

---

# Long Transaction Chains and the Bitcoin Heartbeat

---

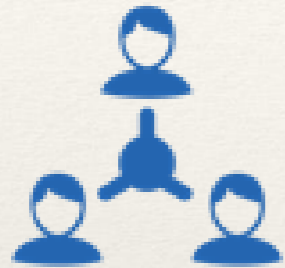
*Di Battista, Di Donato, **Pizzonia**  
Department of Engineering  
Roma Tre University*

*Perugia, Feb 1st, 2018*

---

# Bitcoin

---



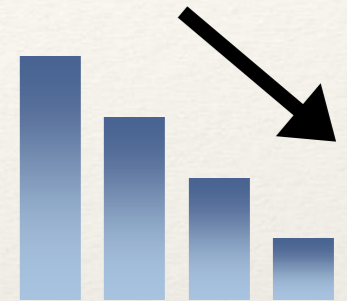
Peer-to-peer  
transactions



No need for  
banks



Worldwide and  
almost immediate

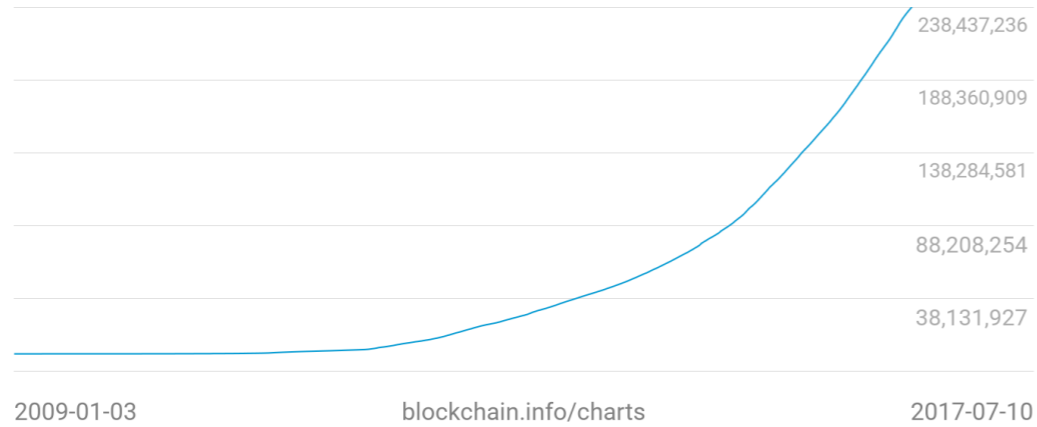


Low  
processing fees

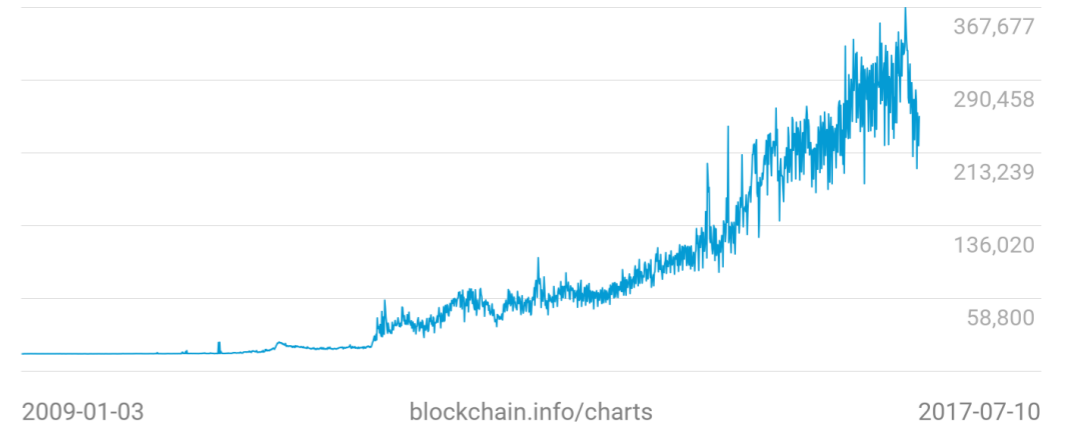
- Bitcoin is the most successful *cryptocurrency* and a digital *payment system*
- 2008: S. Nakamoto. **Bitcoin: A peer-to-peer electronic cash system.** Whitepaper on a popular cryptography mailing list
- 2009: released the first **bitcoin software** that launched the network and the first units of the cryptocurrency

