Journal Pre-proof

Dissecting bitcoin blockchain: Empirical analysis of bitcoin network (2009–2020)

Pranav Nerurkar, Dhiren Patel, Yann Busnel, Romaric Ludinard, Saru Kumari, Muhammad Khurram Khan

PII: S1084-8045(20)30398-2

DOI: https://doi.org/10.1016/j.jnca.2020.102940

Reference: YJNCA 102940

To appear in: *Journal of Network and Computer Applications*

Received Date: 16 May 2020

Revised Date: 5 November 2020

Accepted Date: 23 November 2020

Please cite this article as: Nerurkar, P., Patel, D., Busnel, Y., Ludinard, R., Kumari, S., Khan, M.K., Dissecting bitcoin blockchain: Empirical analysis of bitcoin network (2009–2020), *Journal of Network and Computer Applications* (2021), doi: https://doi.org/10.1016/j.jnca.2020.102940.

Credit Author Statement

1. **Pranav Nerurkar**: Conceptualization, Methodology, Software, Writing – Original Draft, Investigation

2. **Dhiren Patel**: Conceptualization, Methodology, Data curation, Funding acquisition

3. **Yann Busnel**: Conceptualization, Resources, Data curation, Funding acquisition Validation, Formal analysis

4. **Romaric Ludinard**: Conceptualization, Data curation, Funding acquisition, Formal analysis, Writing – Review & Editing

5. **Saru Kumari**: Conceptualization, Writing – Review & Editing, Supervision

6. **Muhammad Khurram Khan**: Methodology, Writing – Review & Editing, Supervision

Dissecting bitcoin blockchain: Empirical Analysis of Bitcoin network (2009-2020)

Pranav Nerurkar[a,b], Dhiren Patel[b], Yann Busnel[c], Romaric Ludinard[c], Saru Kumari[d,*], Muhammad Khurram Khan[e]

[a]*Dept. of Data Science, MPSTME, NMIMS University, Mumbai, India*
[b]*Dept. of CE&IT, VJTI-Mumbai, India*
[c]*SRCD department, IMT Atlantique, Rennes, France*
[d]*Dept. of Mathematics, Ch. Charan Singh University, Meerut, India*
[e]*Center of Excellence in Information Assurance (CoEIA), King Saud University, Saudi Arabia*

Abstract

Bitcoin system (or Bitcoin) is a peer-to-peer and decentralized payment system that uses cryptocurrency named bitcoins (BTCs) and was released as open-source software in 2009. Unlike fiat currencies, there is no centralized authority or any statutory recognition, backing, or regulation for Bitcoin. All transactions are confirmed for validity by a network of volunteer nodes (miners) and after collective agreement is subsequently recorded into a distributed ledger "Blockchain". Bitcoin platform has attracted both social and anti-social elements. On the one hand, it is social as it ensures the exchange of value, maintaining trust in a cooperative, community-driven manner without the need for a trusted third party. At the same time, it is anti-social as it creates hurdles for law enforcement to trace suspicious transactions due to anonymity and privacy. To understand how the social and anti-social tendencies in the user base of Bitcoin affect its evolution, there is a need to analyze the Bitcoin system as a network. The current paper aims to explore the local topology and geometry of the Bitcoin network during its first decade of existence. Bitcoin transaction data from 03 Jan 2009 12:45:05 GMT to 08 May 2020 13:21:33 GMT was processed for this purpose to build a Bitcoin user graph. The characteristics, local and global network properties of the user's graph were analyzed at ten intervals between 2009-2020

*saryusiirohi@gmail.com

with a gap of one year. Small diameter, skewed distribution of transactions, power-law distributed in and out degrees, disconnected graph, and presence of large connected components were the observations from network analysis. Thus, it could be inferred that despite anti-social tendencies, Bitcoin network shared similarities with other complex networks. Network analysis also uncovered twenty types of legal and anti-social entities operating on Bitcoin and provided a path for uncovering these anti-social entities.

Keywords: Bitcoin, Network Science, Graph Algorithms, Exploratory Data Analysis

---

1. Introduction

Originally proposed in 2008 by an unknown individual (or a group of individuals) who used a pseudonym "Santoshi Nakamoto", Bitcoin cryptocurrency has since then emerged as the most successful cryptocurrency amongst its peers, reaching an adoption level unrealized by older digital currencies [1, 2, 3]. As on 19^{th} March 2020, Bitcoin has a market cap of USD$98,584,789,143 with 18,277,112 bitcoins (BTC's) in circulation each with a value of USD$5,393.89. Bitcoin differs from its traditional online banking peers by relying on a decentralized consensus scheme for verifying the correctness and authentic nature of currency transfers between users [4, 5, 6]. The decentralized consensus scheme is made possible by an organized collective of nodes in the Bitcoin system known as "miners". The miners confirm each transaction for authenticity. This increases security in the Bitcoin system and ensures the core philosophy of Bitcoin "Maintain trust in an untrusted environment" without the need for a trusted third party as a reward miners collect transaction fees for the transactions that they confirm.

Illustrating the transaction fundamentals of bitcoin transfers, consider that user i wants to transfer n bitcoins to user j. Then i will need a bitcoin wallet, which holds all his private keys and the wallet address of j (Figure 1). Also, the transaction is valid only if user i signs it using his cryptographic key.

Valid transactions are then broadcast over the Bitcoin network, and all miners are informed. Technically, the transaction is not broadcast to all nodes in the Bitcoin network, as a single node can be connected to a maximum 125 (incoming connections=8, outgoing connections=117) other nodes. However, by recursive broadcasts "gossip protocol," a transaction eventually reaches all

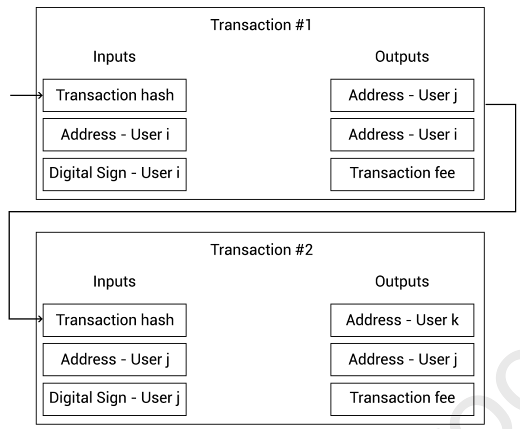


Figure 1: Transfer of bitcoins from user i to j and j to k

nodes [1, 7]. Miners keep all received transactions in their memory pool and combine these transactions to form a "candidate block." Each miner then competes with other miners to add its candidate block to the blockchain. The miner who succeeds gets a reward in BTC's and broadcasts its newly mined block to other miners. Other miners will independently verify the newly mined block before adding it to their blockchain.

Since Bitcoin's inception in 2009, the initial two years saw slow adoption with hardly 1000 unique addresses and less than 10000 transactions per day [1, 8]. However, as bitcoin became financially significant, there was an exponential growth in transactions from 2012-2016, which also saw the entry of serious users, investors, speculators, and independent mining industries. Before the popularity of bitcoin, the users were mostly crypto-enthusiasts. The change in the profile of Bitcoin's user base was also evident from the increase in the transaction values, fluctuations in BTC price, and volumes of BTC's. This phase also saw the emergence of Ponzi schemes, money laundering, frauds [9], embezzlements, extortion [10] and tax evasion [11] practices that used the blanket of secrecy afforded by Bitcoin to mislead the audit trail. There emerged a diversity even amongst the miners in terms of geography and size. When Bitcoin was launched, it was feasible for any participant to become a miner, but as the user base increased, mining became competitive and required specialized hardware. Miners prefer large warehouses with access to cheap electricity [12]. With time, solo miners decreased and gave way to mining pools.

As the scale and complexity of the Bitcoin network increased, research interest too emerged to allow for its better understanding [4, 11, 12, 13, 14]. However, analysis of network properties of Bitcoin graph is an interesting domain, albeit one that has received comparatively less attention. A reason for this could be the complexity of identifying users in the Bitcoin network. In the Bitcoin network, identifying users by wallet addresses (aka accounts, bitcoin addresses, public keys, or other unique identifiers used interchangeably to refer to users' in Bitcoin system) is complicated as these can be generated and discarded multiple times [12]. There is also no upper limit to the identities a single person can create or any limits on the number of transactions or beneficiaries. These factors significantly enhance the hurdles in analyzing the Bitcoin network. To overcome the hurdle caused by multiple identities of a single user, heuristic clustering is applied to the Bitcoin network. With heuristic clustering, multiple identities of a single user are grouped into a single identity. This strategy is used in several Bitcoin network studies [15, 16, 17, 18] and has the advantage of reducing the number of entities of the Bitcoin network.

1.1. Motivation

Based on an oft-quoted maxim in network science, "We will never understand complex systems unless we develop a deep understanding of the networks (graphs) behind them" [19], the current paper proposes to shed light on the network properties of Bitcoin. Bitcoin is a diverse ecosystem inhabited by users (wallets) that could be ordinary people interested solely in the exchange of assets or mining nodes competing to ensure that the transactions in their memory pool get added to the blockchain. Though the interactions behind entities in other large systems such as the internet, wireless sensor networks [20, 21, 22], social networking websites, citation systems, file sharing systems are well studied, However Bitcoin system failed to receive similar attention. Network analysis would also help machine learning based applications of Bitcoin such as illegal transaction detection and forensics improve feature engineering.

1.2. Contributions

- Conducted a comprehensive study of the large-scale Bitcoin system and interactions occurring in it from 2009 to 2020 by constructing a network from the blockchain files.

