Blockchain Data Analysis

Daniel Rincon
@dsrincon

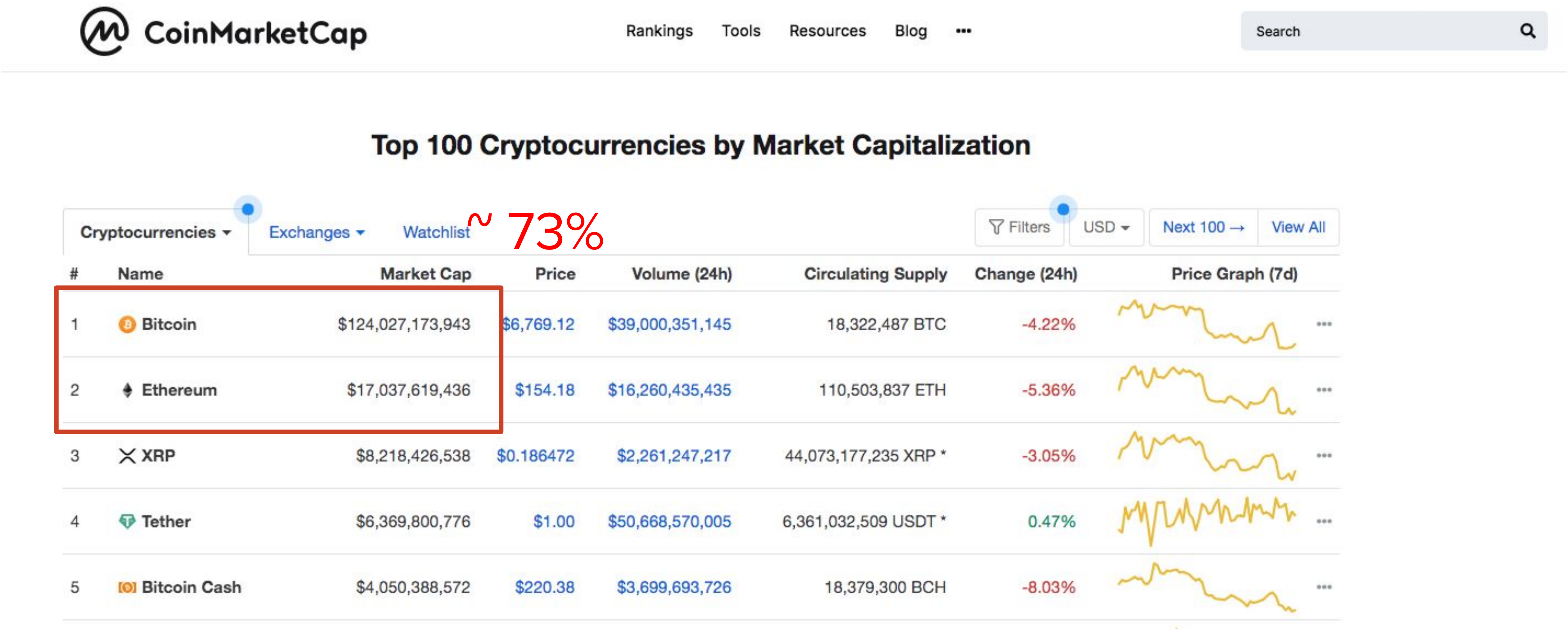
BLOCKCHAIN
AT BERKELEY

LECTURE OUTLINE

- What does Blockchain data tell us?
  - What is Blockchain?
  - Who uses Blockchain?
  - How is Blockchain used?
- Accessing Blockchain data
  - Online explorers
  - Raw data
- Analyzing Blockchain data
  - Defining a graph
  - Basics of graph properties
  - Network analysis libraries
- Demo/Homework

What is a Blockchain?

A digital asset

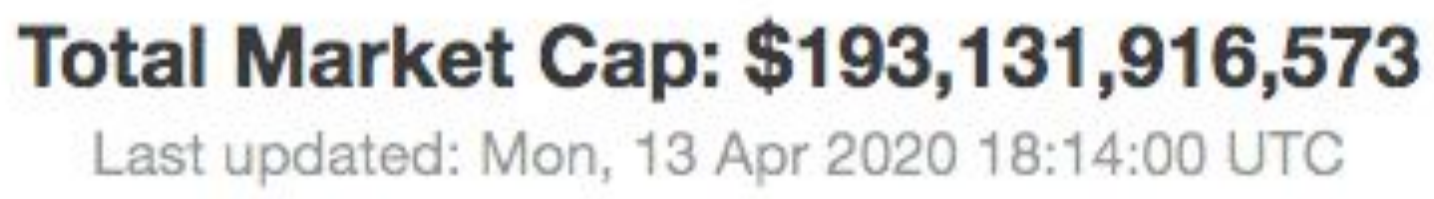


Top 100 Cryptocurrencies by Market Capitalization

~ 73%

#	Name	Market Cap	Price	Volume (24h)	Circulating Supply	Change (24h)	Price Graph (7d)
1	Bitcoin	$124,027,173,943	$6,769.12	$39,000,351,145	18,322,487 BTC	-4.22%	
2	Ethereum	$17,037,619,436	$154.18	$16,260,435,435	110,503,837 ETH	-5.36%	
3	XRP	$8,218,426,538	$0.186472	$2,261,247,217	44,073,177,235 XRP *	-3.05%	
4	Tether	$6,369,800,776	$1.00	$50,668,570,005	6,361,032,509 USDT *	0.47%	
5	Bitcoin Cash	$4,050,388,572	$220.38	$3,699,693,726	18,379,300 BCH	-8.03%	

Source: coinmarketcap.com



Total Market Cap: $193,131,916,573

Last updated: Mon, 13 Apr 2020 18:14:00 UTC

vs.

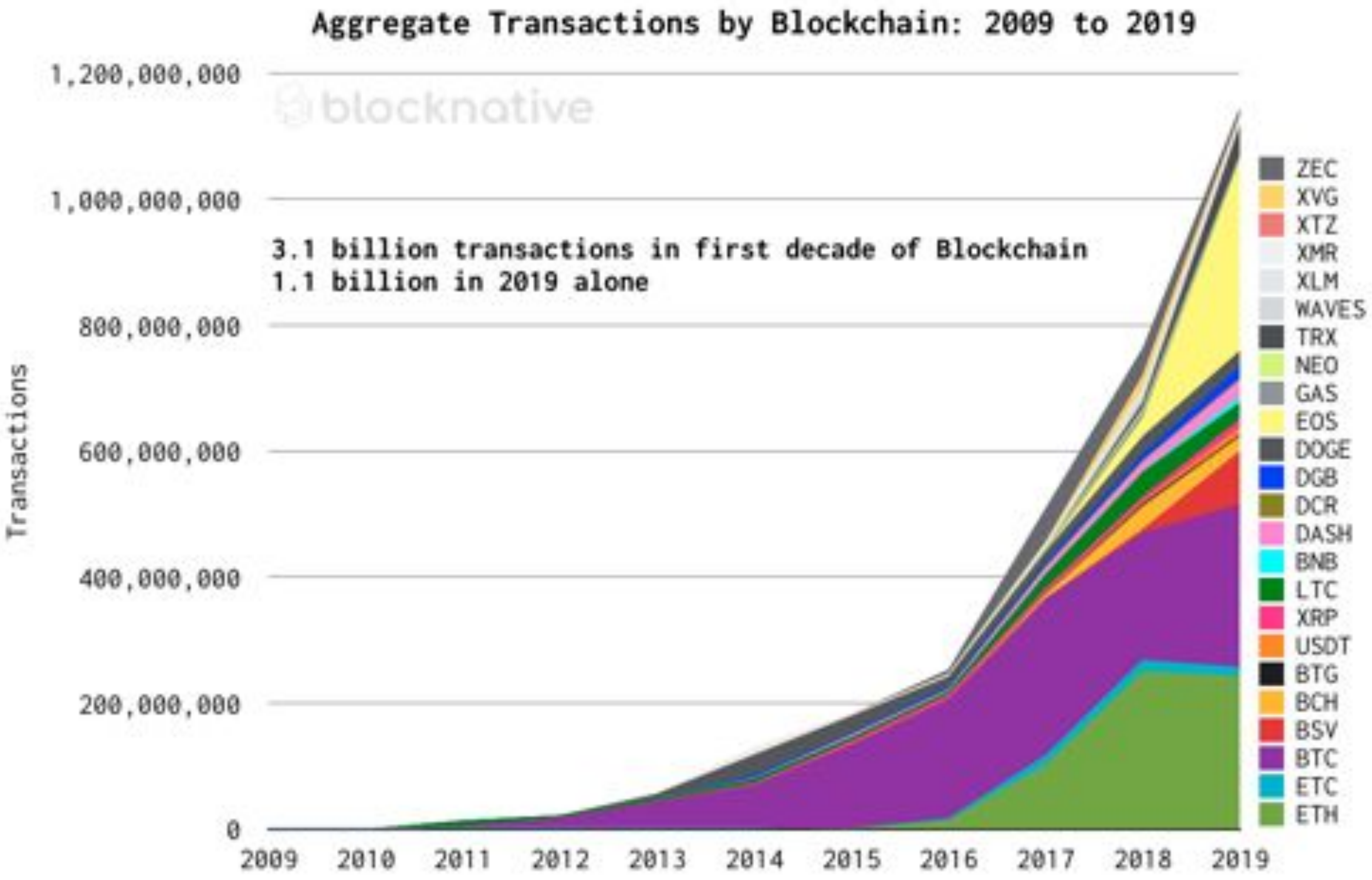
~ 50x

Nasdaq 100: 9.6 Trillion
(13 Apr 2020)

BLOCKCHAIN AT BERKELEY

What is a Blockchain?

A payment network



3.1 billion transactions in first decade of Blockchain
1.1 billion in 2019 alone

Source: blocknative.com

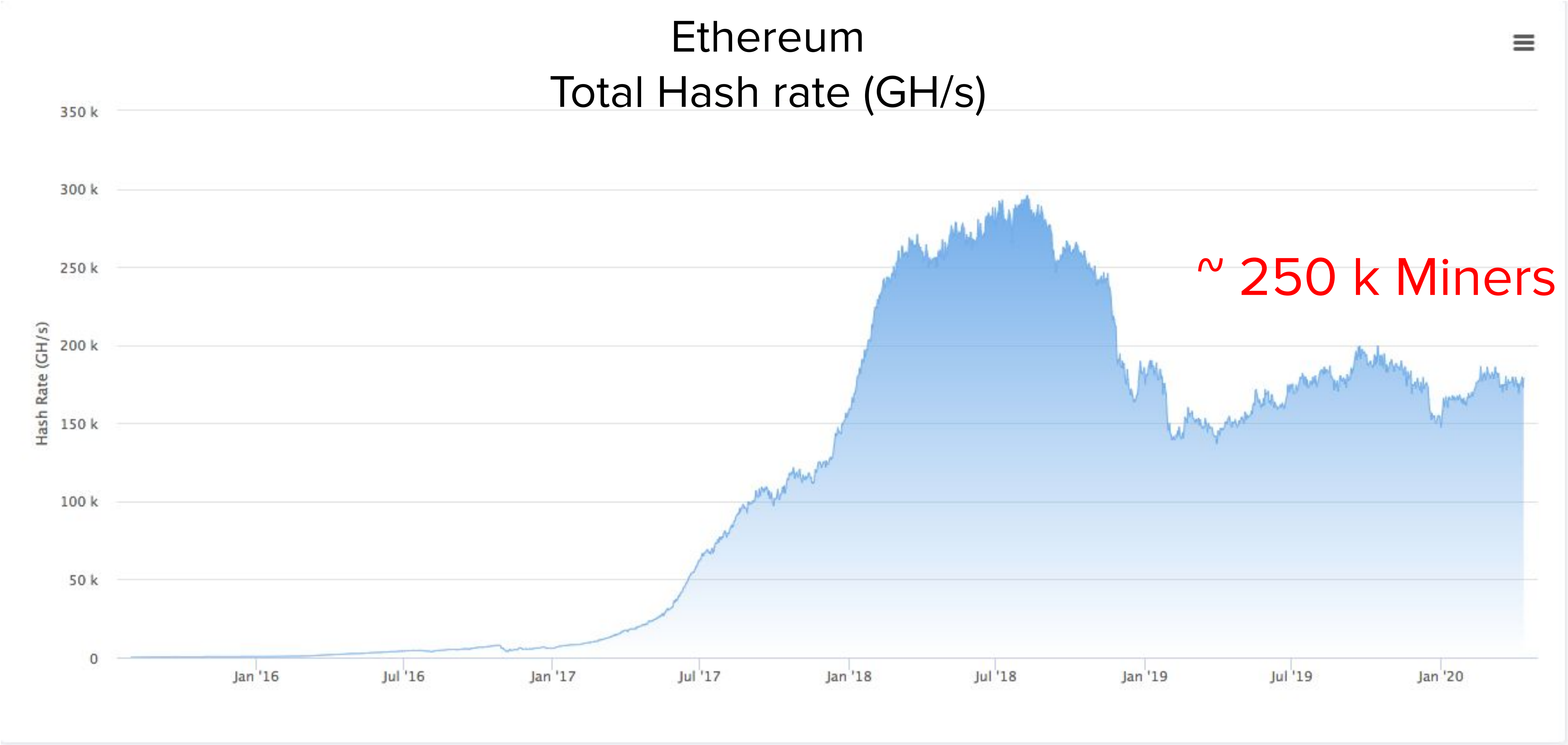
1.1 Billion Blockchain transactions 2019

vs.