# Some statistics...

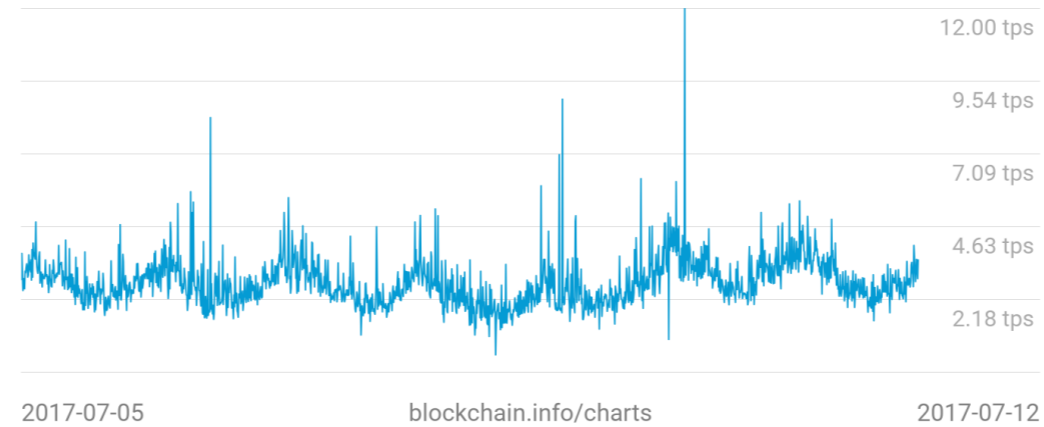
Total Number of Transactions  
**238,458,702**



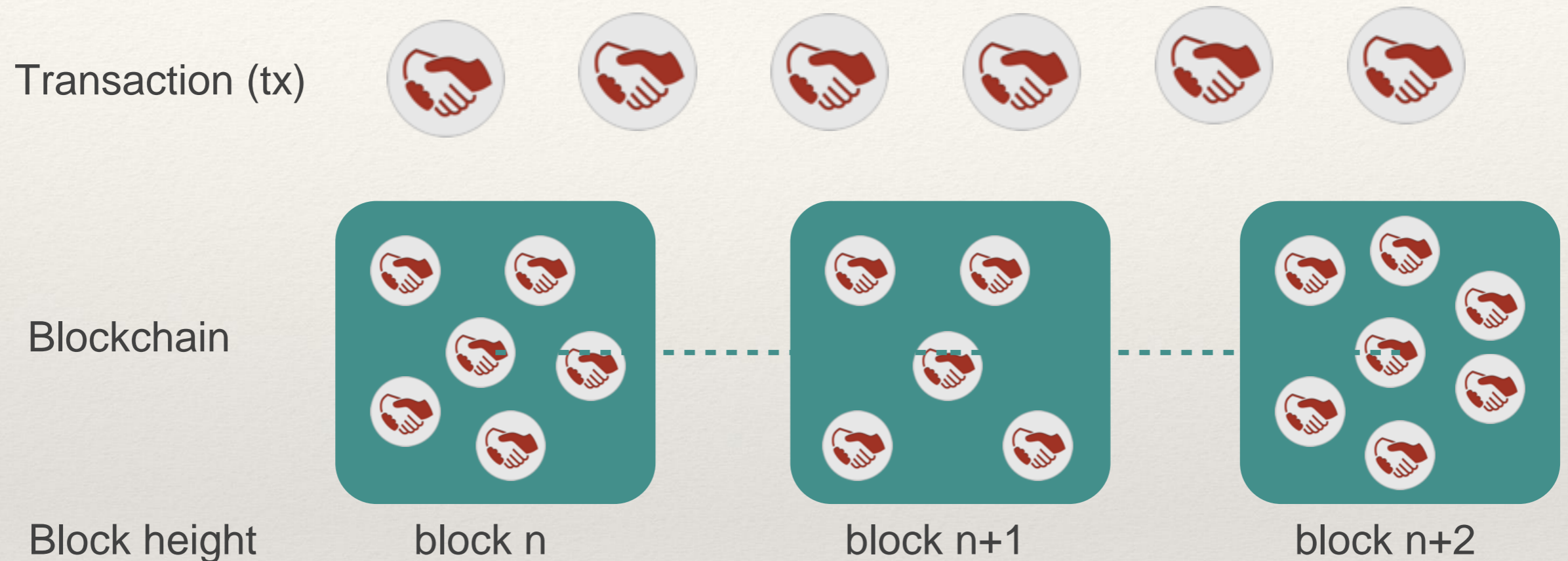
Confirmed Transactions Per Day  
**251,722**