- Studied the Bitcoin network at scale based on local and global graph properties (see Section 3.2).
- Network analysis to uncover types of legal and illegal entities operating on Bitcoin and provide a path for uncovering these entities to aid digital forensic tools.
- Proposed techniques for detection of illegal entities operating in bitcoin network
- Used structural information of Bitcoin network to characterize interactions and evaluate it at scale
- Open sourced the Bitcoin network dataset to motivate independent research
- A time series analysis was performed using previous data obtained of the Bitcoin network. The data for training the machine learning models was from years 2009-2020 and the predictions were made for the year 2021.

So far only I Alqassem *et al.* [12] and X Lee *et al.* [13] have provided a detailed graph-theoretic assessment of Blockchain cryptocurrencies. However, X Lee *et al.* focused on Ethereum blockchain, and I Alqassem *et al.* focused on the time period of 2009-2014 to analyze Bitcoin systems. Although these papers provide a technical foundation for the current work, there is no overlap. Ethereum is not just a crypto-currency but also a platform that enables distributed applications. Analysis cannot be compared between Ethereum and Bitcoin. Bitcoin has higher volumes, users and market cap so affects more users and should therefore receive more attention. I Alqassem *et al.* [12] worked on Bitcoin 2009-2014 so the current papers extended their work to 2020. Additionally, observations and conclusions on future outlook of Bitcoin were made using time series analysis. Time series models are data-driven. So observations and conclusions are obtained after experimentation. The data is allowed to "speak for itself" and used for predicting growth outlook for year 2021.

The rest of the paper is organized as follows: Section 2 gives the related work done on Bitcoin and other cryptocurrencies. The procedure to convert raw data into a processed form is outlined in Section 3, followed with a description of network analysis tools in Section 3.2 and discussion of results in Section 4. The paper concludes in Section 5, mentioning future works for subsequent research.

2. Related work

The related work reviewed can be divided into two categories: First, the work that examined the Bitcoin system itself. Second, work that examined other blockchain-based systems.

2.1. Bitcoin studies

The journey of Bitcoin, which builds upon nearly two decades of ideas proposed in mailing lists, forum posts, blogs [23], wikis, and source code found in cryptographic circles, is described by F Tschorsch *et al.* [14]. However, the authors focused more on framing a tutorial on Bitcoin that includes an outline of selective existing literature. I Alqassem *et al.* have provided a longitudinal network-based analysis of Bitcoin systems from 2009-2014. The authors have commented upon the changing nature of bitcoin users over time and also drew attention to various structural properties of the Bitcoin system viz. longest connected component, network diameter, densification power law, degree assortativity, time-evolving community structure and inequality in the network [12]. The authors agreed that analyzing the Bitcoin system presents challenges due to the anonymity seeking behaviors of the user base. Though the results highlighted key differences between the Bitcoin network and networks of other systems, the continuous developments and fluctuations in the complex cyber-physical Bitcoin systems necessitate another up-to-date review. T Chang *et al.* analyzed the various heuristics that are proposed in the literature to identify all public keys that belong to the same user. The heuristics create an approximation of the original Bitcoin network by merging multiple user identifiers to a single identifier and reducing number of entities in the network. Previous studies on network analysis of cryptocurrencies [12, 13, 11] to have used heuristics and hence, it is a tried and tested method for improving network analysis. S Park *et al.* scanned the live Bitcoin network for 37 consecutive days in 2018 to track the behavior of the miners. The authors commented upon Bitcoin network statistics such as the number of users, the geographic distribution of users, Bitcoin wallet protocols, and messages propagating in the network [1].

2.2. Studies on other blockchain-based systems

Y Li *et al.* used the Ethereum transaction graph (interactions between smart contracts and users) to explore the relationship between the graph structure and crypto-currency price fluctuations [24]. H Sun *et al.* attempt clustering analysis on Ethereum data to segment malicious users from the rest [25]. S Ferratti *et al.* has used global network statistical measures such as the order of the network, degree distribution, distance, clustering coefficient, and the tendency of exhibiting a "small world" effect [26]. Based on the observations from these measures, the authors have speculated about the online behavior of Ethereum users, the geographic distribution of miner nodes, and the characteristics of transactions. While S Ferratti *et al.* argued for the advantages of studying the blockchain structure through a complex network perspective, their focus remained on the Ethereum blockchain structure only. X Lee *et al.* studied the Ethereum blockchain at scale and applied network analysis measures to characterize interactions between users in Ethereum [13]. The authors studied the network characteristics (vertex count, edge count, self-loop count, and edge density), local network properties (degree distribution, correlation of out and indegree, node centrality measures) and global network properties (reciprocity, assortativity, connected component distribution, diameter, path length, adhesion, cohesion). Just like [26], the authors focused on Ethereum blockchain only but have emphasized that a similar line of network analysis could be extended to another web of blockchain networks. The work in the current paper relies on tools and methods given by S Ferratti *et al.* [26] and X Lee *et al.* [13] but targets a longitudinal analysis of Bitcoin network. Table 1 gives the methods and results of network-based studies on blockchain and other real-world systems.

Table 1: Results of published network studies

System under review	**Network theory used**	**Observation**
Twitter [27]	Gini index	Dominant nodes are present
Facebook [28]	Assortativity coefficient	Negative assortativity
Social networking websites [29]	Diameter and Average path length	Small
Social networking websites [29]	Clustering coefficient	High
Social networking websites [30, 31]	Average degree, Edge density	High
World wide web [30, 31]	Degree distribution	In and out degree distribution follow power law
Protein-protein interaction [31]	Degree distribution	Power law
World wide web [32]	Small world effect	19 hops between any two webpages
Facebook [32, 33]	Strongly connected component (SCC)	99.8% - 100% nodes and edges are covered.
Citation networks [32, 33]	Graph structure	Acyclic
Citation network [30]	Degree distribution	In and out degree distribution follow power law
Film actors [30]	Degree distribution	Power law
Company directors [30]	Degree distribution	No power law
Co-authorship network [34]	Degree distribution	No power law
Ethereum network [13]	Vertices, arcs, self-loops, edge density, degree distributions, centrality measures, reciprocity, assortativity, SCC	In and out degree distribution follow power law. Density is low, reciprocity is positive, assortativity is negative. SCC has 98-99% nodes and edges.
D Ding *et al.* [35]	Study topological connectivity and message routability of P2P overlays	Degree and Connectivity Analysis
D Ding *et al.* [36]	Study topological connectivity and message routability of P2P overlays	Degree and Connectivity Analysis

It can be observed from Table 1 that using a unified set of tools and principles, networks of different fields can be studied. This is because, despite variations, networks grow following certain basic principles [37].

3. Bitcoin blockchain to Graph

Bitcoin blockchain dataset in raw form was obtained from full node at VJTI Blockchain lab [1]. The dataset was of size 268GB and consisted of blockchain in the form of blk.data files. All blocks and transactions from 03 Jan 2009 12:45:05 GMT to 08 May 2020 13:21:33 GMT were present in the dataset. This raw data was then converted to CSV files using the blockchain parser built by the VJTI Blockchain lab [2]. The processed dataset, which is in the form of ".csv" files were made available for download [3]. Table 2 shows the four ".csv" files of the processed dataset.

Table 2: Description of processed dataset

Relation	Attributes		
Output	tx_hash:START_ID	wallet_address:END_ID	amount
Address	wallet_address:ID		
Inputs	wallet_address:START_ID	tx_hash:END_ID	amount
Transactions	tx_hash:ID	timestamp	

From the Transactions dataset, it is possible to obtain the count of transactions occurring in that year. Each transaction (tx) is identified in blockchain by a unique hash (tx_hash: ID) and has a timestamp, which is the UNIX time of the transaction. For the year 2009, transactions start from 03 Jan 2009 12:45:05 GMT, and for the year 2020, transaction up to 08 May 2020 13:21:33 GMT is considered. Bitcoin entities were identified using an API[4] [38]. Table 3 and 4 describes the dataset.

Table 3: Distribution of transactions in Bitcoin blockchain network (2009-2015)

	2009	2010	2011	2012	2013	2014	2015
Transactions	32741	185410	1902443	8459093	19645798	25265702	45689861
Inputs	2810	108965	1902443	5716084	15407017	33300547	54564769
Outputs	32643	143863	2595309	5981241	16278420	34586691	57150816
Max BTC's in a tx	22500	96999	550000	158336.30	194993.50	217517.63	172841.81
Max inputs in a tx	320	901	529	673	1757	674	1519
Max outputs in a tx	2	98	2002	2792	3075	5352	13107
Input sending highest amount	COINBASE	COINBASE	CoinJoin Mess	DeepBit.net	DeepBit.net	Unknown	Unknown
Output receiving highest amount	Unknown	Unknown	CoinJoin Mess	DeepBit.net	DeepBit.net	Unknown	Unknown
Total BTCs sent	1978736	22667790	297984085	925215501	429732306	264107039	548006072

[1]https://www.vjti-bct.in/
[2]https://github.com/pranavn91/blockchain
[3]https://drive.google.com/open?id=1pEpBAUXKgQX0BP8ircQgd9yXiucLY14h
[4]https://www.walletexplorer.com

Table 4: Distribution of transactions in Bitcoin blockchain network (2016-2020)

	2016	2017	2018	2019	2020
Transactions	82634637	104081930	81393458	119729415	39978670
Inputs	90773554	128642149	77568478	128768057	52805351
Outputs	95783964	144361281	104780607	133558733	54179450
Max BTCs in a tx	99489.99	87082.81	109735.6	157457.612	182501
Max inputs in a tx	677	1089	1061	1347	1442
Max outputs in a tx	11515	6626	5027	7266	6990
Input sending highest amount	Unknown	Unknown	Unknown	Unknown	Unknown
Output receiving highest amount	Unknown	Unknown	Unknown	Unknown	Unknown
Total BTCs sent	1068404725	896026050.66	290858051.91	515972850.159	128637285.824

By parsing through the Bitcoin blockchain dataset, a transaction graph (representing the exchange of bitcoins between wallet addresses) was built. Each transaction has multiple inputs and outputs, as shown in Figure 2. This transaction graph is refined further by heuristic clustering to obtain the user's graph (see Figure 3). The heuristic used for clustering is called the regular inputs heuristic, i.e., all input addresses in a transaction belong to the same user [5, 15]. The user's graph (payments made between users) leads to meaningful analysis compared to the transaction graph [15, 16, 17, 18]. Additionally, the results from the user's graph of Bitcoin can be compared with social network analysis of other real-world systems viz. web, social networking websites, citation graphs. A similar comparison is not possible if the transaction graph of Bitcoin is considered.