150M Daily Visa transactions

What is a Blockchain?

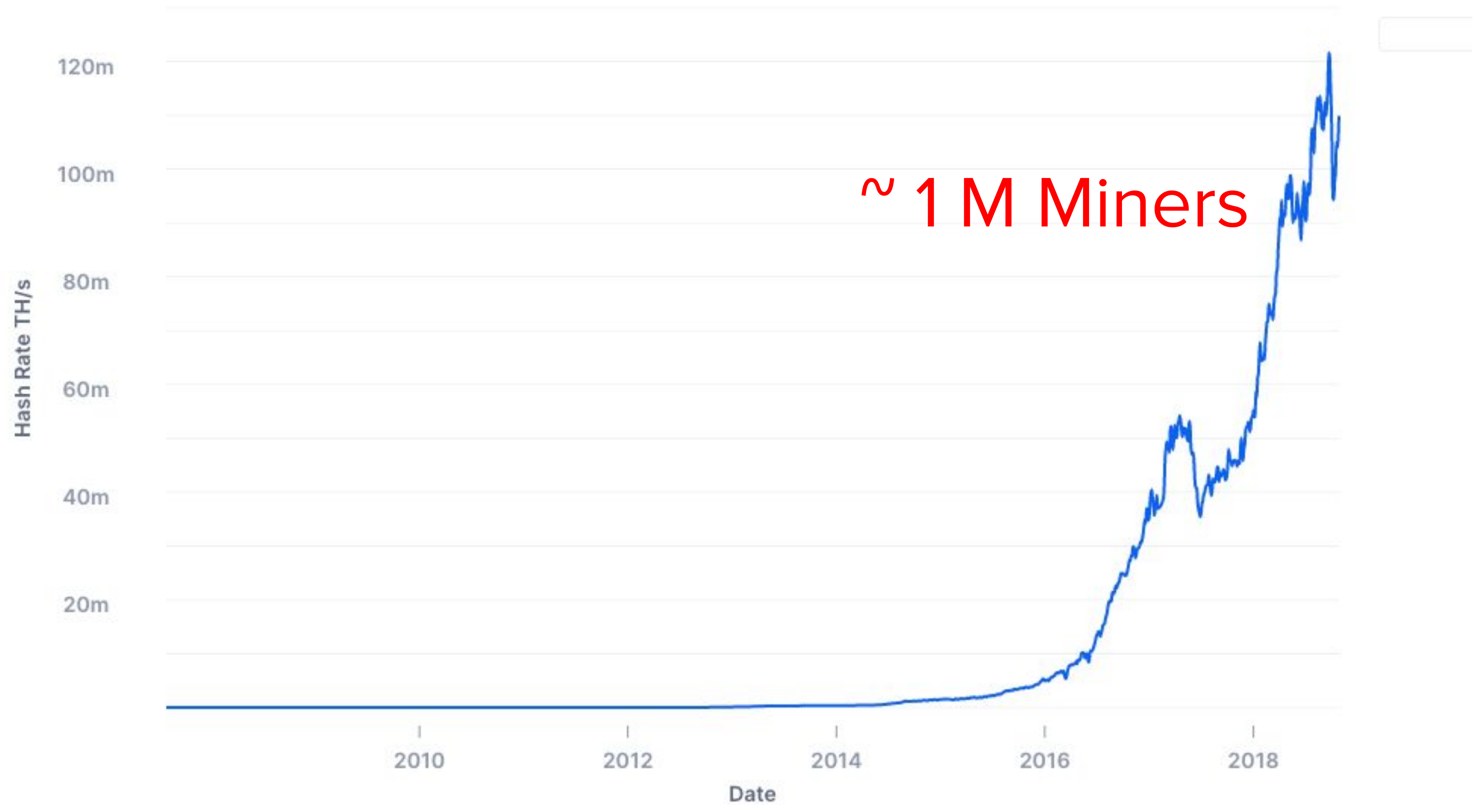
Computer Network and Distributed Database


Ethereum
Total Hash rate (GH/s)

~ 250 k Miners

Source: etherscan.io

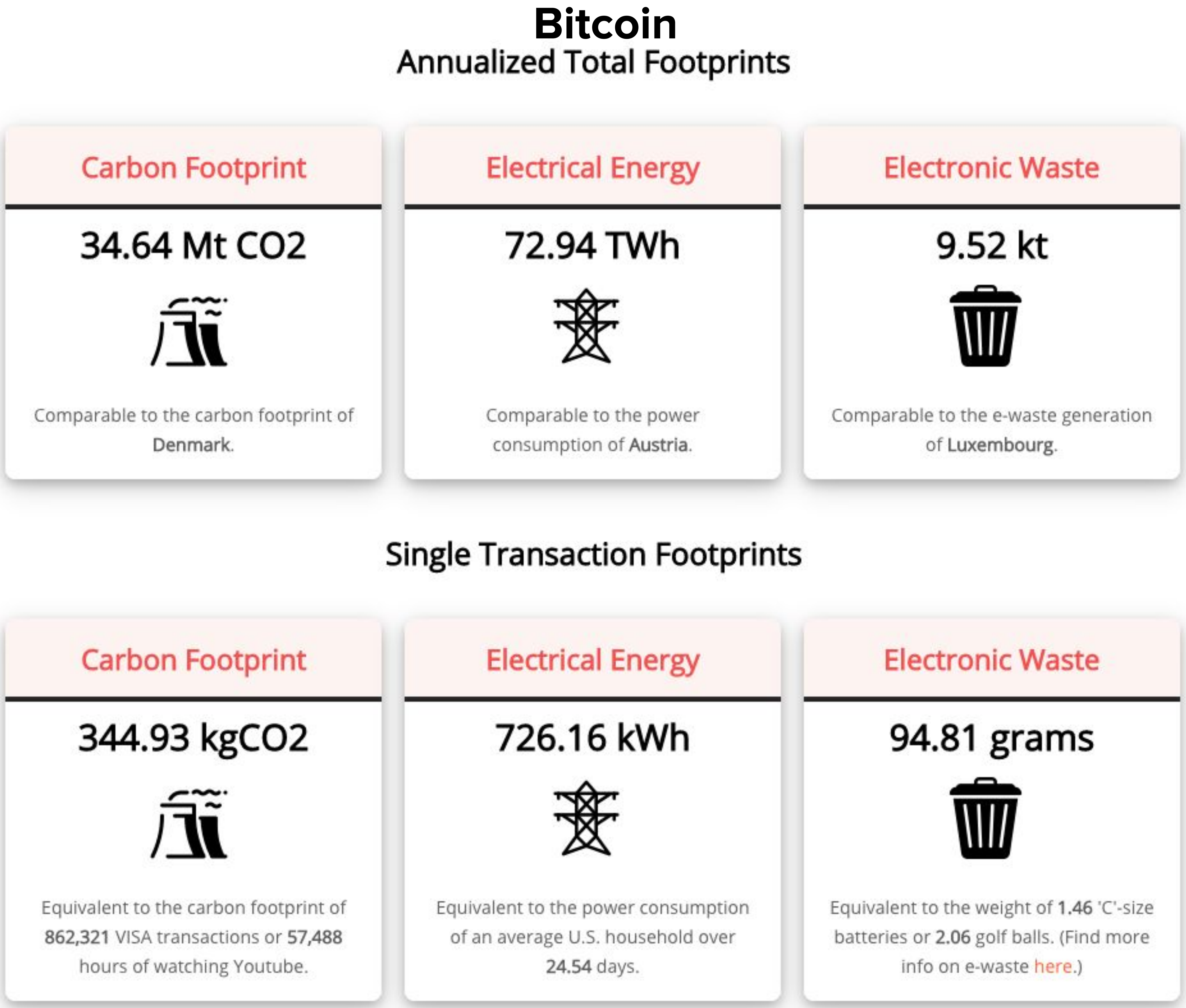
Bitcoin
Total Hash rate (TH/s)


~ 1 M Miners

Source: blockchain.com

What is Blockchain?

A gargantuan energy consumption machine



Bitcoin

Annualized Total Footprints

Carbon Footprint	Electrical Energy	Electronic Waste
34.64 Mt CO2	72.94 TWh	9.52 kt
Comparable to the carbon footprint of **Denmark**.	Comparable to the power consumption of **Austria**.	Comparable to the e-waste generation of **Luxembourg**.

Single Transaction Footprints

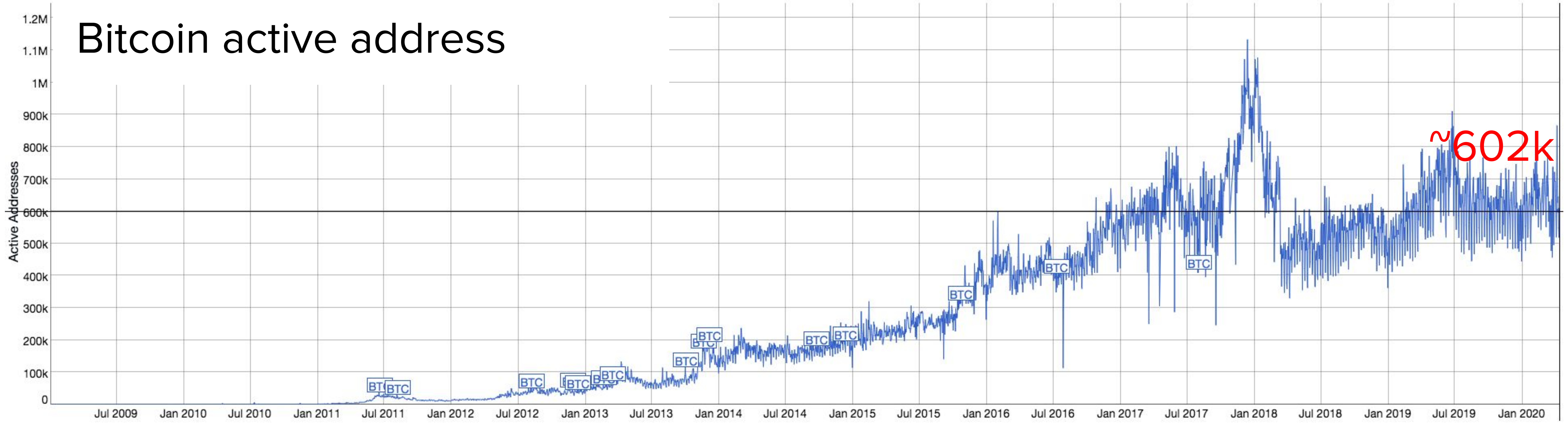
Carbon Footprint	Electrical Energy	Electronic Waste
344.93 kgCO2	726.16 kWh	94.81 grams
Equivalent to the carbon footprint of **862,321** VISA transactions or **57,488** hours of watching Youtube.	Equivalent to the power consumption of an average U.S. household over **24.54** days.	Equivalent to the weight of **1.46** 'C'-size batteries or **2.06** golf balls. (Find more info on e-waste here.)

Source: digiconomist.net

Estimates say Ethereum consumes 25-50% of this.