Transaction Rate  
**3.51 tps**



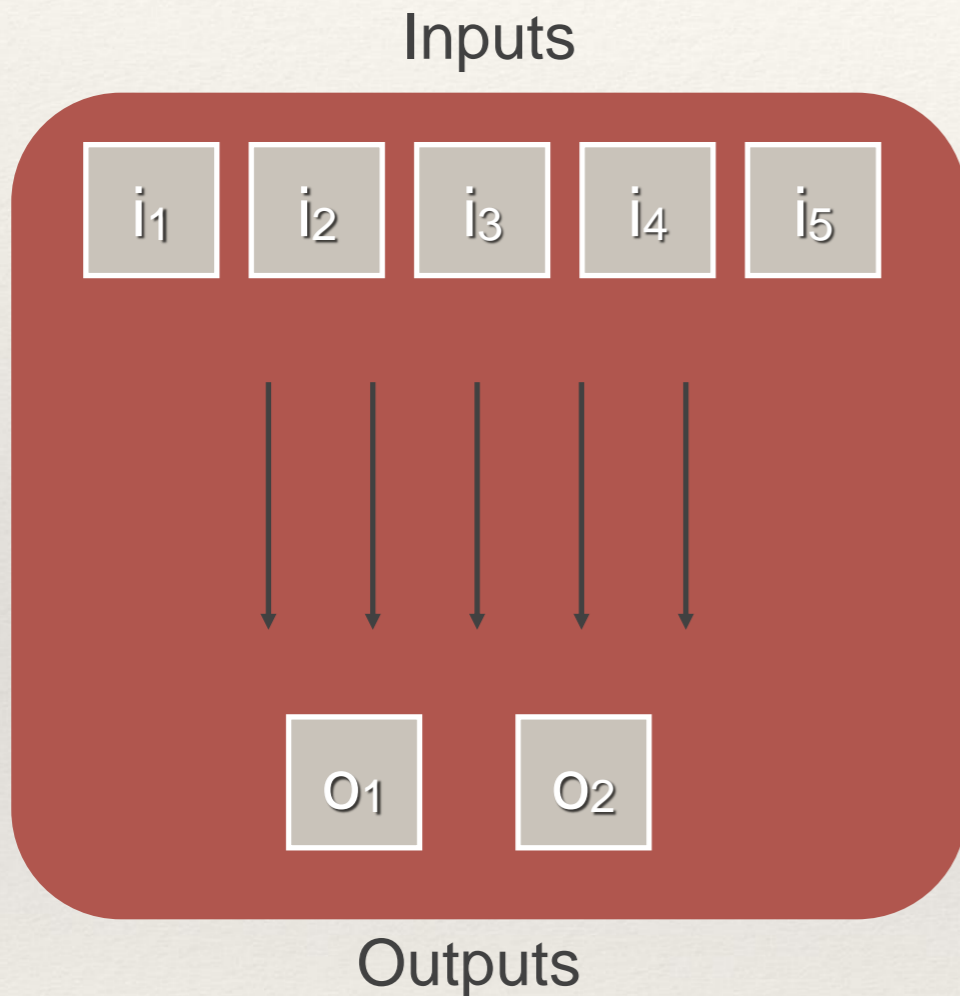
# Bitcoin Blockchain



- Bitcoins are transferred by means of Transactions (txs)
- All transactions are recorded in a public ledger called **Blockchain**
- Each **block** is identified by a number (block **height**)
- A new block is created approximately every 10 minutes

# Bitcoin Transactions

Transaction ID: 83de96b548febec40e9ecaa49a5e092a6019a6fa45513fee36af95bbd380dad5





