University of Galway Research Repository

Measuring node decentralisation in blockchain peer to peer networks

Title	Measuring node decentralisation in blockchain peer to peer networks
Author(s)	Howell, Andrew;Saber, Takfarinas;Bendechach, Malika
Publication Date	2023-02-08
Publication information	Howell, Andrew, Saber, Takfarinas, & Bendechache, Malika. (2023). Measuring node decentralisation in blockchain peer to peer networks. Blockchain: Research and Applications, 4(1), 100109. doi: https://doi.org/10.1016/j.bcra.2022.100109
Publisher	Elsevier
Link to publisher's version	https://doi.org/10.1016/j.bcra.2022.100109
Item record	http://hdl.handle.net/10379/17727

Measuring Node Decentralisation in Blockchain Peer to Peer Networks

Andrew Howell[1,4], Takfarinas Saber[2,4], and Malika Bendechache*[,2,3,4]

[1] Blockdaemon, Los Angeles, California, USA
andrew.howell2@mail.dcu.ie
[2] Lero – The Irish Software Research Centre
[3] ADAPT – Science Foundation Ireland Research Centre
[4] School of Computing, Dublin City University, Ireland
{takfarinas.saber, malika.bendechache}@dcu.ie

Abstract New blockchain platforms are launching at a high cadence, each fighting for attention, adoption, and infrastructure resources. Several studies have measured the peer-to-peer network decentralisation of Bitcoin and Ethereum (i.e., two of the largest used platforms). However, with the increasing demand for blockchain infrastructure, it is important to study node decentralisation across multiple blockchain networks–especially those containing a small number of nodes. In this paper, we propose NodeMaps, a data processing framework to capture, analyse, and visualise data from several popular P2P blockchain platforms such as Cosmos, Stellar, Bitcoin, and Lightning Network. We compare and contrast the geographic distribution, the hosting provider diversity, and the software client variance in each of these platforms. Through our comparative analysis of node data, we found that Bitcoin and its Lightning Network layer 2 protocol are widely decentralised P2P blockchain platforms, with the largest geographical reach and a high proportion of nodes operating on The Onion Router (TOR) privacy-focused network. Cosmos and Stellar blockchain nodes predominantly operate in large cloud providers or well-known data centres and that their increased complexities have led to reduced node participation.

Keywords: Blockchain, Peer-to-peer, Decentralisation, Networks, Bitcoin, Lightning Network, Cosmos, Stellar, TOR

1 Introduction

Distributed ledger technology (DLT) has seen tremendous growth over the last decade. Bitcoin, the first widely deployed blockchain technology, claims to be "A Peer-to-Peer Electronic Cash System" [1]. The system leveraged a novel amalgamation of cryptographic techniques coupled with properties of peer-to-peer (P2P) networking [1]. In its simplest form, a blockchain is an append only ledger of blocks linked together using a chain of cryptographic hashes, computed by hashing the previous block data. Every block added to the chain extends the linked hash list; the integrity of the chain can be validated by computing hashes from the current block all the way back to the genesis block, the first block in the ledger. The addition of new blocks reinforces the preceding blocks resulting in a modification resistant series of ordered records.

Block data is propagated via the P2P node network, each node can store a full copy of the ledger data. A new node joining the network bootstraps the consensus state from its peers by downloading historic ledger data using P2P networking.

Decentralisation is a cornerstone of a P2P blockchain platform, Srinivasan [31] discusses how it is important to quantify the term as it can be used to refer to the subsystems that comprise a blockchain platform. We must consider the distribution of nodes geographically, the distribution of voting power used to mint blocks, the number of entities who operate node infrastructure and the distribution of participants who engage with or develop blockchain systems. In this work, we define decentralisation in the context of infrastructure, the geographical distribution of nodes in the network and the hosting vendor diversity of where nodes appear to be hosted.

It is sufficiently difficult to bootstrap a new decentralised blockchain network with independent node operators without some centralised infrastructure [13]. Incentives must be balanced to promote engagement while taking into consideration network health by reducing centralisation of important node infrastructure [7]. Attention in the blockchain industry is a commodity, the sheer proliferation of blockchain platforms and the complexities of running performant node operations can present barriers to engagement thus diluting attention, creating a challenging system with potential for high participant churn [7].

In a recent study, Shrivas et al. [8] discuss the dramatic rise of DLT technology, the variety of new platforms, tools, programming languages and comment on the sustained year on year growth of employment in the sector for candidates with blockchain skills [8]. Interacting with a blockchain requires a node, it's not fair to assume that each developer or organisation will operate their own infrastructure, similar to how the majority of email users do not operate their own email servers. This leads to an interesting scenario where companies have filled the vacuum and offer specialised blockchain infrastructure services, many organisations defer node operations to service providers. We theorise that node infrastructure is operated by the following cohorts.

- Individuals with a strong believe in a blockchain project
- Foundations for a given blockchain platform subsidising bootstrap node infrastructure
- Researchers performing analysis of a blockchain
- Sufficiently savvy companies interacting with blockchain platforms
- Specialised blockchain infrastructure providers offering complex service
- Entities engaged in the block minting process
- Malicious actors

Several studies have investigated the decentralisation in the blockchain. While some of these studies are theoretical and applicable to a wide range of blockchain platforms (e.g., Kwon et al. [2] proves the lack of decentralisation in permissionless platforms), the rest of these studies are specific to very limited blockchain platforms (mainly Bitcoin or Ethereum blockchain networks) and/or propose intrusive tools and adaptations to the open-source blockchain client software to mimic a functioning blockchain node in peer discovery. For instance, Kim et al. [9] developed *NodeFinder*; an open-source scanning and monitoring of Ethereum's P2P network based on Geth (a command-line interface for running Ethereum node implemented in Go Language) to identify active nodes and periodically retrieves their client information. Gencer et al. [10] built *Falcon Relay Network* to serve as a backbone for ferrying blocks and to measure decentralisation in Bitcoin and Ethereum Networks. Venati [11] adapted the *NodeFinder* proposed by Kim et al. [9] for the purposes of node counting in the Ethereum private network and Ethereum public network, performed connections at a higher rate, and measured their impact on the network.

Rather than rehash similar techniques, our work aims to design an extensible framework for collecting, processing, and analysing snapshots of various blockchain platforms (e.g., Bitcoin, Lightning Network, Cosmos, Stellar) from a combination of public sources (e.g., [14, 15] and managed blockchain nodes in a efficient non-intrusive manner.

Furthermore, the objective of our designed framework is to facilitate:

- Investigation of the geographical distribution of various blockchain platforms.
- Identification of server infrastructure providers hosting the nodes associated with the peer IP address.
- Analysis of the reported client software versions and investigation of the user agent filed to assess the variances in deployed software.