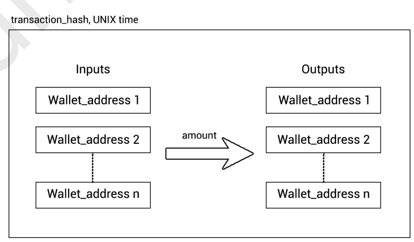


Figure 2: Multi-input multi-output transactions

The heuristic clustering reduces the multi-input multi-output transactions to a form more suited for network analysis. Multiple inputs are clustered, and a single address is used as a starting point for the transaction. The details of the heuristic clustering strategy are given in [15, 16, 17, 18]. Figure 3 graphically shows the information of each attribute and relation in the dataset after heuristic clustering is applied.

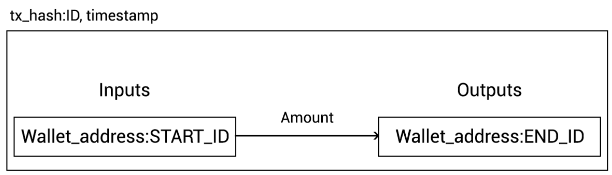


Figure 3: Illustration of attributes of processed dataset

3.1. Experimental setup

The preprocessing code is in Python 3.6, and the code for network analysis is in R. The network analysis functions are from the igraph package of R [39]. The experiments are performed on a single core 1 TB Intel(R) Xeon(R) Silver 4114 CPU@2.20GHz.

3.2. Network measurements of Bitcoin network

For this study, Bitcoin user graph is represented as a network $G = (V, E)$, where V refers to the addresses of users' wallets, while E represents a bitcoin exchanges between these wallets. The timestamp of transaction, tx_hash, and amount are attributes of E. As multiple transactions can occur between wallet_addresses, G is a directed multi-graph. Using tools described in Section 3.3, an analysis of the Bitcoin network G is performed for the period 2009-2020.

3.3. Description of tools for Network analysis

1. Vertex count (order of graph) $|V|$ and edge count (size of graph) $|E|$
2. Graph density (G_D): Number of edges present graph G amongst all possible edges in G. G_D for undirected and directed graphs is given by below equations 1 and 2 respectively.

$$\frac{2|E|}{|V|(|V|-1)} \tag{1}$$

$$\frac{|E|}{|V|(|V|-1)} \tag{2}$$

3. Average degree d

$$d = \frac{1}{|V|}\sum_{u \in V} d(u) = \frac{2m}{n} \tag{3}$$

4. Degree distribution of graph $P(k) = \frac{n_k}{n}$ is fraction of nodes in the network with degree k i.e. n_k where n is the Graph order.
5. Probability distribution
   (a) Power law: $y = k^{-\alpha}$ (k=constant, α=exponent)
   (b) Exponential: $y = e^{-\lambda k}$ (λ= mean time between events)
   (c) Lognormal: $y = \frac{1}{k}e^{-\frac{(logk-\mu)^2}{2\sigma^2}}$ (μ=scale parameter, σ=shape parameter)
   (d) Poisson: $\frac{e^{-\mu}\mu^x}{x!}$
6. Adhesion or edge connectivity E for connected graph G is the minimum number of edges $\lambda(G)$ whose deletion from a graph G disconnects G.

7. cohesion - a minimum number of vertices needed to remove to make the graph not strongly connected
8. Diameter is the length $max_{(u,v)}d(u,v)$ of the "longest shortest path" (i.e., the longest graph geodesic) between any two graph vertices (u, v) of a graph, where $d(u, v)$ is a graph distance.

9. Average path length $L = \sum_1^E(G)\frac{d(u,v)}{E(G)}$

10. reciprocity ρ as given in Eq. 4 is the measure of the likelihood of vertices in a directed network to be mutually linked.

$$\rho = \frac{\sum_{i\neq j(a_{ij}-\overline{a})(i\neq j(a_{ji}-\overline{a})}}{sum_{i\neq j(a_{ij}-\overline{a})^2}} \tag{4}$$

11. Assortativity: level of homophily of the graph.

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \tag{5}$$

   where,
   - q_k number of edges leaving the node, other than the one that connects the pair j, k
   - σ_q standard deviation of q in Eq. 5

- e_{jk} refers to the joint probability distribution of the remaining degrees of the two vertices

12. Number of connected components of a graph G is $c(G)$. A connected component is a set of vertices all of which are connected, and unconnected to the other nodes in the network. The weakly connected components are found by performing breadth-first search. The strongly connected components are implemented by two consecutive depth-first searches.
13. Degree Centrality of a vertex v_i is defined as $deg(v_i)/2|E|$
14. Betweenness centrality $C_B(v)$ of $v \in V$ is the fraction of times v occurs on any shortest path connecting any other pair of vertices $s, t \in V$. Let σ_{st} be the total number of shortest paths connecting vertex s with vertex t. Let $\sigma_{st}(v)$ be the number of these shortest paths containing v. The geodesic centrality of v is:

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{6}$$

15. Size of largest strongly connected component N_s - a set of vertices in a directed graph such that any node is reachable from any other node using a path following only directed edges in the forward direction.

$$N = \max_{F \subseteq \mathcal{C}} |F| \tag{7}$$
$$\mathcal{C} = \{C \subseteq V \mid \forall u, v \in C : \exists w_1, w_2, \ldots \in V : u \sim w_1 \sim w_2 \sim \cdots \sim v\}$$

16. Relative size of the largest connected component (N_{rel}) equals the size of the largest connected component divided by the size of the network

$$N_{\text{rel}} = \frac{N}{n}. \tag{8}$$

17. Number of triangles defined in the following way is independent of the orientation of edges when the graph is directed.

$$t = |\{\{u, v, w\} \mid u \sim v \sim w \sim u\}| \;/\; 6 \tag{9}$$

18. Global clustering of a network is the probability that two incident edges are completed by a third edge to form a triangle

$$c = \frac{|\{u, v, w \in V \mid u \sim v \sim w \sim u\}|}{|\{u, v, w \in V \mid u \sim v \neq w \sim u\}|} \tag{10}$$

Tools for network measurement can be divided into three groups: measures for characteristics (vertex count, edge count, edge density), measures of local network properties (radius, local clustering coefficient, node degree) and measures for global network properties (degree distribution, adhesion, cohesion, components, centralization, k-cores).

4. Experimental study

Bitcoin users graph is studied using the tools given in Section 3.3. The entire Bitcoin network is studied at eleven intervals, as seen in the results. The year in the results corresponds to a Bitcoin users graph built from transaction data considered from 01 Jan 12:00:00 AM GMT to 31 Dec 11:59:59 PM GMT of that year. An exception is the year 2020, which is built using transaction data from 01 Jan 2020 12:00:00 AM GMT to 08 May 2020 13:21:33 GMT.

4.1. Bitcoin Network characteristics

Table 5 gives the bitcoin users graph. Two versions of edge density are indicated by (S) for a simple, undirected version of the user's graph and (D) for the directed user's graph. Multiple directed edges between two users are collapsed to a single undirected edge to obtain edge density (S). Vertex count in Table 5 and 6 gives the total senders and receivers in that calendar year. Bitcoin users have increased till 2017, leading to the price of BTC's reaching its peak in Dec 2017. The following years have seen a decline in both users and the value of BTCs. In 2009, out of 32741 transactions, 32522 were COINBASE transactions. The highest number of BTCs transferred in a single transaction was 22500, and 320 were the highest number of inputs present in a transaction. Limited edges were created as transactions between users were less. The edge density is low in both the directed graph (Edge density (D)) and the undirected graph (Edge density (S)) for the period 2009-2020 compared to social networks. The low density is due to the skewed distribution of transactions amongst the users. 99.8% of the total users in 2009 made almost a single transaction. This declined to 73.24% by 2020.

Table 5: Characteristics of Bitcoin blockchain network (2009-2015)

	2009	2010	2011	2012	2013	2014	2015
Vertex count	32644	143943	2599119	6001831	16337189	34693993	57381025
Edge count	32808	233872	4642054	19710026	49336100	78077032	145496703
Edge density (S)	6.16e-05	2.25e-05	1.28e-07	3.4e-07	0.94e-07	3.7e-08	2.37e-08
Edge density (D)	3.08e-05	1.12e-05	6.87e-07	5.4e-07	1.85e-07	6.48e-08	4.42e-08

Table 6: Characteristics of Bitcoin blockchain network (2016-2020)

	2016	2017	2018	2019	2020
Vertex count	57107986	78724132	53049193	32288199	3160555
Edge count	29365348	625420597	330885984	230911982	24840651
Edge density (S)	5.2e-08	0.49e-07	0.68e-07	1.12e-07	1.18e-06
Edge density (D)	9e-08	1.01e-07	1.17e-07	2.21e-07	2.49e-06

Till the year 2010, Bitcoin was used by crypto-enthusiasts and year 2011 saw the entry of the first mixing service and mining pools. Both these services involve transactions with one or limited inputs and several outputs. Consequentially, the maximum number of outputs in a single transaction increased from 98 in 2010 to 2002 in 2011 and has remained in range of 2000-7000. This leads to observation that "Number of outputs" can be used to discriminate between different types of users in Bitcoin.