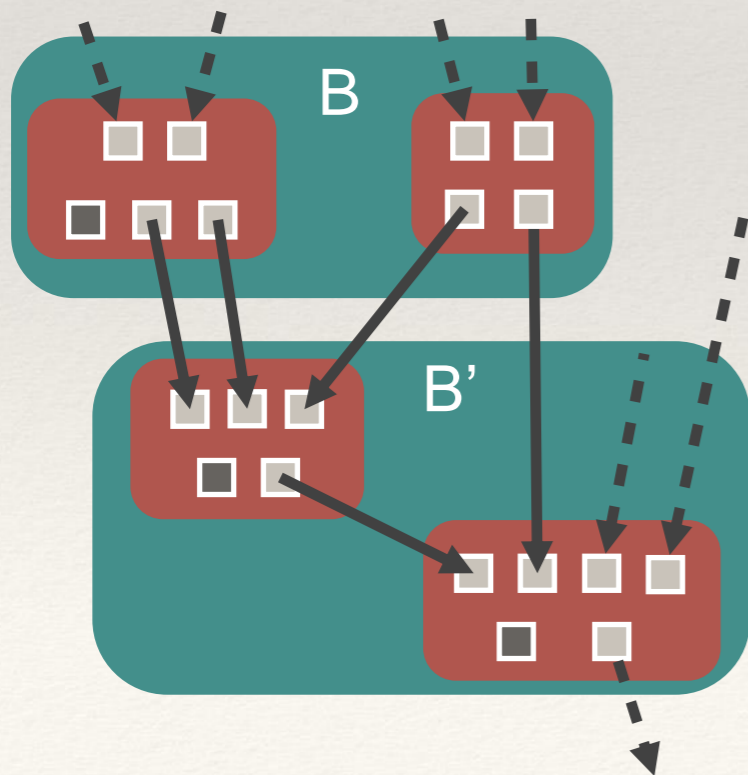
Note that each input contributes to each output

Inputs		
	Address	Amount (BTC)
$i_1$	12eJA61CbTKAENjsHY5wn9jmPKEBRYvDd1	1.5
$i_2$	1N98cnavM1pyTNWRV4QY96Hjpj7GJUDqy	0.01604777
$i_3$	13QcQafdZcHoWKUFGEuJU9M5vUcWwuntKK	0.01066444
$i_4$	1DWgbXbdJQhDdv84T7RfFuKAzoHCCSBmdL	0.15435157
$i_5$	1EKBqoZmJohssgHpw8Qs8DUdqu7bBrWYEO	7.0696295

