EDI: Towards Measuring Blockchain Decentralization

Dimitris Karakostas

Centrally-controlled systems

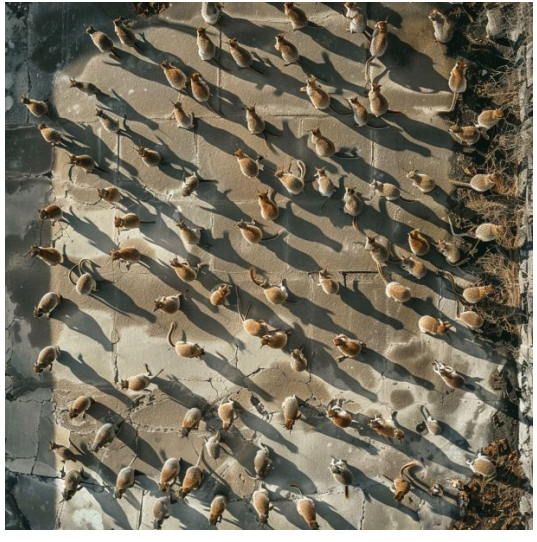
- A single party (node) controls who can read/write/delete data
- If the person/party/node dies/is dishonest/crashes, the system crashes

Controlled-access distributed systems

- Nodes collectively control the system
- If only few nodes faulty, system remains operational
- Controlled participation - only authorized parties

Open-access distributed systems

- Nodes collectively control the system
- If only few nodes faulty, system remains operational
- Anyone can participate, join or leave as they please



What is a blockchain?

- A database...

What is a blockchain?

- A database...
- … that stores financial transactions...

What is a blockchain?

- A database...
- … that stores financial transactions...
- … and programs ("smart contracts")…

What is a blockchain?

- A database...
- … that stores financial transactions...
- … and programs ("smart contracts")…
- … in a distributed manner...

What is a blockchain?

- A database...
- … that stores financial transactions...
- … and programs ("smart contracts")…
- … in a distributed manner...
- … where every node always has the same view of the db ("safe")…

What is a blockchain?

- A database...
- … that stores financial transactions...
- … and programs ("smart contracts")…
- … in a distributed manner...
- … where every node always has the same view of the db ("safe")…
- … and the db data are updated over time ("live")

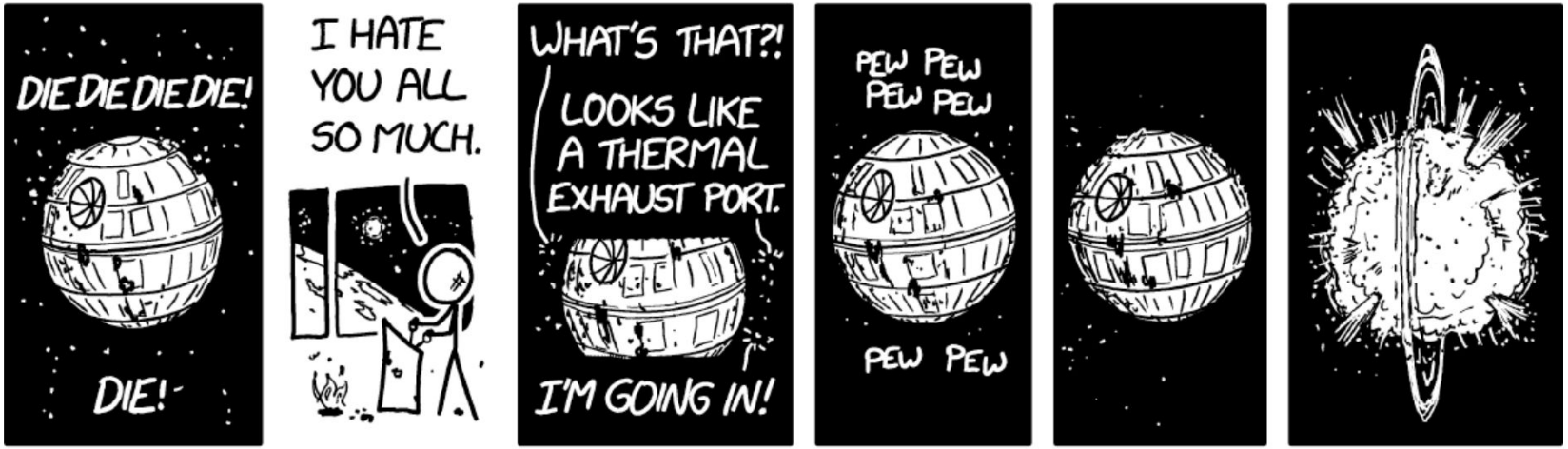
What is a blockchain?

- … in a distributed manner...
  - Anyone can read/write, keep a copy, and maintain the database

A blockchain has the *potential* to be completely *decentralized*

EDI: What does decentralized mean?