The remainder of this paper is organised as follows. Section 2 introduces our chosen blockchain protocol architectures. Section 3 describes the proposed NodeMaps framework. Section 4 details the data collection and processing in our NodeMaps framework. Section 5 presents the results of the data analysis. Section 6 concludes the work.

2 Background

The successful application of Nakomoto's Bitcoin paved the way for a myriad of alternative blockchain platforms, each with a twist on the original blockchain design.

Generation 1, Bitcoin is considered first-generation blockchain technology, it is effectively a P2P transaction settlement system with its own native currency that requires no central entity to operate. Many of Bitcoin's core technical principles remain in newer systems, yet developers have introduced a number of key alterations.

Generation 2, blockchains like Ethereum, innovated on the original P2P ledger system. Buterin discusses in the Ethereum whitepaper that potentially the most important part of Bitcoin was the underlying blockchain technology and its mechanisms for achieving distributed consensus [4] while highlighting the limitations of Bitcoins on-chain scripting language. Buterin proposed a new blockchain platform with a Turing-complete programming language built in. Ethereum's enhancement would lead the way to decentralised smart contracts that reside and execute on the blockchain and can be triggered by state transitions associated with Ethereum accounts [4].

Generation 3, blockchains like Polkadot, are designed to meet the scaling challenges faced with prior iterations of the technology. In order to handle high throughput global scale use cases demand, the core blockchain technologies have been refined or rethought to address future demand [16].

The current blockchain landscape is a complex collection of competing networks mostly operating in their own sandbox with some purported unique differentiating feature. There id also an increasing range of Layer 2 protocols like the Bitcoin Lightning Network (LN), that complement the underlying blockchain and generally add some form of scaling solution. Bitcoin and Ethereum despite their reported scaling issues [18] can still be considered the base layer of the whole blockchain industry due to their huge market share and garnered attention. With new blockchain networks emerging every day and generation of blockchains being developed for scale, we are witnessing the transition from multiple distinct blockchains into an internet of blockchains where we will see cross protocol bridges, asset transfer and decentralised application (DApp) portability [17].

The resulting future could lead to a complex interconnected blockchain fabric, where specialised skills are required to participate in network operations.

2.1 Bitcoin

Bitcoin is a P2P blockchain that leverages Proof of Work (POW) to mint new blocks by hashing the SHA256 algorithm repeatedly; new blocks are generated if the result conforms to a specific signature. Historically any node possessed the capability to mine new blocks using the central processing unit (CPU) of the host device. Driven by the Bitcoin block rewards paid to miners for creating a new block, the mining industry is now dominated by Application Specific Integrated Circuits (ASIC) computers, block production is centralised in a number of mining pools [19] [20]. CPU mining is no longer a viable option for Bitcoin mining.

If a mining node discovers a new block it broadcasts it to its network of peers, each peer then propagates the message to its peers using its gossip protocol, the decentralised consensus process then determines the block validity. If a block is accepted by the network, the miner is rewarded with freshly minted Bitcoin. In a POW system network security is the conversion of large quantities of energy into cryptographic hashes, the incentivisation of the mining function has evolved into a pay-to-play arms race where more hashing power and fervent desire to reduce operational expense equates to greater rewards. Research suggests that over the past decade the complexity of mining has led to the centralisation of mining activity [2] [19] [20]. Mining pools are responsible for multiplexing individual mining resources into a shared pool, cryptographic computation is split between pool participants and rewards are shared proportionally to contributed compute power.

Roughly every two weeks (2,016 blocks) the Bitcoin network automatically re-tunes the difficulty associated with block minting. The network sets a target of difficulty that equates to roughly a ten-minute block time. Block space is also limited to 1MB, transactions have to be at least 250 bytes. Block size coupled with the target block time of 10 minutes means that the network can handle roughly 7 transactions per second [18]. This limitation has led to scaling issues and much debate [21]. Increasing the blocksize and shortening block times might seem like an immediate solution to increasing performance yet there are trade-offs in every decision. Keeping the parameters at the current values means there is some form of predictability in ledger storage and node operating requirements, a Bitcoin full node can operate comfortably on an ARM mini computer like a Raspberry PI [22] [23]. Increasing parameters could have a knock-on effect of increasing hardware requirements to run a node, smaller node operators could be priced out resulting in node centralisation.

2.2 Lightning Network

The Lightning Network (LN) is a Bitcoin Layer 2 transaction scaling solution that leverages off chain payment channels between two parties. Poon et al. [24] describe the Bitcoin network as a gossip protocol, where each ledger state modification is propagated to each node via a gossip mechanism. This node chatter ensures that the node has the required information to form a consensus [24]. This type of network communication is expensive as all nodes must validate the transactions, the solution is vital for consensus but limits transaction scalability. Poon et al. [24] proposed a solution where transactions between two parties could move to an off-chain payment channel where only Alice and Bob know about transactions between each other [24]. In order for a payment channel to be created, each actor must participate in a series of transactions to create and fund the channel using on-chain Bitcoin. The Bitcoin is locked in a 2-of-2 multi signature address with conditions that allow for each party to unlock their respective balance if specific conditions are met. Once funds are locked on the base chain, the LN allows each party to transfer the value of the payment channel between each other without having to broadcast any data on-chain. The channel can be settled if either party wishes to exit, or specific states are detected.

The LN consists of multiple nodes each can have numerous channels, the system is capable of routing multi-hop payments to other system participants by leveraging the network of interconnected payment channels. All transactions are backed by on-chain Bitcoin secured at the base layer, no transaction data is broadcast to the Bitcoin blockchain meaning transactions can happen at lightning speed, fees are kept low as mining is not required, participants only pay the network routing fee defined by the intermediary nodes.

LN is a Layer 2 solution and thus requires access to a Bitcoin node for specific activities that require communication with the base chain. Lightning Nodes typically run a hot wallet, meaning that the nodes indirectly have access to

Bitcoin funds, this also adds a custodial aspect to operating an LN node, which could lead to many participants opting to operate their own node [22] [23]. LN nodes must communicate with other peers, operating a node on the LN with a publicly facing address could lead to an increased security risk as it is trivial to correlate an IP address to the LN nodes balance.

2.3 Cosmos

Cosmos is a blockchain platform built on the Tendermint consensus algorithm [26]. The Cosmos ecosystem consists of many independent blockchain zones, the first of which is called the Cosmos Hub. Each zone is capable of communicating with each other via a novel Inter-Blockchain Communication Protocol (IBC), parallel blockchains can all interact, transferring assets from one zone to another [6].