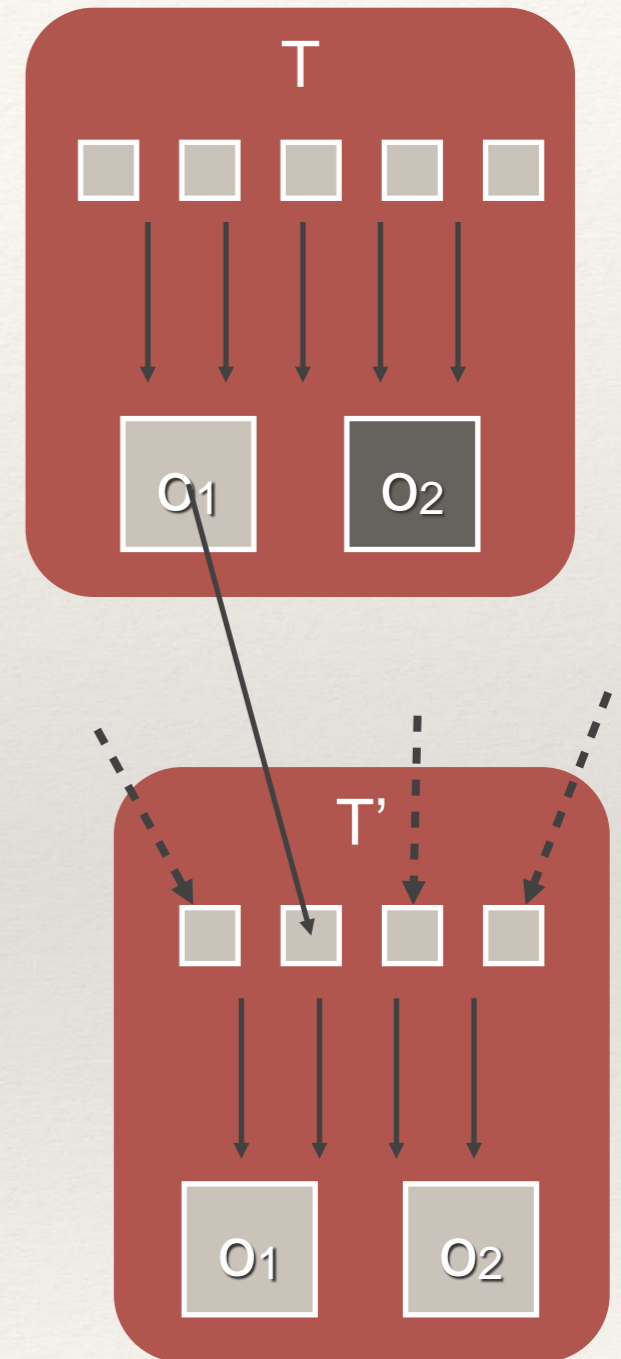
Outputs		
	Address	Amount (BTC)
$O_1$	1Gm23VjzAsbiic7NLPcgq6oeHLtoGVVX2k	0.01032442
$O_2$	1QNX18m2pe5q8p7rTY9AQqck8zxb98Umt	8.73736886

# Bitcoin Transactions (2)

- Once a tx T has been processed, the only way to spend its outputs is to use them as inputs for other txs
- Note that some outputs may be unspent (UTXOs)
-  o<sub>2</sub> is a UTXO whereas  o<sub>1</sub> has been spent in a subsequent tx T'
- Other inputs of T' come from other output txs
- Transactions define a **directed acyclic multi-graph**