The *absence* of a *single point of failure*



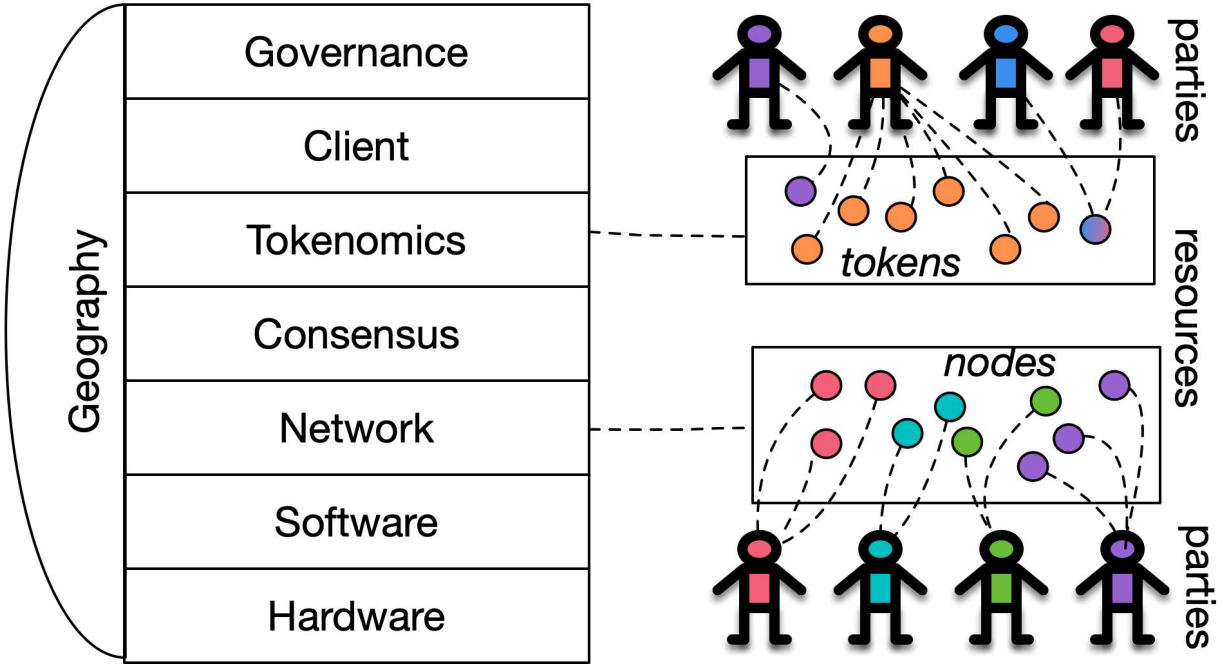
https://xkcd.com/2425

EDI: When is a system decentralized?

- *"Is a system decentralized?"* is the wrong question
- Decentralization is a spectrum
- *Where* can single points of failure come up?
  - Which parts of the system might be compromised?
- *How* decentralized is a system?
  - *How far* is it from a single point of failure

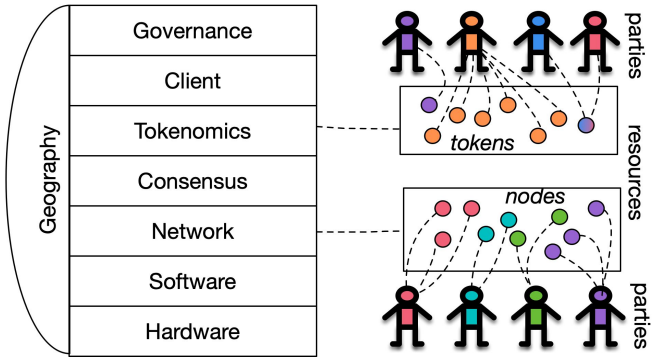
EDI: Blockchain Layers

Where can single points of failure come up?

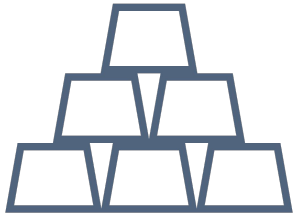


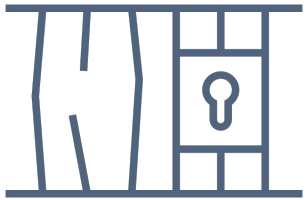
EDI Methodology

How decentralized is a system?



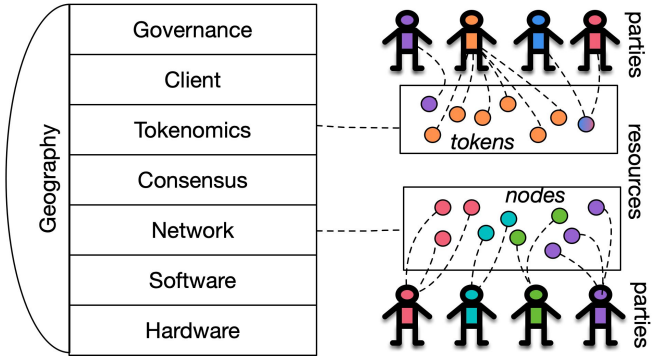
For each layer, we find...



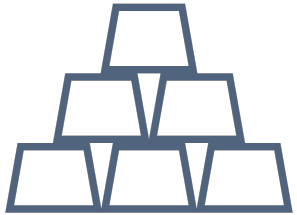


EDI Methodology

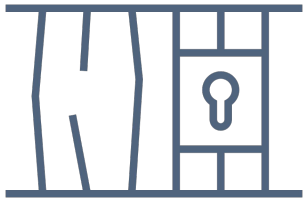
How decentralized is a system?



For each layer, we find...