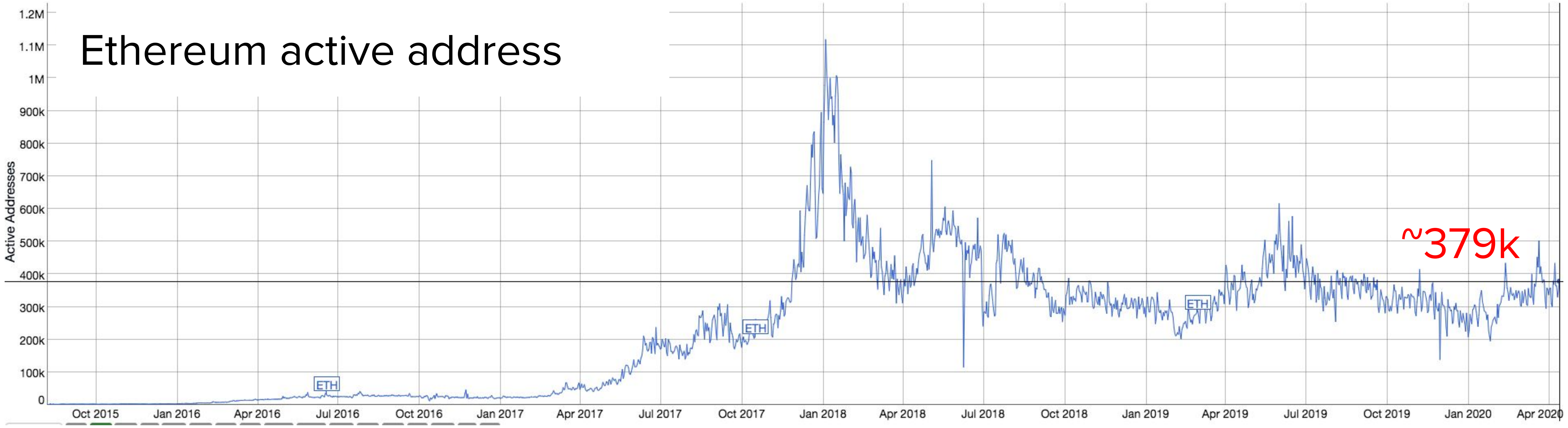
Who uses Blockchain?

Account holders



Bitcoin active address

~602k



Ethereum active address

~379k

Source: bitinfocharts.com

US Crypto Holders:
36.5 Million
(2019)*

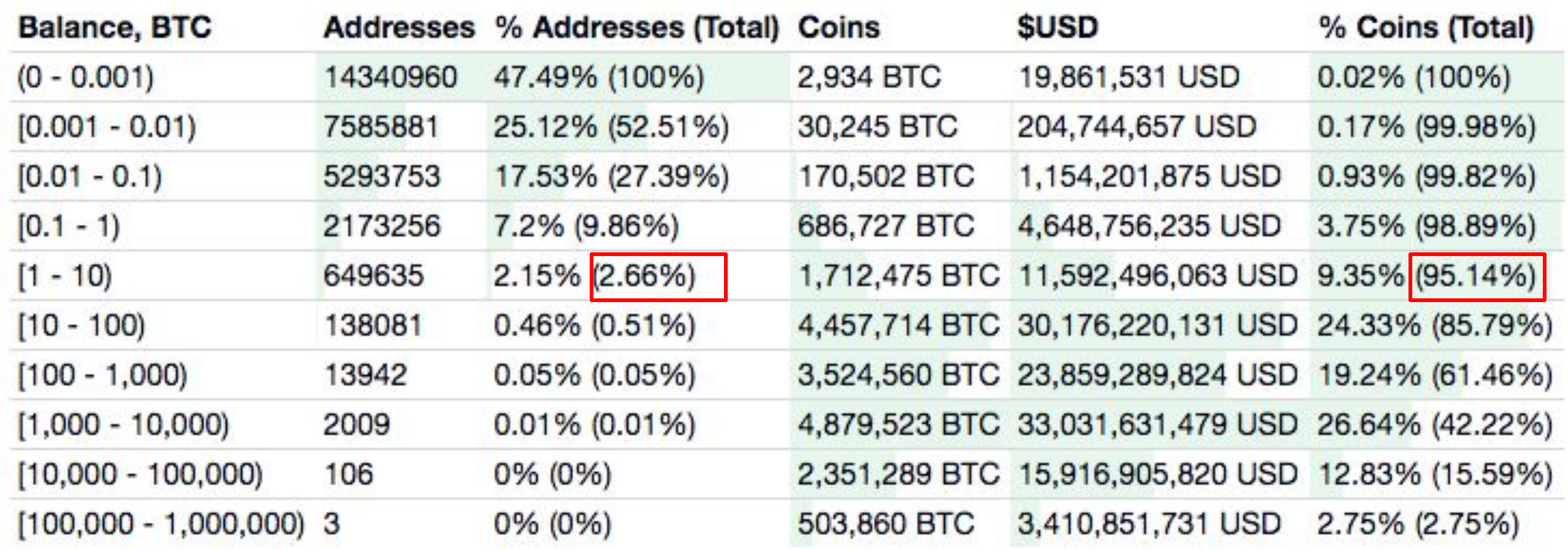
Chase Digital active users:
51 Million
(2019)**

PayPal active users:
305 Million
(2019)***

Who uses Blockchain?

The new banks?

Bitcoin ownership distribution



Balance, BTC	Addresses	% Addresses (Total)	Coins	$USD	% Coins (Total)
(0 - 0.001)	14340960	47.49% (100%)	2,934 BTC	19,861,531 USD	0.02% (100%)
[0.001 - 0.01)	7585881	25.12% (52.51%)	30,245 BTC	204,744,657 USD	0.17% (99.98%)
[0.01 - 0.1)	5293753	17.53% (27.39%)	170,502 BTC	1,154,201,875 USD	0.93% (99.82%)
[0.1 - 1)	2173256	7.2% (9.86%)	686,727 BTC	4,648,756,235 USD	3.75% (98.89%)
[1 - 10)	649635	2.15% (2.66%)	1,712,475 BTC	11,592,496,063 USD	9.35% (95.14%)
[10 - 100)	138081	0.46% (0.51%)	4,457,714 BTC	30,176,220,131 USD	24.33% (85.79%)
[100 - 1,000)	13942	0.05% (0.05%)	3,524,560 BTC	23,859,289,824 USD	19.24% (61.46%)
[1,000 - 10,000)	2009	0.01% (0.01%)	4,879,523 BTC	33,031,631,479 USD	26.64% (42.22%)
[10,000 - 100,000)	106	0% (0%)	2,351,289 BTC	15,916,905,820 USD	12.83% (15.59%)
[100,000 - 1,000,000)	3	0% (0%)	503,860 BTC	3,410,851,731 USD	2.75% (2.75%)

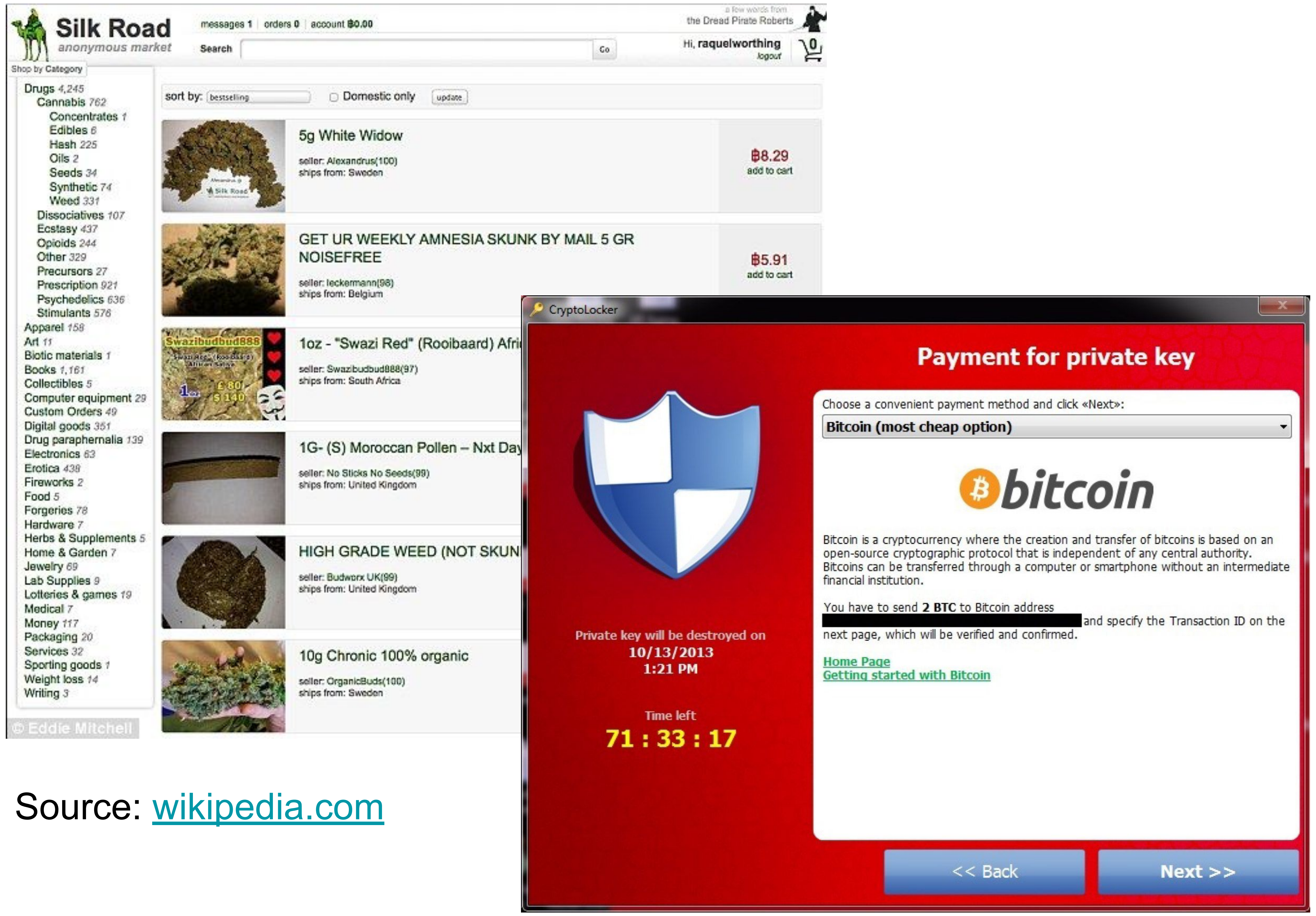
Source: bitinfocharts.com

Less than 3% of addresses own over 95% of all Bitcoins...

..many of them belong to exchanges.

Who uses Blockchain?

Following the money trail



Silk Road
anonymous market

messages 1 | orders 0 | account ฿0.00

Search [ ] Go

a few words from
the Dread Pirate Roberts

Hi, raquelworthing
logout

0

Shop by Category

Drugs 4,245
Cannabis 762
Concentrates 1
Edibles 6
Hash 225
Oils 2
Seeds 34
Synthetic 74
Weed 331
Dissociatives 107
Ecstasy 437
Opioids 244
Other 329
Precursors 27
Prescription 921
Psychedelics 636
Stimulants 576
Apparel 158
Art 11
Biotic materials 1
Books 1,161
Collectibles 5
Computer equipment 29
Custom Orders 49
Digital goods 351
Drug paraphernalia 139
Electronics 63
Erotica 438
Fireworks 2
Food 5
Forgeries 78
Hardware 7
Herbs & Supplements 5
Home & Garden 7
Jewelry 69
Lab Supplies 9
Lotteries & games 19
Medical 7
Money 117
Packaging 20
Services 32
Sporting goods 1
Weight loss 14
Writing 3

sort by: bestselling □ Domestic only update

5g White Widow
seller: Alexandrus(100)
ships from: Sweden
฿8.29
add to cart

GET UR WEEKLY AMNESIA SKUNK BY MAIL 5 GR NOISEFREE
seller: leckermann(98)
ships from: Belgium
฿5.91
add to cart

1oz - "Swazi Red" (Rooibaard) Afri
seller: Swazibudbud888(97)
ships from: South Africa

1G- (S) Moroccan Pollen – Nxt Day
seller: No Sticks No Seeds(99)
ships from: United Kingdom

HIGH GRADE WEED (NOT SKUN
seller: Budworx UK(99)
ships from: United Kingdom

10g Chronic 100% organic
seller: OrganicBuds(100)
ships from: Sweden

© Eddie Mitchell

CryptoLocker

Payment for private key

Choose a convenient payment method and click «Next»:

Bitcoin (most cheap option)

Bitcoin is a cryptocurrency where the creation and transfer of bitcoins is based on an open-source cryptographic protocol that is independent of any central authority. Bitcoins can be transferred through a computer or smartphone without an intermediate financial institution.

You have to send **2 BTC** to Bitcoin address [redacted] and specify the Transaction ID on the next page, which will be verified and confirmed.