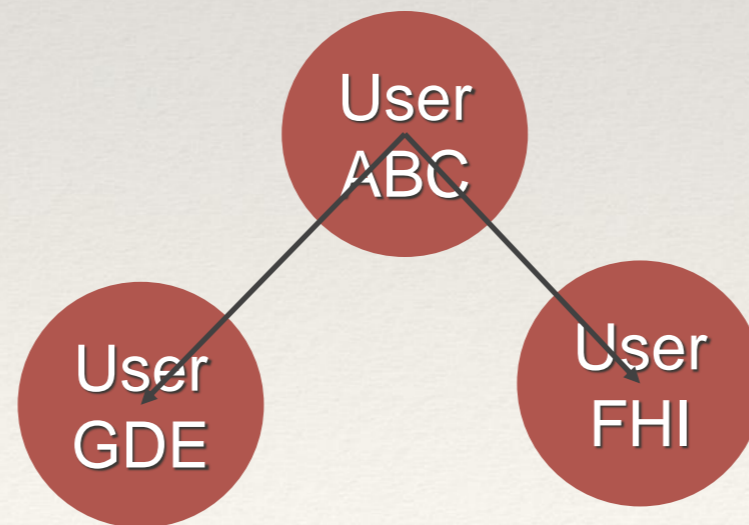
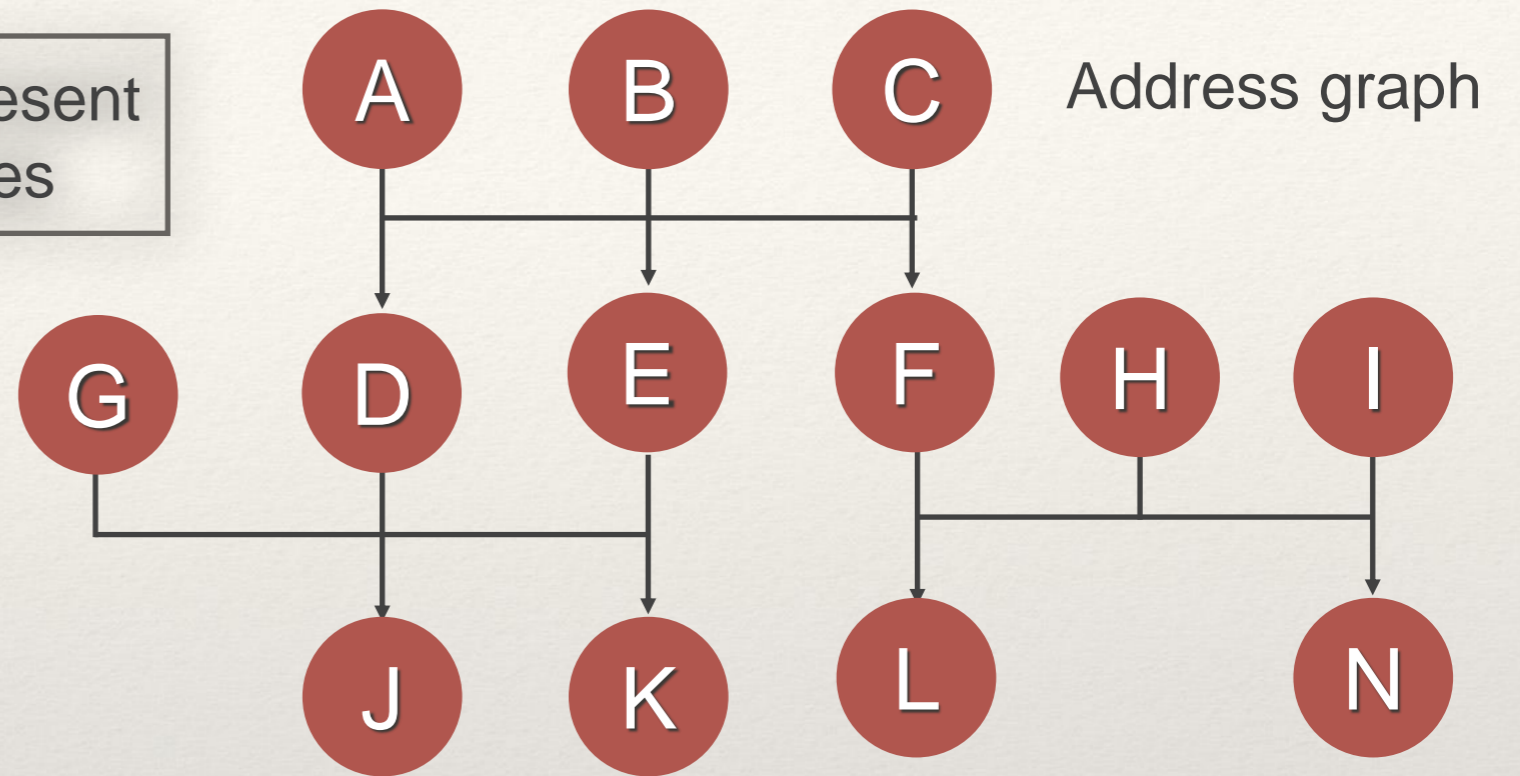
