\beta)_t^{CRIX}||^2.$$

The squared loss function is selected because it heavily penalizes deviations.
The problem stated earlier, even if correct mathematically, is not solvable with the actual data available. In fact, the total market is not really derivable as it is not possible to obtain data of all existing cryptocurrencies. For instance, many may not be listed on reliable exchanges or in general may lack of information. Therefore, is necessary to move from the Total Market, which is a fictitious quantity useful for defining the problem, to something that can actually be calculated.

Definition 8 (Total market index) *The total market index (TMI) contains all the CCs whose price is available (of number k_{max}).*

$$TMI(k_{max})_t = \frac{\sum_{i=1}^{k_{max}} P_{it} Q_{i,t_l^-}}{Divisor(k_{max})_{t_l^-}}.$$

It is used as a proxy for the TM.

The previous minimization problem can now be restated.
Problem: find k and β which achieve

$$\begin{gathered}\min_{k,\beta} ||\hat{\epsilon}(k,\beta)||^2 = \min_{k,\beta} ||\epsilon(k_{max})_t^{TMI} - \epsilon(k,\beta)_t^{CRIX}||^2 \\ \text{s.t. } 1 \leq k \leq k^u, \\ k = k_1 + s, \\ k^u \in [1, k_{max}], \\ s \in [1, k_{max} - k_1], \\ \beta^{1\times k} = (1, ..., 1, \beta_{k_1+1}, ..., \beta_{k_1+s})^T, \\ \beta_{k_1+1}, ..., \beta_{k_1+s} \in (-\infty, +\infty).\end{gathered}$$

It is important to understand well all the quantities involved.

- $\epsilon(k_{max})_t^{TMI}$ are the log returns of the TMI.
- k_1 is the minimum number of constituents of the index, which is specified a priori. The first k_1^{th} CCs are weighted by 1 (or by the liquidity weighting in the case of LCRIX).
- s represents the additional number of cryptocurrencies that could be added to obtain a preferable result. These additional CCs are weighted by a weight β which has to be derived.
- k is the total number of CCs included in the CRIX index.

The CRIX index has the following form:

$$CRIX_t(k,\beta) = \frac{\sum_{i=1}^{k_1} P_{it} Q_{i,t_l^-} + \sum_{j=k_1+1}^{k_1+s} \beta_{j,t_l^-} P_{jt} Q_{j,t_l^-}}{Divisor(k_1)_{t_l^-}}.$$

The choice of s is crucial. Adding additional assets, with proper weights, could improve the tracking ability of the index, while maintaining a number of components as low as possible.
All parameters introduced above must be estimated from the data. The next section explains the chosen procedure.

2.4 CRIX model selection with AIC

In [37] several model selection criteria are introduced and tested. In order to choose the most suitable one for the CRIX problem, multiple aspects can be considered. First, a theoretical one concerns the asymptotic behaviour of the criterion, in terms of consistency and asymptotic efficiency. These properties have been investigated over the years in numerous works which provide precise results. Also the dependency of the criterion on the data could be crucial for the result. Moreover, for instance, there are criteria for which empirical evidence suggests that they choose models which overfit. In the case of CRIX also the sparsity and the stability of the resulting index have been considered as fundamental properties to reach.

In this chapter the introduction of all the possible criteria will be omitted: just the selected one, the Akaike Information Criterion (AIC), is introduced, together with some related results. The AIC is selected because, as shown later, it has good asymptotic properties, it is highly dependant on data and it provides a desirable result.
Akaike information criterion [1] was introduced by the statistician Hirotugu Akaike. It provides an estimator of the relative prediction error of a statistical model with respect to a given set of data. The idea was derived starting from the notion of Kullback-Leibler information or distance between two models, which is a measure of the information lost using a model to approximate another one.

Definition 9 (K-L distance) *The distance between two models f and g, which could be for instance continuous distributions, is:*

$$I(f,g) = \int f(x) log(\frac{f(x)}{g(x|\theta)})dx.$$

$I(f,g)$ is the information lost when g is used to approximate f.

The latter quantity depends on:

- the models to be tested, in terms of their parameters θ,
- the so-called truth, the quantity to be approximated, which can be thought as the distribution which generated the data.

Using the K-L density with real data is usually not possible since the truth is unknown and dependent on a non-specific number of parameters, perhaps infinite. So, the K-L distance can not be used in practice. For more details see [6], chapters 2 and 7.
Akaike found a rigorous way to estimate the expected, relative distance between the fitted model and the unknown true mechanism that actually generated the observed data.

Definition 10 (Akaike information criterion AIC) *Suppose to have a statistical model of some data, dependent on k parameters to be estimated. Let $\hat{L}$ be the maximum likelihood function for the model. Then:*

$$AIC = 2k - 2log(\hat{L}).$$

AIC can be used as a relative measure for model selection, in the sense that, among many models, the one with the lower AIC has to be preferred. The absolute size of the AIC value does not have any meaning, what matters is the relative value of the AIC of one model compared to that of the other tested models.
The formula of the criterion has a very intuitive meaning, given by the two terms that compose it. Indeed, the first one is a penalization term which increases the AIC proportionally to the number of parameters of the model. Instead, the second one improves the measure rewarding the goodness of fit. It is important to notice that, in general, adding more parameters results in an improved goodness of fit. So, considering both terms together with their meaning, the famous bias-variance trade

off, or the equivalent trade-off between over-fitting and under-fitting, stands out.

