THE TREE OF BLOCKCHAIN

A PREPRINT

Florian Spychiger[*]
School of Management and Law
Zurich University of Applied Sciences
Winterthur, Switzerland
spyc@zhaw.ch

Paolo Tasca
Centre for Blockchain Technologies
University College London
London, Great Britain
p.tasca@ucl.ac.uk

Claudio J. Tessone [†]
Business Administration
University of Zurich
Zurich, Switzerland
claudio.tessone@business.uzh.ch

October 7, 2020

ABSTRACT

This study covers the evolutionary development of blockchain technologies over the last 11 years (2009 – 2019) and sheds lights on potential areas of innovation in heretofore unexplored sub-components. For this purpose, we collected and analysed detailed data on 107 different blockchain technologies and studied their component-wise technological evolution. The diversity of their designs was captured by deconstructing the blockchains using the Tasca-Tessone taxonomy (2019) to build what we call the "*tree of blockchain*" composed of blockchain main and sub-components. With the support of information theory and phylogenetics, we found that most design explorations have been conducted within the components in the areas of consensus mechanisms and cryptographic primitives. We also show that some sub-components like *Consensus Immutability and Failure Tolerance*, *Access and Control layer* and *Access Supply Management* have predictive power over other sub-components. We finally found that few dominant design models - the *genetic* driving clusters of Bitcoin, Ethereum and XRP - influenced the evolutionary paths of most of the succeeding blockchains.

***K*eywords** blockchain, distributed ledger technology, taxonomy, information theory, blockchain analytics, innovation, evolution

1 Introduction

Blockchains[3] are composed of a variety of multiple components that ultimately characterise them. The landmark paper of Sathoshi Nakamoto (Nakamoto, 2008) introducing a peer-to-peer electronic cash system – namely Bitcoin – was

[*]Management Associate, UZH Blockchain Center and URRP Social Networks, University of Zurich

[†]Academic Director, UZH Blockchain Center and URRP Social Networks, University of Zurich

[3]In this article we use the term "blockchain technologies" to refer also to the larger family of distributed ledger technologies, i.e., community consensus-based distributed ledgers where the storage of data is not based on chains of blocks.

the starting point for a broad range of blockchain technologies we evince today. The innovative aspect of Bitcoin was an unprecedented combination of pre-existing components. Indeed, blockchains mix technologies and concepts such as triple-entry accounting (Ibañez et al., 2020), cryptographic signatures or consensus algorithms. As such, it is natural to categorise all blockchain technologies based on the specific selection of these components. As of this writing, a multitude of different blockchain technologies exist, therefore researchers have proposed several approaches to classify them in taxonomies (Sarkintudu et al., 2018; Xu et al., 2017; Tasca and Tessone, 2019; Ballandies et al., 2018). While these taxonomies take into account some important components (in some cases with some level of overlap), the most comprehensive approach is given by Tasca and Tessone (2019). Their fine-grained taxonomy can be seen as an overarching framework including many components proposed by other researchers. Its rich specifications allow for a detailed data analysis. Therefore, throughout this paper, we will rely on the Tasca-Tessone (TT) Blockchain Taxonomy.

So far, research has mainly focused on the creation of these classification schemes. However, taxonomies turn into useful instruments only when data of real-world applications is collected and the taxonomy applied on them. This is why we apply the TT Blockchain Taxonomy on a comprehensive dataset of 107 different blockchains and use appropriate methodological tools to unveil how they have evolved over the last 11 years, from 2009 to 2019. Our approach allows us to compare the instantiations of the technology and examine the relative innovation within different blockchain components. The insights derived from our analysis are specifically important given the current variety of blockchain architectures, which - in turn - is a direct consequence of the different technological innovation paths followed by their individual components.

By applying the TT Blockchain Taxonomy, we observe that most of the current blockchains did not stop at the level of continuous combination of previous concepts and technologies; instead, novel components have appeared over time to generate new characteristics and functionalities. Drawing inspiration from the field of biology, taxonomies cannot be just used to classify organisms – akin to technologies, in this context –, but also to explore the evolutionary dynamics that has led to their emergence. For example, the introduction of FORTRAN as first high-level general purpose programming language has fueled not only the invention of a multitude of new programming languages, but also the development of new hardware and software (Zimmermann, 2017). Similar effects may be observable in blockchains: Innovation in a component may cause the emergence of new layouts in other components, thereby creating new classes of blockchain technologies, a problem that we address in this paper from a quantitative point of view. Based on these insights, we create the *tree of blockchain* to shed light on the innovation within different components. This allows us to answer the following research questions: *Which components drive the innovation in blockchain technologies?* and *Have some components co-evolved?*

In order to tackle these questions, we utilise the TT Blockchain Taxonomy and apply it to analyse a comprehensive dataset consisting of 107 blockchain technologies. We resort on methods from information theory to measure the different levels of innovation in the components and further make use of phylogenetic methods to study the evolution of the technology.

The paper is organised as follows: In section 2, we briefly recap the taxonomy. After, in section 3, we introduce our methodology. Section 4 presents the results from the analysis. Section 5 concludes.

2 Taxonomy

The taxonomy introduced by Tasca and Tessone (2019) compartmentalises the blockchain components and establishes the relationships between them in a hierarchical manner. They adopt a reverse-engineering approach to unbundle the blockchains and divide them into *main* (coarse-grained) components. Each main component is then split into more (fine-grained) *sub*components and *sub-sub*components (where necessary). For each of these sub-components (and/or sub-sub-components), some *layouts* are identified and compared. The next eight subsections will resort on the TT

Blockchain Taxonomy and will introduce additional *layouts* for the sub-components (and possibly sub-sub-components that will be subject to our temporal evolution analysis.

2.1 Consensus

The *Consensus* component relates to the set of rules and mechanics that allow the maintenance and the update of the ledger and that guarantee the trustworthiness of the records in it, i.e., their reliability, authenticity and accuracy (Bonneau et al., 2015). It encompasses the following sub-components:

1 **Consensus Network Topology** describes the type of interconnection between the nodes and the type of information flow between them for transaction and/or for the purpose of validation.
— Layouts: *Centralised / Decentralised / Hierarchical*.

2 **Consensus Immutability and Failure Tolerance** encompasses a consensus mechanism to ensure that every node keeps its version of the full transaction history consistent with the other peers.
— Layouts: *dPoS/PoW/DAG/PoS/Hybrid/PoU/BFT/PoW, DAG/dPoW/PoI/PoET/PoA/SCP/other*.

3 **Gossiping** defines how information travels through from one node to another.
— Layouts: *Local/Global*.

4 **Consensus Agreement**

4.1 *Latency* is a sub-sub-component which describes the rule of message propagation in the networks.
— Layouts: *Synchronous / Asynchronous / Not Known*.

4.2 *Finality* describes whether information intended to be stored in a blockchain can be safely considered *perpetually* stored once the recording is performed.
— Layouts: *Deterministic / Non-Deterministic*.

2.2 Transaction Capabilities

The *Transaction Capabilities* component is important to illustrate scalability of transactions and usability in possible applications and platforms. The following list presents its sub-components:

1 **Data Structure in the Blockheader**.

2 **Transaction Model** can be imagined as an accounting ledger that tracks the transaction inputs and outputs and determines how the nodes store and update the user information in the distributed ledger.
— Layouts: *UTXO / Traditional Ledger / Tangle / Message-based*.

3 **Server Storage** can be different among nodes: those which do not store the information fully are "thin clients" connected to the peer-to-peer network (Xu et al., 2018).
— Layouts: *Full Nodes Only / Thin Nodes*.

4 **Block Storage** describes which information is stored in the blockchain.
— Layouts: *Transactional Data/User Balance / Transactional Data and User Balance*.

5 **Limits to Scalability**.

2.3 Native Currency/Tokenisation

Thus far, the financial and monetary features have been the most explored and applied blockchain properties. In particular, cryptocurrencies are generally used as incentive mechanism to encourage the participation in the verification process of the blockchain transactions. The following sub-components belong to the main component *Native Currency/Tokenisation*:

1 **Native Asset** identifies whether a blockchain runs on top of a native asset (i.e., a *digital token*[4]). Tasca (2018) classifies cryptocurrencies or crypto assets (the omni comprehensive family of digital tokens) into "native coins" and "crypto tokens". Native coins, like Bitcoin, represent an alternative asset class of electronic money universally accessible via peer-to-peer payment networks. Instead, crypto tokens are forms of "digital vouchers" that allow the token holders to get access to almost any type of service and assets: from monetary rewards, or commodities to loyalty points to even other cryptocurrencies. Native coins are digital created within a novel or "forked" off a pre-existing blockchain. A native coin "a" exists and operates on the blockchain network "A" which allows peer-to-peer (sometimes, anonymous or pseudo-anonymous) transactions of "a" between different network participants.
— Layouts: *Own Cryptocurrency / Convertible Multiple Assets / None*.