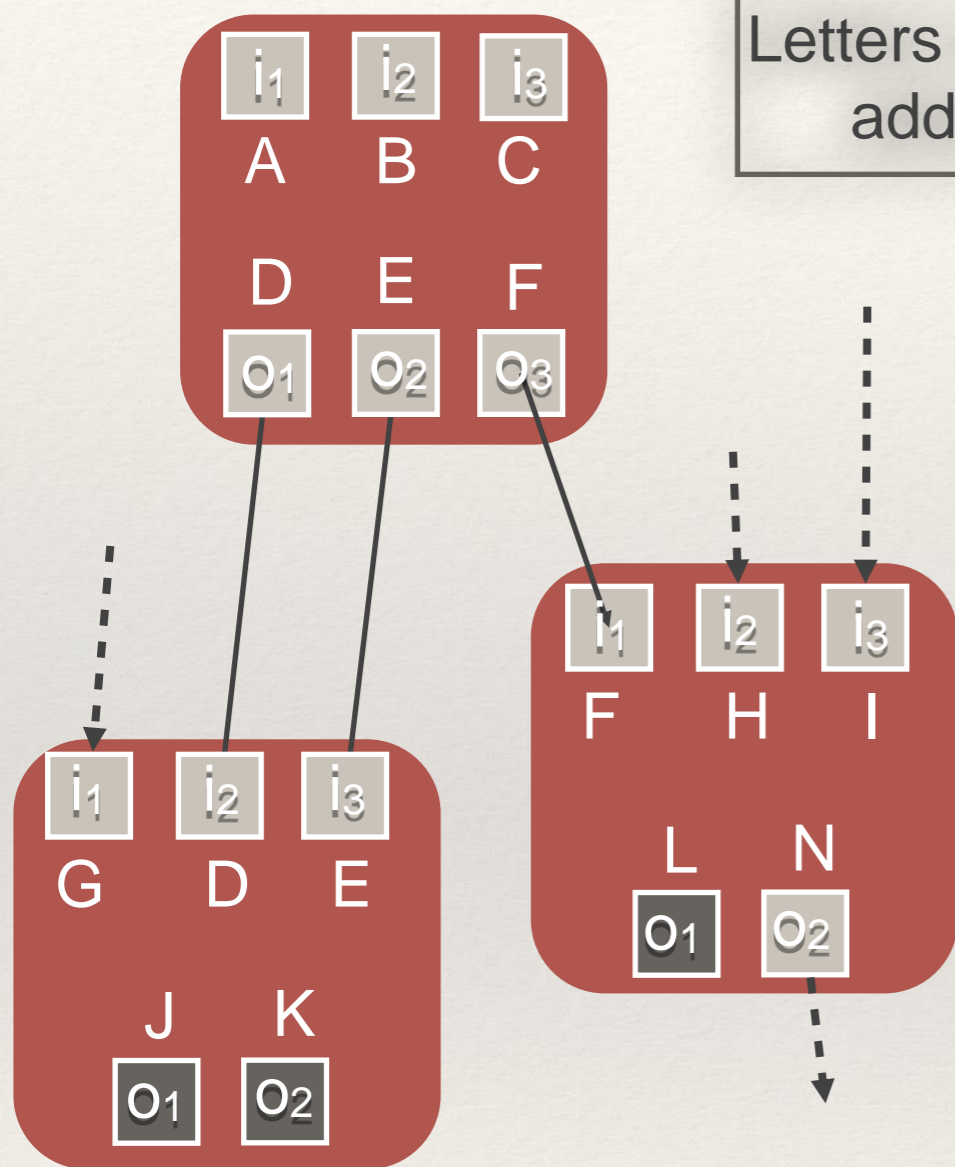