After this brief introduction, the AIC applied to the CRIX estimation problem is the following:

$$AIC(\hat{\epsilon}(k,\beta),s) = -2log(L\{\hat{\epsilon}(k,\beta)\}) + 2s,$$
$$L\{\hat{\epsilon}(k,\beta)\} = \max_{\beta} \prod_{t} f(\hat{\epsilon}(k,\beta)_t),$$

where f is the density of the $\hat{\epsilon}(k,\beta)_t$ over all t.
The other quantities involved, even if already defined, are recalled:

- $\hat{\epsilon}(k,\beta)$ represents the model error to be minimized. It is the difference between the log returns of the TMI and CRIX;
- $s = k - k_1$ is the number of possible additional constituents of the index.

Among the various model selection criteria considered in [37] (Generalized Cross Validation, Generalized Full Cross Validation, Mallows' C_p, Shibata, Final Prediction Error and Akaike Information Criterion), AIC is the most data-dependent, because it takes into account the maximum likelihood and therefore the density of the observations.

In [34] it is shown an important asymptotic efficiency result concerning the AIC. It holds under two assumptions:

1. $\hat{\epsilon}(k,\beta) \sim \mathcal{N}(0, \hat{\sigma}(k,\beta)^2)$;
2. the number of regression variables (in this context the number of constituents) increases with the sample size.

The required hypothesis are bot verified. Indeed:

1. as will be explained later, the distribution of $\hat{\epsilon}(k,\beta)$ is estimated non-parametrically with an Epanechnikov kernel. So, the normality result may not hold. However, it has been shown that the result holds also for a general elliptical distribution. The authors of [37] suppose that the true distribution belongs to this family and exploit the result that the non-parametric estimator tends to it;
2. the second assumption is plausible given the widespread use of cryptocurrencies, which is estimated to increase over time. So, over time there will certainly be new components that will need to be added to monitor an ever growing market.

It is worth commenting on this result in more detail, explaining what is the meaning of asymptotic mean efficiency.
Given a squared loss function L, i.e. a squared norm of the difference between the true model and its estimate, it is possible to provide the following definition.

Definition 11 (Risk Function) *Given a set of estimate models, indexed by j, and the corresponding loss functions L, the risk function is defined as:*

$$R(j) = \mathbb{E}_j[L],$$

for every model j.
$R(j)$ is the expected value of the loss function of model j.

Definition 12 (Mean Efficiency of a model) *Consider a set of estimate models, indexed by j, and the corresponding risk functions $R(j)$.*
Among these models, let j^* be the one such that $R(j^*) \leq R(j)$, $\forall j$.
The mean efficiency for a model $\hat{j}$ is:

$$eff(\hat{j}) = \frac{R(j^*)}{R(\hat{j})}.$$

Finally, it is possible to define the following asymptotic property.

Definition 13 (Asymptotic mean efficiency) *A model indexed by $\tilde{j}$ is asymptotic mean efficient if*

$$\liminf_{n\to\infty} eff(\tilde{j}) = 1.$$

The minimum AIC procedure, under the previously stated conditions, selects a model which is asymptotic mean efficient.
This result is fundamental and explains why AIC is the proper method to solve the problem of finding the components of CRIX.

At this point, several models, each with a different value of s, have to be compared. For each of them, the AIC is computed and the model which achieves its minimum is selected.
However, there is a problem, indeed the density $f(\hat{\epsilon}(k,\beta)_t)$ in the expression of the maximum likelihood in the AIC formula is unknown, so it should be estimated from the data. Density estimation is a very wide topic, greatly used for instance in the field of econometrics. Now, it is introduced with its main results. For more details see [16] and [42].

The approaches commonly used in density estimation are of two types.

1. Parametric: the main assumption is that the density is known up to a finite number of parameters. This could be a simplification, because just specifying the parameters means describing the whole distribution, but also a strong constraint on the shape of the density.

2. Non-parametric: the main assumption is that the density belongs to some infinite collection of curves with some regularity properties. Non-parametric methods, while being generally more complicated, allow great flexibility.

In [37] the density is estimated non-parametrically with an Epanechnikov kernel.
The non-parametric approach is now explored more in depth.
First, a proper form of the estimator should be defined. In [16] it is obtained with a simple procedure, generalizing the theory of density estimation with histograms. The derived estimator is defined as follows.

	K(u)
Uniform	$\frac{1}{2}\mathbb{I}\{\lvert u\rvert \leq 1\}$
Triangle	$(1-\lvert u\rvert)(\mathbb{I}\{\lvert u\rvert \leq 1\})$
Epanechnikov	$\frac{3}{4}(1-u^2)\mathbb{I}\{\lvert u\rvert \leq 1\}$
Quartic	$\frac{15}{16}(1-u^2)^2\mathbb{I}\{\lvert u\rvert \leq 1\}$
Triweight	$\frac{35}{32}(1-u^2)^3\mathbb{I}\{\lvert u\rvert \leq 1\}$
Gaussian	$\frac{1}{\sqrt{2\pi}}exp\{-\frac{1}{2}u^2\}$
Cosine	$\frac{\pi}{4}cos(\frac{\pi}{2}u)\mathbb{I}\{\lvert u\rvert \leq 1\}$

Table 2.1. Different forms of kernel weight.