The Cosmos blockchain utilises Proof-of-Stake (POS) in favour of POW mining, the block minting process is a similar exercise although, in place of physical ASIC miners, there exist validators. Cosmos network participants can bond or delegate ATOM tokens (the native currency of the Cosmos blockchain) to a validator. The validator is a special type of node that has the power to vote on block proposals, its voting rights are proportionally weighted based on the validator cumulative stake. The validator broadcasts signed cryptographic signatures to the network when voting on the next block, in exchange for confirmed validating activities nodes are paid a block reward.

The Tendermint consensus protocol utilised by Cosmos requires a fixed known set of validators [6]. Currently, the network has 125 validator nodes in the active set [27]. In order to become an active validator, a node must have a cumulative stake to reach at least position 125, currently 03/08/2021, this would require 33,000 ATOM [27].

It is commonplace for POS networks to adopt a slashing mechanism to keep nodes honest, effectively a slash involves burning a portion of the node's stake and preventing the node from voting on blocks for some time. If a node double signs a block, the protocol interprets this as an attack on the consensus system and the validate node will be identified and a portion of the node's stake will be slash [28]. Sustained network downtime is also considered a slash-able event, based on this, validator node operators must conform to best practices when deploying node infrastructure. Typically Cosmos validator node infrastructure consists of a public layer of sentry nodes connected by a private link to a protected validator node [29], operating a highly available Cosmos validator deployment has a high cost associated with the entry requirements.

2.4 Stellar

Stellar is an open network for money [5], the platform aims to introduce competition into the international payments markets by leveraging distributed ledger technology to send money around the world quickly and cheaply. Its protocol nativity supports various trading features, such as order books and cross-asset payments [30].

The Stellar Consensus Protocol (SCP), a Byzantine agreement protocol proposed by Mazières et al. [5] introduced a new consensus mechanism that the Stellar blockchain uses to facilitate secure transactions across a network of untrusted intermediaries. Organisations in the Stellar network choose other specific organisations to interact with. The system mirrors the interconnected nature of the traditional financial system where inter-bank relations are commonplace. Stellar network nodes operate in Quorum Sets. Nodes only see others that are part of the quorum, a view of the quorum can be ascertained as each node will learn of all others [30]. The key innovation in SCP is the open-membership approach taken to quorum sets which constantly evolve with new participants [30] joining the system.

3 NodeMaps: the Proposed Data Processing Framework

In this section, we describe the composition of the proposed NodeMaps framework[5], design choices (particularly in terms of data abstraction), and implementation/deployment details.

3.1 NodeMaps Components

Figure 1 shows an overview of the proposed framework (i.e., NodeMaps) to capture, sanitise, store and present node data gathered from blockchain P2P networks.

Scrapers: A set of blockchain specific scrapping services periodically connect to a remote data source to gather data. Each scrapping service performs a pipeline of data processing actions in order to sanitise the gathered data in preparation for storage. The designed scrappers are non-intrusive and keep the captured data to a minimum. In Section 4, we discuss the scraping process for each of our chosen protocols.

Storage: Sanitised node data is written to a Postgres SQL database. A Redis cache is used to store daily snapshots of processed data for each protocol. The cache allows for optimised retrieval of data during analysis. It will also facilitate a future extension of the NodeMaps platform to enable the analysis of different blockchain networks over time.

REST API: REST API retrieves data from the storage systems, the API is utilised in our analysis process.

[5] The source code of our NodeMaps framework will be made publicly available upon acceptance of this manuscript.

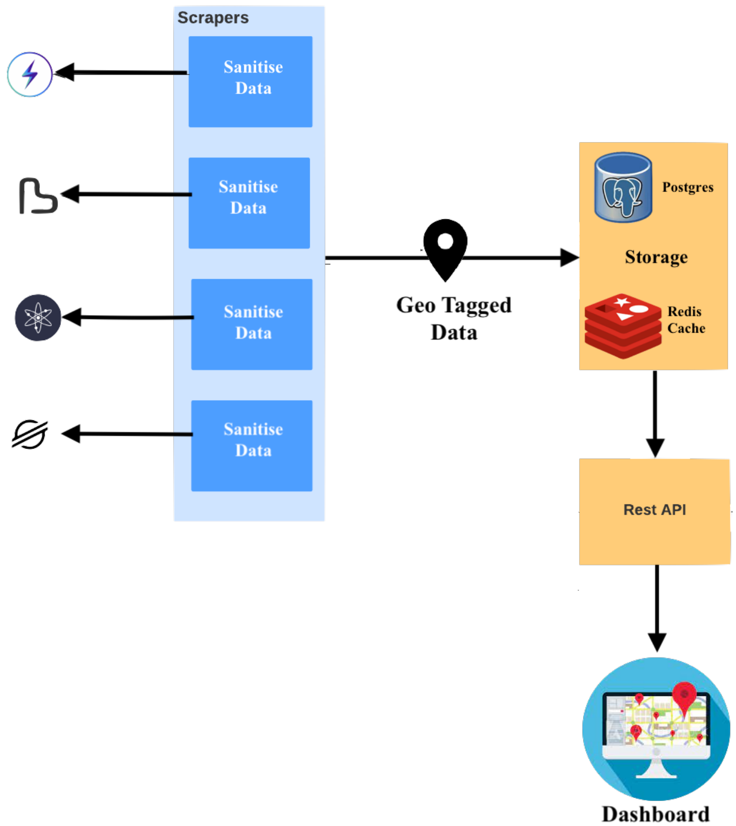

Figure 1. Overview of the Proposed Data Processing Framework Architecture

Dashboard: A basic dashboard details the project's goals and features charts demonstrating the analysis. The dashboard is built with ReactJS and Highmaps modules. It provides a visual representation of nodes that have been Geo-tagged, filters allow different datasets to be toggled on and off so that multiple blockchain networks overlap on the same world map providing additional decentralisation contexts.

3.2 Common Data Points Schema

The objective of our work is to develop an extensible framework where multiple blockchain platforms can be plugged with ease, thus enabling us to handle/analyse data across several/different blockchains. Therefore, we built NodeMaps in the form of a generic framework that abstracts data scraping and sensitisation processes from the data storage, retrieval and presentation process.

We define a common schema that is capable of satisfying the data points required for analysis activities. The scheme is inspired by the Bitnodes API [14], which exposes a data object capable of storing information relating to a Bitcoin node, its geographical and Autonomous System Number (ASN) properties. We extend this data model with additional fields, identified after performing a gap analysis of available node data from our chosen blockchain platforms.

3.3 Implementation/Deployment

The NodeMaps framework is implemented as a set of GoLang services that are based on data models and objects defined with the Swagger Interface Description Language. The node data model defines the common data model and it is codified in a YAML syntax.

Open-source libraries are used to auto-generate GoLang packages and API stubs based on the node data model. The choice of tooling allows for easy extension of the node data model to new fields by simply updating the Swagger definition and running a code-generator.