Home Page
Getting started with Bitcoin

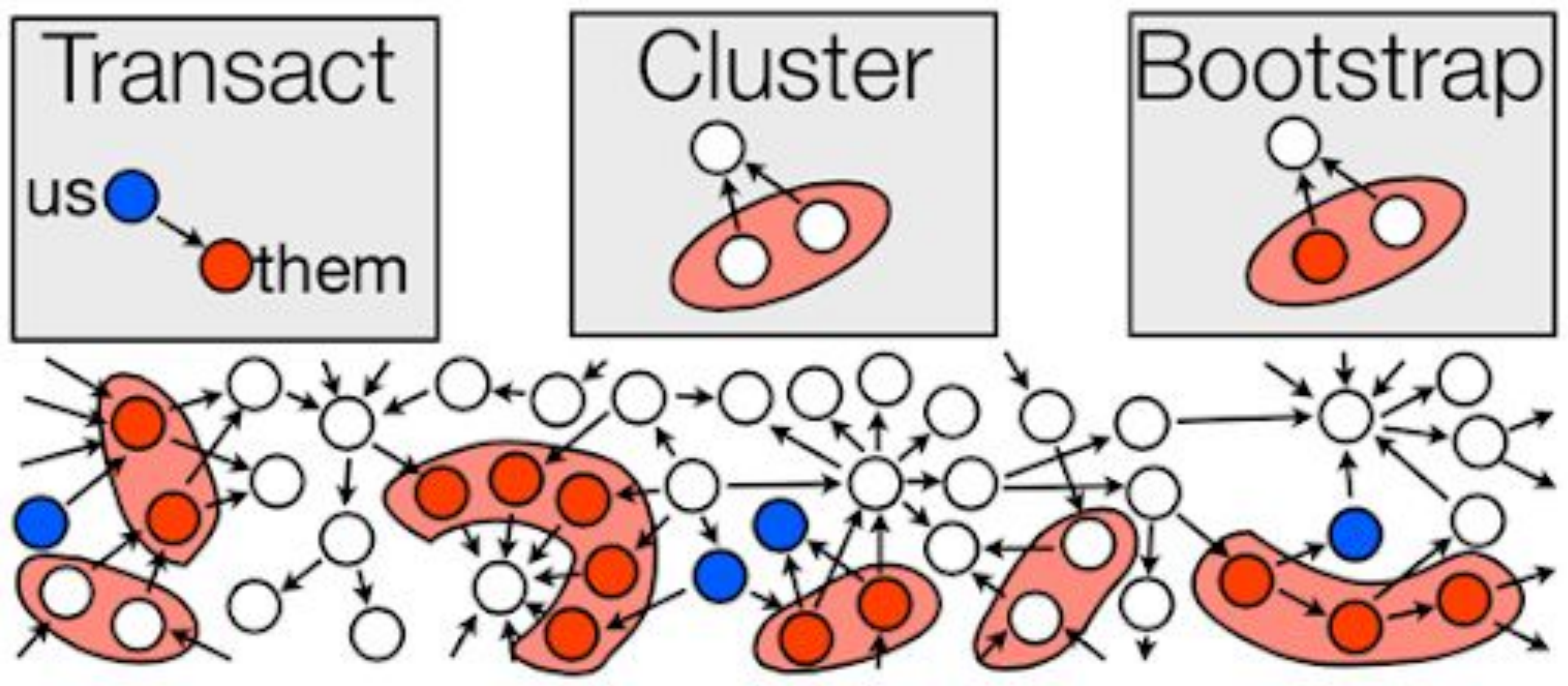
Private key will be destroyed on
10/13/2013
1:21 PM

Time left
71 : 33 : 17

Source: wikipedia.com

Source: bitinfocharts.com

Clustering Heuristics



Source: oreilly.com

Who uses Blockchain?

Financial fingerprint

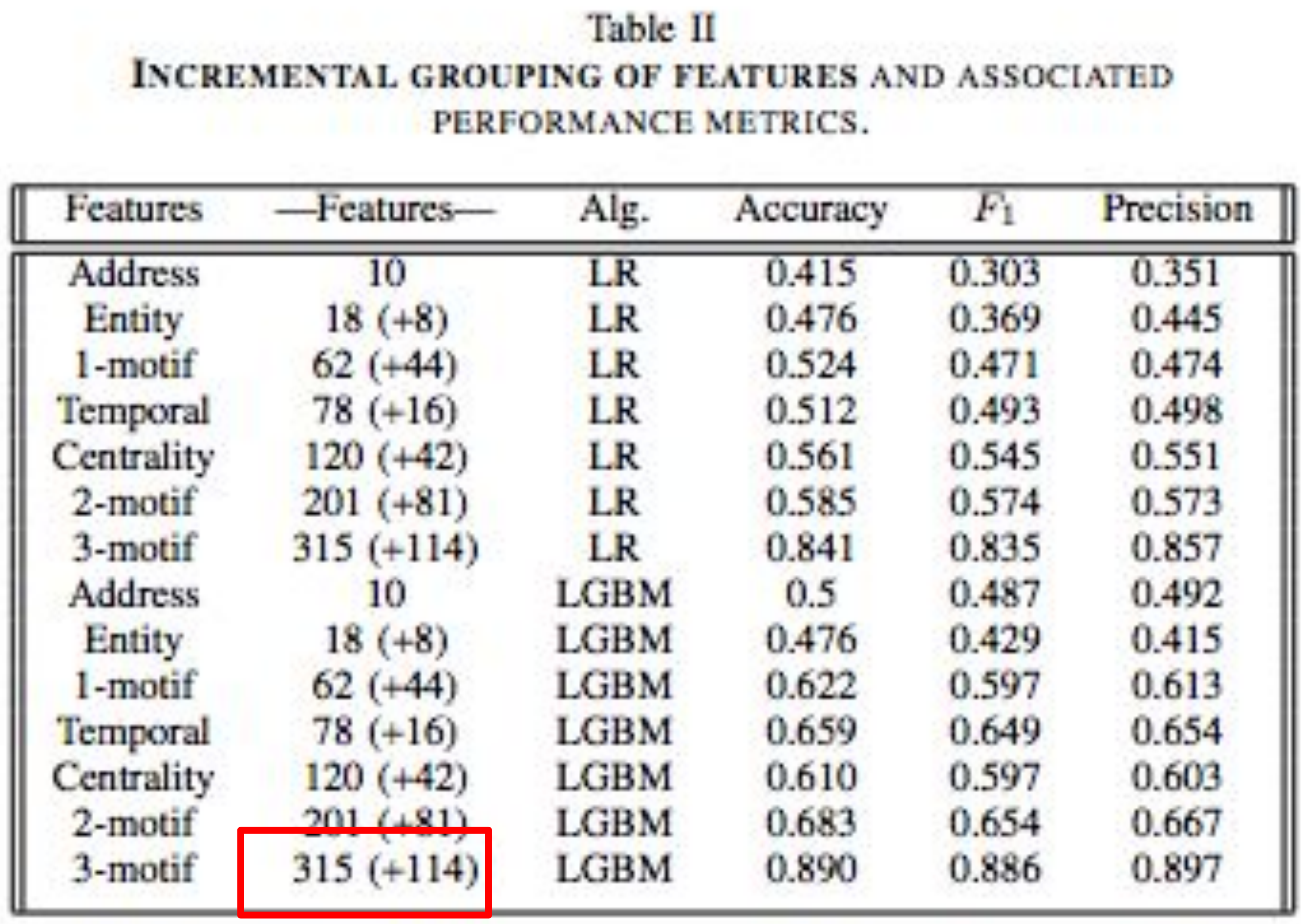
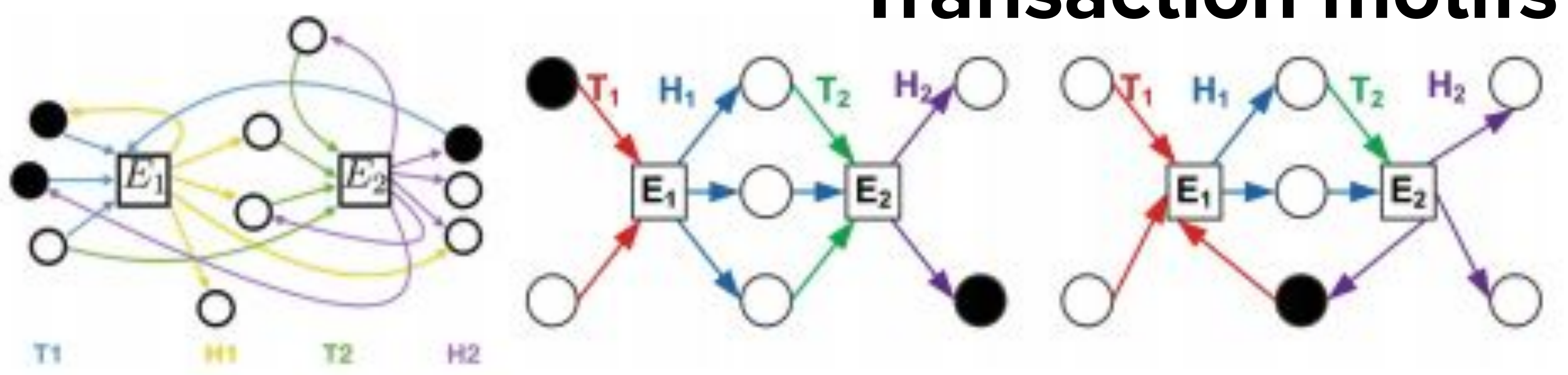


Table II

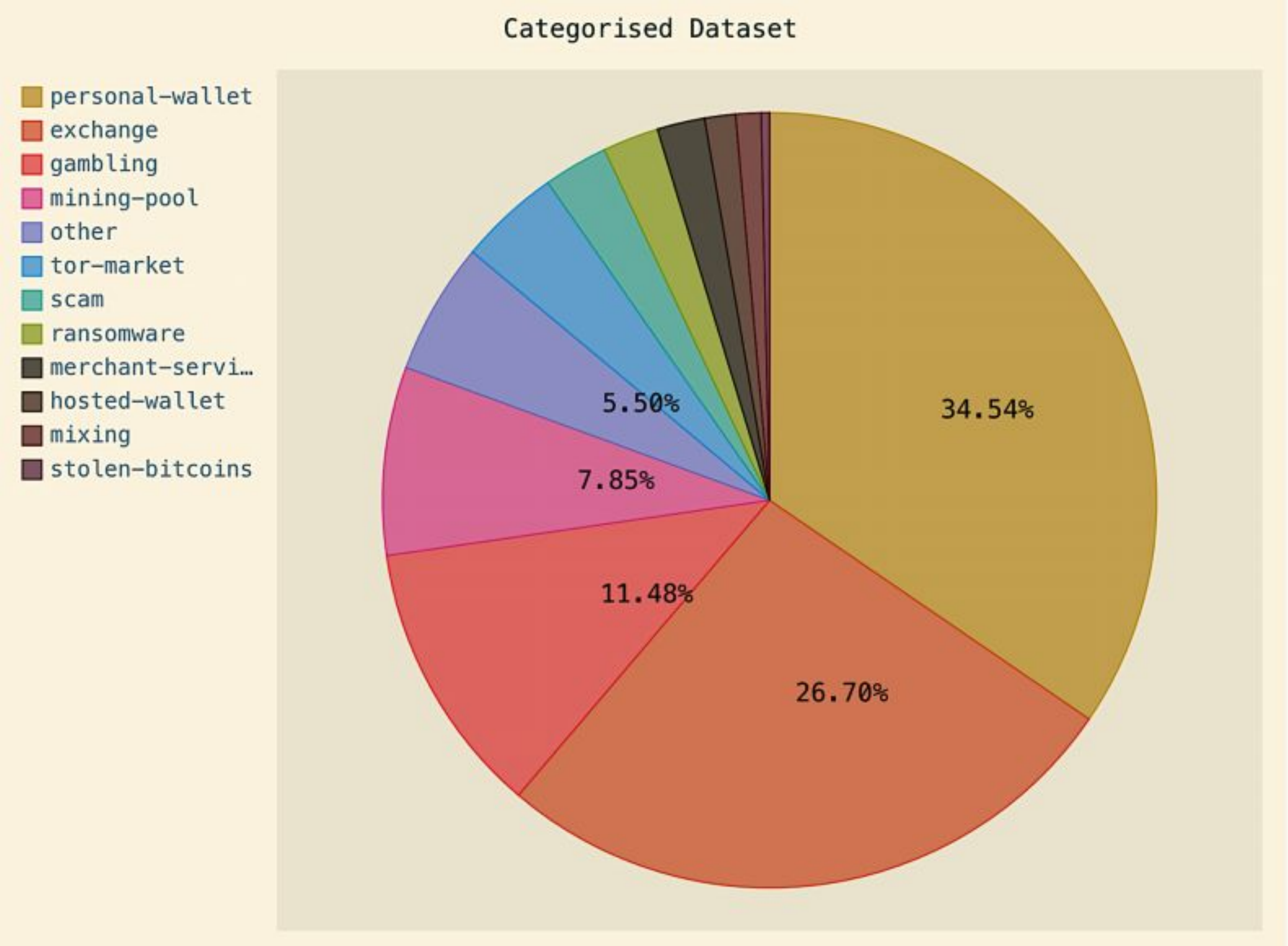
INCREMENTAL GROUPING OF FEATURES AND ASSOCIATED PERFORMANCE METRICS.