Definition 14 (Kernel density estimator) *Given an observed random variable X, with unknown distribution $f(x)$, then it is possible to define the following estimator for the density:*

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}).$$

- *K is called kernel weight and is a probability density function. This implies that also $\hat{f}_h(x)$ is a pdf.*
  Moreover, in general, it is chosen in such a way that:
  $$\int uK(u)du = 0.$$
- *n is the number of observations.*
- *h is called bandwidth or smoothing parameter. It is used in the derivation of the formula. The basic idea, taken from the histogram construction method, is to count, for every x, how many observations fall in the interval $[x-h, x+h]$. Then, the estimated density evaluated at x is:*
  $$\hat{f}(x) = \frac{\#obs.in[x-h, x+h]}{nh}.$$
  Starting from this first estimator it is possible to derive the formula stated above. In the suggested papers the detailed procedure is explained.

The kernel weight function can have various forms, reported in the Table 2.1, whose graphs are shown in Figure 2.5 from [42].

The choices of the kernel weight and of the bandwidth parameter are both extremely influential on the result.
There are several different procedures for the selection which can be found in the suggested papers, a full explanation is not in the scopes of this thesis. Just the choices made for CRIX are developed later. In general, it is worth noticing how much the choice of h could influence the result in terms of the smoothness of the estimated density. In Figure 2.6, from the Wikipedia page about Kernel density estimation, the

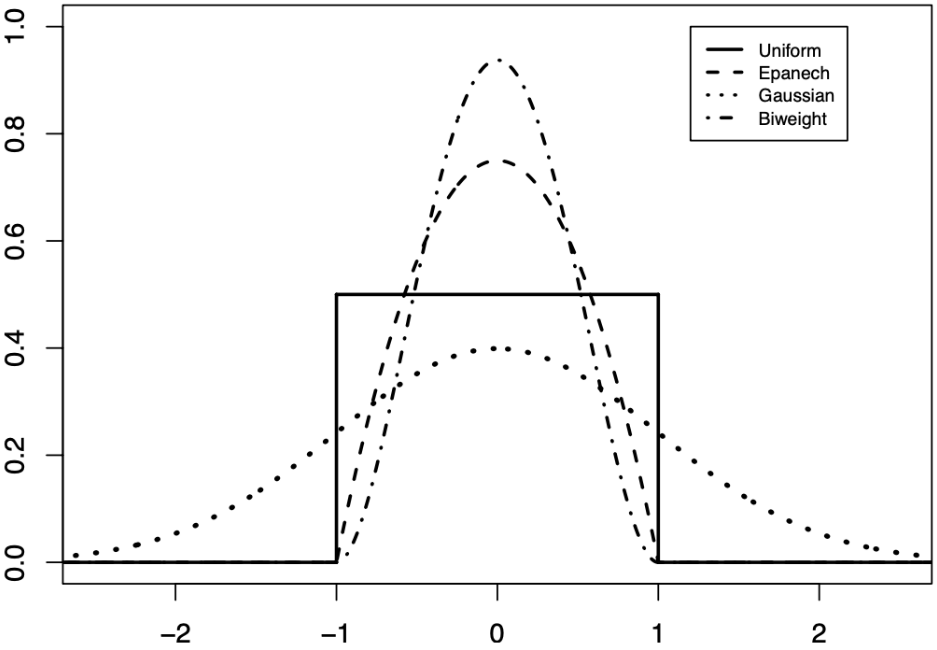


Figure 2.5. Representation of some kernel weights [42].

same density is estimated with a fixed kernel weight, but changing the values of the bandwidth parameter h. The smoothness of the estimated curve changes drastically just slightly modifying h.

This review on kernel density estimate ends with a focus on the methods used in the derivation of CRIX, starting from the choice of the kernel weight.
In order to do this it is necessary to introduce some quantities, that will be useful also in the explanation of the choice of the smoothing parameter.
In general, the evaluation of an estimator can be done by computing its Mean Squared Error:

$$MSE(\hat{f}_h(x)) = \int (\hat{f}_h(x) - f(x))^2 dx = Var(\hat{f}_h(x)) + Bias(\hat{f}_h(x))^2.$$

Now, all the quantities included in the latter formula are derived, starting from the bias.

$$\mathbb{E}[\hat{f}_h(x)] = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{E}(K(\frac{x - X_i}{h})) = \int K(y) f(x - yh) dy.$$

Under some regularity assumptions on f, it is possible to do a second order Taylor expansion of $f(x - yh)$ about x:

$$f(x - yh) = f(x) - hyf'(x) + \frac{1}{2}h^2 y^2 f''(x) + o(h^2).$$

Substituting this quantity in the expression of the expected value, the resulting bias is:

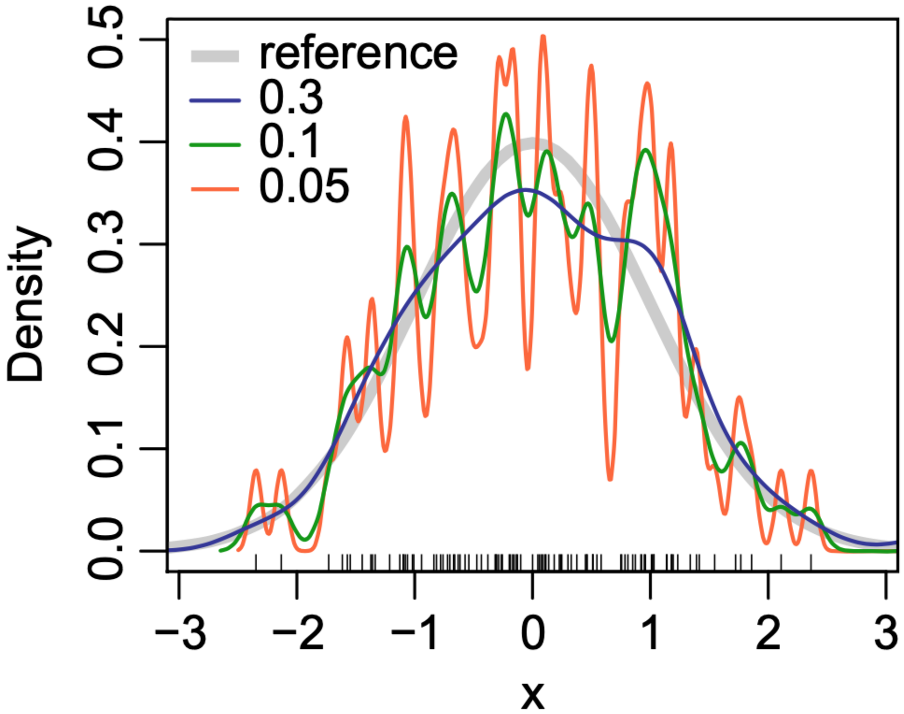


Figure 2.6. Estimate of the same density with different values of h (Wikipedia/Kernel density estimation).

$$Bias(\hat{f}_h(x)) = \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2),$$

where $\mu_2(K) = \int u^2 K(u)du$.
Instead, the expression of the variance is:

$$Var(\hat{f}_h(x)) = \frac{1}{nh} \int K^2(y)f(x-hy)dy - \frac{1}{n}(\mathbb{E}(\hat{f}_h(x)))^2 = \frac{1}{nh}R(K)f(x) + o(\frac{1}{nh}),$$

where $R(K) = \int K(u)^2 du$.
Finally, from the definition of MSE:

$$MSE(\hat{f}_h(x)) = Var(\hat{f}_h(x)) + Bias(\hat{f}_h(x))^2 = \frac{1}{nh}R(K)f(x) + \frac{h^4}{4}f''(x)^2\mu_2(K)^2 + \\ + o(\frac{1}{nh}) + o(h^4).$$

It is straightforward to see that, in order to have consistency, it should be that for $n \to \infty$: $h \to 0$ and $nh \to \infty$. These two conditions on the bandwidth imply that $MSE(\hat{f}_h(x)) \to 0$ for $n \to \infty$.

In the density estimation framework the quantity generally used to evaluate an estimator is the expected value of the MSE, called Mean Integrated Squared Error.

Definition 15 (Mean Integrated Squared Error) *Given a density $f(x)$ and its kernel estimator $\hat{f}_h(x)$, the Mean Integrated Squared Error (MISE) is defined as:*

$$MISE(\hat{f}_h(x)) = \int MSE(\hat{f}_h(x))dx = \frac{R(K)}{nh} + \frac{h^4\mu_2(K)^2R(f'')}{4} + o(\frac{1}{nh}) + o(h^4).$$

It is a L^2 risk function.

From the MISE, it is possible to define the following quantity.

Definition 16 (Asymptotic Mean Integrated Squared Error) *The Asymptotic Mean Integrated Squared Error (AMISE) is an approximate formula of the MISE, obtained ignoring higher order terms.*

$$AMISE(\hat{f}_h(x)) = \frac{R(K)}{nh} + \frac{h^4 \mu_2(K)^2 R(f^{''})}{4}.$$

It is minimized by the following value of the bandwidth parameter:

$$h_{opt} = (\frac{R(K)}{\mu_2^2(K) R(f'')})^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

When h_{opt} is selected, using a proper density function, the Epanechnikov kernel achieves the minimum AMISE among all kernel weights.
For this reason, and for its good numerical performance, the Epanechnikov kernel is chosen in [37] for the derivation of CRIX.

Instead, h is found using a plug-in approach for bandwidth selection.
The plug-in selector solves a problem in deriving the optimal bandwidth. Indeed, the value of h_{opt} depends on the unknown quantity $f^{''}$. Plug-in methods refer to those that find a pilot estimate of $R(f^{''})$, starting from a pilot estimate of h, and "plug it in" the estimation of AMISE. In this way the optimal bandwidth can be effectively computed. There are several possibilities, all exploring various ideas to estimate $R(f^{''})$. In the derivation of CRIX, the plug-in method proposed by Sheather and Jones [33] is used. First, it extends the estimation of h_{opt} by adding another term in the Taylor expansion of MISE. Then, it selects an estimator of $R(f^{''})$, as explained in [16] and [42].
There are several other methods to derive the optimal value of h, without one that is always better than the others. The choice of the proper method to select the smoothing parameter is strongly case dependent. In this case it is verified that the plug-in approach gives the best result.