2 **Tokenisation** means the possibility of tokens acting as a digital bearer bond whose ownership is determined by the data embedded in the blockchain enabling a range of possible use cases outside the purely financial world (Tsukerman, 2015; Adhami et al., 2018; Conley, 2017). There are four main token classes: (1) *Payment tokens* which are used as a means of payment for acquiring goods or services or as a means of money or value transfer; (2) *Utility tokens* intended to be the only way to provide digital access to applications and/or services (generally) built on the top of blockchain-based infrastructures; (3) *Asset/Debt tokens* that have a similar role as a share and for the investor they represent assets such as a debt or equity security owned; (4) *Hybrid tokens* which are characterised by a mixture of the previous three features. See Tasca (2018).
— Layouts: *Tokenisation present / Tokenisation through third-party addons / No tokenisation*.

3 **Asset Supply Management** means the digital asset creation often being a pillar of the incentive scheme that users have to participate in (or not) as part of the validation process (Tessone and Garcia, 2018).
— Layouts: *Limited-Deterministic / Unlimited-Determininistic / Non-Deterministic / Pre-Mined*.

2.4 Extensibility

The future ecosystem of the blockchain network and the integration possibilities of variety of blockchain related technologies is determined by the following sub-components forming the component *Extensibility*:

1 **Interoperability** illustrates the overall capability of blockchains to exchange information with other systems, outside of blockchains.
— Layouts: *Explicit Interoperability / Implicit Interoperability / None*.

2 **Intraoperability** illustrates the overall capability of blockchains to exchange information with other blockchains.
— Layouts: *Explicit Intraoperability / Implicit Intraoperability / None*.

3 **Governance** rules are crucial for the successful implementation of the blockchains and for their capability to adapt, change and interact.
— Layouts: *Open-source Community / Alliance/Technical Leading House*.

4 **Script Language** describes the flexibility of the scripting language to modify the conditions under which certain information (e.g. transactions) will be included into the public record (smart contracts).
— Layouts: *Turing Complete / Generic Non-Turing Complete / Application-specific Non-Turing Complete / Non-Turing Complete + External*.

[4]Digital tokens can either be fungible or non-fungible.

2.5 Security and Privacy

Security and privacy principles apply to any ICT system containing or processing PII, including blockchain systems. The *Security and Privacy* component consists of the following sub-components:

1 **Data Encryption**

- *Hashing* is used all over in blockchain technologies, e.g. for chaining blocks together, in the consensus mechanism and in address generation.
  — Layouts: *Equihash / SHA3 / SHA2 / SHA2 + RIPEMD160 / Scrypt, CryptoNight, SHA3 + BLAKE / BLAKE / X11 / SHA256 + RIPEMD160 / Groestl / Kerl / CryptoNight + SHA3 / SHA3 + Skein / SHA2 + Scrypt / SHA2 + BLAKE / Combination*.
- A *Signature* is necessary for participants of blockchain systems to authorise transactions.
  — Layouts: *Ed25519 / ECDSA / ECDSA + Ed25519 / Schnorr / BLS, W-OTS, RingCT / EC-KCDSA / ECDH / Redjubjub / Combination*.

2 **Data Privacy** involves several alternative solutions to balance the trade-off between a decentralised peer-validate system and the security and privacy of information.
— Layouts: *Built-in Data Privacy / Add-on Data Privacy / Data Privacy by Third Party Systems / No Data Privacy*.

2.6 Codebase

The codebase delivers information about the challenges developers could face and about possible changes of the underlying programming language. *Codebase* is structured in three sub-components:

1 **Coding Language**.

2 **Code License** illustrates the possibility of changes to the source code of the underlying technology.
— Layouts: *Open Source / Closed Source*.

3 **Software Architecture**.

2.7 Identity Management

The component *Identity Management* ensures secure access to sensitive data to establish a suitable governance model for the blockchain. It consists of two sub-components:

1 **Access and Control Layer** refers to Blockchains having different permissions according to which access and control to data is allowed.
— Layouts: *Public Blockchain / Permissioned Private Blockchain / Permissioned Public Blockchain*.

2 **Identity Layer** describes the fact that the on-boarding and off-boarding of nodes / entities to the blockchain networks are handled differently by the various software solutions.
— Layouts: *Anonymous / Pseudonymous / KYC/AML.*

2.8 Charging and Rewarding System

Blockchain systems incur operational and maintenance costs that are generally absorbed by the network participants. The *Charging and Rewarding System* main component is structured in:

1 **Reward System** which illustrates the rewarding mechanisms designed to compensate active members contributing to data storage or transaction validation and verification.
— Layouts: *Lump-Sum Reward / Block + Security Reward.*

2 **Fee System**:

2.1 *Fee Reward* describes the kind of rewards provided directly by the users to other participants for any request in the network for storage, data retrieval, or computation and validation.
— Layouts: *Optional Fees / Mandatory Fees / No Fees*.

2.2 **Fee Structure** describes the nature of the fees that users are required to contribute when using a blockchain.
— Layouts: *Variable Fees / Fixed Fees*.

3 Methodology

3.1 Data

The dataset includes 107 technologies (cf. Table 2 in the Appendix). The sample contains a variegated sample of blockchain technologies introduced in the period 2009 – 2019. Each technology data set contains 25 sub-components (or sub-sub-components), where we could find 84.34% of the overall data. For a detailed description of the dateset and the sub-components (sub-sub-components), we refer to Figure 9 in the Appendix.

The data collection was crowd-sourced, and each technology was randomly assigned to students from the University of Zurich, Zurich University of Applied Sciences and École Polytechnique Fédérale de Lausanne. To cross-check the results, some of the technologies were assigned more than once. Eventually, the quality and correctness of the whole dataset was diligently checked and validated by ourselves.

3.2 Information Theoretic Analysis

In order to analyse the information contained in the data, we apply Shannon's information theory (Shannon, 1948). We calculate the entropy of each sub-component (resp. sub-sub-component) defined here with S. The entropy measures the amount of information present in the realisations of a random variable. If a high-probability event occurs, little is leart about the random variable and the entropy is low. If a rare event occurs, the amount of information (surprisal) is high. In Biology, researchers call the entropy Shannon-Index and use it to measure biodiversity (Spellerberg and Fedor, 2003). Instead of probabilities, they use the relative frequency of a species. Similarly, if we calculate the entropy using the relative frequency of the sub-components' realised layouts in our sample, we can measure innovation. When a new layout emerges, the entropy of the sub-component will increase, since new information is conveyed. For a sub-component S with n realised layouts $x \in X$, the entropy is defined as

$$H(S) = -\sum_{x \in X} p(x) \log_2 \left(p(x) \right) \tag{1}$$

where $p(x)$ is the probability mass of layout x. We normalise the entropy by dividing it through the maximum entropy $\log_2(n)$.

We also calculate the mutual information between the sub-components to measure how they are related to each other. The mutual information measures the amount of information about a variable contained in another. It is a more general measure than correlation capturing also non-linear dependencies. In our specific context, it is able to determine whether two layouts from different components tend to occur jointly (or also in an anti-correlated fashion) in blockchain systems.