Features	—Features—	Alg.	Accuracy	F_1	Precision
Address	10	LR	0.415	0.303	0.351
Entity	18 (+8)	LR	0.476	0.369	0.445
1-motif	62 (+44)	LR	0.524	0.471	0.474
Temporal	78 (+16)	LR	0.512	0.493	0.498
Centrality	120 (+42)	LR	0.561	0.545	0.551
2-motif	201 (+81)	LR	0.585	0.574	0.573
3-motif	315 (+114)	LR	0.841	0.835	0.857
Address	10	LGBM	0.5	0.487	0.492
Entity	18 (+8)	LGBM	0.476	0.429	0.415
1-motif	62 (+44)	LGBM	0.622	0.597	0.613
Temporal	78 (+16)	LGBM	0.659	0.649	0.654
Centrality	120 (+42)	LGBM	0.610	0.597	0.603
2-motif	201 (+81)	LGBM	0.683	0.654	0.667
3-motif	315 (+114)	LGBM	0.890	0.886	0.897

Transaction motifs



Source: Jourdan et al. 2018

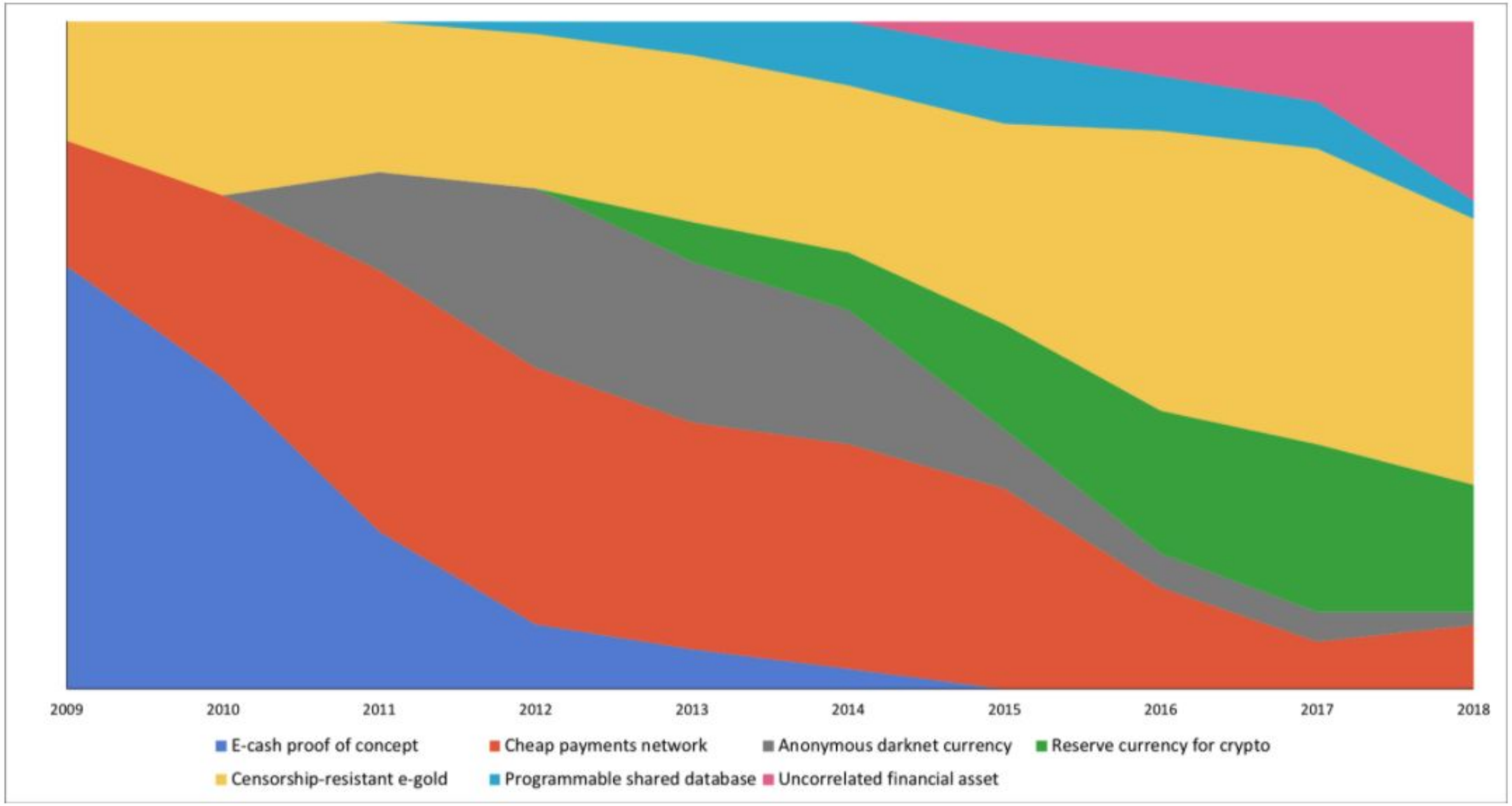


Source: Sun Yin et al. 2017

How is Blockchain used?

What people say

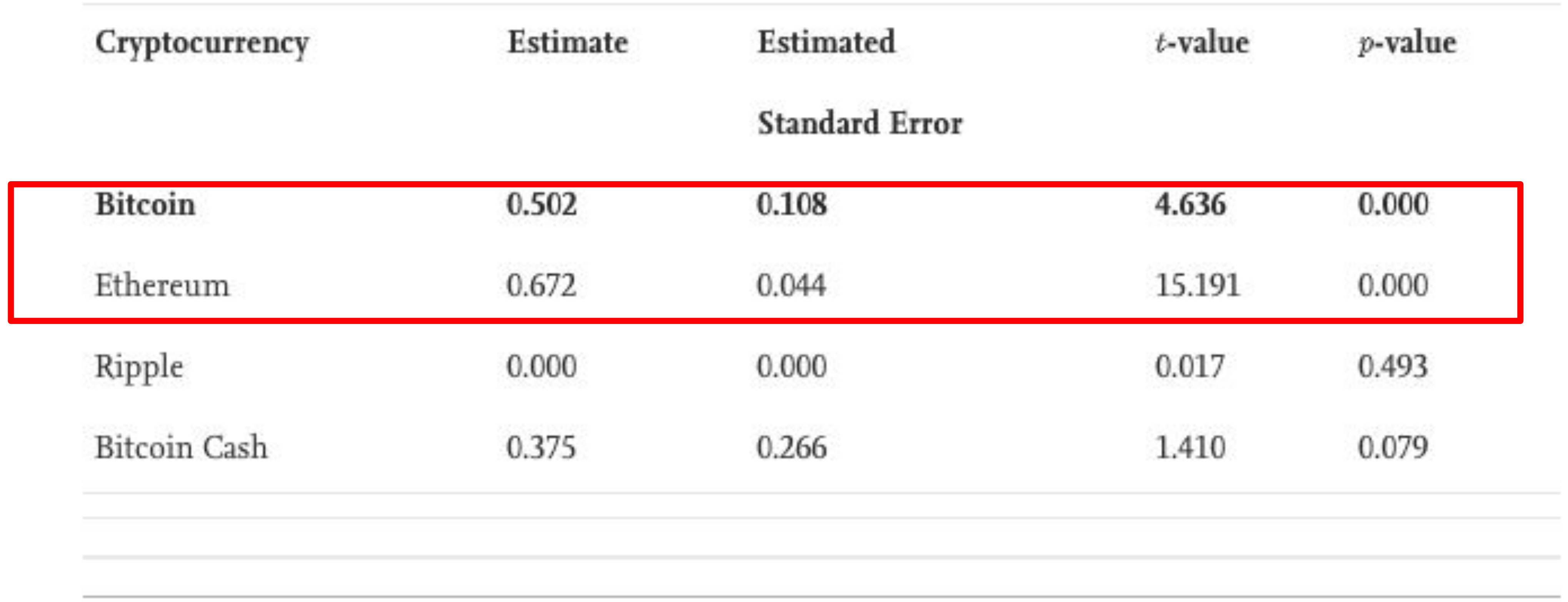
Post categorizations Bitcoin Talk



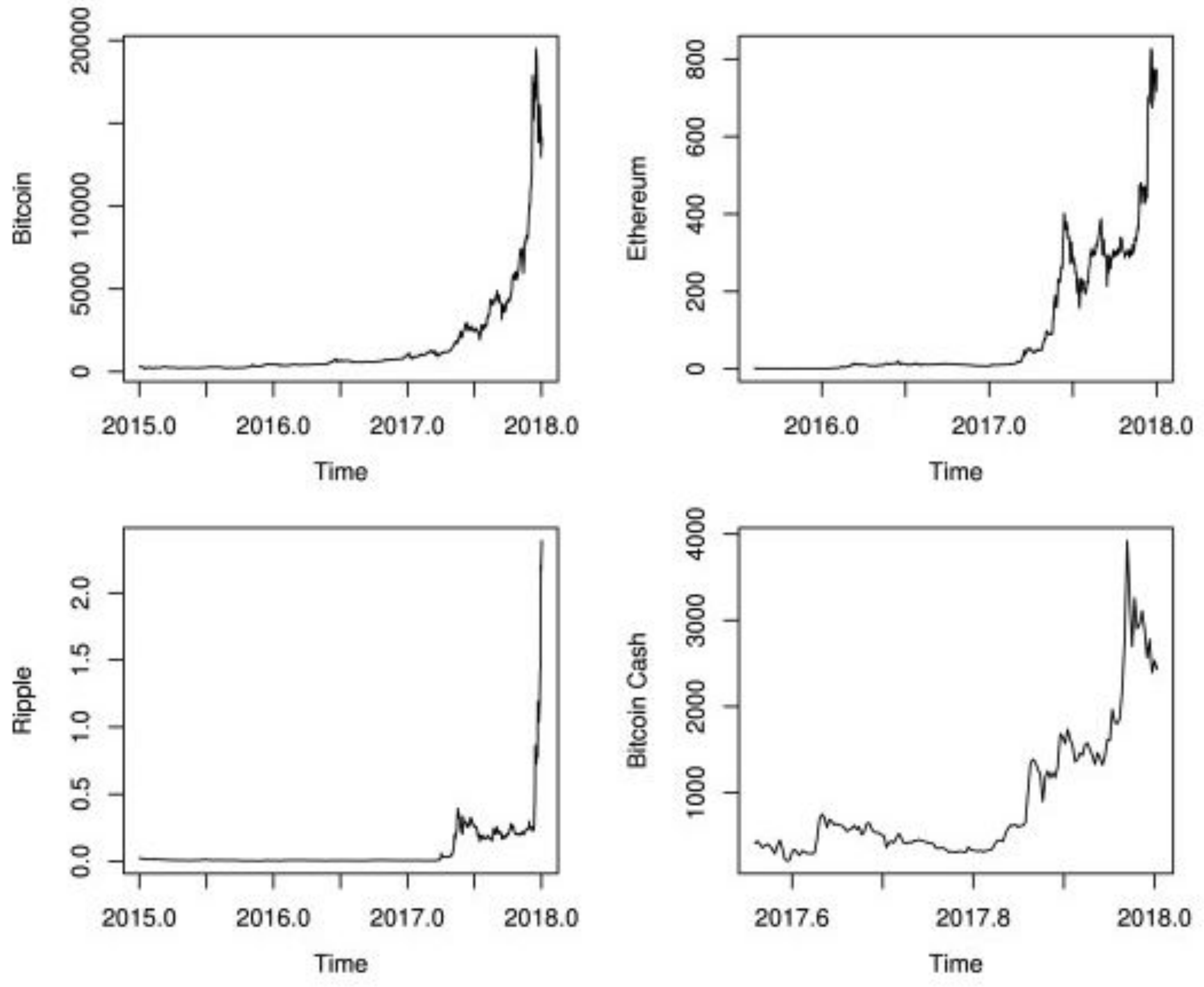
Source: Carter et al. 2018

How is Blockchain used?

Speculation?