4.2. Vertex degree distribution

The procedure mentioned by C Gillespie [40] was followed to understand the distribution of in (see Table 7 and 8) and out degrees (Table 9 and 10) of users graph. In 2009, for the distribution of in degrees, the minimum value from which the power-law distribution was fitted i.e., (x_{min}) was 4 and for exponential x_{min} was 1, log-normal x_{min} was 1 and poission x_{min} was 5. For 2010, x_{min} was 31 for power law, 183 for exponential, 29 for log-normal and 4351 for poisson. In 2011, x_{min} was 397 for power law, 279 for exponential, 359 for log-normal and 8079 for poisson. In 2012, x_{min} was 621 for power law, 72053 for exponential, 608 for log-normal and 5352 for poisson. In 2013, x_{min} was 987 for power law, 76728 for exponential, 1151 for log-normal and 4751 for poisson. In 2014, x_{min} was 1615 for power law, 99867 for exponential, 1702 for log-normal and 154 for poisson. In 2015, x_{min} was 2994 for power law, 99891 for exponential, 1950 for log-normal and 359 for poisson.

Table 7: Likelihood ratio tests for comparing in degree distribution (2009-2015)

Distributions	Parameters	2009	2010	2011	2012	2013	2014	2015
Power law	α	1.99	1.54	2.35	1.86	1.88	1.98	2.12
Exponential	λ	0.11	0.001	0.011	0.004	0.002	0.002	0.0001
Log-normal	μ	1.79	2.59	-26.61	-52.63	-29.818218	-21.38	2.62
	α	1.01	2.65	5.06	8.42	6.50	5.55	2.61
Poisson	μ	13.83	4992.6	26133.67	43568.6	43778.7	7764.21	8610.67

In 2016, x_{min} was 2318 for power law, 99549 for exponential, 1510 for log-normal and 5 for poisson. In 2017, x_{min} was 3118 for power law, 99671 for exponential, 99671 for log-normal and 6294 for poisson. In 2018, x_{min} was 1862 for power law, 96500 for exponential, 2179 for log-normal and 11175 for poisson. In 2019, x_{min} was 2674 for power law, 97258 for exponential, 97258 for log-normal and 1 for poisson. In 2020, x_{min} was 2588 for power law, 95384 for exponential, 1939 for log-normal and 1 for poisson. From Table 7 it is observed that power-law and log-normal are better fit to data than exponential or poisson. Moreover, X_{min} values indicate that tail of the distribution follows power law. α value indicates inverse relationship between degree and frequency of such nodes. High degree nodes such as mixing services and pools would form LSCC/LWCC making it easy for identifying them on Bitcoin.

Table 8: Likelihood ratio tests for comparing in degree distribution (2016-2020)

Distributions	Parameters	2016	2017	2018	2019	2020
Power law	α	2.1	2.11	1.92	2.4	2.2
Exponential	λ	0.001	0.001	0.001	0.001	0.003
Log-normal	μ	5.15	-194.65	-17.11	-398.36	-7.01
	α	2.06	12.1	5.29	15.85	3.78
Poisson	μ	7918	29039.39	63050.8	5095.25	4054.3

In 2009, for the distribution of out degrees, the minimum value from which the power-law distribution was fitted i.e., (x_{min}) was 4 and for exponential x_{min} was 3, log-normal x_{min} was 1 and poission x_{min} was 12. For 2010, x_{min} was 14 for power law, 5136 for exponential, 15 for log-normal and 42 for poisson. In 2011, x_{min} was 520 for power law, 42350 for exponential, 145 for log-normal and 252 for poisson. In 2012, x_{min} was 667 for power law, 93316 for exponential, 562 for log-normal and 2210 for poisson. In 2013, x_{min} was 1073 for power law, 94828 for exponential, 94828 for log-normal and 2244 for poisson. In 2014, x_{min} was 1540 for power law, 98344 for exponential, 1544 for log-normal and 2334 for poisson. In 2015, x_{min} was 2251 for power law, 98992 for exponential, 2214 for log-normal and 300 for poisson.

Table 9: Likelihood ratio tests for comparing out degree distribution (2009-2015)

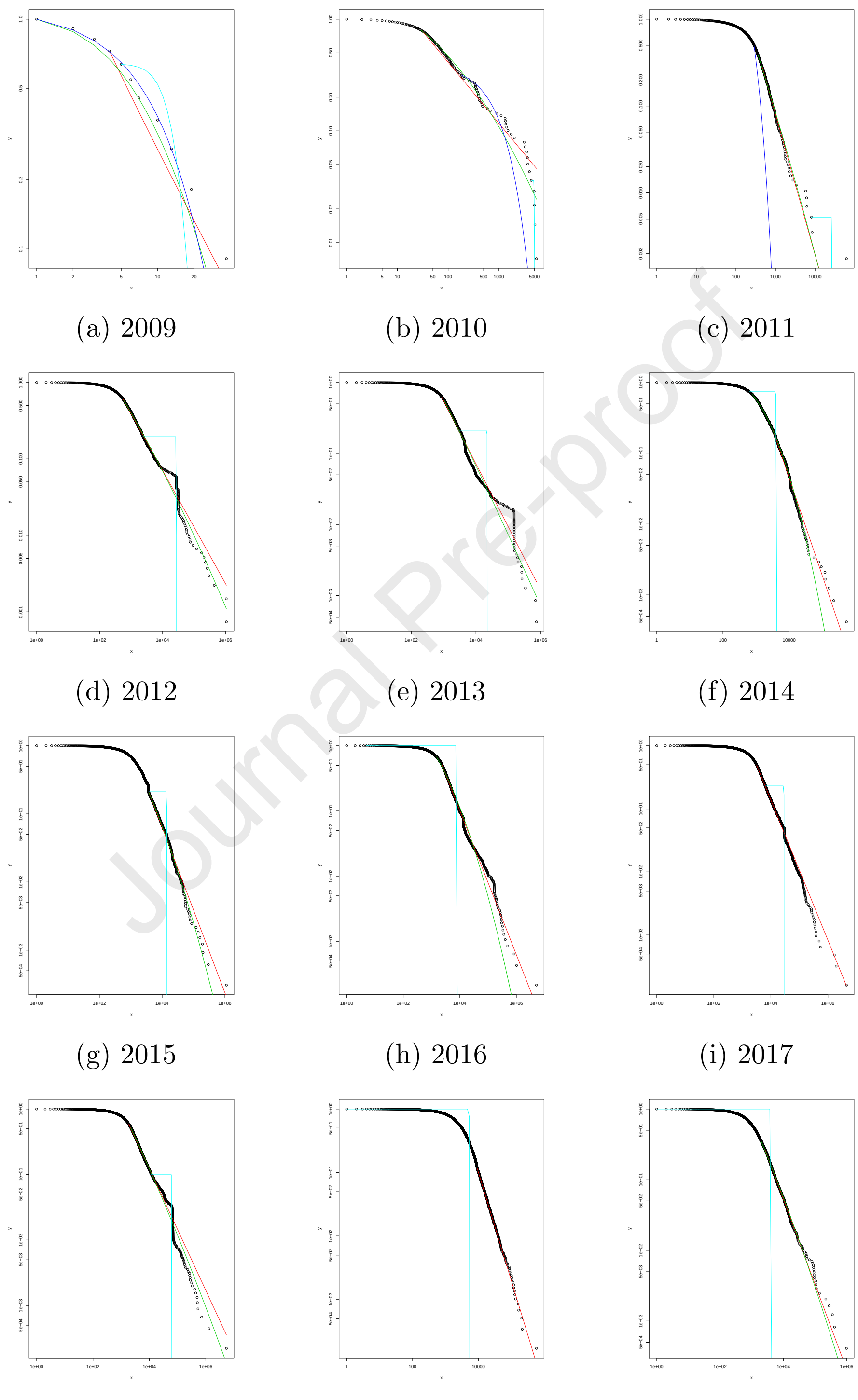
Distributions	**Parameters**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Power law	α	1.33	1.42	1.73	1.74	1.85	1.86	1.87
Exponential	λ	0.25	0.06	0.013	0.005	0.002	0.002	0.001
Log-normal	μ	-7.27	-4.52	-52.81	-7.835970	-137.41132	-18.89	1.25
	α	6.10	5.14	8.31	4.77	10.3	5.73	3.14
Poisson	μ	10851.33	3754.7	4516.74	27558.8	24466.7	25145.02	14322.95

In 2016, x_{min} was 2224 for power law, 99977 for exponential, 1722 for log-normal and 2314 for poisson. In 2017, x_{min} was 5338 for power law, 96639 for exponential, 2820 for log-normal and 1 for poisson. In 2018, x_{min} was 4308 for power law, 97340 for exponential, 6600 for log-normal and 10649 for poisson. In 2019, x_{min} was 9124 for power law, 98154 for exponential, 98154 for log-normal and 1 for poisson. In 2020, x_{min} was 842 for power law, 84442 for exponential, 456 for log-normal and 69 for poisson.

Table 10: Likelihood ratio tests for comparing out degree distribution (2016-2020)

Distributions	**Parameters**	**2016**	**2017**	**2018**	**2019**	**2020**
Power law	α	1.77	2.58	2.34	2.7	2.07
Exponential	λ	0.001	0.001	0.0006	0.0007	0.0051
Log-normal	μ	7.3	7.76	4.8	-338.17	5.56
	α	1.8	1.13	2.02	11.65	1.67
Poisson	μ	15859.95	5967.4	28175.95	5362.98	2580.6

Figure 4 and 5 show the fitting of four heavy-tailed distributions to in-degree and out-degree distribution of users graph respectively. Four distributions considered are discrete power law (red), exponential (dark blue), log-normal (green), and Poisson (light blue). Distribution is fit as per protocol specified by C Gillespie [40].