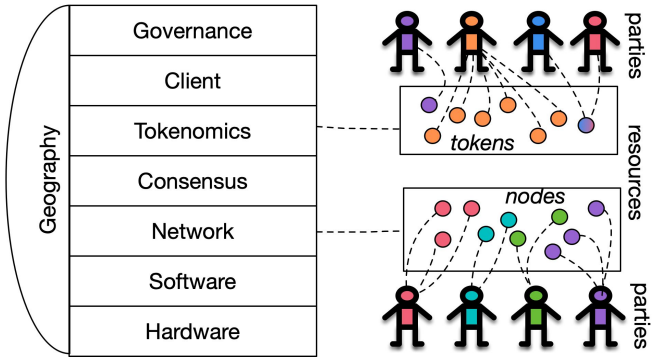


resource

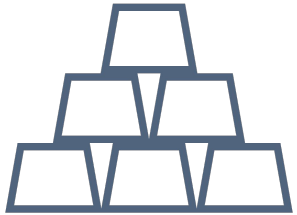


EDI Methodology

How decentralized is a system?

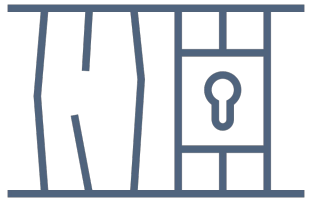


For each layer, we find...



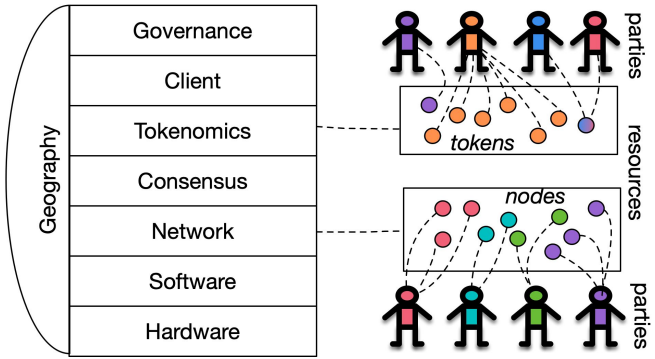
resource

relevant parties

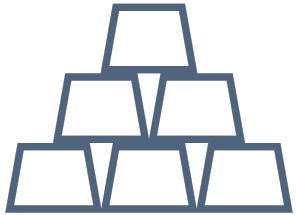


EDI Methodology

How decentralized is a system?

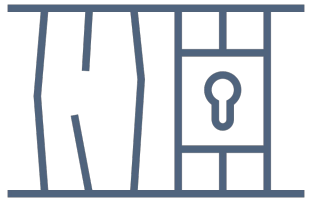


For each layer, we find...



resource

relevant parties



properties at risk

How to measure the distribution?

- Blockchain systems are typically pseudonymous
- Measuring the distribution of resources among *real-world entities* is not always feasible
  - A user can create multiple independent pseudonyms (e.g., addresses)
    - *Clustering* of pseudonyms to the same entity can help with this problem
  - Multiple users may collectively control the same pseudonym (e.g., via threshold signatures)

How to measure the distribution?

- Blockchain systems are typically pseudonymous
- Measuring the distribution of resources among *real-world entities* is not always feasible
  - A user can create multiple independent pseudonyms (e.g., addresses)
    - *Clustering* of pseudonyms to the same entity can help with this problem
  - Multiple users may collectively control the same pseudonym (e.g., via threshold signatures)
- The distribution's *structure* may affect the results
  - A long tail (pseudonyms that control tiny amounts of resources) can skew the decentralization analysis towards either direction
    - *Thresholding* (considering only the top part of the distribution) has been used for this

How to measure decentralization of the distribution?

How to measure decentralization of the distribution?

- Nakamoto coefficient
  - The *minimum* number of parties that control *half* of all resources
  - *Completely ignores* distribution tail

How to measure decentralization of the distribution?

- Nakamoto coefficient
- Herfindahl-Hirschman Index (HHI)
  - Σ [market share]^2
  - Lower HHI $\rightarrow$ Better decentralization
  - *Not too sensitive* of distribution tail

How to measure decentralization of the distribution?

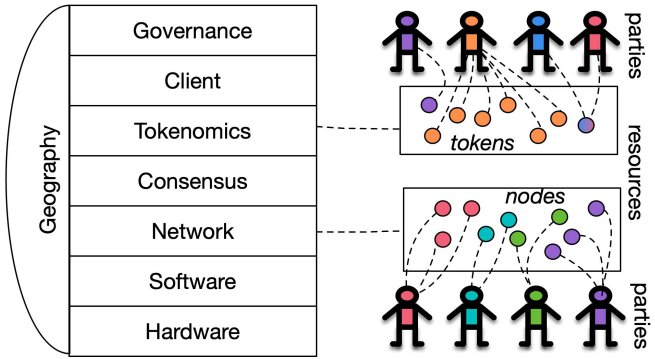
- Nakamoto coefficient
- Herfindahl-Hirschman Index (HHI)
- Gini coefficient
  - “The ratio of the area between the line of equality and the Lorenz curve over the total area under the line of equality”
  - Lower Gini → Better decentralization
  - *Very sensitive* of distribution tail

How to measure decentralization of the distribution?