Cryptocurrency	Estimate	Estimated Standard Error	*t*-value	*p*-value
Bitcoin	**0.502**	**0.108**	**4.636**	**0.000**
Ethereum	0.672	0.044	15.191	0.000
Ripple	0.000	0.000	0.017	0.493
Bitcoin Cash	0.375	0.266	1.410	0.079



Source: [Fry 2018]

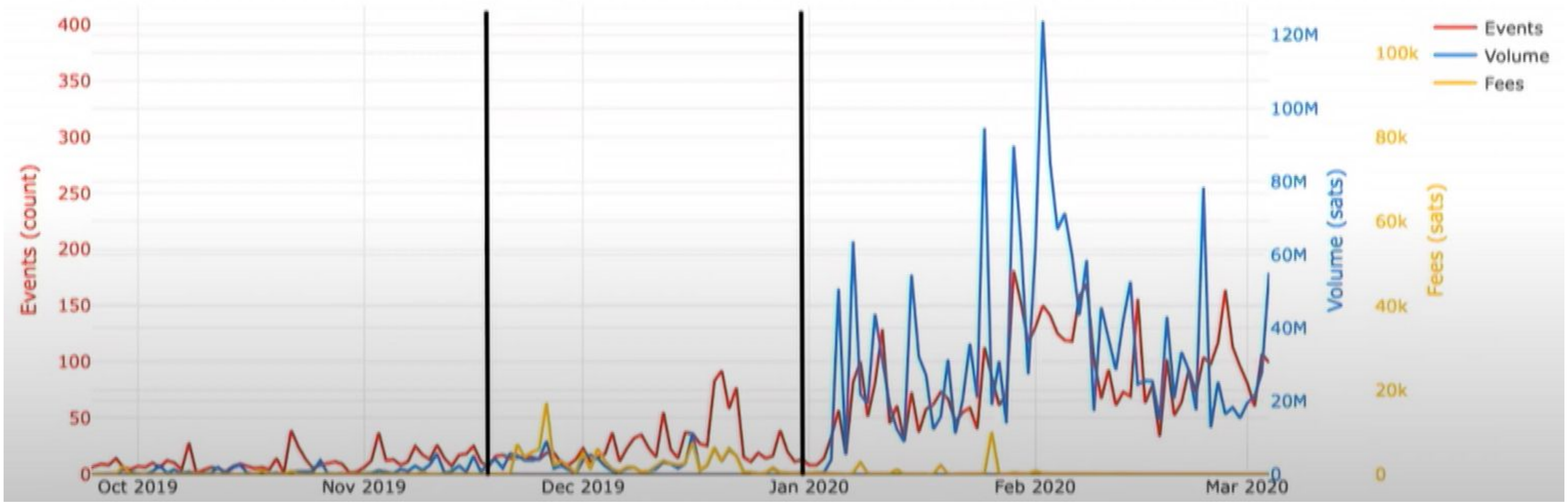
Statistical evidence of Bubbles present in Bitcoin and Ethereum up to 2018

How is Blockchain used?

Payments?

OpenNode Lightning Network payments routed

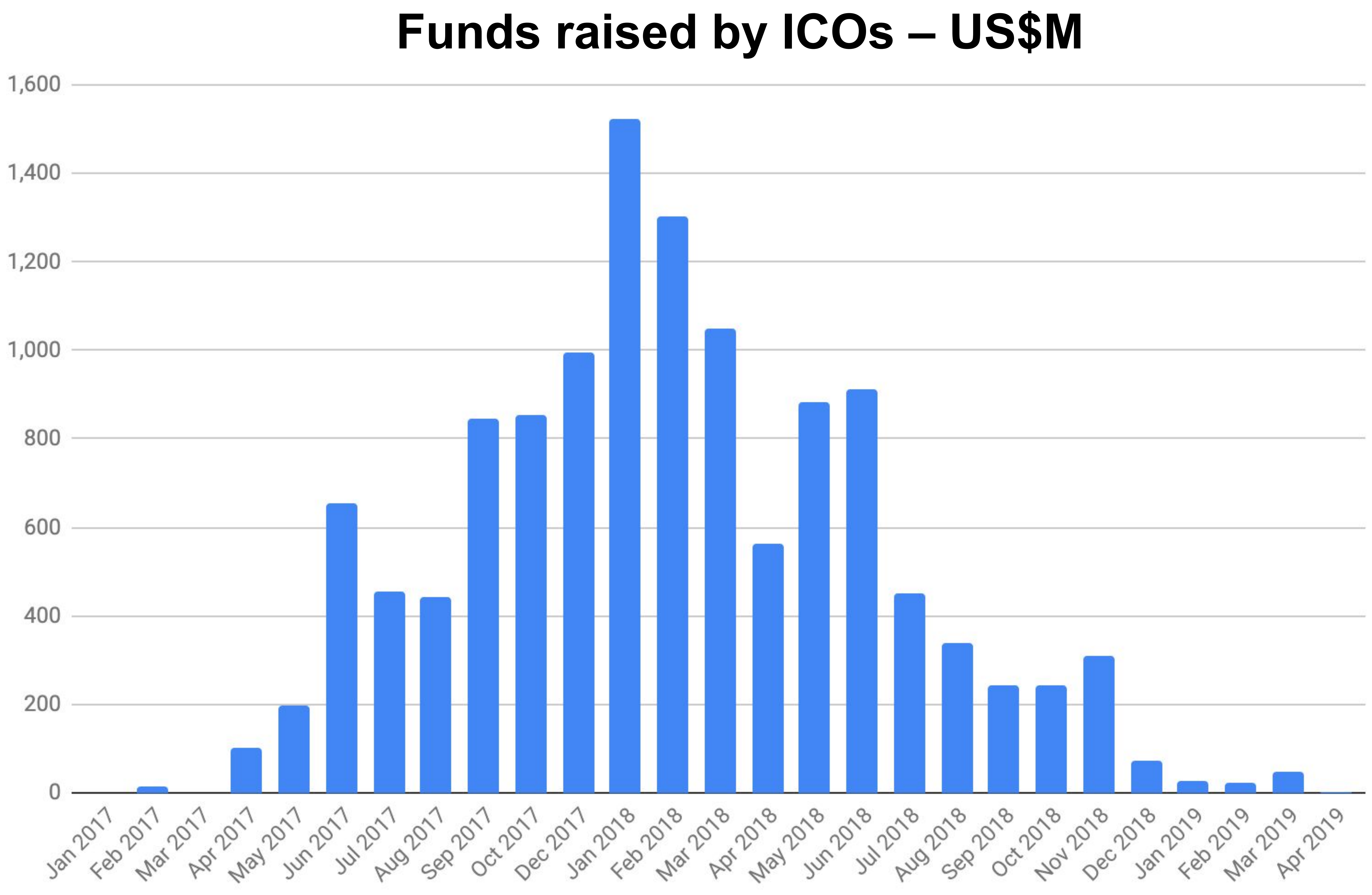
~ 10k USD



Source: OpenNode

How is Blockchain used?

Tokenizing / Fundraising

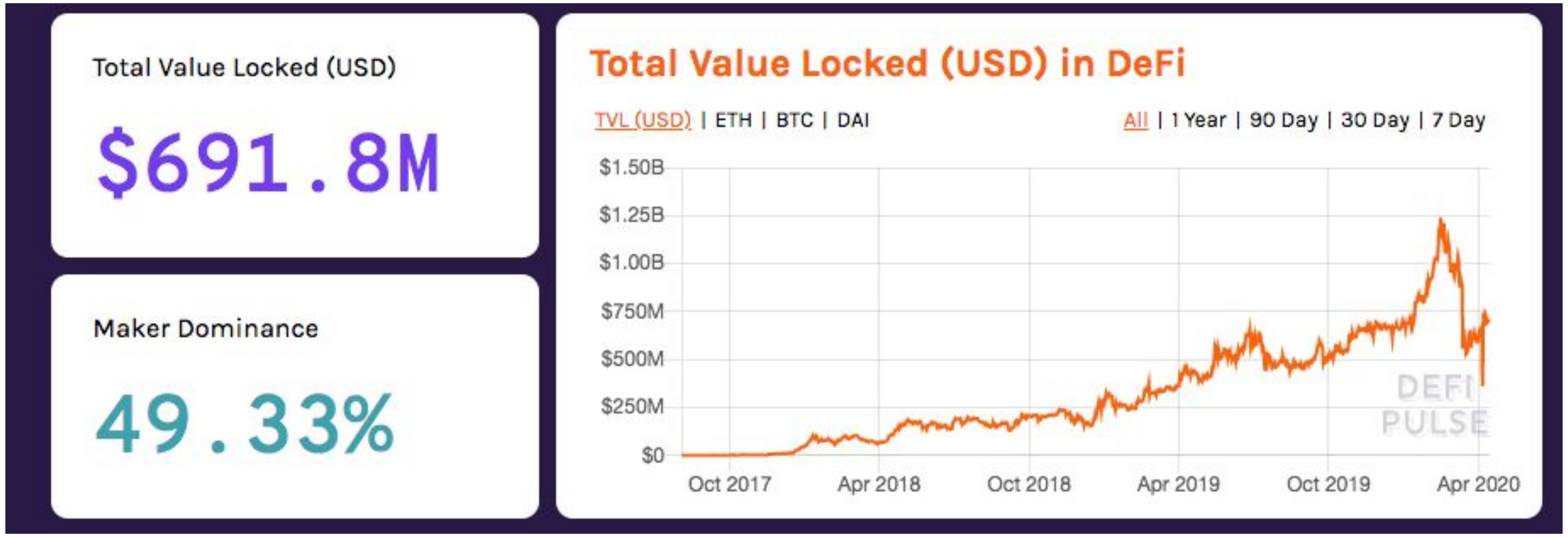


Source: bitmex.org

How is Blockchain used?

Finance

DeFi Market (2020)



Total Value Locked (USD)