We deploy NodeMaps on a single server. All NodeMaps services (including scrapers, data processors, and visualisation) execute as micro services on docker containers within the same server.

4 Data Collection/Processing

In this section, we will discuss the varying approaches we took to capture data snapshots of node data from Bitcoin, Lightning Network, Cosmos, and Stellar blockchain platforms.

4.1 Data Scraping From Bitcoin

Past research has proposed novel ways of extracting P2P node information from the Bitcoin network [10], we choose not to rehash these techniques, in favour of utilising a network snapshot available via the Bitnodes API [14]. Bitnodes [33] operate a network crawler that recursively sends `getaddr` messages to network peers. Starting with a set of seed nodes the system recursively crawls all peers.

Bitnodes maintains frequent snapshots and makes them available via the REST API. The Bitnodes scraper connects to the Bitnodes REST API and calls the snapshots endpoint. This endpoint returns a list of available Bitcoin node data snapshots denoted by a Unix Timestamp. The scraper identifies the latest snapshot and issues another REST call to retrieve the node snapshot JSON data.

The next stage of the data pipeline involves processing each node record returned from Bitnodes. The application iterates over each node recorded in the retrieved snapshot and performs data sanitisation (i.e., mapping Bitnodes fields to those defined in the framework data model and setting default values for empty fields).

Once each record has been sanitised, we store the data in the Postgres database and update the Redis cache with a timestamped snapshot containing all processed records which can be used for further analysis.

4.2 Data Scraping from Lightning Network (LN)

We use a similar technique to that described by Romiti et al. [34] to scrape LN data. We deploy a Lightning Network Daemon (LND) node with an infrastructure provider. Over time the node builds up a graph of all other nodes that it has learned about during its P2P operations. LND exposes an API endpoint that returns a JSON file which describes the LN node graph.

The scraper starts by connecting to the endpoint exposed on our LND node retrieve the built graph data and commence the sanitisation process. Firstly we parse the graph data and we detect if the node record being processed uses Tor (i.e., has a `.onion` address). Some nodes have no address or a private IP. Since the aim of our work is to only process nodes that are exposed publicly to the internet, we filter out nodes that do not have an `IPV4`, `IPV6` or a `.onion` address. However, the framework could easily be adapted to keep all nodes for processing.

Following the pre-checks, we perform data sanitisation to build our final data model. First, we search for the ASN associated with node IP address using MaxMind's `GeoLite2-ASN` database [35]. We also search for the geographical location (i.e., city, country_code, timezone, latitude, and longitude) associated with the node IP address by means of a lookup against the MaxMind `GeoLite2-city` database [35]. Finally, the sanitised data set is saved to a Postgres database and a snapshot is stored in Redis.

4.3 Data Scraping from Cosmos

We run a node with an infrastructure provider for our Cosmos P2P data gathering. We set the configuration of the `gaiad` daemon to function as a seed node (i.e., a special configuration of a Cosmos node to share its peer info and discover as many peers as possible) and populated its initial seeds with the known nodes from the infrastructure provider. We have also adjusted the number of peers to which the node can connect to 1000 at a time. Furthermore, we have increased the maximum number of connections and open file handlers on the host operating system to 90000 in anticipation of a high connection count.

The Cosmos scraper is similar in design to those described in the previous sections. However, the captured data requires recursive processing to reveal additional data about the detected P2P nodes. The process we use is similar to the open-source project developed by Chainlayer (i.e., cosmos-crawler [36]). We query each known node that exposes its network information to collect its details and the list of peers to which it is connected. We continue this process recursively until no more peers are newly discovered.

The first stage of data processing involves removing duplicate peer records that might exist as peers can be the source of multiple connections with different destinations. The next stage involves data sanitisation: peers with no public IP addresses or those that could not be reached on their IP/port details are discarded. Analysis of the address book data suggests that many peers are not contactable after thousands of attempts. The final stage performs ASN and Geo-Tagging analysis on the IP addresses of contactable peers as described in the LN node data pipeline and persists the resulting data into the database.

Data Scraping from Stellar The Stellar data scraper performs a similar technique as used for bitcoin data capture. The process requests a network snapshot from a REST API provided by Stellar Beats [15]. The API returns a comprehensive set of Stellar network data points, information about the Quorum sets, high level statistics and an array of all discovered nodes. Stellar Beat states that their data is gathered every 3 minutes from the network using a network scraper that gathers information about all nodes [15]. There is little reason to re-implement the Stellar Beat system for our research, instead, we process a snapshot and perform our data sanitisation pipeline. To achieve consistency in our results we ignore the Stellar Beat Geo-Location and ASN data and leverage the common scraper utilities as discussed in the previous sections. Once all data mapping has been completed the snapshot is saved.

5 Analysis of Findings

In this section, we will review a snapshot of captured data for Bitcoin, LN, Cosmos, and Stellar blockchain networks collected on 07/08/2021. We perform an analysis of the data to ascertain the geographical distribution of nodes in each network and hosting vendor diversity.

It has to be noted that it is common for nodes in a P2P network to join and exit at any moment. Network latency, maintenance, service disruption and human intervention are all reasons a node may exit a network or appear offline. The view of peers in a network constantly changes and to our knowledge, there is no 100% guaranteed way to discover all nodes at a given time.

5.1 Top Autonomous System Numbers

Our investigation of yielded 1172 unique ASNs across the four blockchain platforms we investigated. Table 1 shows the top ASNs in terms of aggregated number of nodes.

Table 1 highlights that of all the detected nodes, a large portion, 54.28% operate on the TOR network. TOR has the capability of executing programs as hidden services, shielding the source IP address of the server running the application. This provides a form of anonymity to the service operator and the server hosting the applications, making it difficult to perform IP address analysis of the target. Attempting to de-anonymise TOR services is beyond the scope of this research. For the remainder of this section, we will treat TOR as a provider. The private nature of TOR means it is possible that some unknown percentage of the nodes could operate on any of the other ASNs identified.

When we look at the aggregated ASN data across all the investigated blockchains in Table 1, we see that Hetzner Online GmbH ASN is the second largest provider of servers for blockchain platforms, with 5.15% of all nodes. OVH SAS comes in third place of all detected ASNs and have just under half the amount of nodes in Hetzner.