(a) 2009 (b) 2010 (c) 2011

(d) 2012 (e) 2013 (f) 2014

(g) 2015 (h) 2016 (i) 2017

(j) 2018 (k) 2019 (l) 2020

Figure 4: In-degree distribution of Bitcoin users graph (2009-2020)

(a) 2009 (b) 2010 (c) 2011

(d) 2012 (e) 2013 (f) 2014

(g) 2015 (h) 2016 (i) 2017

(j) 2018 (k) 2019 (l) 2020

Figure 5: Out-degree distribution of Bitcoin users graph (2009-2020)

As claimed for most complex networks, even bitcoin users graph followed the ”scale-free” property as power-law exponent ranged from 1.54-2.4 for in-degree distribution and from 1.42-2.7 for out-degree distribution. x_{min} indicated that the tail of the in and out-degree distributions fit the power law. High degree entities such as mixing services, gambling websites and pools will occupy the tail of the degree distribution. Whereas, ordinary users shall be at the other end of the spectrum. Thus, the location of the entity on the degree distribution curve could reveal its nature.

4.3. Bitcoin: Global networks properties

Table 11 and 12 give the global network properties of bitcoin users graph. Measures marked with # could not be computed on the current configuration of the system. $^+$ indicates approximation used for computation as given by M Jackson *et al.* [41]. In 2009, as transactions were infrequent, adhesion and cohesion were zero indicating a sparsely connected graph where information transfer was slow due to long diameter. As the majority were COINBASE transactions in 2009, the graph had high centralization tendency, low reciprocity, girth, and assortativity. Till 2010, crypto-enthusiasts dominated the transactions, and transactions were less, and diameter increased. In 2011, mixing services and miner pools entered, and the DeepBit.net mining pool had 61897 incoming and 120756 outgoing connections. CoinJoin Mess, a mixing service, had 903 incoming and 1800 outgoing connections in 2011. The presence of mining pools and mixing services decreased the diameter and average path length while leading an increase in reciprocity. In 2012, SantoshiDice.com, a gambling website, saw 810474 incoming and 1055385 outgoing connections. In 2013 too SantoshiDice.com continued to get the highest incoming and outgoing connections. In 2014, SantoshiDice.com had the maximum incoming connections (1592352), whereas CoinJoin Mess had the maximum outgoing (2256302). In 2015, another online gambling site LuckyBit.it had the highest incoming connections at 1655881, and CoinJoin-Mess had the highest outgoing connections at 2256344.

Table 11: Global network properties (2009-2015)

	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Adhesion	0	0	0	0	0	0	0
Cohesion	0	0	0	0	0	0	0
Diameter	7	5525	0.03^{+}	0.06^{+}	0.06^{+}	0.05^{+}	0.05^{+}
Average path	1.01	748.54	0.03^{+}	0.06^{+}	0.06^{+}	0.05^{+}	0.05^{+}
Radius	6	1	#	#	#	#	#
Reciprocity	6.11e-05	0.02	0.008	0.2	0.16	0.03	0.019
Girth	3	3	3	3	3	3	3
Assortativity	-0.55	-0.31	0.17	0.12	0.06	0.04	0.17
Centralization	0.99	1	0.99	0.99	0.99	1	1
C_d	0.5	0.23	0.04	0.15	0.05	0.03	0.02
C_c	0.99*	2.1e-06	#	#	#	#	#

In 2016, with 300120 outgoing connections, Faucetbox.com (bitcoin reward site) was very active. In 2017 highest connections were recorded by Poloniex.com, a crypto exchange with 4473190 incoming and 445628 outgoing connections. In 2019, Huobi.com-2, a bitcoin exchange platform, had the highest outgoing connections. Due to anonymity, the identity of an entity with the highest incoming and outgoing connections in 2018 was not found.

Table 12: Global network properties (2016-2020)

	2016	**2017**	**2018**	**2019**	**2020**
Adhesion	0	0	0	0	0
Cohesion	0	0	0	0	0
Diameter	0.09^{+}	0.11^{+}	0.1^{+}	0.11^{+}	0.13^{+}
Average path	0.09^{+}	0.11^{+}	0.1^{+}	0.11^{+}	0.13^{+}
Radius	#	#	#	#	#
Reciprocity	0.016	0.003	0.0016	0.0009	0
Girth	3	3	3	3	3
Assortativity	-0.026	-0.005	-0.022	0.28	0.09
Centralization	0.99	0.99	0.99	1	0
C_d	0.044	0.031	0.05	0.02	0.15
C_c	#	#	#	#	#

Reciprocity is close to 0 indicating that Bitcoin is majorly for payments or investments and not for exchange of BTC's between account owners. Assortativity in range $-1-0$ indicates that low degree nodes (ordinary users, enthusiasts, small investors) are linked to high degree nodes (gambling hubs, exchanges, pools, mixers). Due to the high transactions received by such entities the centralization remained close to 1. Based on these observations, transaction based features would be key in discriminating entities. These features would be - Total transactions in which wallet has participated (T_x), Total incoming transactions to the wallet (${T_x}^{in}$), Total outgoing transactions from the wallet (${T_x}^{out}$), Average number of incoming transactions received by an address of a wallet (A_v), Total number of addresses sending BTC to the wallet (T) and Ratio of Transaction count and address count (R) gives the average number of times an address of the wallet was reused for a transaction.

4.4. Community structure

Usually, triangles, transitivity, and clustering coefficient are higher in social networks than non-social networks [13]. These parameters indicate the tendency of entities in the network to form dense communities. In 2009, the Largest Weakly Connected Component (LWCC) was the entire graph, and Largest Strongly Connected Component (LSCC) was minimal. Triangles and clustering coefficients were also negligible. In 2010, WCC was 25, and SCC was 108482. In 2011, WCC was 1400, and SCC were 2029127. In 2012, WCC was 6165, and SCC were 3149100. In 2013, WCC was 15122, and SCC was 9888167. DeepBit.net formed the largest SCC and largest WCC in 2011. SantoshiDice.com formed the largest SCC and largest WCC in 2012 (see Table 13).

Table 13: Community structure (2009-2012)

		2009	**2010**	**2011**	**2012**
LSCC	Triangles	0	9580	104368	3797352
	Nodes	2 (0%)	34709 (24.1%)	567144 (21.8%)	2846171 (47%)
	Edges	5 (0%)	75367 (32.2%)	1345036 (28.9%)	13908941 (70%)
	Articulation pt.	0	72	638	1389
	C	NaN	0.003	0.003	9.1e-05
LWCC	Triangles	9	18708	3102649	4267711
	Nodes	32644 (100%)	143880 (100%)	2593961 (100%)	5979901 (100%)
	Edges	32808 (100%)	233829 (100%)	4638181 (100%)	19693726 (100%)
	Articulation pt.	79	20774	496060	1440988
	C	2.4e-05	1.11e-05	0.0005	0.0001
Full network	Triangles	9	18709	3102700	4267910
	Articulation pt.	79	20784	497641	1447747
	C	2.4e-05	1.11e-05	0.0005	0.0001

In 2013, 2014 and 2015 too the largest SCC and WCC were formed by SantoshiDice.com (see Table 14). In 2014, there were a total of 40508 WCC and 24516983 SCC in the network. In 2015, WCC was 253244, and SCC were 35766309 in the network.

Table 14: Community structure (2013-2015)

		2013	**2014**	**2015**
LSCC	Triangles	7751768	5140336	21461343
	Nodes	6437119 (39.4%)	10157747 (29.6%)	17445491 (30.2%)
	Edges	32501745 (65.8%)	41139689 (52.3%)	85078065 (58.9%)
	Articulation pt.	9270	14777	14790
	C	0.0002	0.0008	0.0004
LWCC	Triangles	7751768	6832830	25928531
	Nodes	16282225 (100%)	34556782 (100%)	57084066 (100%)
	Edges	49292728 (100%)	77961419 (100%)	145254102 (100%)
	Articulation pt.	4282322	7775376	13682985
	C	0.0002	0.0001	0.0002
Full network	Triangles	9102472	6834251	25931343
	Articulation pt.	4297982	7809891	13771043
	C	0.0002	0.0001	0.0002

In 2016, unknown wallets had formed the largest WCC and SCC. In 2017, Bittrex.com, a crypto trading exchange, formed the largest SCC. In 2019, the largest SCC was formed by Bitcoin exchange service Huobi.com-2. In 2016, WCC was 871640, and SCC was 46385054 in the network. In 2017, WCC was 1476165, and SCC were 69375203. In 2018, WCC was 1032588, and SCC were 30074974. In 2019, WCC were 967845 and SCC were 26896674 (see Table 15).

Table 15: Community structure (2016-2020)

		2016	**2017**	**2018**	**2019**	**2020**
LSCC	Triangles	125423937	95674389	62367145	24089648	0
	Nodes	10698736 (18.7%)	9306342 (3%)	3242666 (6.1%)	844423 (2.7%)	1
	Edges	120658573 (41.1%)	169589795 (15.07%)	62330136 (18.8%)	18010394 (8.2%)	0
	Articulation pt.	1259	2206	717	522	0
	C	0.0015	0.0009	0.0004	0.004	0
LWCC	Triangles	213985326	210765433	214016097	88648952	0
	Nodes	53556287 (93.7%)	74366786 (94.4%)	47785524 (90.7%)	26470992 (85.5%)	123583 (0.03%)
	Edges	287695383 (93.7%)	618579809 (98.9%)	325783461 (98.4%)	212922543 (97.8%)	403262 (0.01%)
	Articulation pt.	5333181	6854728	4535938	3167225	4785
	C	0.0005	0.0003	0.0001	0.0004	0
Full network	Triangles	214055511	287646955	214094259	88721557	0
	Articulation pt.	6212728	6987676	5488866	4060330	351463
	C	0.0005	0.0003	0.0001	0.0004	0

The LSCC increased from 2009-2012 to close to 47% of all nodes of the graph in 2012 and then has declined to $2-3\%$ of all nodes by 2019. LWCC has remained in a range of $97-98\%$ of the total nodes. LWCC and LSCC were formed mainly because of mixing services, gambling services, and crypto exchanges. The LSCC formed in past years (see Table 16) confirms this. Reuse of addresses for transferring BTCs led to the compromise of anonymity of bitcoin users. Thus, another feature to discriminate entities is suggested - Ratio of Transaction count and address count (R). This feature gives the average number of times an address of the wallet was reused for a transaction.