The setup is almost complete, there is just another result left, which is fundamental for the numerical computation.

Theorem 1 *The rate of improvement of the AIC depends on the relative value of an asset in the market.*

This result says that cryptocurrencies can be added in the model using a top-down approach based on their market capitalization. It is proved in [37] under the assumption of normally distributed errors.

At this point the whole mathematical procedure has been presented.
The next section introduces some rules for choosing the composition of the index and for reallocating its components.

2.5 Composition and reallocation rules

In this section some rules concerning index composition and reallocation are introduced. As already pointed out, the procedure to build CRIX is almost totally statistical: in this way there is not a problem of trust in the entity which created it.
Of course, some a priori choices have to be made, because the index should be not only a good approximation of the market, but also sparse and investable.
The composition rules are three:

1. the first regards the parameter k_1, the minimum number of components, which is set equal to 5;

2. the second regards the number of components. This rule imposes to have an index with a number of components evenly divisible by 5, which is common in the financial industry. So, the AIC is computed for the models with $k = 5, 10, 15, ...$;

3. finally, a stopping rule is proposed, because the goal of obtaining a sparse index could be achieved just by finding a local optimum of the AIC. There is no need of finding the global minimum, since this could mean including too many components and increasing the numerical complexity. In other terms, the research for the optimal model stops at level j whenever:

$$AIC(\hat{\epsilon}(k_j, \beta), k_j - k_1) < AIC(\hat{\epsilon}(k_{j+1}, \beta), k_{j+1} - k_1).$$

   The performance of the resulting index is still very good and, considering just the models with $k_i \leq k_j$, the choice given by AIC is still asymptotically optimal.

The number of components of CRIX, once determined, can not remain the same for a too long period. A frequent reallocation is necessary to keep track of the strong dynamism of the CCs market. Indeed, just in few weeks new coins can be developed and go at the top of the market while others may disappear.
For CRIX, on the last day of every month, the constituents are revised, while keeping their number. This is done by checking the market capitalization of the CCs. Every three months also the number of index constituents is recomputed.
So, CRIX is treated like a monthly re-balanced portfolio which number of constituents is reviewed quarterly.

Finally, it may happen that some data of the analyzed time series are missing. If an isolated missing value occurs alone in the dataset, then Missing At Random (MAR) is assumed. The Last-Observation-Carried-Forward (LOCF) method is then applied to fill the gap. The idea behind the application of LOCF, instead for instance of a regression method, is to alter as little as possible the direction of data.
If there are several consecutive missing observations then the corresponding cryptocurrency is omitted.

This ends the explanation of the construction of the index. The next section is dedicated to the performance analysis.

2.6 Performance analysis

The performance analysis is a central task to evaluate if the index actually works well with real data.
The main focus is in understanding if the composition rules allow the achievement of a satisfactory result. To do this, the idea is to build other indices using new composition rules and then compare the results. The comparison is made through:

1. the sparsity of the index,
2. the Mean Squared Error (MSE)

$$MSE(CRIX(k)) = \frac{1}{t_l^+ - t_l^-} \sum_{t_l^-}^{t_l^+} (CRIX(k)_t - TMI(k_{max})_t)^2,$$

3. the Mean Directional Accuracy (MDA)

$$\begin{aligned} MDA(CRIX(k)) = \frac{1}{t_l^+ - t_l^-} \sum_{t_l^-}^{t_l^+} \mathbb{I}\{sign(TMI(k_{max})_t - TMI(k_{max})_{t-1}) = \\ = sign((CRIX(k)_t - (CRIX(k)_{t-1})\}. \end{aligned}$$

Moreover, also other model selection criteria different from the AIC are tested. The reason of this is to understand if the theoretical motivation that led to the choice of AIC is supported by the empirical results.
This section will not cover the performance analysis in full details, since the main focus of this chapter is to describe the statistical procedure. However, the main results are provided and commented. Their full derivation can be found in [37].

The new indexes obtained by changing composition rules are now presented.

- ECRIX: this index, whose name stands for Exact-CRIX, is built changing the first and the second composition rule. Indeed, the number of its constituents is chosen in steps of 1, starting from $k_1 = 1$. The stopping criterion is the same as for CRIX.
- EFCRIX: this index, whose name stands for Exact-Full-CRIX, follows the same rules of ECRIX however, without considering the stopping criterion. Indeed, the global minimum of AIC is computed. The intention is to obtain an index which may approach the number of components of the TMI, but only if even small assets help improve the approximation on the total market.

A comparison of the number of constituents selected by different model selection criteria shows that the AIC performs better in obtaining a more sparse result. Moreover, it also achieves good results for the MSE and the MDA. This confirms the choice of AIC as model selection criterion.