$691.8M

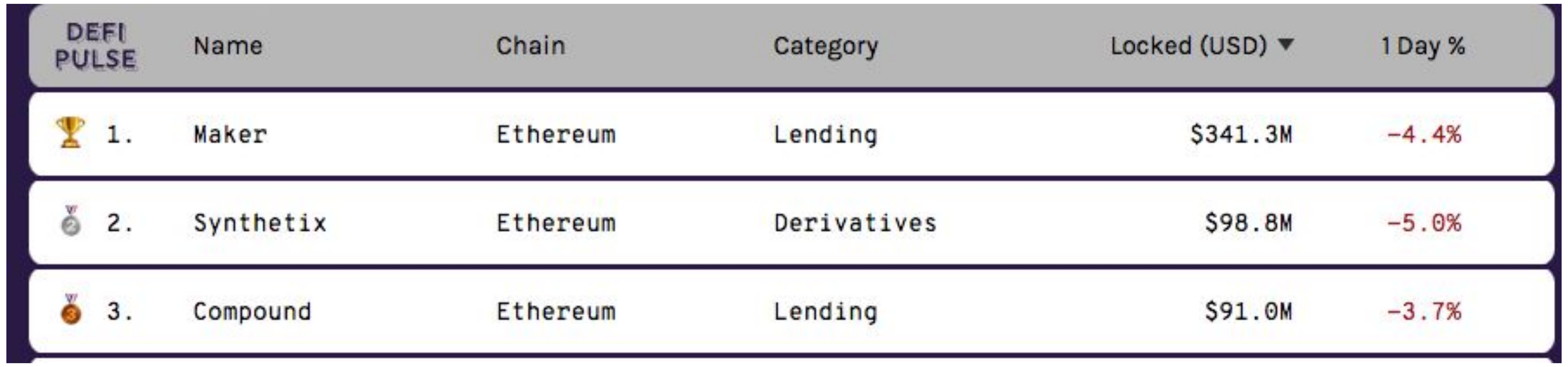
Maker Dominance

49.33%

Total Value Locked (USD) in DeFi

TVL (USD) | ETH | BTC | DAI

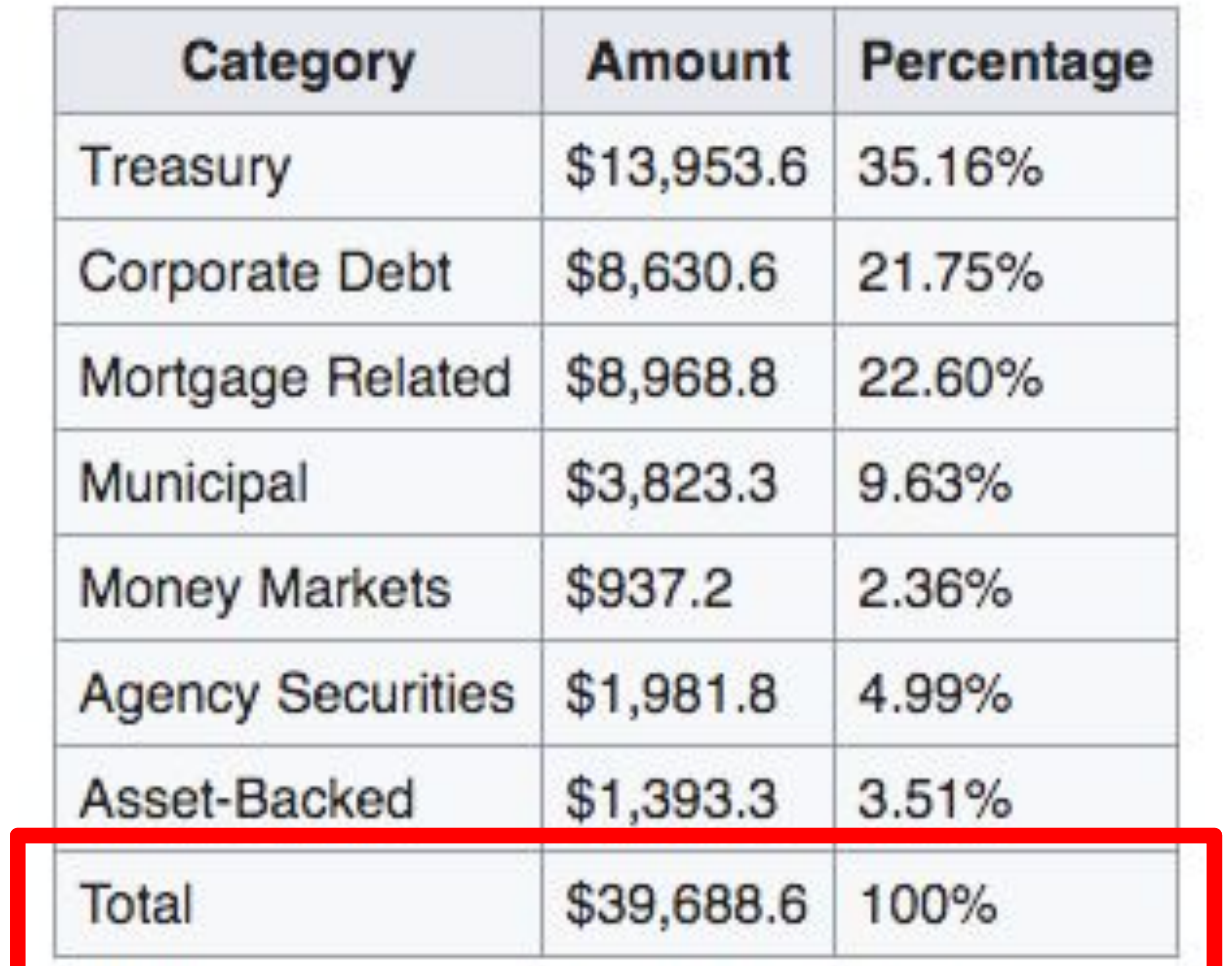
All | 1 Year | 90 Day | 30 Day | 7 Day



DEFI PULSE	Name	Chain	Category	Locked (USD) ▼	1 Day %
🏆 1.	Maker	Ethereum	Lending	$341.3M	-4.4%
🥈 2.	Synthetix	Ethereum	Derivatives	$98.8M	-5.0%
🥉 3.	Compound	Ethereum	Lending	$91.0M	-3.7%

Source: defipulse.com

US Bond Market (2017)



Category	Amount	Percentage
Treasury	$13,953.6	35.16%
Corporate Debt	$8,630.6	21.75%
Mortgage Related	$8,968.8	22.60%
Municipal	$3,823.3	9.63%
Money Markets	$937.2	2.36%
Agency Securities	$1,981.8	4.99%
Asset-Backed	$1,393.3	3.51%
Total	$39,688.6	100%

39T USD

Source: wikipedia.com

BLOCKCHAIN AT BERKELEY

Accessing Blockchain data

Online explorers

- Bitcoin: https://www.blockchain.com/
- Ethereum: https://etherscan.io/
- Bitcoin: https://txstats.com/
- Multiple Blockchains: https://bitinfocharts.com/
- Multiple Blockchains: https://app.santiment.net/
- Defi: https://defipulse.com/
- Ethereum apps: https://amberdata.io/dashboards/applications
- Lightning Network: https://1ml.com/
- Paid services: coinmetrics, chainanalysis, kaiko

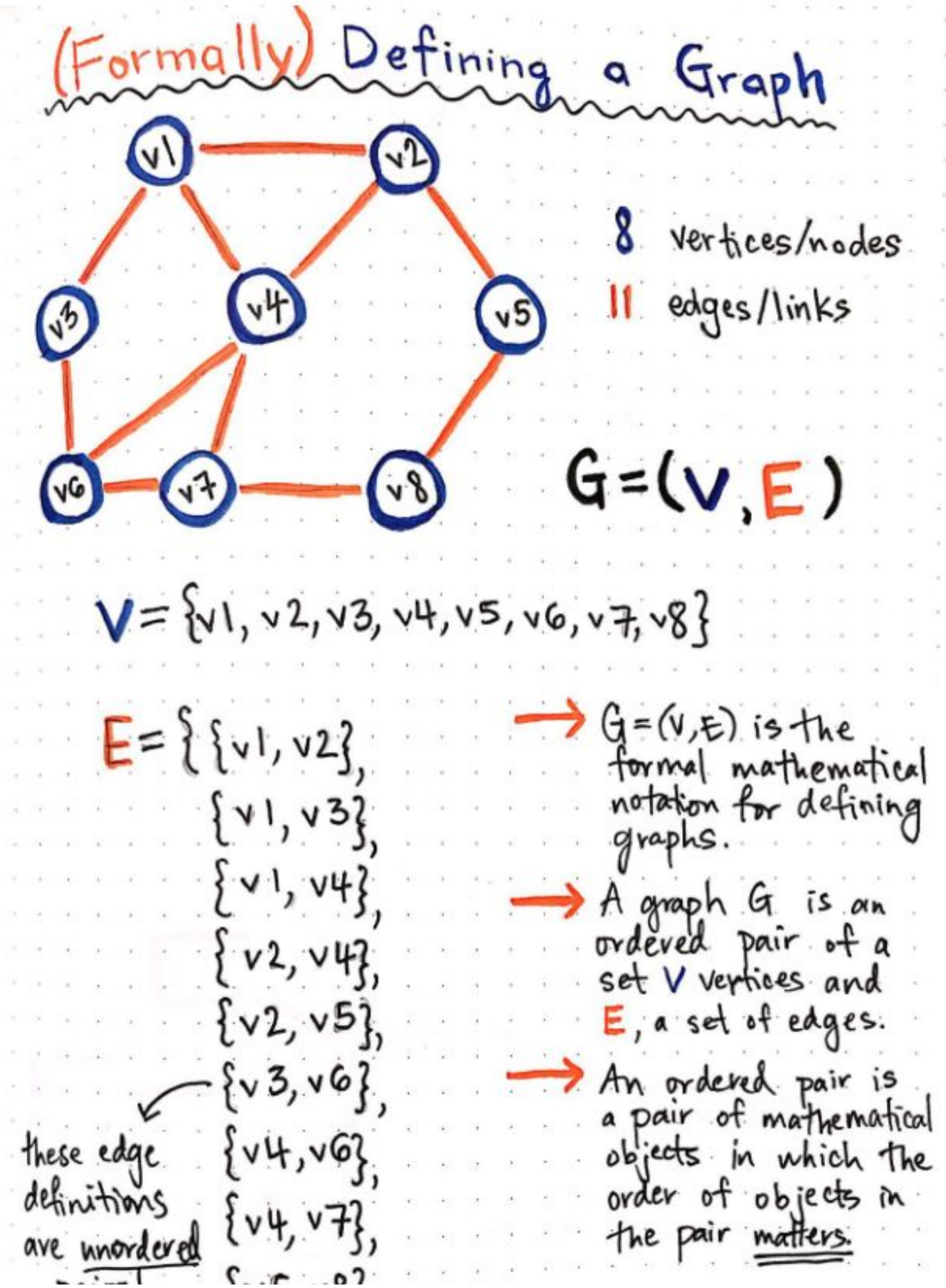
Accessing Blockchain Data

Access raw data

- Run a node: Ethereum, Bitcoin, Lightning
- Bitcoin: BlockSci
- SQL - Data Multiple Blockchains: Google BigQuery
- Lightning Network: https://ln.bigsun.xyz/

Analyzing Blockchain data

Defining a Graph



(Formally) Defining a Graph

v1
v2
v3
v4
v5
v6
v7
v8

8 vertices/nodes

11 edges/links

G=(V,E)

V={v1, v2, v3, v4, v5, v6, v7, v8}

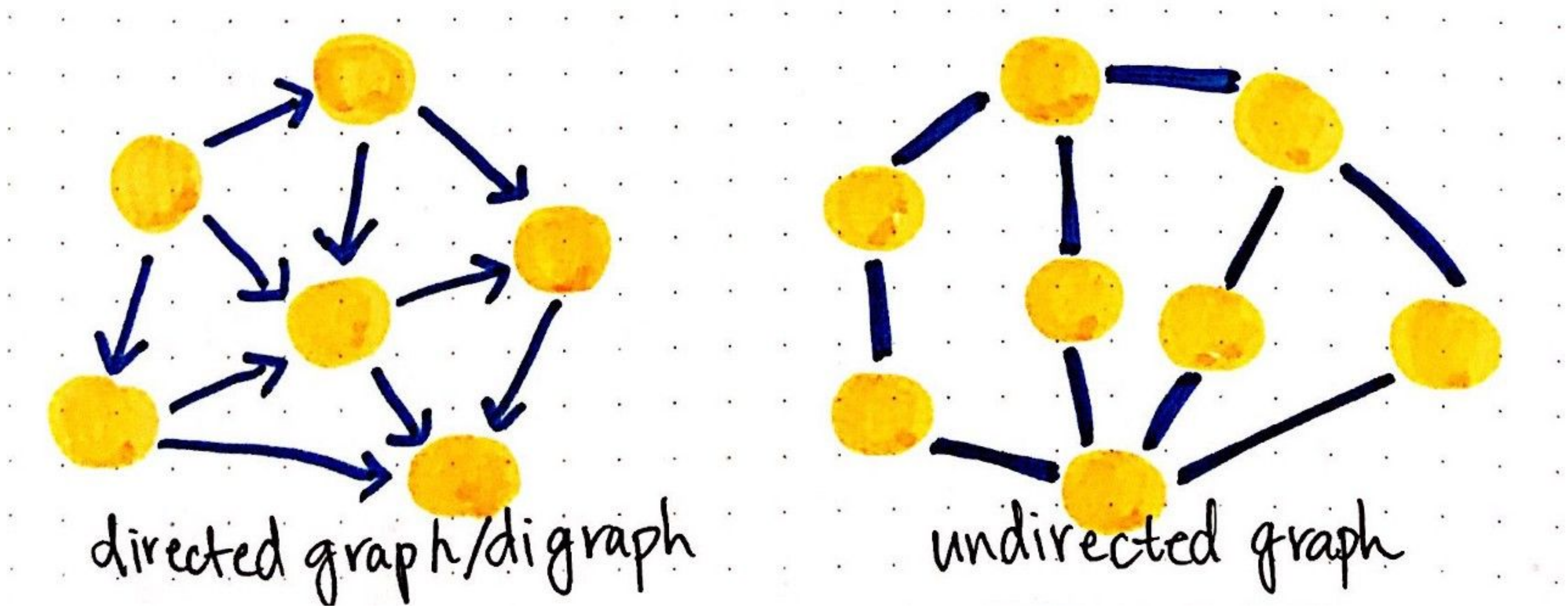
E={{v1, v2},
{v1, v3},
{v1, v4},
{v2, v4},
{v2, v5},
{v3, v6},
{v4, v6},
{v4, v7},

these edge definitions are unordered

→ G=(V,E) is the formal mathematical notation for defining graphs.

→ A graph G is an ordered pair of a set V vertices and E, a set of edges.

→ An ordered pair is a pair of mathematical objects in which the order of objects in the pair matters.