ASNs	Node Count	Percentage
Tor network	14085	54.28
Hetzner Online GmbH	1337	5.15
OVH SAS	648	2.50
AMAZON-02	591	2.28
DIGITALOCEAN-ASN	535	2.06
AMAZON-AES	479	1.85
COMCAST-7922	418	1.61
Online S.a.s.	327	1.26
SHRD SARL	322	1.24
GOOGLE	309	1.19
Contabo GmbH	281	1.08
COGENT-174	247	0.95
ATT-INTERNET4	172	0.66
UUNET	169	0.65
Vodafone GmbH	143	0.55
Deutsche Telekom AG	136	0.52
AS-CHOOPA	106	0.41
Vodafone Libertel B.V.	93	0.36
Alibaba US Technology Co., Ltd.	88	0.34
Hangzhou Alibaba Advertising Co.,Ltd.	86	0.33
Others	5379	20.73

Table 1. Top ASNs for All Blockchain Platforms

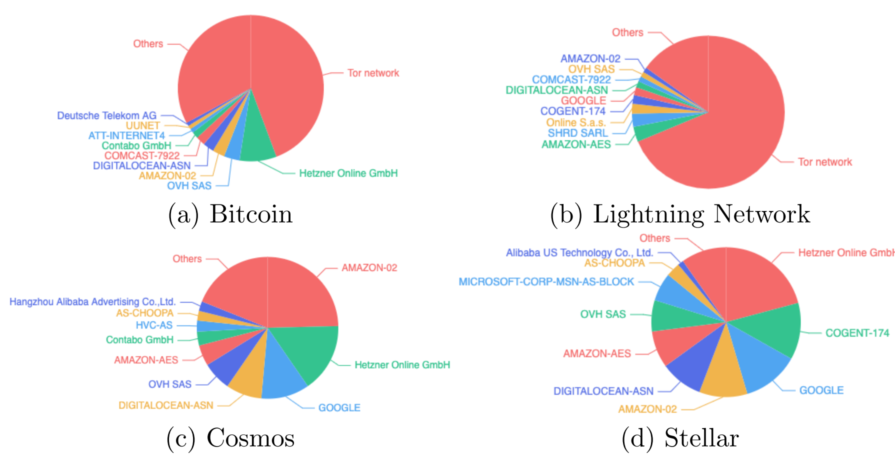

Figure 2. Distribution of Nodes Per ASN in Each Blockchain Platform

Figure 2 shows that Hetzner Online Gmb is also the top detected provider (other than TOR) in both Stellar and Bitcoin, also coming in second for Cosmos nodes and eleventh in LN. We will discuss the breakdown of each blockchain platform in the following sections.

5.2 ASNs Per Country

Excluding TOR nodes, Table 2 places the United States (US) as the top node location globally, followed by Germany (DE) and France (FR).

Country	n/a	US	DE	FR	CA	NL	GB	FI	RU	CH	SG	Other
#Nodes	14085	3272	2281	1237	649	607	412	255	236	219	214	2484
%Nodes	54.28	12.61	8.79	4.77	2.50	2.34	1.59	0.98	0.91	0.84	0.82	9.57

Table 2. Top Countries for Node Location

Table 3 provides a summary of the top 20 ASNs identified per country aggregated from each blockchain network snapshot. From this data, we can see that Hetzner Online GmbH in Germany has the largest concentration of nodes per provider and per country at 4.32%. When we compare the top node countries and the top ASNs per country, we can correlate the number of different ASNs per region, Table 3 shows that there are 7 ASNs in the US, 4 in Germany and 3 in France where the majority of nodes are deployed, indicating that blockchain node infrastructure tends to centralise around a few key providers in specific regions.

5.3 Findings in Bitcoin Platform

We analyse the finding in the Bitcoin platform in terms of ASNs, geographical locations, and software versions.

ASNs: A total of 14129 Bitcoin nodes were identified in the network snapshot. Table 4 shows a breakdown of the top 10 ASNs detected when analysing the data. TOR commands the highest overall slice of the network with 6290 nodes equating to 44.52% which indicates that a major part of Bitcoin users is privacy-focused.

The remaining ASNs are a mix of cloud, bare metal server data centers, and Internet Service Providers (ISP's). The ability to easily run a node on basic hardware and the culture surrounding the project are many factors that explain why the node proliferation is so high.

Hetzner Online GmbH is the largest detected ASN with 6.99% (i.e., 987) of publicly identifiable Bitcoin nodes. The remaining network nodes result in 910 different detected ASNs, consisting of a varied mix of cloud providers and global ISP's with 13.84% of ASNs hosting less than the 10 nodes and 464 unique ASNs hosting only a single node.

ASN	Country Code	Node Count	Percentage
TOR	n/a	14085	54.28
Hetzner Online GmbH	DE	1121	4.32
AMAZON-AES	US	479	1.85
OVH SAS	FR	422	1.63
COMCAST-7922	US	418	1.61
SHRD SARL	FR	322	1.24
Online S.a.s.	FR	310	1.19
Contabo GmbH	DE	281	1.08
GOOGLE	US	242	0.93
COGENT-174	CA	238	0.92
DIGITALOCEAN-ASN	US	225	0.87
Hetzner Online GmbH	FI	215	0.83
AMAZON-02	US	178	0.69
ATT-INTERNET4	US	172	0.66
UUNET	US	169	0.65
Vodafone GmbH	DE	143	0.55
Deutsche Telekom AG	DE	136	0.52
OVH SAS	CA	123	0.47
DIGITALOCEAN-ASN	DE	109	0.42
AMAZON-02	IE	93	0.36
Others		6470	24.93

Table 3. Top ASNs Per Country

ASN	Country	#Nodes	%Nodes
TOR	n/a	6290	44.52
Hetzner Online GmbH	DE	987	6.99
OVH SAS	FR	347	2.46
COMCAST-7922	US	289	2.05
Contabo GmbH	DE	205	1.45
Hetzner Online GmbH	FI	164	1.16
DIGITALOCEAN-ASN	US	155	1.10
ATT-INTERNET4	US	129	0.91
UUNET	US	119	0.84
Deutsche Telekom AG	DE	105	0.74
AMAZON-02	US	94	0.67
Other		5245	37.12

Table 4. Bitcoin ASNs Per Country

Geographical Locations: Table 5 provides a geographic breakdown of node locations by country. 6290 nodes operate on TOR meaning the location is unknown. 13.79% of nodes are located in the United States followed closely by Germany at 12.82%. China where Bitcoin miners were recently pressured to shut down [37] still operates 1.06% of nodes. The remaining 13.61% of publicly identifiable nodes are spread across 81 countries where 42 of those countries feature less than 10 nodes.

Country	TOR	US	DE	FR	NL	CA	GB	RU	FI	CN	Others
#Nodes	6290	1948	1811	575	429	319	271	219	194	150	1923
%Nodes	44.52	13.79	12.82	4.07	3.04	2.26	1.92	1.55	1.37	1.06	13.61

Table 5. Bitcoin Nodes Per Country

Software Versions: Table 6 presents the top ten software versions in the network snapshot. The Bitcoin `user agent` property indicates the software version of the node (i.e., the field exchanged during the node peering process). 43% of nodes (i.e., 6184) reported the latest Bitcoin Core client software version `Satoshi:0.21.1`, while 21.23% of nodes featured the previous release. 77% of nodes detected reported releases from within the last year and a half [38] (i.e., versions 0.20.* and 0.21.*), indicating high engagement from node operators to keep nodes up to date.