Between the three constructed indexes the CRIX is preferred.
All the three variants are close in terms of MDA, but CRIX outperforms ECRIX and EFCRIX in terms of MSE. Moreover, in the various time periods considered for the analysis, the CRIX has a number of constituents which is more stable.
It is interesting to point out that, in general, the ECRIX has more than one component under the AIC. This means that Bitcoin alone, even if it accounts for the most of the market capitalization, is not able to describe the whole market: also other CCs are necessary.

Conclusions

The *main achievement* of this thesis is to provide a structured and formal picture of cryptographic tokens, although, inevitably, not yet complete, and a description of the mathematical techniques used to construct the CRIX index.
The thesis encompasses and seeks to develop some of the most recent results presented in numerous papers, articles and conferences on tokens and blockchain, adding also several examples of real projects. This material is carefully selected: it was crucial in the choice of sources to opt for reliable and consistent ones. The latter work is easier for topics with an already well-developed and clear theory, but tokens are not among them. Indeed, as tokens are an emerging topic, their theory is still uneven and sometimes inconsistent: there are many small contributions but a shared overview is still lacking. For this reason, the problem of selecting sources required special attention.
Moreover, the material is also linked, commented on and provided with some new ideas and reasoning. For example, the definition of cryptographic token is developed on the basis of several works conducted by authoritative authors or agencies, such as ISO [18] or NIST [28]. The section on classification also contains some comments and additions to the available theory.
The structure in which the topics are presented is designed to make them easier to understand. This was very challenging, especially for the tokens classification part. In fact, given the many ways to classify tokens, it was necessary to think about how to introduce the arguments so as not to run into the problem of providing an unstructured and confusing explanation. The order in which the main tokens features are presented is structured in an attempt to avoid the latter issue.
The CRIX index represents an additional financial focus on cryptocurrencies, as its structure can provide important information, for instance on which cryptocurrencies are most representative of the market. All the techniques used to construct the index are explained, and reference material for further studies is provided.

The analysis of tokens shows how versatile and at the same time complex these objects are. The many aspects that characterise tokens are strongly related, as they influence each other, and at the same time multidisciplinary: in the most complex cases, mathematics, finance, computer science, game theory, economics, law and social sciences are involved.
However, starting from the abundant specific characteristics, it is apparent that there are generic properties and aspects that all tokens possess and that can be considered in a high-level analysis.

1. Each token generates some form of *interaction.*

This aspect is implicit in the actual definition of a cryptographic token provided in this thesis, which defines a token as a tool to "exchange value, exchange verifiable data and achieve coordination between users".
In general, the interaction generated is of one, or more simultaneously, of the following types:

- between users and the blockchain itself. This is, for example, the case for native or work tokens, which are used for network governance through a mechanism that encourages users to have a certain type of interaction with the ecosystem;
- between user and user. This is the most common case as most tokens, with the exception of some NFTs, can be given away, exchanged or sold to other users. In general, all fungible tokens allow this kind of interaction. Native tokens also generate interaction between users, as they compete to generate the next block and get a reward;
- between the blockchain and the outside world. As already pointed out, there are tokens created on a blockchain but representing the possession of a real (tokenized) asset (asset-backed tokens) or of a share (security tokens). There is also the possibility of tokens on a blockchain that give the possibility of using a service in the real world (usage tokens) and many others.

2. Each token has some *economic and financial aspects.*
   These concern, for example, how the token is issued to users, its life cycle, its conversion rules and exchange mechanism and in general everything that could impact its value.
   In the case of tokens that offer trading possibilities, it is essential to consider what their value might depend on and whether to manage it in some way (for example, as already mentioned, with bonding curves) or to make it depend only on supply and demand.

3. Each token has some *legal aspects.*
   When a token is designed, in every case, it should always be considered into which existing legal category it may be placed after release. If a token is designed for a certain use that is not intended for the legal category into which it is inserted, then the whole project is rendered useless.
   The issue of legal predictability is fundamental and linked to the evolution of regulations.

When creating a token, it is convenient to first reason in general terms, for example thinking about what kind of interaction one wants it to generate, what its value should depend on, or which legal category would be most appropriate for its purpose. Subsequently, one can proceed to a more specific characterisation such as that shown in the dedicated chapter.
The CRIX index can be useful for several applications, that go beyond simply describing the cryptocurrencies market, and range from finance to further studies on tokens. Some of them are:

- build portfolios with cryptocurrencies: the index can be used, for instance, to understand how a specific cryptocurrency, or a set of them, performs with respect to the whole market, or to build a portfolio that tracks the CCs market;

- compare the CCs market with more established ones. This comparison could help, for example, in the creation of mixed portfolios composed of traditional assets and cryptocurrencies, to improve diversification and decrease risk exposure;

- get new insights on the unique characteristics, among those presented in this thesis, that drive the price of cryptocurrencies, or find new ones.
  In general, having the ability to consult a market index can be useful in understanding how to develop the various financial aspects of a cryptocurrency.

Moreover, the mathematical procedure is very interesting and innovative compared to that used for other indices because it relies on an almost purely data-driven choice of constituents. It can be applied also to build indices for other markets which, like the one of cryptocurrencies, change rapidly [37].