Table 16: Categories and address forming LSCC

Year	**Address**	**Category**	**Entity name**
2010	1Bw1hpkUrTKRmrwJBGdZTenoFeX63zrq33	Unclassified	0091107f8aaff711
2011	1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Miner	DeepBit.net
2012	1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Miner	DeepBit.net
2013	1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Miner	DeepBit.net
2013	1P49eoo8YgWrdYmMJwo7KYAvyhJYtDfWBg	Mixer	BitcoinFog
2014	1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Miner	DeepBit.net
2014	1P49eoo8YgWrdYmMJwo7KYAvyhJYtDfWBg	Mixer	BitcoinFog
2015	1VayNert3x1KzbpzMGt2qdqrAThiRovi8	Miner	DeepBit.net
2015	1P49eoo8YgWrdYmMJwo7KYAvyhJYtDfWBg	mixer	BitcoinFog
2016	1NxaBCFQwejSZbQfWcYNwgqML5wWoE3rK4	Gambling	LuckyB.it
2016	1changeGhAXKoTEkMntbAe1VHh52jFQhh	Gambling	BitZillions.com
2016	19DhUuwoywejreRPhW9XWXKZTmSRNwud8x	Mixer	HelixMixer-old3
2016	184S3jPkbwS7UJbCUYgL7VKeye5aqSKinF	Darkmarket	AlphaBayMarket
2019	1HckjUpRGcrrRAtFaaCAUaGjsPx9oYmLaZ	Exchange	Huobi.com-2

4.5. k-core decomposition

Table 17 and 18 give the core decomposition of bitcoin users graph. The k-core of a graph is the maximal subgraph in which every vertex has at least degree k. The core decomposition is a set of all k-cores of a graph. Core decompositions are used to study the resilience or robustness of a network [42]. Due to the existence of single entities that captured the majority of all incoming connections, the k-cores had single nodes from 2011-2019. These single nodes were DeepBit.net (2011), SantoshiDice.com (2012-2015), Unknown wallets (2016,2018), Bittrex.com (2017), and Huobi.com-2 (2019).

Table 17: Core decomposition (2009-2015)

	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Cores in LSCC	5	9930	120262	1065542	347630	333420	601493
Cores in LWCC	24	10964	120262	1065542	347630	333420	601493
Cores in full graph	24	10964	120262	1065542	347630	333420	601493

Table 18: Core decomposition (2016-2020)

	2016	**2017**	**2018**	**2019**	**2020**
Cores in LSCC	146836	72718	272896	1154252	0
Cores in LWCC	112356	72718	272896	1154252	704
Cores in full graph	375513	72718	272896	1154252	109080

4.6. Time series analysis of Bitcoin network

Figure 6 gives the fluctuations in the characteristics of Bitcoin network from 2009-2020. To predict the future outlook of the network, time series analysis is performed. The objective of the analysis is to predict the outlook of Bitcoin network for year 2021. Four models were selected for the analysis, the settings are listed:

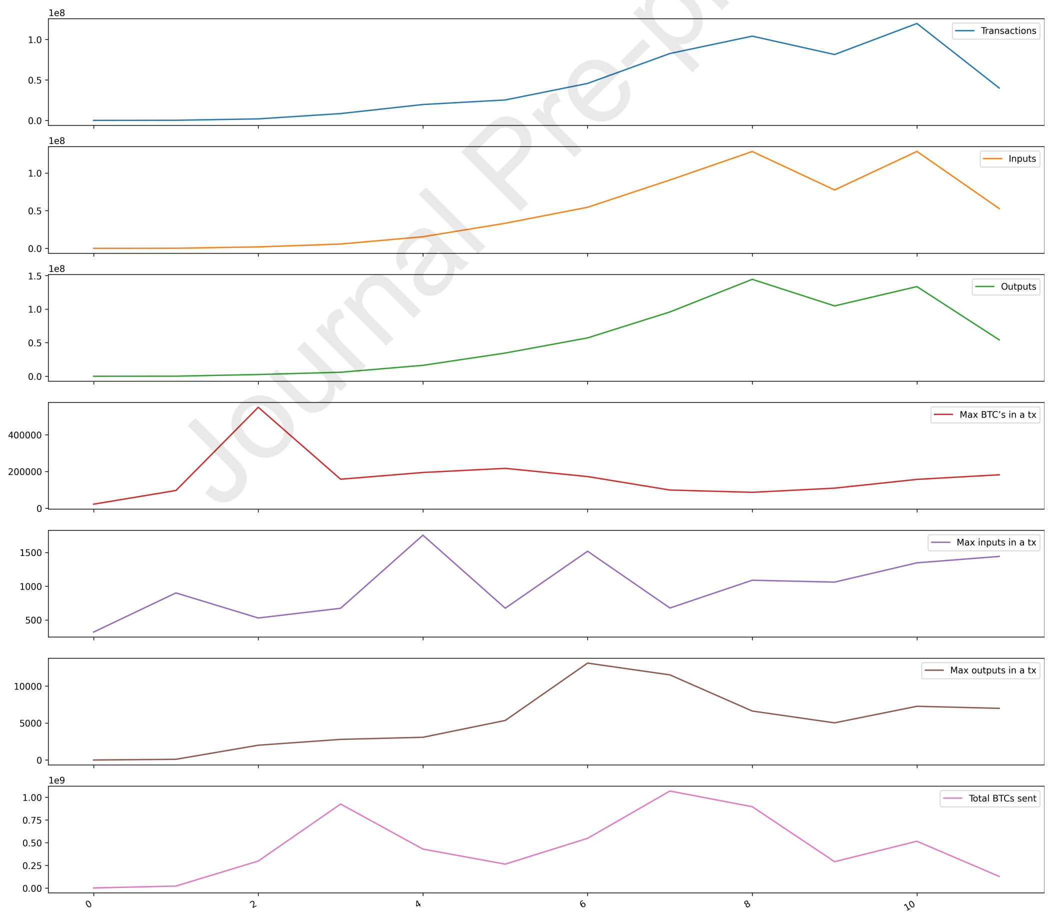


Figure 6: Distribution of transactions in Bitcoin blockchain network (2009-2020)

- Linear regression
- Neural network: Two layers NN (units=64, activation=none)
- Convolutional neural network: Two layers (Filter=32, size=1, stride=1, padding=0)
- LSTM: Single layer (units=32, activation=none)

The four models were trained on a single step, single output time series prediction task on the dataset of Bitcoin network characteristics from 2009-2020 viz. data mentioned in Tables 3-10 and 13-18. Results of four models on validation and test set are illustrated in Figure 7. Comparatively, dense models are better suited for the time series prediction although all four models have mean absolute error close to 0.

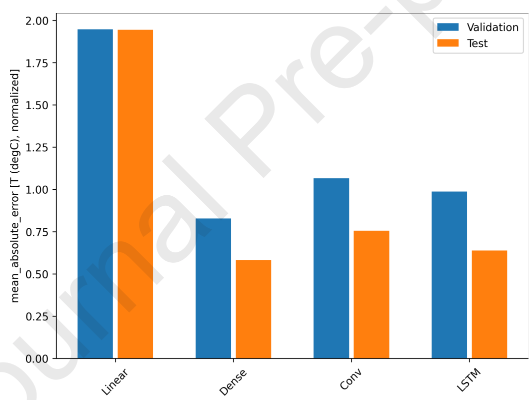


Figure 7: Performance of models on Validation and Test set

Dense model was used to predict the characteristics of the Bitcoin model for the Year 2021 and results of the prediction are given in Table 19. Transactions, inputs, outputs and Max BTC's in a Tx may continue a downward trend seen in Bitcoin networks since 2019. Degree distributions could not be predicted using past data; However, centralization measures, assortativity and reciprocity were in range of previous years. Assortativity shall remain negative and reciprocity low which conforms to standard notions of Bitcoin networks. The LSCC and LWCC in Bitcoin network shall continue to dominate reaching 81% and 99% of the total network size respectively. Cores in full graphs will see a decline to 2018 levels. Overall, it can be concluded that data-driven time series analysis observes normalcy will be restored in the Bitcoin network in the year 2021 from the 2019 all time highs.