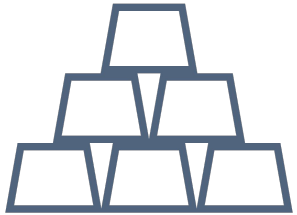
- Nakamoto coefficient
- Herfindahl-Hirschman Index (HHI)
- Gini coefficient
- τ-decentralization
- Shannon entropy
- Theil index

Case study: Tokenomics

EDI Methodology - Tokenomics

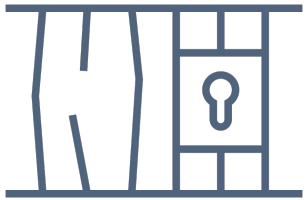


For tokenomics, we find...



tokens

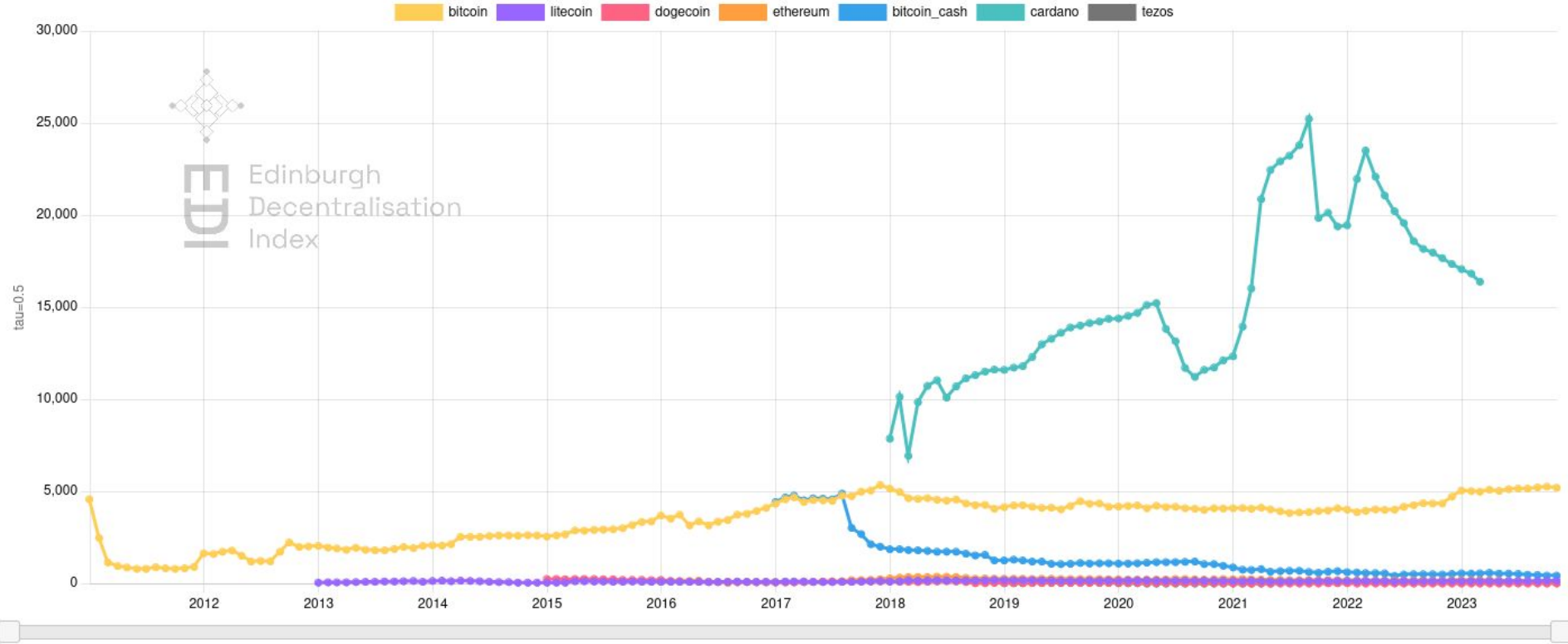
token
holders



safety
liveness
stability

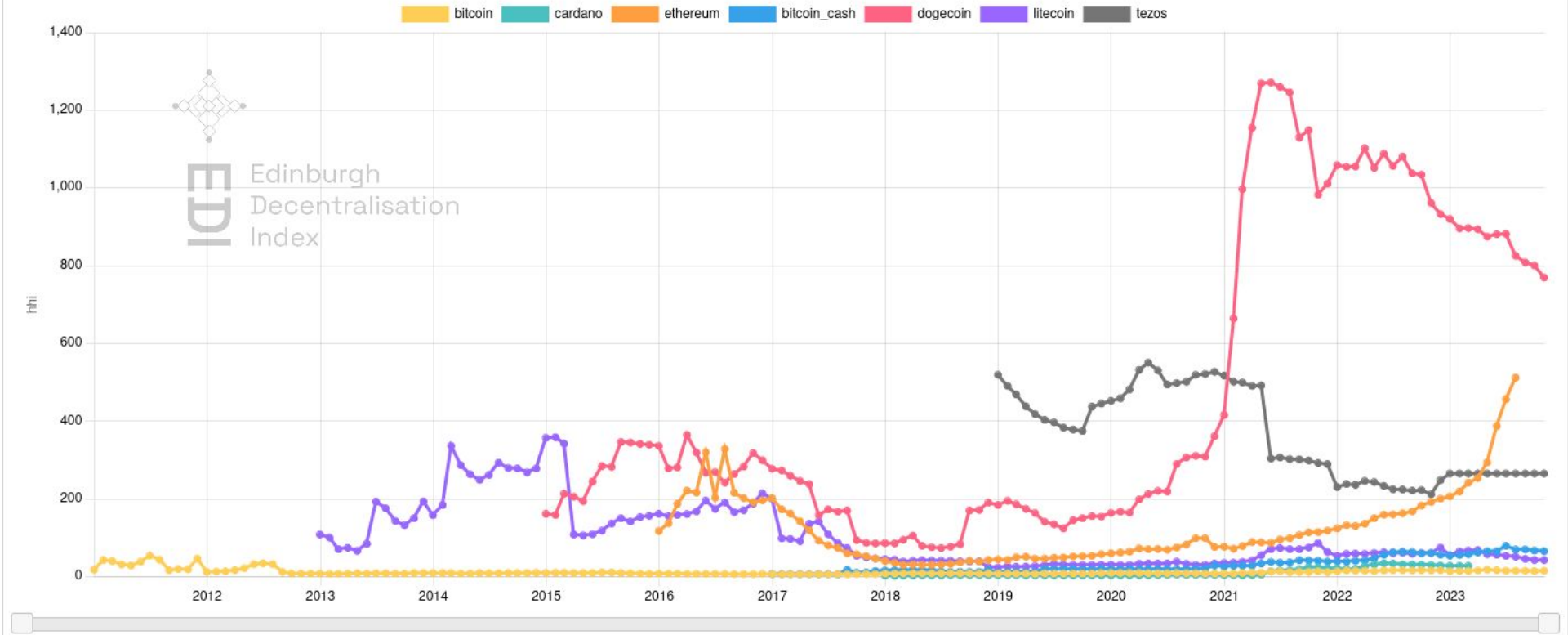
Nakamoto coefficient

The Nakamoto coefficient represents the minimum number of entities that collectively control