It worth noting that 7 nodes reported `Satoshi:0.8.1` which was released in 2013 [38] and it is unclear if these nodes operate as expected. 10.49% (1482) of reported node user agents contain a varying mix of reported software clients and versions. 116 nodes appear to have manually modified user agent strings and utilise the field as a form of P2P digital graffiti, where the field contains a personalised message.

5.4 Findings in Lightning Network Platform

We analyse the finding in the LN platform in terms of ASNs and geographical locations. Note that we do not analyse software versions in the LN platform as we were unable to gather such information from LN nodes.

ASNs: A total of 11340 LN nodes were identified in the network snapshot, Table 7 shows a breakdown of the top 10 ASNs detected when analysing LN graph data. Similar to Bitcoin, TOR commands the highest overall slice of the network, with 7795 nodes equating to a substantial 68.74%. The remaining ASNs combined add up to 16.13% of the network. As with Bitcoin, the providers are a mix of cloud, bare metal server data centers and Internet Service Providers (ISP's). The remaining 15.13% of the network nodes result in 500 different detected ASNs, consisting of a varied mix of cloud providers and global ISP's. 462 (7.81%) of ASNs host less than 10 nodes, 278 unique ASNs host only a single node.

Version	#Nodes	%Nodes
Satoshi:0.21.1	6184	43.77
Satoshi:0.21.0	3000	21.23
Satoshi:0.20.1	961	6.80
Satoshi:0.20.0	757	5.36
Satoshi:0.18.0	397	2.81
Satoshi:0.18.1	389	2.75
Satoshi:0.15.0.1	285	2.02
Satoshi:0.19.1	265	1.88
Satoshi:0.19.0.1	229	1.62
Satoshi:0.17.1	180	1.27
Others	1482	10.49

Table 6. Bitcoin User Agent Version

AMAZON-AES has the second-highest node count after TOR with 3.20% of LN nodes, on further investigation of the captured data we identify 318 nodes that share 33 IP addresses. In some cases, there are over ten unique LN node public keys reporting the same IP address and port `9735`, LN's P2P port. To our knowledge, it is not possible to run multiple LND instances on the same port on the same machine with the same IP address at any given time. Each peer record appears to share a similar signature for the alias field suggesting they are created by an automated process. Attempting to connect an LND node to a set of peers fails. We assume these anomaly peers are stale data that has not been pruned from the LN graph.

SHRD SARL makes up 2.78% of the LN nodes, analysing the alias field for these nodes, it appears that the majority are operated by an LN node as a service provider, Nodl. On further inspection of the alias field of all nodes in the LN `graph.json` snapshot, Nodl also appears to operate a large portion of nodes in Online S.a.s. (2.07%), the third major ASN detected in the snapshot data, suggesting that SHRD SARL and Online S.a.s. are their primary hosting partners.

Geographical Locations: Table 8 provides a geographic breakdown of LN node locations by country. 7795 or 68.74% of nodes operate on TOR meaning the location is unknown. 10.53% of nodes are located in the United States, although many of the node records appear to be stale and uncontactable as discussed previously. 5.71% of LN nodes operate from France primarily in SHRD SARL and Online S.a.s, followed by Germany at 3.32%. According to the LN graph snapshot, 31.26% of nodes can be detected in 76 different countries.

ASN	Country	#Nodes	%Nodes
TOR	n/a	7795	68.74
AMAZON-AES	US	363	3.20
SHRD SARL	FR	315	2.78
Online S.a.s.	FR	229	2.02
COGENT-174	CA	205	1.81
GOOGLE	US	164	1.45
COMCAST-7922	US	128	1.13
Hetzner Online GmbH	DE	80	0.71
Contabo GmbH	DE	66	0.58
OVH SAS	FR	65	0.57
AMAZON-02	US	59	0.52
Other		1871	16.50

Table 7. Lightning Network ASNs Per Country

Country	TOR	US	FR	DE	CA	NL	GB	CN	IT	JP	Others
#Nodes	7795	1194	648	376	282	157	134	67	43	43	601
%Nodes	68.74	10.53	5.71	3.32	2.49	1.38	1.18	0.59	0.38	0.37	5.30

Table 8. Lightning Nodes Per Country

5.5 Findings in Cosmos Platform

We analyse the finding in the Cosmos platform in terms of ASNs, geographical locations, and software versions.

ASNs: A total of 317 Cosmos nodes were identified in the network snapshot. Identified nodes were hosted in a total of 39 unique ASNs, with the top ten ASNs hosting 58.04% of them.

Table 9 shows a breakdown of the top 10 identified ASNs with the number (percentage) of nodes running in their infrastructure. In contrast to Bitcoin and LN, there were no nodes detected to be operating on the TOR network. AMAZON-02 has top 10 ASNs in 3 different countries, hosting 62 Cosmos nodes in total (or 19.55%), followed by Hetzner Online GmbH with 15.46% and Google with 7.89%. Cosmos continues to diverge from Bitcoin and LN data as there are no ISP's in the top ten, only server infrastructure providers.

Cosmos proof of stake blockchain has some complexities in its design as a central hub of an "Internet of Blockchains". Node operators must stake ATOM tokens to a validator node in order to become a block producer. The value of stake requires operating a node in the active set which has the effect of limiting node operations to entities that are well capitalised.

ASN	Country	#Nodes	%Nodes
Hetzner Online GmbH	DE	33	10.41
AMAZON-02	IE	31	9.78
GOOGLE	US	25	7.89
AMAZON-02	US	21	6.62
Hetzner Online GmbH	FI	16	5.05
AMAZON-AES	US	15	4.73
OVH SAS	CA	12	3.79
DIGITALOCEAN-ASN	US	11	3.47
Contabo GmbH	DE	10	3.15
AMAZON-02	SG	10	3.15
Others		133	41.96

Table 9. Cosmos ASNs Per Country

Geographical Locations: Table 10 provides a geographic breakdown of node locations by country. 90 or 28.39% of nodes operate in the United States, 68 (21.45%) in Germany, and 31 (9.78%) from Ireland. In total Cosmos nodes were detected in 23 countries, 20.19% of those countries operate less than 10 nodes. These numbers are not indicative of the overall Cosmos network as a secure validator network would consist of public sentry nodes and private validator nodes not publicly addressable as described in Section 2.3.