For sub-components S_1 with n layouts $x \in X$ and S_2 with m layouts $y \in Y$, the mutual information is given by

$$I(S_1, S_2) = \sum_{x \in X} \sum_{x \in X} p_{X,Y}(x, y) \log_2 \left(\frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} \right). \tag{2}$$

We further use the normalised version given by

$$MI(S_1, S_2) = \frac{2 \times I(S_1, S_2)}{H(S_1) + H(S_2)}. \tag{3}$$

3.3 Temporal Evolution

Our analysis of the temporal evolution of blockchain technologies borrows methods from pyhologenetics: a branch of biology that studies the evolutionary relationships between individuals or group of organisms. We construct the *tree of blockchain* with the R-packages *metacoder* (Foster et al., 2017) and *taxa* (Zachary et al., 2018) used in the evolutionary analysis of microbiota – microorganisms hosted by humans, animals and plant. Similarly, blockchain technologies also "host" (rather consist of) several (micro-) components. As a consequence, these frameworks are suitable to visualise the blockchain components and their layouts. Another pivotal tool to show the formation of species already used by Charles Darwin (Darwin, 1859) is an evolutionary tree. There exists several types of evolutionary trees. We make use of a chronogram tree and a dendrogram. For constructing the chronogram, we derive the ancestors of a blockchain technology and the timing of branching from the data. While some blockchain technologies are novel inventions created from scratch, many others have "forked" off pre-existing blockchain architectures. Taking this fact into consideration, a chronogram tree where even the internal taxonomic units (nodes) can be annotated can easily be recovered and plotted with the R-packages *treeio* (Wang et al., 2019) and *ggtree* (Yu et al., 2017). For the construction of the dendrogram, we take the genetic similarities into account. We construct a hierarchial clustering dendrogram from the data. In the dataset, each row represents a blockchain technology and each column a sub-component. From this, we calculate a dissimilarity matrix of genetic distances. As we have nominal variables, we use the algorithm of Gower (1971). The dissimilarity d_{ij} between two rows i and j is calculated as follows:

$$d_{ij} = \frac{\sum_{s=1}^{S} \delta_{ij}^s d_{ij}^s}{\sum_{s=1}^{S} \delta_{ij}^s} \tag{4}$$

where δ_{ij}^s is 0 or 1, and only 0 if either one or both layouts in rows i or j are missing. The dissimilarity contribution d_{ij}^s is 1 if the layouts of the two rows are different, otherwise 0. The resulting dissimilarity matrix with the entries $d_{ij} \in [0, 1]$ can be used to construct a dendrogram. We use the UPGMA (unweighted pair group method with arithmetic mean) algorithm – a simple, yet effective hierarchical clustering method. Starting with the $N \times N$ dissimilarity matrix, we combine the two nearest blockchain technologies into a new high-level cluster. Afterwards, we eliminate the two corresponding rows in the dissimilarity matrix and add a new row corresponding to the newly formed cluster. The new dissimilarities between the new cluster and the other blockchain technologies are calculated as the proportional averages of the two eliminated dissimilarities rendering a $(N-1) \times (N-1)$ matrix. These steps are repeated until when we remain with a single cluster – the root of the dendrogram. In the following, we illustrate the procedure in a simple example with four elements. After three steps, we arrive at the dendrogram shown in Fig. 1.

1. Step

$$D_1 = \begin{pmatrix} 0.00 & 0.50 & 0.75 & 0.25 \\ 0.50 & 0.00 & 1.00 & 0.50 \\ 0.75 & 1.00 & 0.00 & 1.00 \\ 0.25 & 0.50 & 1.00 & 0.00 \end{pmatrix} \implies D_2 = \begin{pmatrix} 0.00 & 1.00 & 0.50 \\ 1.00 & 0.00 & 0.875 \\ 0.50 & 0.875 & 0.00 \end{pmatrix} \quad H_1 = 0.25$$

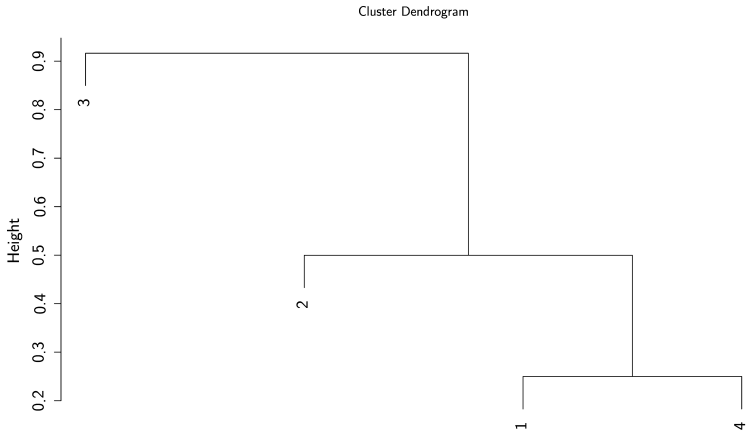


Figure 1: Example of a dendrogram. Elements 1 and 4 are clustered together, followed by 2 and then 3.

2. Step

$$D_2 = \begin{pmatrix} 0.00 & 1.00 & 0.50 \\ 1.00 & 0.00 & 0.875 \\ 0.50 & 0.875 & 0.00 \end{pmatrix} \implies D_3 = \begin{pmatrix} 0.00 & 0.91\overline{6} \\ 0.91\overline{6} & 0.00 \end{pmatrix} \quad H_2 = 0.5$$

3. Step

$$D_3 = \begin{pmatrix} 0.00 & 0.91\overline{6} \\ 0.91\overline{6} & 0.00 \end{pmatrix} \implies H_3 = 0.91\overline{6}$$

4 Results

4.1 Innovation Dynamics

Blockchain technologies have undergone an extensive innovation during the last few years, but not all their components have benefited from the same rates of innovation. Fig. 2 shows the *tree of blockchain*. The tree nicely illustrates how some sub-components seem to follow quite stable designs. For example, not many technologies have experimented with the network topology, the latency or the codebase sub-components. By contrast, other sub-components are in an exploratory state. In particular, many innovation have been carried out for the *immutability and failure tolerance* sub-component. Some technologies have also innovated on the cryptographic building blocks, even though SHA-2 (resp. SHA-3) based hashing and ellipic curve digital signature algorithms are still the most used schemes. On the level of the components, it is not clear where the most innovation has happened as the dynamics in the sub-components seem quite heterogenuous.

The technological innovation of blockchains is mainly driven by consensus-, security- and supply-related sub-components while some basic principles have remained unchanged. The entropy – and thereby the surprisal effect – of the sub-components is shown in Fig. 3. The highest innovation activity took place within the *immutability and failure tolerance* sub-component followed by the hashing algorithm. Many cryptocurrencies have also experimented with the total monetary supply, even though there is usually no clear economic foundation behind these monetary policies. The high entropy of the asset supply mechanism sub-component indicates that there is not yet a preferred solution. Similarly, the consensus mechanism is an active sub-component where innovation is still ongoing (Cachin and Vukolić, 2017; Mingxiao et al., 2017). The original ideas of the proof-of-work algorithm have been adjusted and many new layouts such as proof-of-stake, proof-of-elapsed-time or byzantine-fault tolerance have been applied in