more than 50% of the resources (in this case, the majority of circulating tokens at a given point in time).



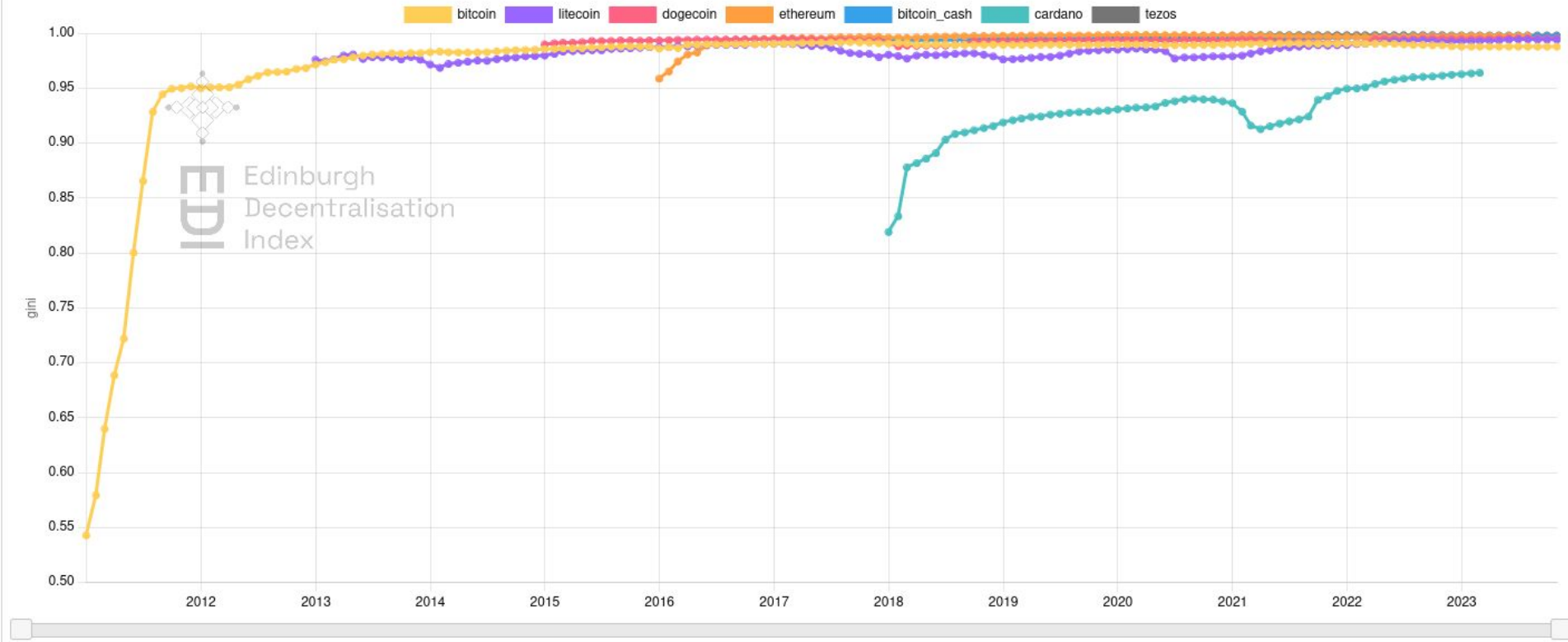
HHI

The Herfindahl-Hirschman Index (HHI) is a measure of market concentration. It is defined as the sum of the squares of the market shares (as whole numbers, e.g. 40 for 40%) of the entities in the system. Values close to 0 indicate low concentration (many entities hold a similar number of tokens) and values close to 10,000 indicate high concentration (one entity controls most or all tokens).