Some possible future *developments* of this thesis may concern both the academic and the professional environment. Moreover, the work carried out may be used by those interested in learning more about the discipline as a basis, with the opportunity to explore many topics more in details through the cited works.
In the business environment this thesis might help in the design of a token, as it provides various information and comments, particularly on classification, which can be used as a basis to understand what the correct choices might be.
In academic setting, this thesis might provide a starting point for future works, where one may wish to investigate a particular aspect of tokens among the various presented, or for example provide a tool for tokens classification based on the features introduced in this thesis.
Of course, all the possible uses of CRIX listed above, and certainly many others, can be the basis for future research works in finance and token analysis.

To conclude, it is important to underline that this work explores only a part of what concerns cryptographic tokens, seeking to provide a reliable representation. Indeed, it would be impossible to try to cover all aspects, which are too vast and involve too many disciplines. Moreover, the rapid development of studies on the topics presented in this thesis would make the attempt to do so also pointless.
In the future, however, this work could be the basis for further enrichments and updates, which could keep track of the developments in this field that is so vast and destined to attract more and more attention.

Bibliography

[1] Hirotugu Akaike. “Statistical predictor identification.” In: *Ann. Inst. Statist. Math. 22 (1), 203–217.* (1970).

[2] Adam Back; Matt Corallo; Luke Dashjr; Mark Friedenbach; Gregory Maxwell; Andrew Miller; Andrew Poelstra; Jorge Timón; Pieter Wuille. *Enabling Blockchain Innovations with Pegged Sidechains.* 2014.

[3] Srinivasan Balaji S. *Thoughts on Tokens.* 2017. URL: https://news.earn.com/thoughts-on-tokens-436109aabcbe.

[4] Bitcoin.org. *Developer guides.* URL: https://developer.bitcoin.org/devguide/index.html.

[5] BitcoinWiki. *ERC20.* URL: https://en.bitcoinwiki.org/wiki/ERC20#How_Does_ERC-20_Token_Contract_Work.3F.

[6] Kenneth P. Burnham; David R. Anderson. *Model selection and multimodel inference.* Springer, 2002.

[7] Vitalik Buterin. *Ethereum Whitepaper.* 2013. URL: https://ethereum.org/en/whitepaper/.

[8] Token Alliance - Chamber of digital commerce. *UNDERSTANDING DIGITAL TOKENS: Market Overviews and Proposed Guidelines for Policymakers and Practitioners.* URL: https://digitalchamber.org/initiatives/token-alliance/.

[9] Yan Chen. “Blockchain tokens and the potential democratization of entrepreneurship and innovation”. In: *Business Horizons, Elsevier, vol. 61(4), pages 567-575.* (2018).

[10] Patrick Collins. *What Is a Blockchain Oracle?* 2020. URL: https://betterprogramming.pub/what-is-a-blockchain-oracle-f5ccab8dbd72.

[11] EUROPEAN COMMISSION. *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Markets in Crypto-assets, and amending Directive (EU) 2019/1937.* 2020. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0593.

[12] EFFORCE. *White Paper.* 2020. URL: https://efforce.io/WP_ENG_V1.pdf.

[13] Thomas Euler. *The Token Classification Framework: A multi-dimensional tool for understanding and classifying crypto tokens.* 2018. URL: http://www.untitled-inc.com/the-token-classification-framework-a-multi-dimensional-tool-for-understanding-and-classifying-crypto-tokens/.

[14] EU Blockchain Observatory & Forum. *EU Blockchain Ecosystem Developments.* URL: https://www.eublockchainforum.eu/sites/default/files/reports/EU%20Blockchain%20Ecosystem%20Report_final_0.pdf.

[15] S&P Global. *Index Mathematics Methodology.* URL: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-index-math.pdf.

[16] W. Härdle; M. Müller; S. Sperlich; A. Werwatz. *Nonparametric and Semiparametric Models.* 2006. URL: https://www.researchgate.net/publication/43185083_Nonparametric_and_Semiparametric_Models.

[17] Garrick Hileman; Michel Rauchs. *Global blockchain benchmarking study.* 2017. URL: http://dx.doi.org/10.2139/ssrn.3040224.

[18] IS0. *ISO 22739:2020(en) Blockchain and distributed ledger technologies — Vocabulary.* 2020. URL: https://www.iso.org/obp/ui/#iso:std:iso:22739:ed-1:v1:en.

[19] Julien Klepatch. *ERC20 & ERC721: How Do They Work?* 2019. URL: https://blog.coinmarketcap.com/2019/02/25/erc20-erc721-how-do-they-work/.

[20] Paul Kohlhaas. *Token Bonding Curve Design Parameters.* 2019. URL: https://medium.com/molecule-blog/token-bonding-curve-design-parameters-95d365cbec4f.

[21] Patrick Laurent; Thibaut Chollet; Michael Burke; Tobias Seers. "The tokenization of assets is disrupting the financial industry. Are you ready?" In: *Inside magazine issue 19* (2018). URL: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/financial-services/lu-tokenization-of-assets-disrupting-financial-industry.pdf.

[22] Jose Maria Macedo. *Work tokens as a breakthrough in financial instrument design.* 2019. URL: https://medium.com/amazix/work-tokens-as-a-breakthrough-in-financial-instrument-design-e7d3ad805ee6.