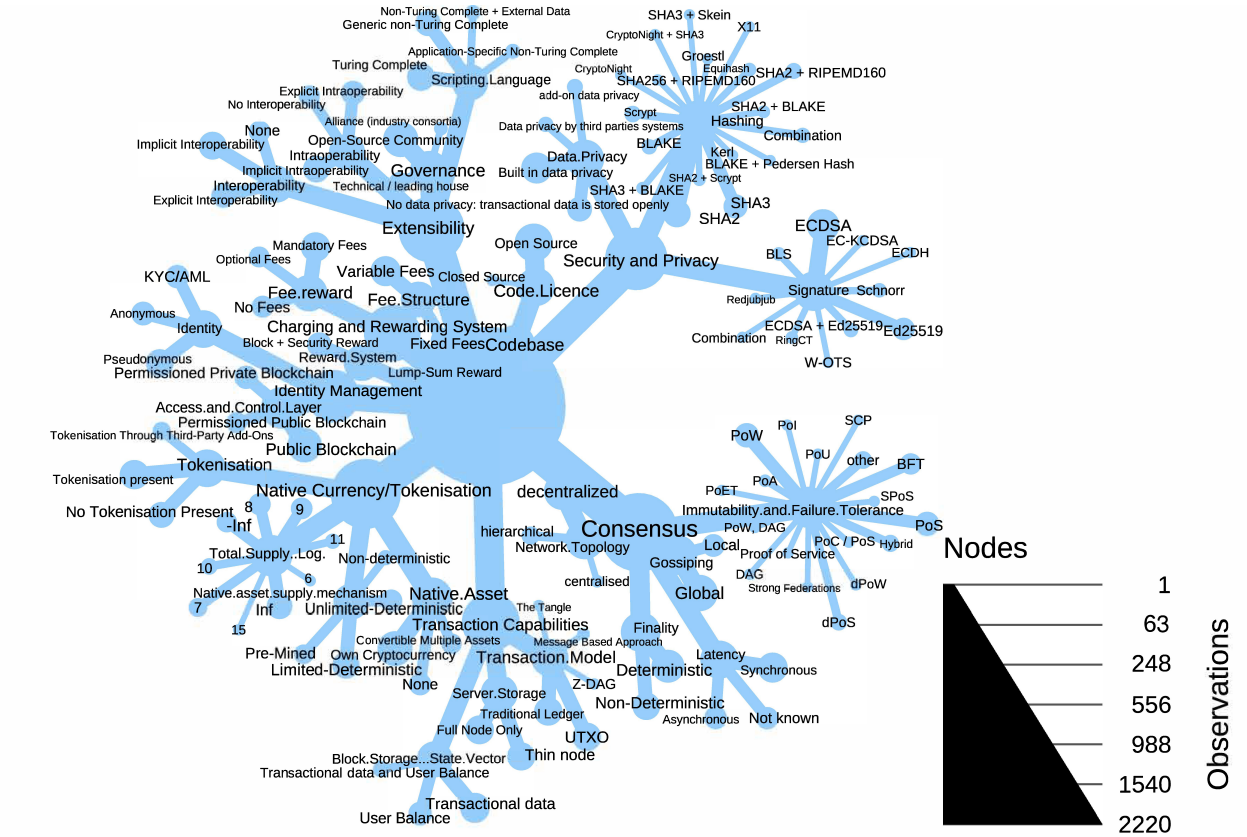


Figure 2: The *tree of blockchain* illustrates the varying innovation activity within the components. The nodes represent the observations in the (sub)components and the links the connection to the sub-components resp. layouts.

blockchains. Most other sub-components have experienced only moderate innovation: This suggests that many of the basic design choices of the original Bitcoin architecture have been inherited. In fact, some of the most important blockchain layouts, such as the *decentralised* network topology and the *open source* code license, have barely been challenged by alternative solutions. Going back to the asset supply mechanism, we can observe from Fig. 4 that from 2009 to 2012 the *limited-deterministic* supply layout was the only one. However, after 2012 other layouts (pre-minded, non-deterministic, unlimited-deterministic) started to become popular pushing to a higher entropy towards the end of 2019. Also the *immutability and failure tolerance* sub-component followed an innovation path similar to the asset supply mechanism. It started in 2009 with a single layout (proof-of-work) and soon after alternative layouts (e.g., proof-of-stake, DAG, etc) did evolve. The consensus is a central part of each blockchain system, and the current high entropy suggests that there is no dominant design yet – even though proof-of-work is still the most used algorithm. See Fig. 4.

The design of a sub-component contains information about the design of the other sub-components. This implies that certain design choices might emerge jointly. To measure this mutual dependency, we analyse the mutual information of the sub-components. In Fig. 5 the sub-components are sorted by the sum of mutual information they share with other sub-components. This gives us a hint about the predictive power of a sub-component in a blockchain system. Again, the consensus, the security and the total supply contain a lot of information on other sub-components. This means that if we know the layouts of these information-carrying sub-components, we are able to infer the design of other sub-components. Similarly, the access and control layer holds a lot of information about the other sub-components. This comes not much as a surprise because there exist important fundamental differences between *public* and *permissioned* blockchains. In general, *permissioned* blockchains do not have native assets and as such also use different consensus

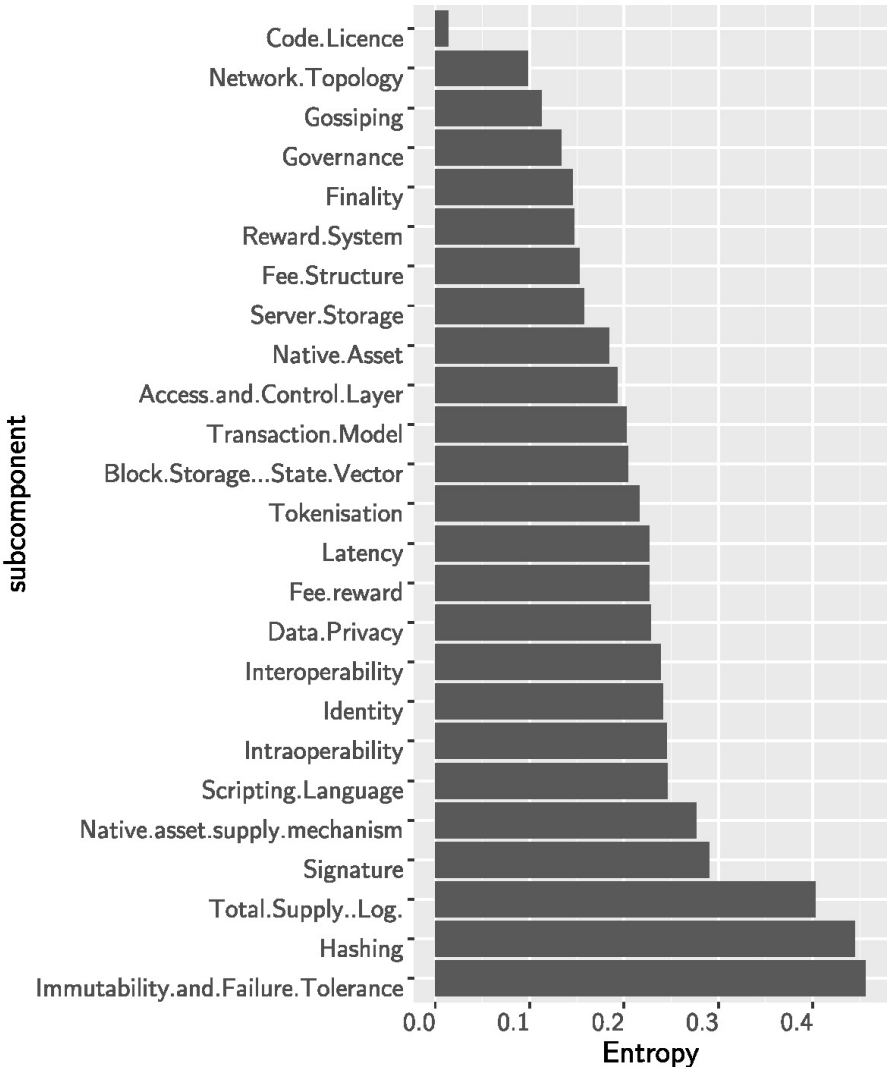


Figure 3: The entropy for the (sub)components: *immutability and failure tolerance*, *security* and *total supply* show the largest entropy. Within these two, many new designs have emerged along the sample period.

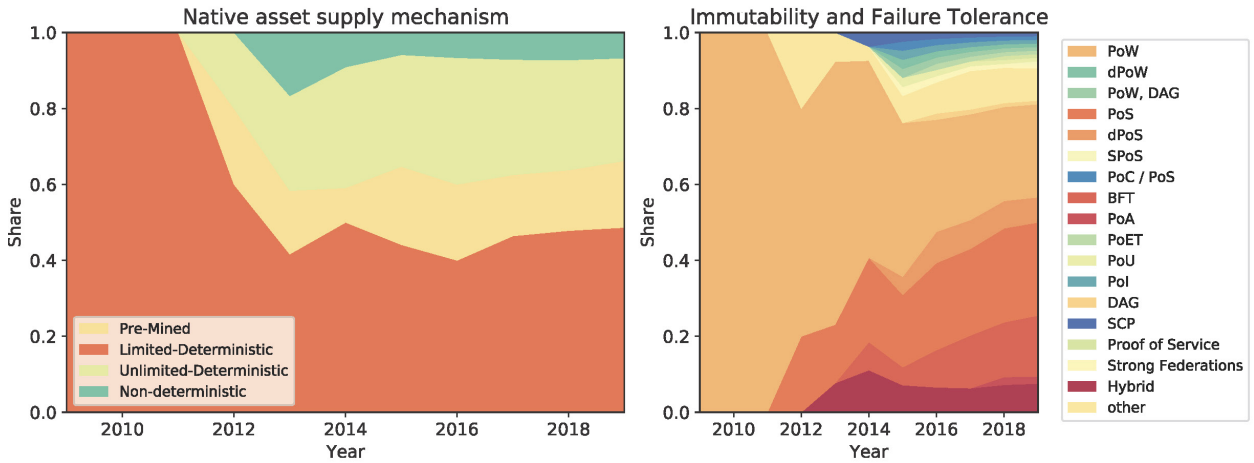


Figure 4: New supply and consensus mechanisms have emerged over time.