Table 19: Prediction of Bitcoin network for Year 2021

Year	**Transactions**	**Inputs**	**Outputs**	**Max BTCs in a tx**
2021	17916462.0	19343784	134251.34	15666966.0
Max inputs in a tx	**Max outputs in a tx**	**Total BTCs sent**	**Vertex count**	**Edge count**
1176.0	2485	9928711	269887	2283282
Edge density (S)	**Edge density (D)**	**Power law \alpha in**	**Exp lambda in**	**Lognormal \mu in**
4.39e-06	3.64e-06	0.034	0.13	0.83
Lognormal alpha in	**Poisson in**	**Power law \alpha out**	**Exp lambda out**	**Lognormal \mu out**
0.46	0.53	1.15	-0.11	0.2
Lognormal alpha out	**Poisson out**	**Diameter**	**Avg path length**	**Reciprocity**
-0.07	-1.01	6.7e-02	4.4e-02	3.8e-02
Assortativity	**Centralization**	**Cd**	**Triangles (LSCC)**	**Nodes (LSCC)**
-0.2	0.99	4.7e-02	8.6e+06	1.5e+05
Edges (LSCC)	**AP (LSCC)**	**C (LSCC)**	**Triangles (LWCC)**	**Nodes (LWCC)**
5.1e+06	5.9e-03	5.2e+04	2.9e+07	1.8e+07
Edges (LWCC)	**AP (LWCC)**	**C (LWCC)**	**Triangles (Full)**	**Nodes (Full)**
6.5e+07	3.7e+06	1.4e-04	1.8e+07	2.4e+06
Edges (Full)	**Cores (LSCC)**	**Cores (LWCC)**	**Cores (Full)**	
7.3e+05	2.5e+05	1.65e+05	4e+05	

4.7. Summary of Results with Discussion and lessons learnt

- The edge density is low in both the directed graph (Edge density (D)) and the undirected graph (Edge density (S)) for the period 2009-2020 compared to social networks

- 99.8% of the total users in 2009 made at the most a single transaction this declined to 73.24% by 2020.

- Even bitcoin users graph followed the "scale-free" property as power-law exponent ranged from 1.54-2.4 for in-degree distribution and from 1.42-2.7 for out-degree distribution

- LWCC and LSCC were formed mainly because of mixing services, gambling services, and crypto exchanges.

- k-cores had single nodes from 2011-2019

Comparing complex networks with bitcoins users graph, it is seen that it shares certain features with the Ethereum network. Unlike social networks (Twitter, Facebook, Actors, Directors, Co-authorship, citation), it has no giant LSCC but follows properties of "scale-free" networks.

Table 20: Comparison with other complex networks

Complex network	Hubs?	Assortativity	Small diameter?	C	Degree distribution	Giant LSCC	Edge density
Bitcoin	Yes	(-)	Yes	Low	Power law	No	Low
Citation	NA	(-)	NA	Low	Power law	NA	Low
WWW	Yes	(+)	Yes	Low	Power law	Yes	Low
Social networking	Yes	(-)	Yes	High	Power law	Yes	High
Protein-Protein	NA	(+)	NA	Low	Power law	NA	Low
Co-authorship	NA	(+)	NA	Low	No power law	NA	Low
Ethereum	Yes	NA	Yes	Low	Power law	Yes	Low
Film actors	NA	NA	NA	NA	Power law	NA	Low
Company directors	NA	NA	NA	NA	No power law	NA	Low

With the use of deanonymizing and network analysis, common types of services on Bitcoin network datasets were able to be identified. These are listed as follows:

- Exchanges: Allow trading of BTC to fiat currencies
- Pools: Individual users combine their processing power for mining blocks
- Gambling: Allow placing of bets using BTCs
- Wallets: Store BTC private keys and balance
- Payment gateways: Allow accepting payment for services in BTCs
- Miner: Organizations competing to mine blocks
- Darknet markets: Selling and buying goods using BTCs
- Mixers: Remove traceability of BTCs from source
- Trading sites: Purchase equities using BTCs
- P2Plenders: Crowdsourcing BTCs for loans
- Faucets: Reward in BTCs to subscribers
- Explorer: Educational websites provide API to explore Bitcoin
- P2PMarket: Marketplace for second-hand goods where buyers can contact sellers, payments in BTCs
- Bond markets: Buying bonds or debt instruments in BTC

- Affiliate marketers: Pay per click in BTC
- Video sharing: Payment in BTCs for viewing videos
- Money launderers: Convert fiat currencies to BTC
- Cyber-security providers: Provide cybersecurity products for BTC
- Cyber-criminals: Blacklisted by governments
- Ponzi: High yield investment scams

To build a system for detection of these entities in Bitcoin network and aid forensic tools, network analysis conducted in the current paper identified discriminating features. Feature list is given in Table 21. These features can be used to build a classifier for detecting or identifying illegal activities or users in Bitcoin.

Table 21: List of Features

Feature symbol	Feature description
T_x	Total transactions in which wallet has participated
B	Current BTC present in the wallet
${T_x}^{in}$	Total incoming transactions to the wallet
${T_x}^{out}$	Total outgoing transactions from the wallet
L	Total active life of the wallet
A_w	Total addresses of the wallet
A_v	Average number of incoming transactions received by an address of a wallet
T	Total number of addresses sending BTC to the wallet
R	Ratio of Transaction count and address count. Gives the average number of times an address of the wallet was reused for a transaction.

5. Conclusion and Future works

Since its launch in 2009, Bitcoin has seen a steady increase in its user base and transactions, both volume and value. As it aims to promote the exchange of value without reliance on a trusted third party, it could be speculated that the network form of the Bitcoin system should be decentralized and disconnected without any giant connected component. This would mean a robust structure. However, in reality, there are connected components in the bitcoin users graph. These components have emerged due to gambling websites, mixing services, crypto trading exchanges, and mining pools. These services have been easier to identify due to the high incoming and outgoing connections they have with other bitcoin users. From 2011, these entities have created giant connected components in bitcoin users graph. A result of their presence was a reduction in diameter, average path length, and radius. Additionally, "scale-free" property, was observed in bitcoin users graph as preferential attachment occurred.

The blanket of anonymity and secrecy provided by Bitcoin has made it difficult to label each and every address with a label. However, network analysis can shed light on this confidentiality and reveal the nature of the bitcoin user. There is no straightforward application of network analysis on bitcoin data as bitcoin users are identified by addresses, and a single user can have multiple addresses. This issue of multiple identities is not seen in other networks. Heuristic clustering, such as combining multi-inputs to a single transaction as a single entity, can reduce this issue to some extent and hence is commonly used in bitcoin network studies.

Even with clustering and network analysis without labeled datasets, limited progress can be made in tracing entities on the Bitcoin network. To overcome this drawback, features related to each entity can be extracted from the blockchain to train a supervised learning technique for identifying unknown wallets.

Bitcoin scenario has changed drastically in the last 3 months - e.g. Feb 20, 2020 - BTC @10k USD, March 12, 2020 - BTC@4k USD, April 2020 - BTC@6k-9k, May 8 - BTC again @10k (reward halving will be happening on 11 May 2020). BTC is detaching itself from linearity of cryptocurrency market (i.e. Since last 3 months, BTC and ETH were going neck to neck in terms of percentage pricing variation). This detachment may be because of the following considerations: Pandemic Work From Home culture created opportunity for people to shift focus on stock markets and cryptocurrency markets. BTC is reemerged as a parking heaven (hedging / protection against inflation) - due to USD influx of 7 Trillion - COVID 19 stimulus printing of money - and other bailouts by governments across the World. India legalized crypto currencies from March 2020 first week (after a ban of about 2 years) - and market started buzzing with large number of new players/small investors. Steady emergence of Internet of Trusted Things - which sees blockchain as a platform to build trust.

Acknowledgment

This work was supported in part by the Raman Charpak Fellowship of the Indo-French Centre for the Promotion of Advanced Research Grant no: IFC/4132/RCF 2019/716. The authors thank VJTI Mumbai and IMT Atlantique, France for providing the lab resources. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

[1] S. Park, S. Im, Y. Seol, J. Paek, Nodes in the bitcoin network: comparative measurement study and survey, IEEE Access 7 (2019) 57009–57022.

[2] Q. Feng, D. He, S. Zeadally, M. K. Khan, N. Kumar, A survey on privacy protection in blockchain system, Journal of Network and Computer Applications 126 (2019) 45 – 58.

[3] L. Wang, X. Shen, J. Li, J. Shao, Y. Yang, Cryptographic primitives in blockchains, Journal of Network and Computer Applications 127 (2019) 43 – 58.

[4] M. Rahouti, K. Xiong, N. Ghani, Bitcoin concepts, threats, and machine-learning security solutions, IEEE Access 6 (2018) 67189–67205.

[5] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Technical Report, Manubot, 2019.

[6] S. Aggarwal, R. Chaudhary, G. S. Aujla, N. Kumar, K.-K. R. Choo, A. Y. Zomaya, Blockchain for smart communities: Applications, challenges and opportunities, Journal of Network and Computer Applications 144 (2019) 13 – 48.

[7] A. A. Monrat, O. Schelén, K. Andersson, A survey of blockchain from the perspectives of applications, challenges, and opportunities, IEEE Access 7 (2019) 117134–117151.

[8] A. Ghosh, S. Gupta, A. Dua, N. Kumar, Security of cryptocurrencies in blockchain technology: State-of-art, challenges and future prospects, Journal of Network and Computer Applications 163 (2020) 102635.

[9] R. Böhme, N. Christin, B. Edelman, T. Moore, Bitcoin: Economics, technology, and governance, Journal of economic Perspectives 29 (2015) 213–38.

[10] V. G. Reyes-Macedo, M. Salinas-Rosales, G. G. Garcia, A method for blockchain transactions analysis, IEEE Latin America Transactions 17 (2019) 1080–1087.

[11] K. Toyoda, P. T. Mathiopoulos, T. Ohtsuki, A novel methodology for hyip operators' bitcoin addresses identification, IEEE Access 7 (2019) 74835–74848.

[12] I. Alqassem, I. Rahwan, D. Svetinovic, The anti-social system properties: Bitcoin network data analysis, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2018).

[13] X. T. Lee, A. Khan, S. S. Gupta, Y. H. Ong, X. Liu, Measurements, analyses, and insights on the entire ethereum blockchain network (2019).

[14] F. Tschorsch, B. Scheuermann, Bitcoin and beyond: A technical survey on decentralized digital currencies, IEEE Communications Surveys & Tutorials 18 (2016) 2084–2123.

[15] D. D. F. Maesa, A. Marino, L. Ricci, The bow tie structure of the bitcoin users graph, Applied Network Science 4 (2019) 56.