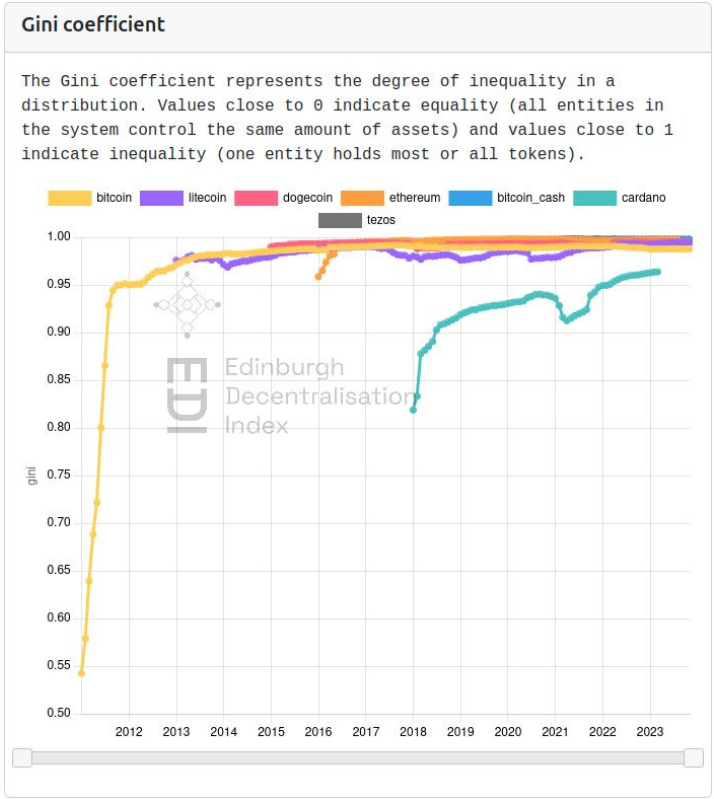
Gini coefficient

The Gini coefficient represents the degree of inequality in a distribution. Values close to 0 indicate equality (all entities in the system control the same amount of assets) and values close to 1 indicate inequality (one entity holds most or all tokens).

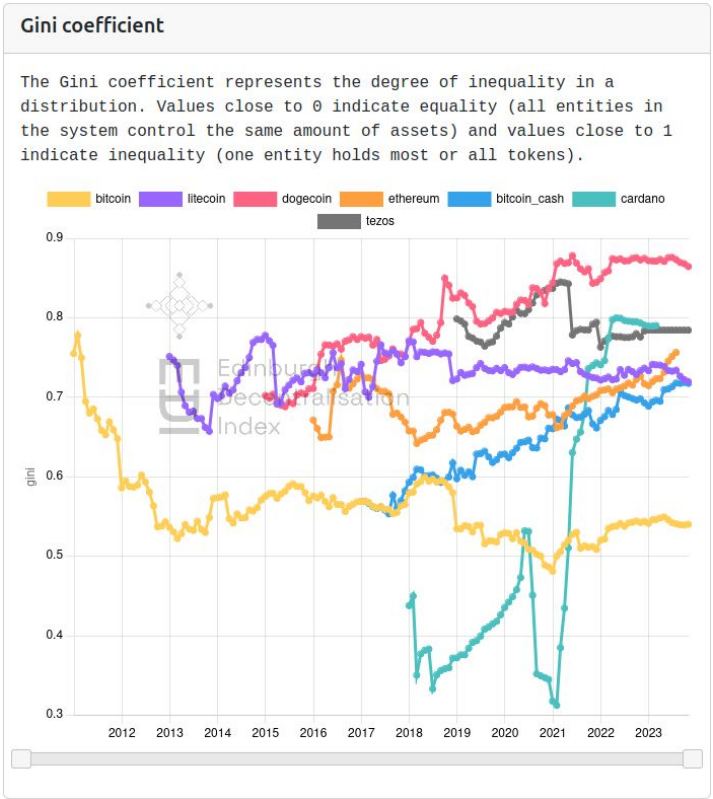


How thresholding affects the results

Without thresholding

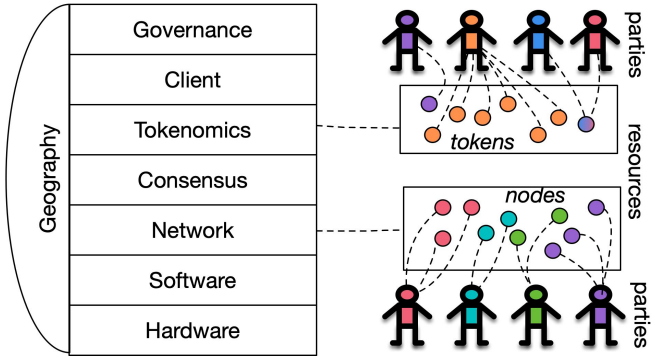


Top 1000

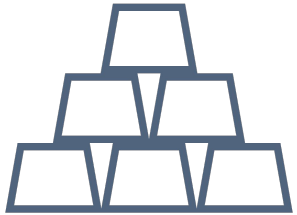


Case study: Consensus

EDI Methodology - Consensus

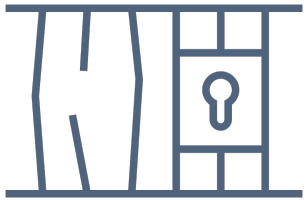


For consensus, we find...