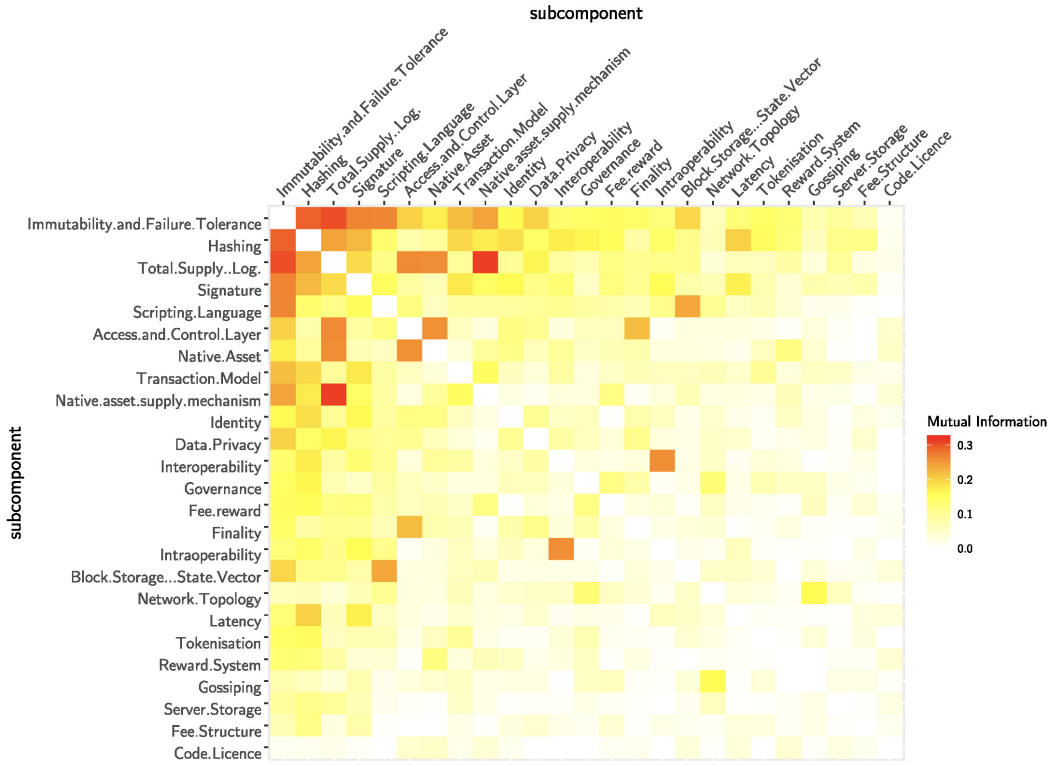


Figure 5: The normalised mutual information measures the dependencies between the sub-components. The sub-components are ordered by the total amount of mutual information.

mechanisms – not based on monetary incentives – than *public* infrastructures. The low cumulative mutual information of the native asset supply mechanism (despite the high entropy) consolidates the impression that the monetary policy of blockchain technologies is not a design choice based on fundamentals. Interestingly, the scripting language seems also quite indicative for the design of other sub-components. The ability of the scripting language defines whether a blockchain is able to run smart contracts which in turn implies many specific design choices. To continue, we observe a strong dependency between the access and control layer and the native asset sub-components, mainly due to the fact that the lack of a native cryptocurrency implies a permissioned infrastructure. The strong dependency between total supply and consensus is also influenced by permissioned blockchains: a total supply of zero usually implies *BFT* consensus. But the dependency between total supply and consensus is amplified by some design choices of permissionless blockchains. For example, a supply of 21 million as in Bitcoin generally implies *PoW*. Similar trivial effects are at play between the total supply and the supply mechanism (e.g. an *infinite* supply requires a *unlimited* supply mechanism). Finally, we observe a rather strong dependency between the transaction model and the supply mechanism. We argue that this is not an obvious relationship and it deserves to be investigated in further studies. Differently, intra- and interoperability tend to be implemented together but this relationship seems to be more obvious.

As shown in Table 1, the results of our analysis are robust. Kendall's τ is calculated between the full sample and the bootstrapped versions for the ranking of the entropy and the ranking of the summed mutual information of the sub-components. Kendall's τ is high even for half the sample size and Kendall's test shows a significant dependence for all sub-sample sizes.

4.2 Evolutionary Analysis

In this Section we examine the temporal evolution of the blockchain components over the last 11 years (2009–2019).

robustness					
sub-sample size	50%	60%	70%	80%	90%
entropy Kendall's τ	0.90	0.92	0.94	0.95	0.97
mutual information Kendall's τ	0.81	0.85	0.89	0.91	0.94

Table 1: Kendall's τ between the full sample and the bootstrapped sub-samples for different sizes. Each value is the average of 100 bootstrapped sub-samples.

Fig. 6 helps us to map over time the technology life cycle of the blockchain architectures. We observe two phases. The first phase of "technological discontinuity" (2009-2013) is characterised by revolutionary breakthrough innovations: Bitcoin, Litecoin[5], XRP[6], Peercoin[7], Novacoin[8]. Bitcoin was the first-ever blockchain innovation which originated in October 2008 when the Satoshi Nakamoto whitepaper appeared in the cypherpunk mailing list. However, the genesis block was not mined until the 3rd of January 2009. As shown in the chronogram tree (Fig. 6), Bitcoin was the only implemented blockchain technology for the first few years. In the meantime, the community started to think about some alternative innovations that for a while remained at the idea level only. In 2011, two Bitcoin software forks were implemented and deployed in the market, namely, Litecoin and Namecoin[9]. But we had to wait until 2012 to see the deployment of the first-ever Bitcoin-independent blockchain technology: XRP. While Namecoin did not lead to any further developpment, both Litecoin and XRP inspired further technologies as Dogecoin[10] (Litecoin spinoff) or Stellar[11] (XRP spinoff) for example. The next large innovation wave called "era of ferment" (2014-Today), ignited by Ethereum[12], is caracterised by technological rivalry, competitions and technological uncertainty. The Ethereum smart contract concept led to many descendants, but also to the development of a wide range of independent platforms with smart contract capabilities. Only recently, alternative architectures have started to come up (IOTA[13] being the early exception) such as Tendermint[14], Byteball[15] and Hedera Hashgraph[16]. In particular, the first permissioned blockchains emerged in 2016, mainly driven by the Hyperledger[17] initiative but also Corda[18]. Interestingly, the practice of software forks does not seem common in permissioned frameworks (or at least they are not publicly communicated). An exception are the private forks of Ethereum, for example Quorum[19].

If we zoom into the taxonomy of Tasca and Tessone (2019), we could replicate the same analysis of the technology life cycle for all the blockchain sub-components. As an example, we take into consideration the sub-component *immutability and failure tolerance*. Fig. 7 helps us to map over time its technology life cycle. In particular, we can observe three phases. Also in this case, we observe a first initial phase of "technological discontinuity" (2009-2013) characterised by revolutionary breakthrough innovations: the proof-of-work deployed in January 2009, the Ripple Consensus Algorithm (RPCA) in early 2012, the proof-of-stake mechanism deployed with Peercoin in mid 2012 and the hybrid consensus of Novacoin in 2013. Differently from the previous analysis, the second phase of technological rivalry seems to be already concluded (2014-2017). This phase reached a peak in 2015 with the larger number of new

[5] https://litecoin.org/
[6] https://ripple.com/xrp/
[7] https://www.peercoin.net/
[8] http://novacoin.org/
[9] https://www.namecoin.org/
[10] https://dogecoin.com/
[11] https://www.stellar.org/
[12] https://ethereum.org/en/
[13] https://www.iota.org/
[14] https://tendermint.com/
[15] https://obyte.org/
[16] https://www.hedera.com/
[17] https://www.hyperledger.org/
[18] https://www.corda.net/
[19] https://www.goquorum.com/

consensus mechanisms brought to the market (dPOS, DAG, PoC, etc.). Since 2018 we entered the phase of "dominant design" (2018-Today) characterised by less innovation and the emergence of consensus industry standards.