Country	US	DE	IE	SG	CA	FI	NL	FR	CN	KR	Others
#Nodes	90	68	31	17	17	16	14	9	9	7	39
%Nodes	28.39	21.45	9.78	5.36	5.36	5.05	4.42	2.84	2.84	2.21	12.30

Table 10. Cosmos Nodes Per Country

Software Versions: Table 11 shows the reported software versions in the Cosmos network snapshot. 66.88% (212) of nodes reported version `v0.34.11`, while `v0.34.9` was detected 2.21% (7) of the time. Due to the node scraping technique used we were unable to find data associated with 30.91% (98) of the known network. Data related to peers contained in the `gaiad` address book file does not have much-identifying data. Similar to LN software versions, further development would be required to probe the remote peer during the P2P process. Alternatively running a larger set of seed nodes may provide a deeper view of the network.

Analysis of the captured versions suggests high engagement by node operators to keep nodes up to date. Many newer POS protocols like Cosmos often require nodes to all run similar software versions as there are often network-wide upgrades that have to break changes (e.g., the recent Stargate update [39]).

Version	#Nodes	%Nodes
v0.34.11	212	66.88
n/a	98	30.91
v0.34.9	7	2.21

Table 11. Cosmos Software Versions

5.6 Findings in Stellar Platform

We analyse the finding in the Stellar platform in terms of ASNs, geographical locations, and software versions.

ASNs: A total of 165 Stellar nodes were identified in network snapshot. All detected nodes are hosted in a total of 26 unique ASNs.

Table 12 shows a breakdown of the top 10 ASNs detected. Similar to Cosmos there were no nodes identified as operating on the TOR network. Hetzner Online GmbH is the top ASN with 34 nodes clocking in at 20.61% of the public addressable network. COGENT-174 features 21 nodes or 12.73% and Google US with 11 nodes or 6.67% of the network.

It is worth noting that all ASNs hosting more than one node are all tier 1 cloud platforms or data centres (hosting a total of 90.30% of all detected Stellar nodes). The remaining 9.70% of nodes are hosted on 15 providers (one node per provider).

ASN	Country	#Nodes	%Nodes
COGENT-174	CA	21	12.73
Hetzner Online GmbH	DE	21	12.73
Hetzner Online GmbH	FI	13	7.88
AMAZON-AES	US	13	7.88
GOOGLE	US	11	6.67
OVH SAS	CA	5	3.03
OVH SAS	FR	5	3.03
DIGITALOCEAN-ASN	NL	5	3.03
DIGITALOCEAN-ASN	US	4	2.42
MICROSOFT-CORP-MSN...	US	4	2.42
Others		63	38.18

Table 12. Stellar ASNs Per Country

Geographical Locations: Table 13 provides a geographic breakdown of node locations by country. 40 (24.24%) of nodes operate in the United States, 31

(18.79%) in Canada, and 26 (15.76%) operate in Germany. In total Stellar nodes were detected in 19 countries, 5 of which operate 73.33% of the public-facing nodes.

Overall, the data shows that the network underpinning Stellar does not have the same geographical node distribution and hosting provider decentralisation as Bitcoin. This is probably due to the consensus design choices which centralise validation activities to a low number of node operators.

Country	US	CA	DE	FI	SG	NL	JP	FR	BE	GB	Others
#Nodes	40	31	26	13	11	7	5	5	4	4	19
%Nodes	24.24	18.79	15.76	7.88	6.67	4.24	3.03	3.03	2.42	2.42	11.52

Table 13. Stellar Nodes Per Country

Software Versions: Table 14 shows the reported software versions in the Stellar network snapshot. A strong 39.39% of nodes reported the latest release at the time of writing this paper (i.e., `stellar-core 17.3.0`). In total 66.67% of nodes reported releases from within the last six months, indicating a high engagement from Stellar node operators.

It is worth noting that 19 nodes reported a software version that appears to be custom-built (denoted by the `-dirty` postfix in the captured version string) which we assume to be based on Stellar Core to some degree. 12 nodes reported operating on unique software versions.

Version	#Nodes	%Nodes
stellar-core 17.3.0 (0b4c12a...)	65	39.39
stellar-core 17.1.0 (fbc0325...)	25	15.15
stellar-core 17.2.0 (e47d483...)	20	12.12
v17.1.0	9	5.45
stellar-core 17.0.0 (096f6a7...)	8	4.85
v16.0.0-129-gb0671b82-dirty	4	2.42
stellar-core 17.1.0 (6c86d89...)	4	2.42
e0ae42ee-dirty	3	1.82
ee87cdcb-dirty	3	1.82
c848b944-dirty	3	1.82
Others	21	12.73

Table 14. Stellar Software Versions

6 Conclusion

In this paper, we proposed NodeMaps, an extensible framework to capture, analyse, and visualise decentralisation data from several popular blockchain platforms such as Bitcoin, Lightning Network, Cosmos, and Stellar. Leveraging NodeMaps, we also performed an IP address analysis a snapshot of each of these blockchain platforms to compare and contrast the geographic, ASN and version distributions of their nodes.

Our analysis showed the decentralisation in Bitcoin and Lightning Network with identifiable nodes hosted by several ASNs in numerous countries and using a wide range of software versions. However, it also highlighted the user focus on privacy as a large percentage of nodes run on TOR. Our analysis also showed that Stellar (a competing network which claims similar open money principles) does not have the same geographical and ASN decentralisation (the majority of nodes run out of just ten providers). It also highlighted the significant number of custom-built nodes that are active in Stellar. Furthermore, our analysis showed that Cosmos has a limited number of nodes that are operated by only 39 small ASNs.

In the future, we would like to extend the NodeMaps data scraping pipelines to handle more modern blockchain platforms like Substrate and Cosmos SDK. Moreover, we would like to introduce a time perspective with P2P node tracking to assess the evolution of blockchain platforms over time and how they respond to various real-life events.

Acknowledgement

This work was supported in part by the Science Foundation Ireland grants 13/RC/2094_P2 (Lero) and 13/RC/2106_P2 (Adapt).