[23] Rudolf Marcin. *Economics of Entangled Tokens.* 2017. URL: https://blog.neufund.org/economics-of-entangled-tokens-9fc5b084e2d2.

[24] Amani Moin; Emin Gün Sirer; Kevin Sekniqi. "A Classification Framework for Stablecoin Designs." In: *Eprint arXiv preprint arXiv:1910.10098* (2019). URL: https://arxiv.org/pdf/1910.10098.pdf.

[25] William Mougayar. *Tokenomics — A Business Guide to Token Usage, Utility and Value.* 2017. URL: https://medium.com/@wmougayar/tokenomics-a-business-guide-to-token-usage-utility-and-value-b19242053416.

[26] Philip Nadler; Yike Guo. "The fair value of a token: How do markets price cryptocurrencies?" In: *Research in International Business and Finance, Volume 52, April 2020, 101108* (2019).

[27] Satoshi Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System.* 2008.

[28] NIST National Institute of Standards & Technology. *Blockchain Networks: Token Design and Management Overview.* 2020. URL: https://doi.org/10.6028/NIST.IR.8301.

[29] OECD. *The Tokenisation of Assets and Potential Implications for Financial Markets, OECD Blockchain Policy Series.* 2020. URL: http://www.oecd.org/finance/The-Tokenisation-of-Assets-and-Potential-Implications-for-Financial-Markets.htm.

[30] Luis Oliveira; Zavolokina Liudmila; Bauer Ingrid; Schwabe Gerhard. "To Token or not to Token: Tools for Understanding Blockchain Tokens". In: *International Conference of Information Systems (ICIS 2018), San Francisco, USA, 12 December 2018 - 16 December 2018.* (2018).

[31] Joseph Poon; Thaddeus Dryja. *The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments.* 2016.

[32] Philipp Sandner. *Will Blockchain Replace Clearinghouses? A Case Of DVP Post-Trade Settlement.* 2020. URL: https://www.forbes.com/sites/philippsandner/2020/12/02/will-blockchain-replace-clearinghouses-a-case-of-dvp-post-trade-settlement/.

[33] S.J. Sheater; M.C. Jones. "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." In: *Journal of the Royal Statistical Society. Series B. Methodological 53, 683–690.* (1991).

[34] Ritei Shibata. "Asymptotic mean efficiency of a selection of regression variables." In: *Annals of the Institute of Statistical Mathematics volume 35, pages 415–423* (1982).

[35] Amritraj Singh; Kelly Click; Reza M. Parizi; Qi Zhang; Ali Dehghantanha; Kim-Kwang Raymond Choo. "Sidechain technologies in blockchain networks: An examination and state-of-the-art review." In: *Journal of Network and Computer Applications* (2020).

[36] SYGNA. *MiCA: A Guide to the EU's Proposed Markets in Crypto-Assets Regulation.* 2020. URL: https://www.sygna.io/blog/what-is-mica-markets-in-crypto-assets-eu-regulation-guide/.

[37] Simon Trimborn; Wolfgang Karl Härdle. "CRIX an Index for cryptocurrencies." In: *Journal of Empirical Finance* (2018).

[38] Shermin Voshmgir. *Token ecomony: How the Web3 reinvents the Internet.* BlockchainHub Berlin, 2020.

[39] Petro Wallace. *Layer 1 vs Layer 2 : What you need to know about different Blockchain Layer solutions.* URL: https://medium.com/the-capital/layer-1-vs-layer-2-what-you-need-to-know-about-different-blockchain-layer-solutions-69f91904ce40.

[40] Wikipedia. *Bitcoin.* URL: https://en.wikipedia.org/wiki/Bitcoin#cite_note-19.

[41] Wikipedia. *Ethereum.* URL: https://en.wikipedia.org/wiki/Ethereum.

[42] A.Z. Zambom; R. Dias. *A Review of Kernel Density Estimation with Applications to Econometrics.* 2012. URL: https://arxiv.org/abs/1212.2812v1.

[43] Shijie Zhang; Jong-Hyouk Lee. "Analysis of the main consensus protocols of blockchain." In: *ICT Express Volume 6, Issue 2, June 2020, Pages 93-97* (2019).

Appendix A

Table of acronyms

ACRONYM	MEANING
AIC	Akaike information criterion
AMISE	Asymptotic mean integrated squared error
AML	Anti-money laundering
BTC	Bitcoin
CC	Cryptocurrency
CFT	Countering the financing of terrorism
CRIX	Cryptocurrencies Index
DLT	Distributed Ledger Technology
ECRIX	Exact cryptocurrencies index
EFCRIX	Exact full cryptocurrencies index
ERC	Ethereum request for comments
ETH	Ethereum
ICO	Initial coin offering
IPO	Initial public offering
ITO	Initial token offering
LCRIX	Liquidity cryptocurrencies Index
MiCAR	Market in crypto-assets regulation
MISE	Mean integrated squared error
MSE	Mean squared error
MDA	Mean directional accuracy
NFT	Non-fungible token
PDF	Probability density function
PoW	Proof of Work
PoS	Proof of Stake
P2P	Peer to peer
SHA-256	Secure hash algorithm 256
TM	Total market
TMI	Total market index
UTXO	Unspent transaction output

Table A.1. Table of some selected acronyms.