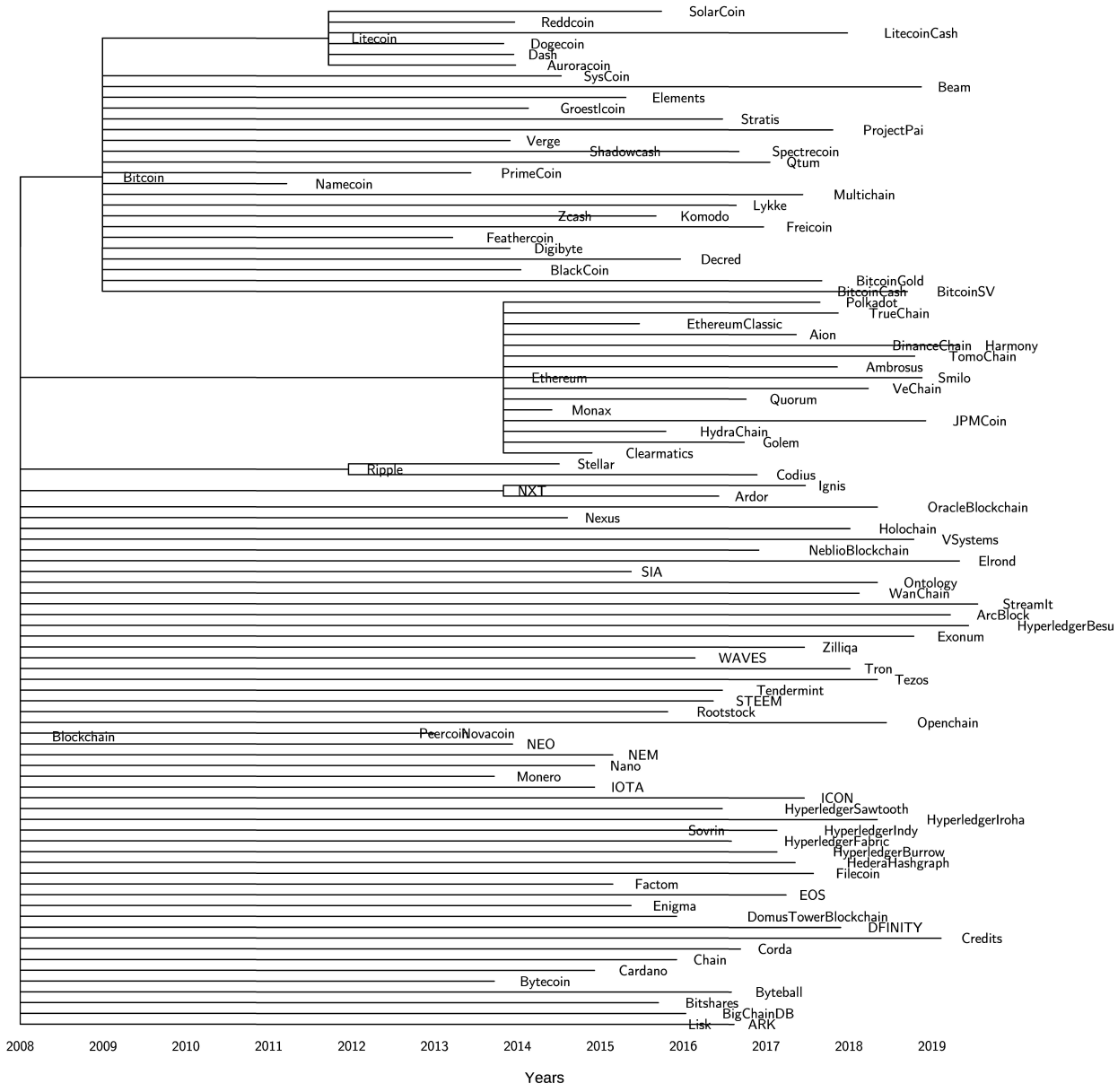


Figure 6: The chronogram tree of the blockchain technologies.

Another interesting observation we can make from our analysis is about the different evolutionary paths followed by public and permissioned blockchains (Fig. 6). Although this finding seems to be quite intuitive and linked to the different governance models that characterise the two classes of blockchains, we argue that there is a clear genetic difference between public and permissioned blockchains. The application of the hierarchical clustering algorithm (UPGMA) yields the dendrogram shown in Fig. 8. Starting from the root (top of the figure), the tree branches into two main clusters (blue and green). The green cluster on the right primarily includes permissioned blockchains. It is obvious that the different Hyperledger frameworks (grey) are genetically very closed to each other. Their distance (as indicated by the height on the y-axis) is very low. This cluster is again part of a larger cluster (red) consisting of

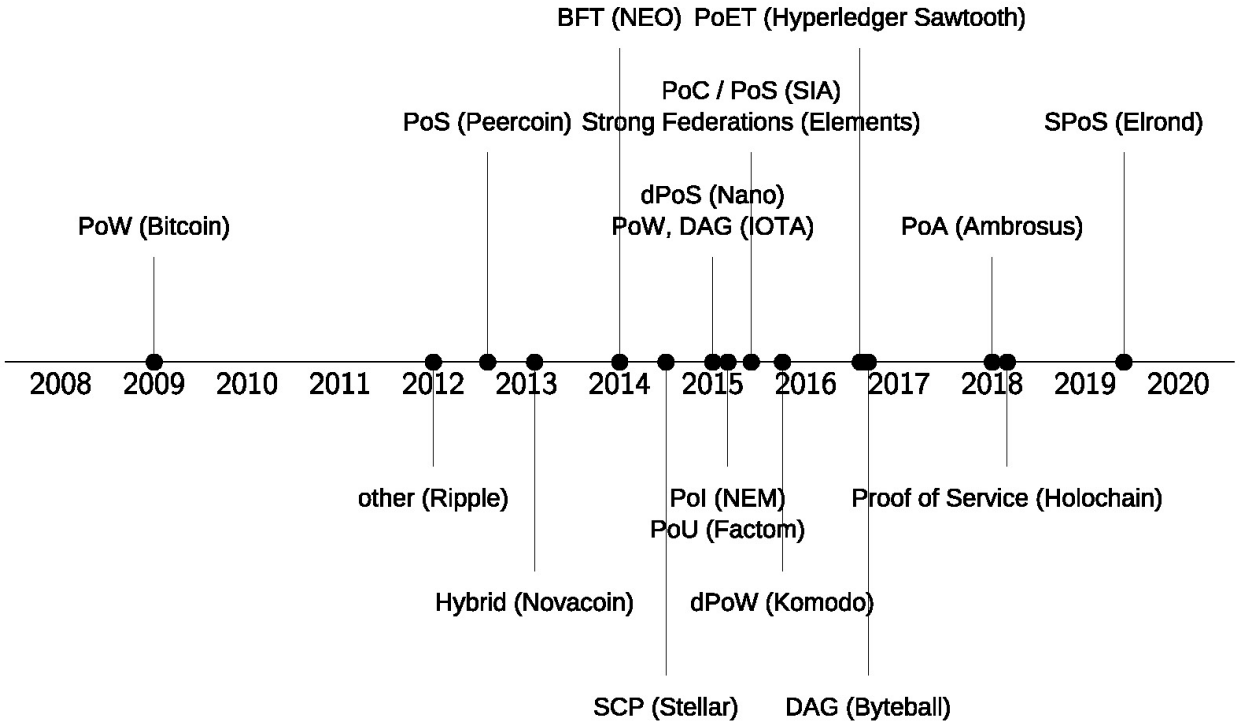


Figure 7: The emergence of new consensus layouts over time.

almost all permissioned technologies. Of particular interest is the very yellow cluster on the right which is mainly composed of public blockchains. That cluster contains technologies such as Stellar and Filecoin[20] which exhibit both features of public and permissioned blockchains. Enigma[21], however, being a second-layer technology occupies an isolated space in the right cluster. In the large blue cluster on the left, we can identify a dense subcluster (violet) around Bitcoin containing both Bitcoin Gold[22] and Bitcoin Cash[23] (forks of the original Bitcoin protocol). Many of the early cryptocurrencies are within or close to this subcluster, whereas more recent technologies such as Tron[24], IOTA, VeChain[25] and EOS[26] are further away indicating the adoption of breakthrough features developed within these new technologies.

5 Conclusions

Since the introduction of Bitcoin in 2009, we have witnessed a Cambrian explosion of blockchain architectures. This expansion combined with the fact that blockchain design allows for many degrees of freedom, makes it difficult to both understand the blockchain innovation path(s) and to have an early detection of emergent technological patterns.