[16] D. D. F. Maesa, A. Marino, L. Ricci, The graph structure of bitcoin, in: International Conference on Complex Networks and their Applications, Springer, 2018, pp. 547–558.

[17] D. D. F. Maesa, A. Marino, L. Ricci, Data-driven analysis of bitcoin properties: exploiting the users graph, International Journal of Data Science and Analytics 6 (2018) 63–80.

[18] D. D. F. Maesa, A. Marino, L. Ricci, Uncovering the bitcoin blockchain: an analysis of the full users graph, in: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2016, pp. 537–546.

[19] A.-L. Barabási, et al., Network science, Cambridge university press, 2016.

[20] X. Fu, H. Yao, O. Postolache, Y. Yang, Message forwarding for wsn-assisted opportunistic network in disaster scenarios, Journal of Network and Computer Applications 137 (2019) 11–24.

[21] X. Fu, G. Fortino, W. Li, P. Pace, Y. Yang, Wsns-assisted opportunistic network for low-latency message forwarding in sparse settings, Future Generation Computer Systems 91 (2019) 223–237.

[22] X. Fu, G. Fortino, P. Pace, G. Aloi, W. Li, Environment-fusion multipath routing protocol for wireless sensor networks, Information Fusion 53 (2020) 4–19.

[23] N. Szabo, Bit gold, 1970. URL: https://unenumerated.blogspot.com/2005/12/bit-gold.html.

[24] Y. Li, U. Islambekov, C. Akcora, E. Smirnova, Y. R. Gel, M. Kantarcioglu, Dissecting ethereum blockchain analytics: What we learn from topology and geometry of ethereum graph, arXiv preprint arXiv:1912.10105 (2019).

[25] H. Sun, N. Ruan, H. Liu, Ethereum analysis via node clustering, in: International Conference on Network and System Security, Springer, 2019, pp. 114–129.

[26] S. Ferretti, G. D'Angelo, On the ethereum blockchain structure: A complex networks theory perspective, Concurrency and Computation: Practice and Experience (2019) e5493.

[27] P. Nerurkar, M. Chandane, S. Bhirud, Empirical analysis of synthetic and real networks, International Journal of Information Technology (2019) 1–13.

[28] P. Nerurkar, M. Chandane, S. Bhirud, Understanding structure and behavior of systems: a network perspective, International Journal of Information Technology (2019) 1–15.

[29] J. Leskovec, A. Krevl, SNAP Datasets: Stanford large network dataset collection, http://snap.stanford.edu/data, 2014.

[30] M. E. Newman, The structure and function of complex networks, SIAM review 45 (2003) 167–256.

[31] M. Golosovsky, Preferential attachment mechanism of complex network growth:" rich-gets-richer" or" fit-gets-richer"?, arXiv preprint arXiv:1802.09786 (2018).

[32] J. MO, Social and economic networks, Princeton university press, 2010.

[33] S. Fortunato, D. Hric, Community detection in networks: A user guide, Physics Reports 659 (2016) 1–44.

[34] L. A. N. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, Classes of small-world networks, Proceedings of the national academy of sciences 97 (2000) 11149–11152.

[35] D. Ding, M. Conti, R. Figueiredo, Impact of country-scale internet disconnection on structured and social p2p overlays, in: 2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE, 2015, pp. 1–9.

[36] D. Ding, M. Conti, R. Figueiredo, Wide-scale internet disconnection: impact and recovery on social-based p2p overlays, IEEE Transactions on Network Science and Engineering 6 (2018) 734–747.

[37] A.-L. Barabási, Network science, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (2013) 20120375.

[38] A. Janda, Walletexplorer. com: Smart bicoin block explorer, 2016.

[39] G. Csardi, T. Nepusz, et al., The igraph software package for complex network research, InterJournal, complex systems 1695 (2006) 1–9.

[40] C. S. Gillespie, Fitting heavy tailed distributions: the powerlaw package, arXiv preprint arXiv:1407.3492 (2014).

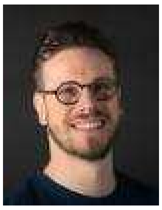
[41] M. O. Jackson, Social and economic networks, Princeton university press, 2010.

[42] F. D. Malliaros, C. Giatsidis, A. N. Papadopoulos, M. Vazirgiannis, The core decomposition of networks: theory, algorithms and applications, The VLDB Journal 29 (2020) 61–92.

Pranav Nerurkar is a PhD research scholar in Dept of CE & IT, VJTI, and Assistant Professor in NMIMS University. He worked as Raman charpak fellow in SRCD Department IMT Atlantique, France from March 2020 – September 2020. His research topics are mainly related to Deep learning and Data mining.

Dhiren Patel is Professor in Dept of CE & IT, VJTI and Director of VJTI. His research topics are mainly related to Blockchain, Cryptography and Cyber security in networks.

Romaric Ludinard is currently associate professor at IMT Atlantique. Before that, was associate professor at ENSAI, the national school for Statistics and Information Analysis during two years. was head of the Data scientist specialization, and head of the computer science team. Before that, was research and teaching assistant at University of Rennes during one year. His research mainly focuses on models, safety and performance of distributed systems. He felt in these fields just after having received my engineering degree in computer science from University of Rennes 1 in 2006. At this time, He has started working on fault tolerance in the specific field of intrusion detection systems in the context of DADDi project. Then, worked on byzantine fault tolerance in large scale distributed systems in the context of P2PIm@ges project. In 2010, joined Supelec for one year to work on anomaly based intrusion detection systems in the context of DALI project. In 2011, was funded by Inria and Technicolor to prepare my phD on fault characterization in large scale networks. Currently focusing research on Blockchain protocols and their guarantees.



Yann Busnel Full Professor at IMT Atlantique, Bretagne-Pays de la Loire. Since January 1st 2017, He is also head of SRCD department (Network Systems, Cybersécurity and Digital Law), on Rennes Campus. After PhD in Computer Sciences at University of Rennes (France) in November 2008, he spent one year at Université "La Sapienza" in Roma, Italy (2008-2009). Then, He has been assistant professor (Maître de Conférences) at Université de Nantes, in the computer sciences department at UFR Sciences et Techniques (2009-2014). Following, He has been associate-professor at ENSAI (National Graduate School on Statistics and Data Analysis), where He has acted as Head of the Computer Sciences department (2014-2016). He obtained his "Habilitation à diriger les recherches" (HDR) in Computer Sciences from École Normale Supérieure de Rennes (France) in December 2016. His research topics are mainly related to Data stream analysis over Big Data and Model for large-scale distributed systems and networks.

Saru Kumari is currently an Assistant Professor with the Department of Mathematics, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India. She received her Ph.D. degree in Mathematics in 2012 from Chaudhary Charan Singh University, Meerut, UP, India. She has published more than 200 research papers in reputed International journals and conferences, including more than 180 research papers in various SCI-Indexed Journals. She is on the editorial board of more than a dozen of International Journals, of high repute, under Elsevier, Springer, Wiley and others including SCI journals such as IEEE Systems, AEÜ - International Journal of Electronics and Communications, Elsevier (SCI); International Journal of Communication Systems, Wiley (SCI-E); Telecommunication Systems, Springer (SCI); Human Centric Computing and Information sciences, Springer (SCI-E); Transactions on Emerging Telecommunications Technologies; Wiley (SCI-E), Information Technology and Control, Kaunas University of Technology, Lithuania (SCI-E); KSII Transactions on Internet and Information Systems (SCI-E), published from Taiwan; Information Security: A Global Perspective, Taylor & Francis (ESCI, Scopus); International Journal of Wireless Information Networks (ESCI, Scopus), Springer; Journal of Reliable Intelligent Environments, Springer (ESCI, Scopus); etc. She has served as the Guest Editor of many special issues in SCI Journals under IEEE, Elsevier, Springer and Wiley. She has been involved in the research community as Technical Program Committee (TPC) member or PC chair for more than a dozen of International conferences of high repute. She is also serving as a reviewer of dozens of reputed Journals including SCI-Indexed of IEEE, Elsevier, Springer, Wiley, Taylor & Francis, etc. Her current research interests include information security and applied cryptography.

Muhammad Khurram Khan is currently a Full Professor with the Center of Excellence in Information Assurance (CoEIA), King Saud University, Saudi Arabia. He is one of the founding members of CoEIA and has served as the Manager R&D, from 2009 to 2012. He has published more than 325 research papers in the journals and conferences of international repute. In addition, he is an inventor of ten US/PCT patents. He has edited seven books/proceedings published by Springer-Verlag and IEEE. His research interests include cybersecurity, digital authentication, biometrics, multimedia security, and technological innovation management. He has secured an outstanding leadership award at IEEE International Conference on Networks and Systems Security 2009, Australia. He received the Gold Medal for the Best Invention & Innovation Award at 10th Malaysian Technology Expo 2011, Malaysia. Moreover, his invention recently got a Bronze Medal at '41st International Exhibition of Inventions', Geneva, Switzerland, in 2013. In addition, he received the Best Paper Award from the Journal of Network & Computer Applications (Elsevier), in 2015. He was a recipient of the King Saud University Award for Scientific Excellence (Research Productivity), in 2015. He was also a recipient of the King Saud University Award for Scientific Excellence (Inventions, Innovations, and Technology Licensing), in 2016. He is the Editor-in-Chief of a well-esteemed international journal Telecommunication Systems (Springer-Verlag, since 1993) with an impact factor of 1.542 (JCR 2017). He is also the full-time Editor/Associate Editor of several international journals/magazines. He has also played role of the guest editor of several international ISI-indexed journals of Springer-Verlag and Elsevier Science. Moreover, he is one of the organizing chairs of more than five dozen international conferences and member of technical committees of more than ten dozen international conferences.