blocks

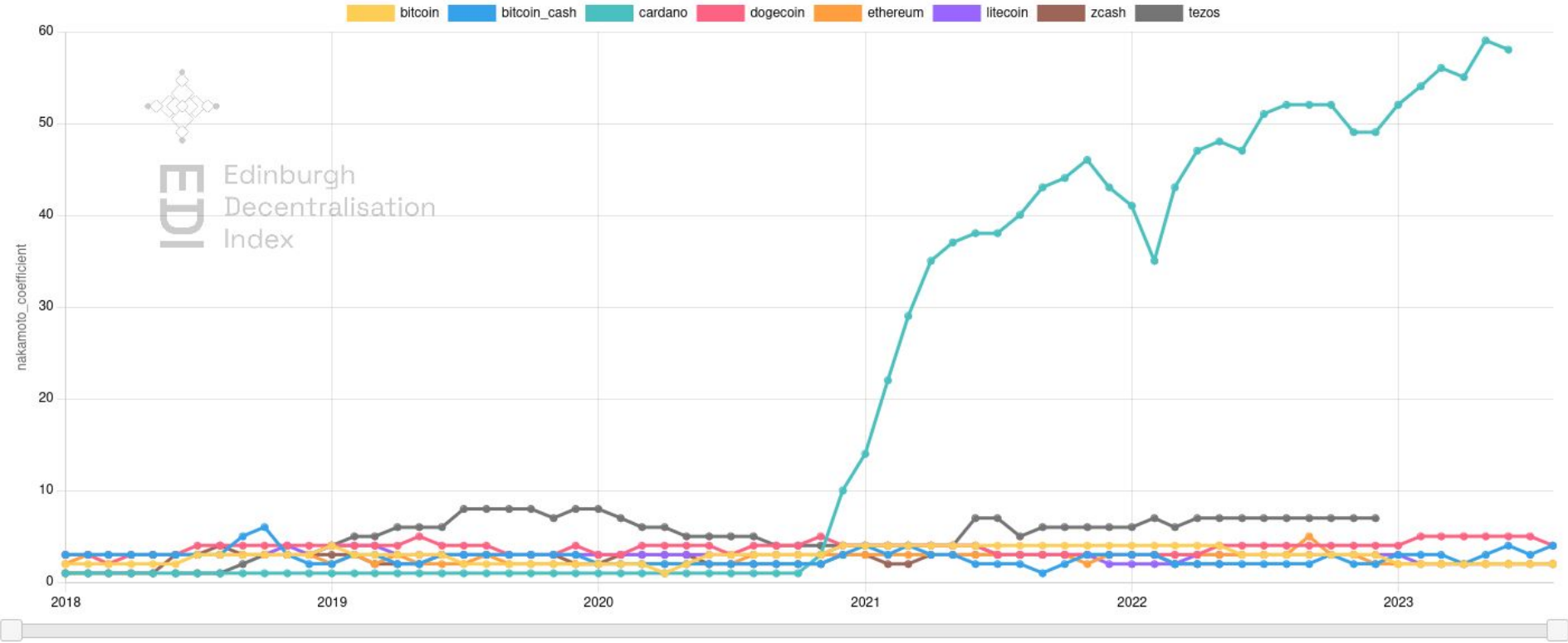
block producers



safety liveness

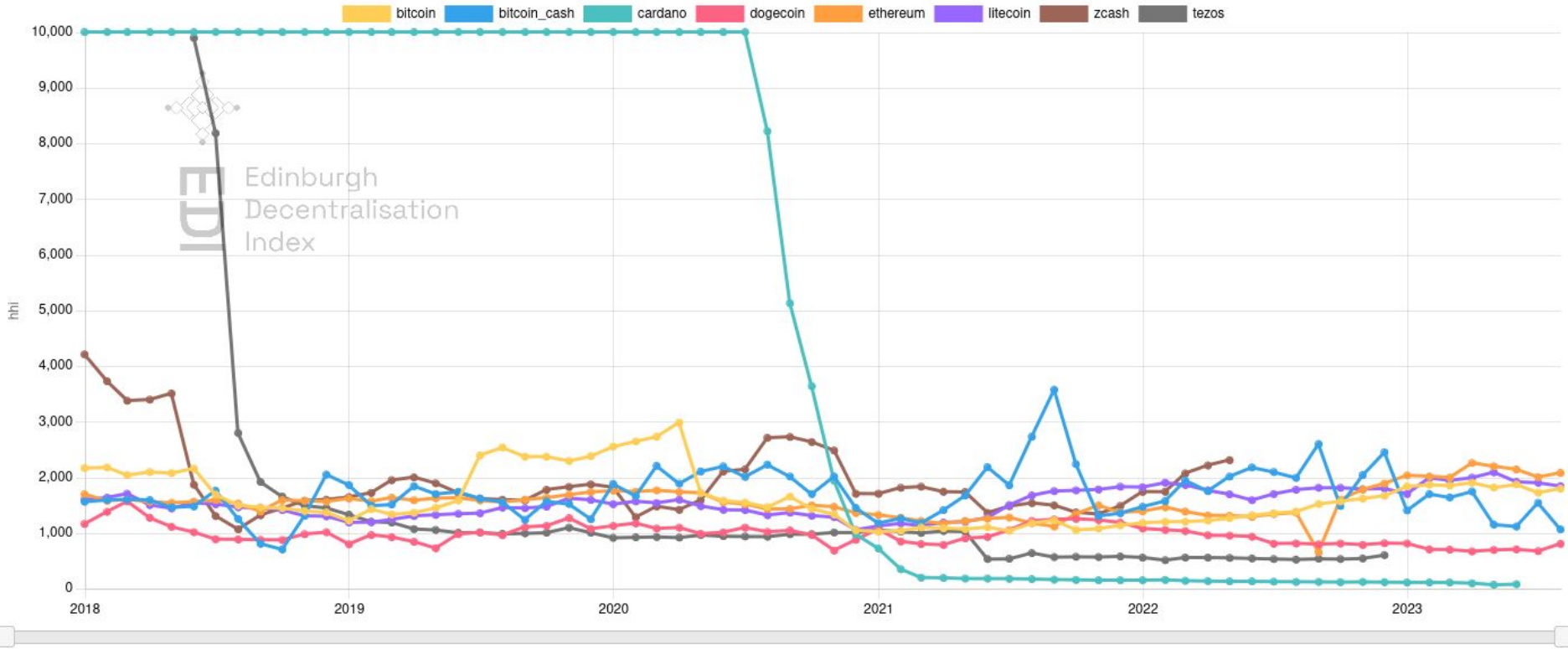
Nakamoto coefficient

The Nakamoto coefficient represents the minimum number of entities that collectively control more than 50% of the resources (in this case, the majority of mining / staking power).



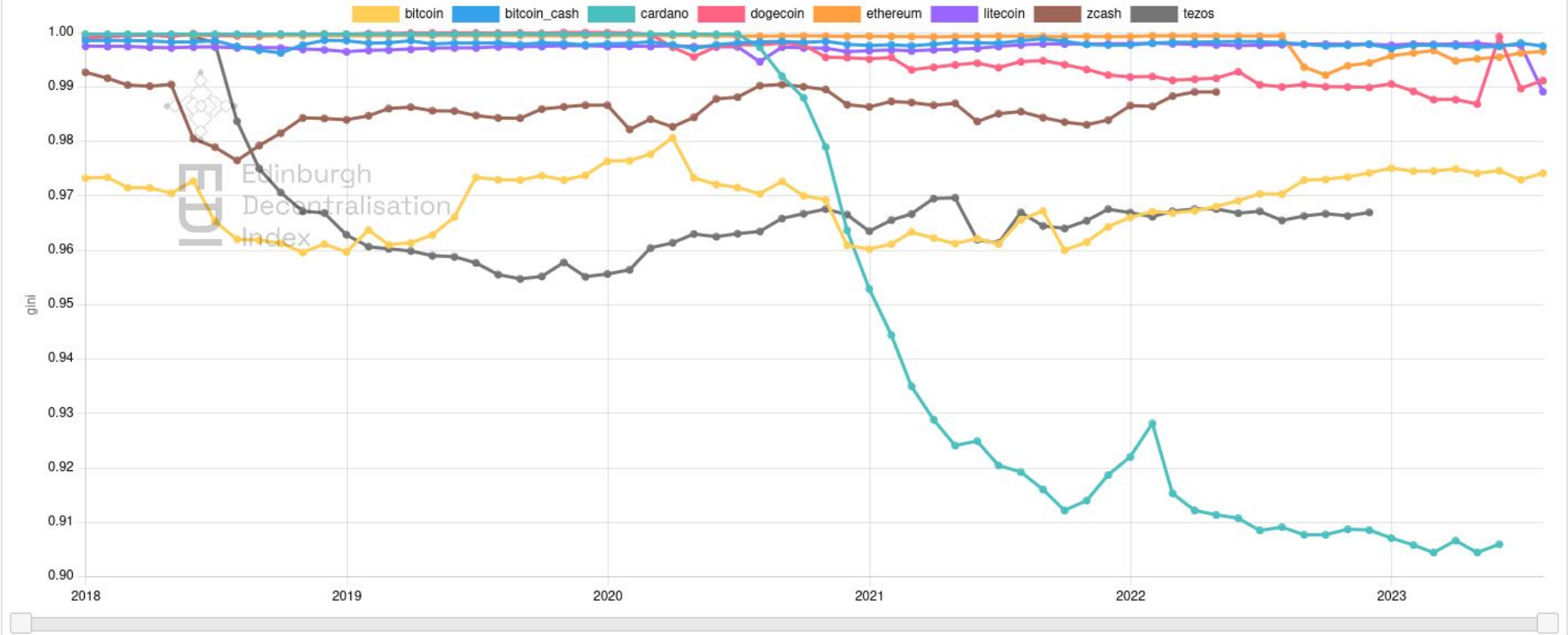
HHI

The Herfindahl-Hirschman Index (HHI) is a measure of market concentration. It is defined as the sum of the squares of the market shares (as whole numbers, e.g. 40 for 40%) of the entities in the system. Values close to 0 indicate low concentration (many entities produce a similar number of blocks) and values close to 10,000 indicate high concentration (one entity produces most or all blocks).



Gini coefficient

The Gini coefficient represents the degree of inequality in a distribution. Values close to 0 indicate high equality (in our case, all entities in the system produce the same number of blocks) and values close to 1 indicate high inequality (one entity produces most or all blocks).



Conclusion

- Decentralization is a spectrum
  - How close is a system to a single point of failure?
- A system may be decentralized in one layer but not others
  - In general, the consensus layer is more centralized than tokenomics
- Data pre-processing choices may affect the results
  - Clustering can help counter the effects of pseudonymity
  - Thresholding can change the results qualitatively
- No one metric can perfectly express decentralization - yet
  - The choice of metrics may affect a system’s classification
  - Some metrics are more sensitive to long distribution tails

Website

Dashboard