In this paper, tackled this problem by using the taxonomy of Tasca and Tessone (2019) to explore innovation patterns within blockchain sub-components. We have demonstrated the usefulness of this approach by applying the taxonomy

[20] https://filecoin.io/
[21] https://www.enigma.co/
[22] https://bitcoingold.org/
[23] https://www.bitcoincash.org/
[24] https://tron.network/
[25] https://www.vechain.org/
[26] https://eos.io/

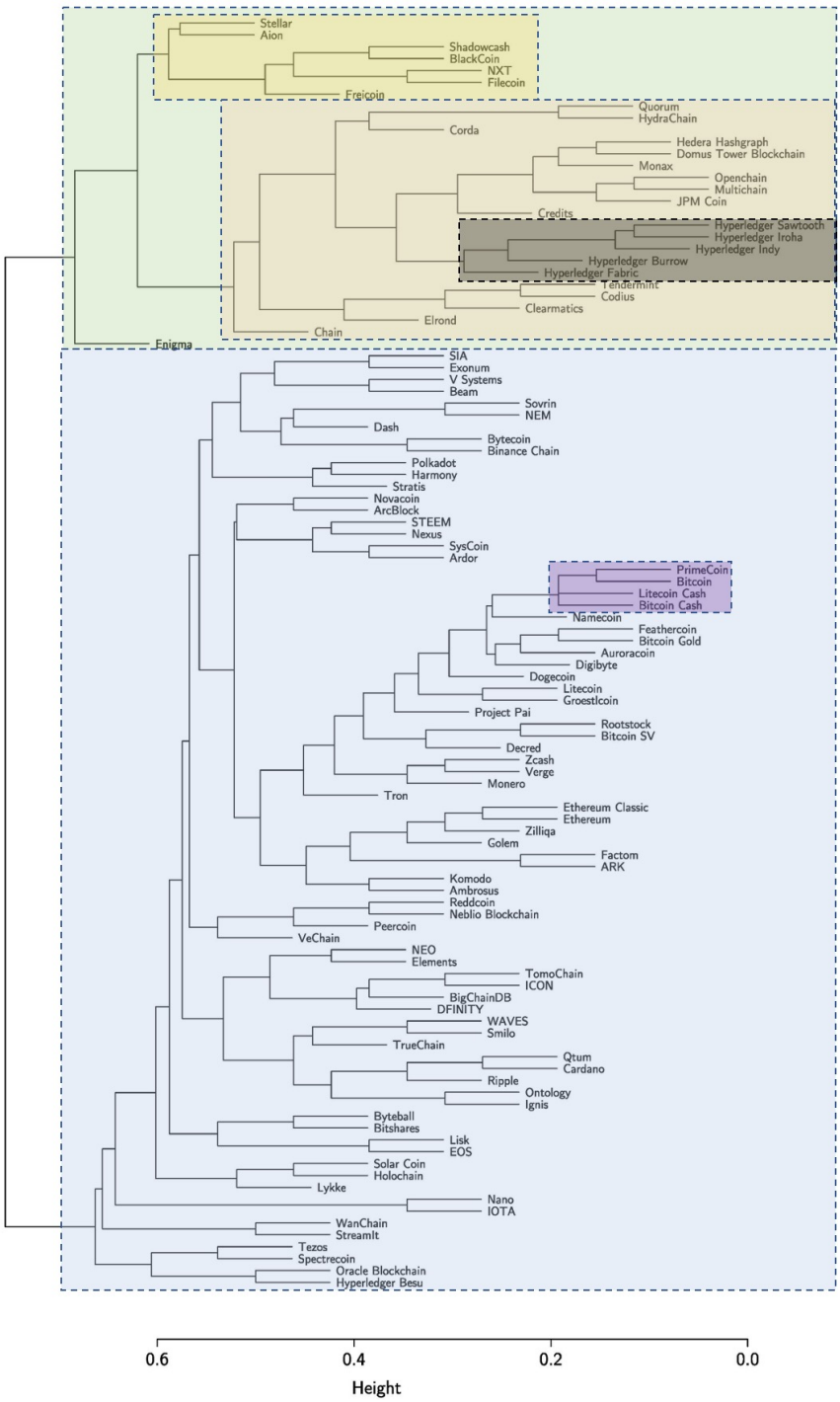


Figure 8: The dendograms shows several subclusters of blockchain technologies.

to a sample of 107 blockchain technologies. Our results provides a unique, comprehensive understanding of the (r)evolutionary and incremental changes that these technologies undertook over the last 11 years (2009–2019) and explores the connection between different design choices.

By analysing the dependencies between the sub-components with methods from information theory and phylogenetics, we find that the *consensus* mechanism, the *security* and the *asset supply* components explain most of of the variability of blockchain technologies. Interestingly, these components tend to induce certain layout choices in other sub-components. Furthermore, the *access and control* layer also has some predictive power with respect to the other sub-components. The chronogram analysis shows that the approach we have taken identifies differentiated clusters of blockchain technologies centred around Bitcoin (and a sub-cluster around Litecoin), Ethereum, XRP, while other technologies have departed much more from previous ones.

Further, our study sheds light on the architectural divergence between *public* and *permissioned* blockchains. This reflects the different field of applications for which these technologies have been designed. Even within the *public* and *permissioned* clusters, several genetic subgroups of blockchain technology have emerged, e.g. a cluster that is directly related to Bitcoin and a cluster consisting of the Hyperledger family.

Since blockchain is still undergoing its “era of ferment" (see Section 4.2) our work here lays the foundation for a continuous observation of the technological development of the platforms. We plan to continue this project and make the results available on a public webpage, where we also plan to augment our sample with additional technologies and to update the current ones. This should contribute to a better understanding of the design choices taken in blockchain technologies and at the same time inspire researchers and developers to experiment with sub-components that until now have remained technologically under-developed.

Acknowledgments

We thank all students from UZH and ZHAW who helped in the data collection used in this paperr. The authors thank Jiahua (Java) Xu and Gaspard Peduzzi for their majestic coordination of the work run by the Master students at the École Polytechnique Fédérale de Lausanne. CJT acknowledges financial support of the University of Zurich through the University Research Priority Programme (URPP) Social Networks.

Author Contributions

The idea for this paper was conceived by all three authors, FS conducted the analysis, wrote the bulk of the text and developed together with CJT the methodology, all authors commented, polished and agreed on the final manuscript.

Data Availability Statement

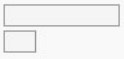
The datasets generated for this study can be found in the `https://theblockchaintree.com/`.

A Data Sample

Blockchain Technologies		
Aion	Factom	Ontology
Ambrosus	Feathercoin	Openchain
ArcBlock	Filecoin	Oracle Blockchain
Ardor	Freicoin	Peercoin
ARK	Golem	Polkadot
Auroracoin	Groestlcoin	PrimeCoin
Beam	Harmony	Project Pai
BigChainDB	Hedera Hashgraph	Qtum
Binance Chain	Holochain	Quorum
Bitcoin	HydraChain	Reddcoin
Bitcoin Cash	Hyperledger Besu	Ripple
Bitcoin Gold	Hyperledger Burrow	Rootstock
Bitcoin SV	Hyperledger Fabric	Shadowcash
Bitshares	Hyperledger Indy	SIA
BlackCoin	Hyperledger Iroha	Smilo
Byteball	Hyperledger Sawtooth	Solar Coin
Bytecoin	ICON	Sovrin
Cardano	Ignis	Spectrecoin
Chain	IOTA	STEEM
Clearmatics	JPM Coin	Stellar
Codius	Komodo	Stratis
Corda	Lisk	StreamIt
Credits	Litecoin	SysCoin
Dash	Litecoin Cash	Tendermint
Decred	Lykke	Tezos
DFINITY	Monax	TomoChain
Digibyte	Monero	Tron
Dogecoin	Multichain	TrueChain
Domus Tower Blockchain	Namecoin	V Systems
Elements	Nano	VeChain
Elrond	Neblio Blockchain	Verge
Enigma	NEM	WanChain
EOS	NEO	WAVES
Ethereum	Nexus	Zcash
Ethereum Classic	Novacoin	Zilliqa
Exonum	NXT	