Source: Medium

Graphs can also have weighted nodes and/or edges

Analyzing Blockchain data

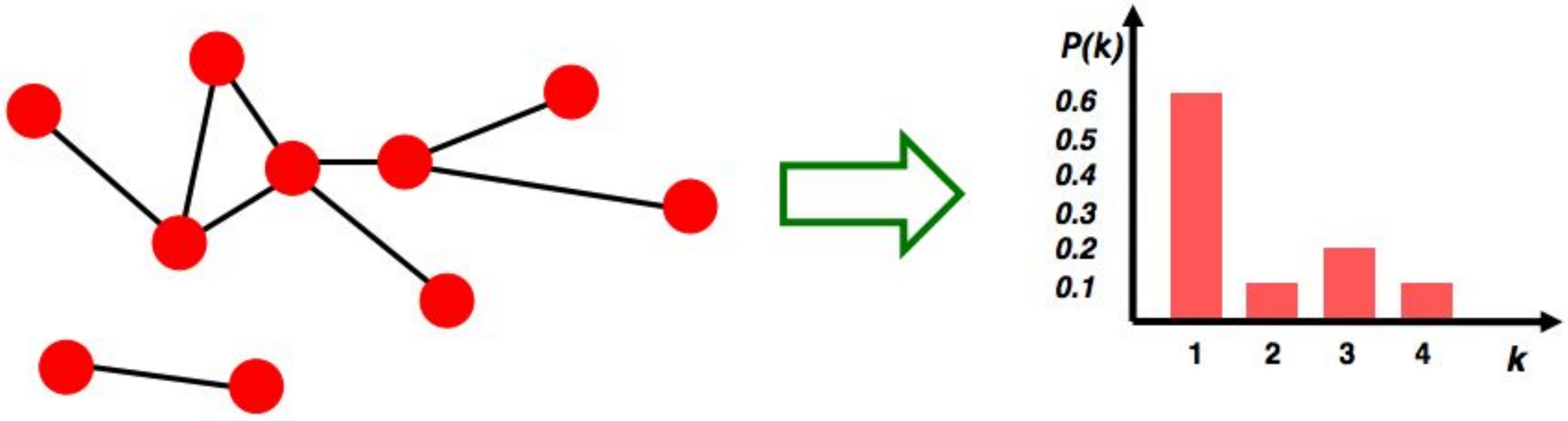
Graph properties

- **Degree distribution $P(k)$:** Probability that a randomly chosen node has degree k

  N_k = # nodes with degree k

- Normalized histogram:

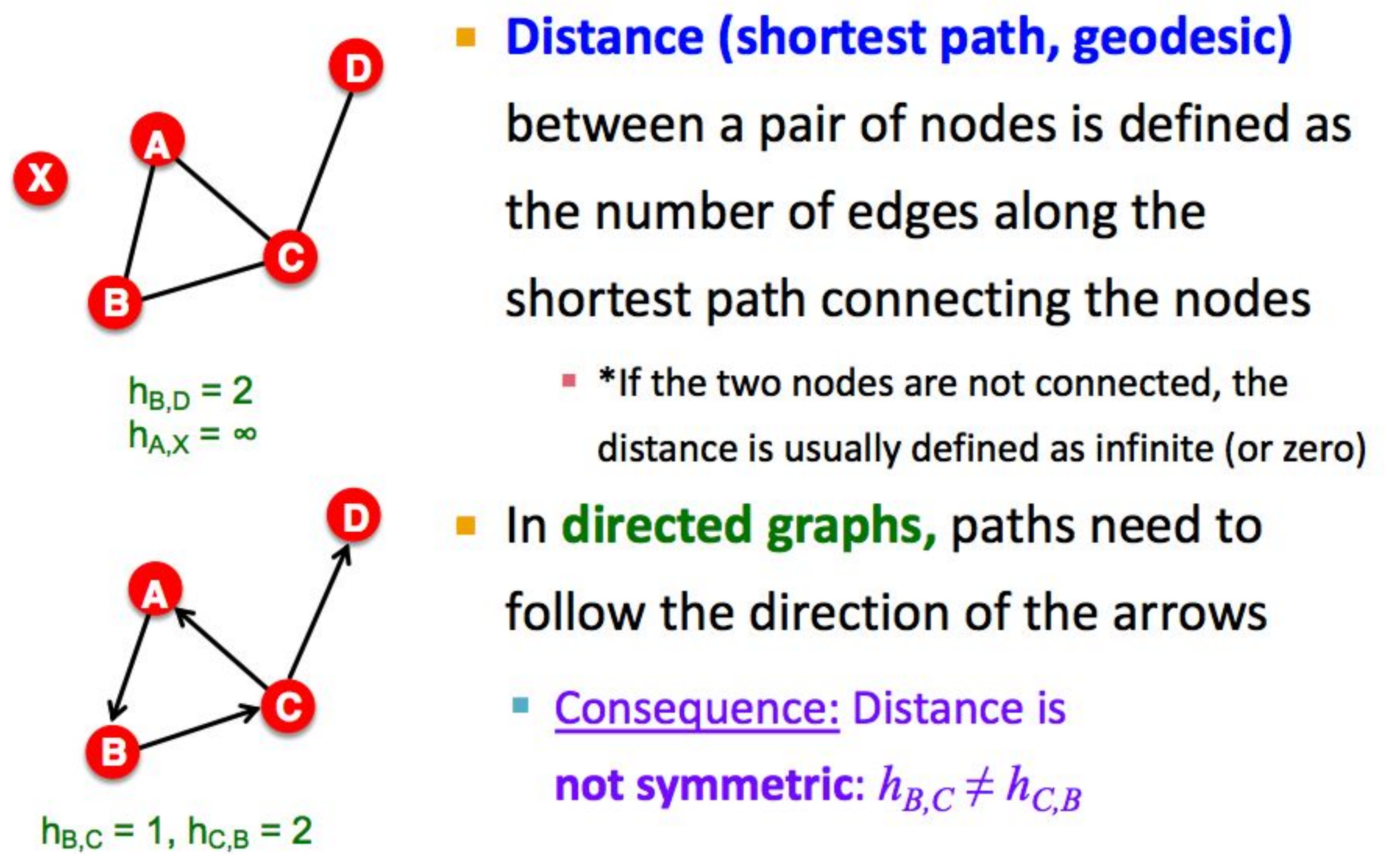
  $P(k) = N_k / N$ → plot



Source: Stanford CS224W

Analyzing Blockchain data

Graph properties



$h_{B,D} = 2$
$h_{A,X} = \infty$

$h_{B,C} = 1$, $h_{C,B} = 2$

- **Distance (shortest path, geodesic)** between a pair of nodes is defined as the number of edges along the shortest path connecting the nodes
    - *If the two nodes are not connected, the distance is usually defined as infinite (or zero)
- In **directed graphs,** paths need to follow the direction of the arrows
    - Consequence: Distance is **not symmetric:** $h_{B,C} \neq h_{C,B}$

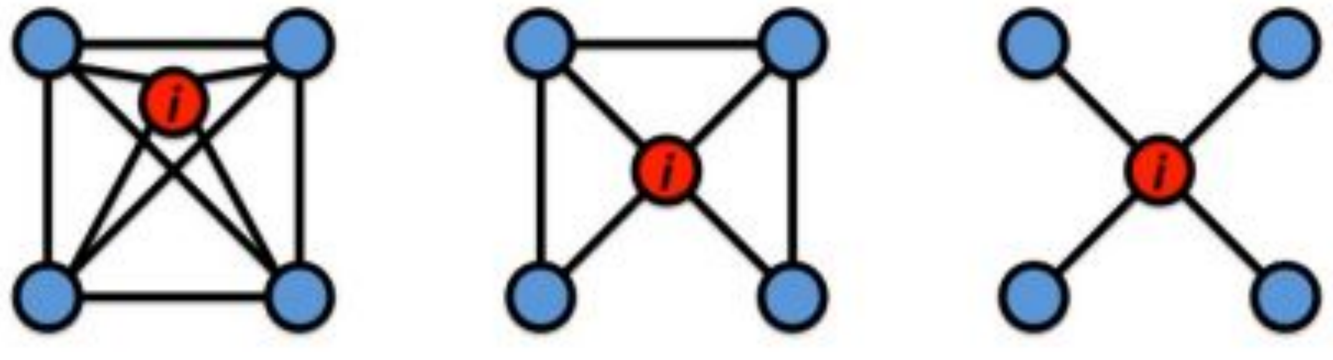
Source: Stanford CS224W

Analyzing Blockchain data

Graph properties

- **Clustering coefficient (for undirected graphs):**
  - How connected are ***i***'s neighbors to each other?
  - Node ***i*** with degree $\boldsymbol{k_i}$
  - $C_i \in [0,1]$
  - $C_i = \dfrac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of node i

Note $k_i(k_i-1)$ is max number of edges between the k_i neighbors



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

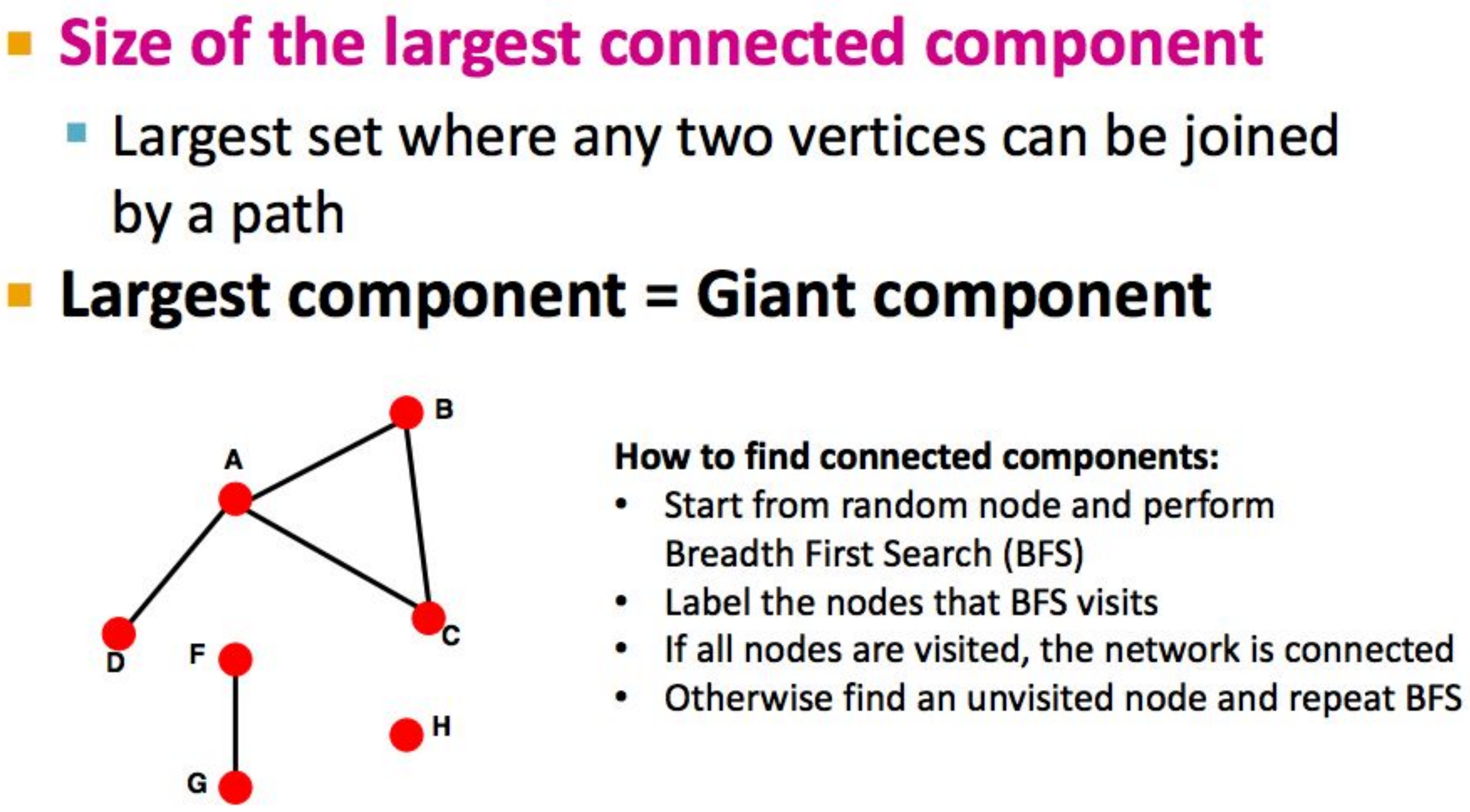
Clustering coefficient is undefined (or defined to be 0) for nodes with degree 0 or 1

- **Average clustering coefficient:** $C = \dfrac{1}{N}\sum_{i}^{N} C_i$

Source: Stanford CS224W

Analyzing Blockchain data

Graph properties



- **Size of the largest connected component**
  - Largest set where any two vertices can be joined by a path
- **Largest component = Giant component**

How to find connected components:

- Start from random node and perform Breadth First Search (BFS)
- Label the nodes that BFS visits
- If all nodes are visited, the network is connected
- Otherwise find an unvisited node and repeat BFS

Source: [Stanford CS224W]

Analyzing Blockchain Data

Network Analysis Libraries

- NetworkX: http://networkx.github.io/
- Snapy: https://snap.stanford.edu/snappy/
- GraphX:

  https://spark.apache.org/docs/2.1.0/graphx-programming-guide.html

Analyzing Blockchain Data

DEMO / HOMEWORK TIME!

https://github.com/BerkeleyBlockchain/Dev-DeCal-Spring-2020/tree/master/hw9-Blockchain%20Data%20Analysis