References

1. Nakamoto, S. (n.d.). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from www.bitcoin.org
2. Kwon, Y., Liu, J., Kim, M., Song, D., & Kim, Y. (2019). Impossibility of full decentralization in permissionless blockchains. AFT 2019 - Proceedings of the 1st ACM Conference on Advances in Financial Technologies, 110–123.
3. Mazières, D. (2015). The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus.
4. Ethereum Whitepaper — ethereum.org. (n.d.). Retrieved August 2, 2021, from https://ethereum.org/en/whitepaper/
5. Stellar - an open network for money. (n.d.). Retrieved August 1, 2021, from https://www.stellar.org/?locale=en
6. Kwon, J., & Buchman, E. (n.d.). Cosmos Whitepaper A Network of Distributed Ledgers. Retrieved from https://cosmos.network/resources/whitepaper
7. No Incentive? Algorand Blockchain Sparks Debate at Cryptography Event - CoinDesk. (n.d.). Retrieved August 2, 2021, from https://www.coindesk.com/no-incentive-algorand-blockchain-sparks-debate-cryptography-event

8. Shrivas, M. K., & Yeboah, T. (2018). The Disruptive Blockchain: Types, Platforms and Applications Blockchain powered platform for social enterprises: A case of Entreprise Sociale de Marché Commun (ESMC) View project Banker's Draft Management and Verification System using Distributed Ledger Technology View project The Disruptive Blockchain: Types, Platforms and Applications. ©TWCS. https://doi.org/10.21522/TIJAR.2014.SE.19.01.Art003
9. Kyun Kim, S., Ma, Z., Murali, S., Mason, J., Miller, A., & Bailey, M. (2018). Measuring Ethereum Network Peers, 14. https://doi.org/10.1145/3278532.3278542
10. Gencer, A. E., Basu, S., Eyal, I., Van Renesse, R., & Sirer, E. G. (n.d.). Decentralization in Bitcoin and Ethereum Networks.
11. Reddy Venati, Y. (2019). COUNTING NODES IN ETHEREUM BLOCKCHAIN.
12. Kim, S. (2017). Measuring Ethereum's peer-to-peer network. Retrieved from https://www.ideals.illinois.edu/handle/2142/99409
13. Dinger, J., & Waldhorst, O. P. (2009). Decentralized Bootstrapping of P2P Systems: A Practical View. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5550 LNCS, 703–715. https://doi.org/10.1007/978-3-642-01399-7_55
14. Global Bitcoin Nodes Distribution - Bitnodes. (n.d.). Retrieved August 1, 2021, from https://bitnodes.io/
15. API - Stellarbeat.io. (n.d.). Retrieved August 2, 2021, from https://stellarbeat.io/api
16. The Blockchain Generations — Ledger. (n.d.). Retrieved August 2, 2021, from https://www.ledger.com/academy/blockchain/web-3-the-three-blockchain-generations
17. From Ethereum to Binance Smart Chain - BSC Explained. (n.d.). Retrieved August 2, 2021, from https://defiprime.com/binance-smart-chain
18. The problem of scalability in the Bitcoin network — Bitpanda Academy. (n.d.). Retrieved August 2, 2021, from https://www.bitpanda.com/academy/en/lessons/the-problem-of-scalability-in-the-bitcoin-network/
19. Cong, L. W., He, Z., Li, J., Thank, W., Baldimtsi, F., Bonneau, J., … Zhang, X. (2021). Decentralized Mining in Centralized Pools. The Review of Financial Studies, 34, 1191–1235. https://doi.org/10.1093/rfs/hhaa040
20. Tovanich, N., Soulié, N., Heulot, N., & Isenberg, P. (2021). Interactive demo: visualization for Bitcoin mining pools analysis. https://doi.org/10.1109/ICBC51069.2021.9461124ï
21. Atluri, R. (2018). A Survey into the Bitcoin Scaling Dilemma.
22. rootzoll/raspiblitz: Get your own Bitcoin & Lightning Node running - on a RaspberryPi with a nice LCD. (n.d.). Retrieved August 2, 2021, from https://github.com/rootzoll/raspiblitz
23. How to Set Up a Bitcoin Node, With Lightning - CoinDesk. (n.d.). Retrieved August 2, 2021, from https://www.coindesk.com/becoming-self-sovereign-how-to-set-up-a-bitcoin-node-lightning
24. Poon, J., & Dryja, T. (2016). The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments.
25. Let's Get Ready to Wumbo! - by Lightning Labs - The Lightning Lab. (n.d.). Retrieved August 3, 2021, from https://lightninglabs.substack.com/p/lets-get-ready-to-wumbo
26. Kwon, J. (n.d.). Tendermint: Consensus without Mining.
27. Mintscan - COSMOS Validators. (n.d.). Retrieved August 3, 2021, from https://www.mintscan.io/cosmos/validators
28. Slashing Overview — Cosmos SDK. (n.d.). Retrieved August 3, 2021, from https://docs.cosmos.network/master/modules/slashing/

29. Validators — Tendermint Core. (n.d.). Retrieved August 3, 2021, from https://docs.tendermint.com/master/nodes/validators.html
30. Lokhava, M., Losa, G., Mazières, D., Hoare, G., Barry, N., Gafni, E., Jove, J., Malinowsky, R., McCaleb, J., & Mc-Caleb, J. (n.d.). Fast and secure global payments with Stellar. Proceedings of the 27th ACM Symposium on Operating Systems Principles, 17. https://doi.org/10.1145/
31. Quantifying Decentralization. We must be able to measure blockchain... — by Balaji S. Srinivasan — news.earn.com. (n.d.). Retrieved August 4, 2021, from https://news.earn.com/quantifying-decentralization-e39db233c28e
32. Crosby, S. A., & Wallach, D. S. (n.d.). An Analysis of BitTorrent's Two Kademlia-Based DHTs. Retrieved August 4, 2021, from http://azureus.sourceforge.net
33. ayeowch/bitnodes: Bitnodes is currently being developed to estimate the size of the Bitcoin network by finding all the reachable nodes in the network. (n.d.). Retrieved August 4, 2021, from https://github.com/ayeowch/bitnodes
34. Romiti, M., Victor, F., Moreno-Sanchez, P., Nordholt, P. S., Haslhofer, B., & Maffei, M. (n.d.). Cross-Layer Deanonymization Methods in the Lightning Protocol. https://github.com/MatteoRomiti/lightning_study
35. GeoLite2 Free Geolocation Data — MaxMind Developer Portal. (n.d.). Retrieved August 4, 2021, from https://dev.maxmind.com/geoip/geolite2-free-geolocation-data
36. chainlayer/cosmos-crawler. (n.d.). Retrieved August 4, 2021, from https://github.com/chainlayer/cosmos-crawler
37. China to shut down over 90% of its Bitcoin mining capacity after local bans - Global Times. (n.d.). Retrieved August 9, 2021, from https://www.globaltimes.cn/page/202106/1226598.shtml
38. Releases · bitcoin/bitcoin. (n.d.). Retrieved August 10, 2021, from https://github.com/bitcoin/bitcoin/releases?after=v0.21.0rc4
39. Cosmos Upgrades to Stargate: Another 2017 ICO Very Nearly Completes Its Vision - CoinDesk. (n.d.). Retrieved August 10, 2021, from https://www.coindesk.com/cosmos-upgrades-to-stargate-another-2017-ico-very-nearly-completes-its-vision