Table 2: The blockchain technologies used in this study

Dimensions: 107 x 25
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
1	Immutability.and.Failure.Tolerance [factor]	1. BFT 2. DAG 3. dPoS 4. dPoW 5. Hybrid 6. other 7. PoA 8. PoC / PoS 9. PoET 10. PoI [8 others]	17 (16.0%) 1 (0.9%) 7 (6.6%) 1 (0.9%) 8 (7.5%) 9 (8.5%) 2 (1.9%) 1 (0.9%) 1 (0.9%) 1 (0.9%) 58 (54.7%)		1 (0.93%)
2	Gossiping [factor]	1. Global 2. Local	78 (78.8%) 21 (21.2%)		8 (7.48%)
3	Latency [factor]	1. Asynchronous 2. Not known 3. Synchronous	21 (23.9%) 20 (22.7%) 47 (53.4%)		19 (17.76%)
4	Finality [factor]	1. Deterministic 2. Non-Deterministic	60 (61.9%) 37 (38.1%)		10 (9.35%)
5	Network.Topology [factor]	1. centralised 2. decentralized 3. hierarchical	2 (2.1%) 81 (87.1%) 10 (10.8%)		14 (13.08%)
6	Transaction.Model [factor]	1. Message Based Approach 2. The Tangle 3. Traditional Ledger 4. UTXO 5. Z-DAG	2 (2.4%) 1 (1.2%) 40 (47.1%) 41 (48.2%) 1 (1.2%)		22 (20.56%)
7	Server.Storage [factor]	1. Full Node Only 2. Thin node	42 (51.8%) 39 (48.1%)		26 (24.3%)
8	Block.Storage...State.Vector [factor]	1. Transactional data 2. Transactional data and Us 3. User Balance	46 (60.5%) 7 (9.2%) 23 (30.3%)		31 (28.97%)

11	Total.Supply..Log. [factor]	1. -Inf 2. 6 3. 7 4. 8 5. 9 6. 10 7. 11 8. 15 9. Inf	15 (14.0%) 1 (0.9%) 11 (10.3%) 19 (17.8%) 18 (16.8%) 7 (6.5%) 5 (4.7%) 1 (0.9%) 30 (28.0%)		0 (0%)
12	Tokenisation [factor]	1. No Tokenisation Present 2. Tokenisation present 3. Tokenisation Through Thir	31 (36.5%) 44 (51.8%) 10 (11.8%)		22 (20.56%)
13	Access.and.Control.Layer [factor]	1. Permissioned Private Bloc 2. Permissioned Public Block 3. Public Blockchain	23 (21.5%) 16 (14.9%) 68 (63.5%)		0 (0%)
14	Identity [factor]	1. Anonymous 2. KYC/AML 3. Pseudonymous	24 (28.2%) 23 (27.1%) 38 (44.7%)		22 (20.56%)
15	Reward.System [factor]	1. Block + Security Reward 2. Lump-Sum Reward	26 (33.8%) 51 (66.2%)		30 (28.04%)
16	Fee.reward [factor]	1. Mandatory Fees 2. No Fees 3. Optional Fees	44 (55.7%) 19 (24.1%) 16 (20.2%)		28 (26.17%)
17	Fee.Structure [factor]	1. Fixed Fees 2. Variable Fees	27 (36.5%) 47 (63.5%)		33 (30.84%)
18	Interoperability [factor]	1. Explicit Interoperability 2. Implicit Interoperability 3. None	26 (31.7%) 38 (46.3%) 18 (21.9%)		25 (23.36%)
19	Intraoperability [factor]	1. Explicit Intraoperability 2. Implicit Intraoperability 3. No Interoperability	25 (32.9%) 34 (44.7%) 17 (22.4%)		31 (28.97%)

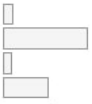
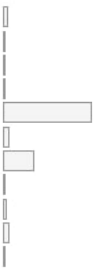
20	Governance [factor]	1. Alliance (industry consor 2. Open-Source Community 3. Technical / leading house	3 (3.3%) 72 (79.1%) 16 (17.6%)		16 (14.95%)
21	Code.Licence [factor]	1. Closed Source 2. Open Source	1 (1.1%) 88 (98.9%)		18 (16.82%)
22	Scripting.Language [factor]	1. Application-Specific Non- 2. Generic non-Turing Comple 3. Non-Turing Complete + Ext 4. Turing Complete	5 (6.4%) 30 (38.5%) 5 (6.4%) 38 (48.7%)		29 (27.1%)
23	Data.Privacy [factor]	1. add-on data privacy 2. Built in data privacy 3. Data privacy by third par 4. No data privacy: transact	6 (6.7%) 51 (57.3%) 5 (5.6%) 27 (30.3%)		18 (16.82%)
24	Hashing [factor]	1. BLAKE 2. BLAKE + Pedersen Hash 3. Combination 4. CryptoNight 5. CryptoNight + SHA3 6. Equihash 7. Groestl 8. Kerl 9. Scrypt 10. SHA2 [8 others]	5 (4.9%) 1 (1.0%) 5 (4.9%) 1 (1.0%) 1 (1.0%) 3 (2.9%) 1 (1.0%) 1 (1.0%) 7 (6.9%) 37 (36.3%) 40 (39.2%)		5 (4.67%)
25	Signature [factor]	1. BLS 2. Combination 3. EC-KCDSA 4. ECDH 5. ECDSA 6. ECDSA + Ed25519 7. Ed25519 8. Redjubjub 9. RingCT 10. Schnorr 11. W-OTS	3 (3.1%) 1 (1.0%) 1 (1.0%) 1 (1.0%) 58 (60.4%) 4 (4.2%) 20 (20.8%) 1 (1.0%) 2 (2.1%) 4 (4.2%) 1 (1.0%)		11 (10.28%)

Figure 9: The data sample

References

Adhami, S., Giudici, G., and Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin offerings. *Journal of Economics and Business*, 100(C):64–75.

Ballandies, M., Dapp, M. M., and Pournaras, E. (2018). Decrypting distributed ledger design - taxonomy, classification and blockchain community evaluation. *ArXiv*, abs/1811.03419.

Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J. A., and Felten, E. W. (2015). Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 104–121.

IEEE.

Cachin, C. and Vukolić, M. (2017). Blockchain consensus protocols in the wild. *arXiv preprint arXiv:1707.01873*.

Conley, J. P. (2017). Blockchain and the Economics of Crypto-tokens and Initial Coin Offerings. Vanderbilt University Department of Economics Working Papers 17-00008, Vanderbilt University Department of Economics.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London. or the Preservation of Favored Races in the Struggle for Life.

Foster, Z., Sharpton, T., and Grünwald, N. (2017). Metacoder: An r package for visualization and manipulation of community taxonomic diversity data. *PLOS Computational Biology*, 13(2):1–15.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871.

Ibañez, J. I., Bayer, C. N., Tasca, P., and Xu, J. (2020). Rea, triple-entry accounting and blockchain: Converging paths to shared ledger systems. *Available at SSRN*.

Mingxiao, D., Xiaofeng, M., Zhe, Z., Xiangwei, W., and Qijun, C. (2017). A review on consensus algorithm of blockchain. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2567–2572.

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

Sarkintudu, S. M., Ibrahim, H. H., and Abdwahab, A. B. (2018). Taxonomy development of blockchain platforms: Information systems perspectives. *AIP Conference Proceedings*, 2016(1):020130.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Spellerberg, I. and Fedor, P. (2003). A tribute to claude shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'shannon–wiener' index. *Global Ecology & Biogeography*, 12:177–179.

Tasca, P. (2018). Token-based business models. In *Disrupting Finance*. Part of the Palgrave Studies in Digital Business & Enabling Technologies.

Tasca, P. and Tessone, C. (2019). A taxonomy of blockchain technologies: Principles of identification and classification. *Ledger*, 4(0).

Tessone, C. J. and Garcia, D. (2018). Bitcoin: The centralisation of a decetralised economy. working paper.

Tsukerman, M. (2015). The block is hot: A survey of the state of bitcoin regulation and suggestions for the future. *Berkeley Tech. LJ*, 30:1127.

Wang, L.-G., Lam, T., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C., Jones, B., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., and Yu, G. (2019). treeio: an r package for phylogenetic tree input and output with richly annotated and associated data. *Molecular biology and evolution*.

Xu, Q., Aung, K. M. M., Zhu, Y., and Yong, K. L. (2018). A blockchain-based storage system for data analytics in the internet of things. In *New Advances in the Internet of Things*, pages 119–138. Springer.

Xu, X., Weber, I., Staples, M., Zhu, L., Bosch, J., Bass, L., Pautasso, C., and Rimba, P. (2017). A taxonomy of blockchain-based systems for architecture design. In *2017 IEEE International Conference on Software Architecture (ICSA)*, pages 243–252.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36.

Zachary, F., Scott, C., and Niklaus, G. (2018). *Taxa: An R package implementing data standards and methods for taxonomic data.*

Zimmermann, K. A. (2017). History of computers: A brief timeline. `https://www.livescience.com/20718-computer-history.html`. Accessed: 2020-07-16.