Note that dashed lines connect inputs/outputs of txs that are not drawn here



# Different Types of Graphs

Letters represent addresses



---

# State of the Art in Bitcoin Blockchain Analysis

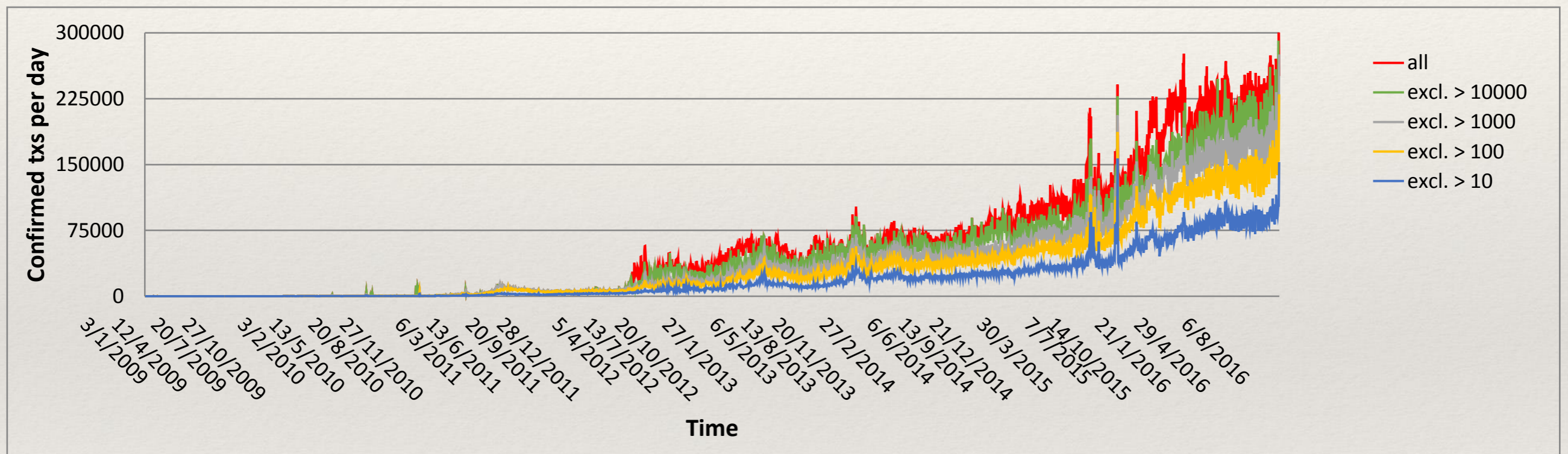
---

- 2011: Reid F., Harrigan M. - Analyze the degree of anonymity provided by Bitcoin. They work with the **transaction graph** and with the **entity graph**
- 2013: Ron D., Shamir A. - Analyze a variety of questions about the typical behaviour of users, how they acquire and how they spend their bitcoins etc. They work with the **address graph** and with the **entity graph**
- 2013: Ober M., Katzenbeisser S., Hamacher K. - Discusses anonymity aspects of the Bitcoin protocol. They work with the **address graph**
- 2015: Di Battista G., Di Donato V., et al. - Design a system for the visual analysis of flows in the bitcoin transaction graph. They work with the **transaction graph**
- 2016 and 2017: Di Francesco M.D., Marino A., Ricci L. - Analyze a variety of questions about the time evolution of Bitcoin and its users behaviour. They work with the **entity graph**



# Number of Daily Transactions and Long Chains

Starting point: “There are many legitimate reasons to create long transaction chains; however, they may also be caused by coin mixing or possible attempts to manipulate transaction volume”. [www.blockchain.info](http://www.blockchain.info)



Observations:

1. **Persistent growth** of the number of confirmed txs per day
2. Slower growth when excluding txs belonging to **long chains** forming in 24 h

Question: is it possible to distinguish between **human** and **non-human** activity?

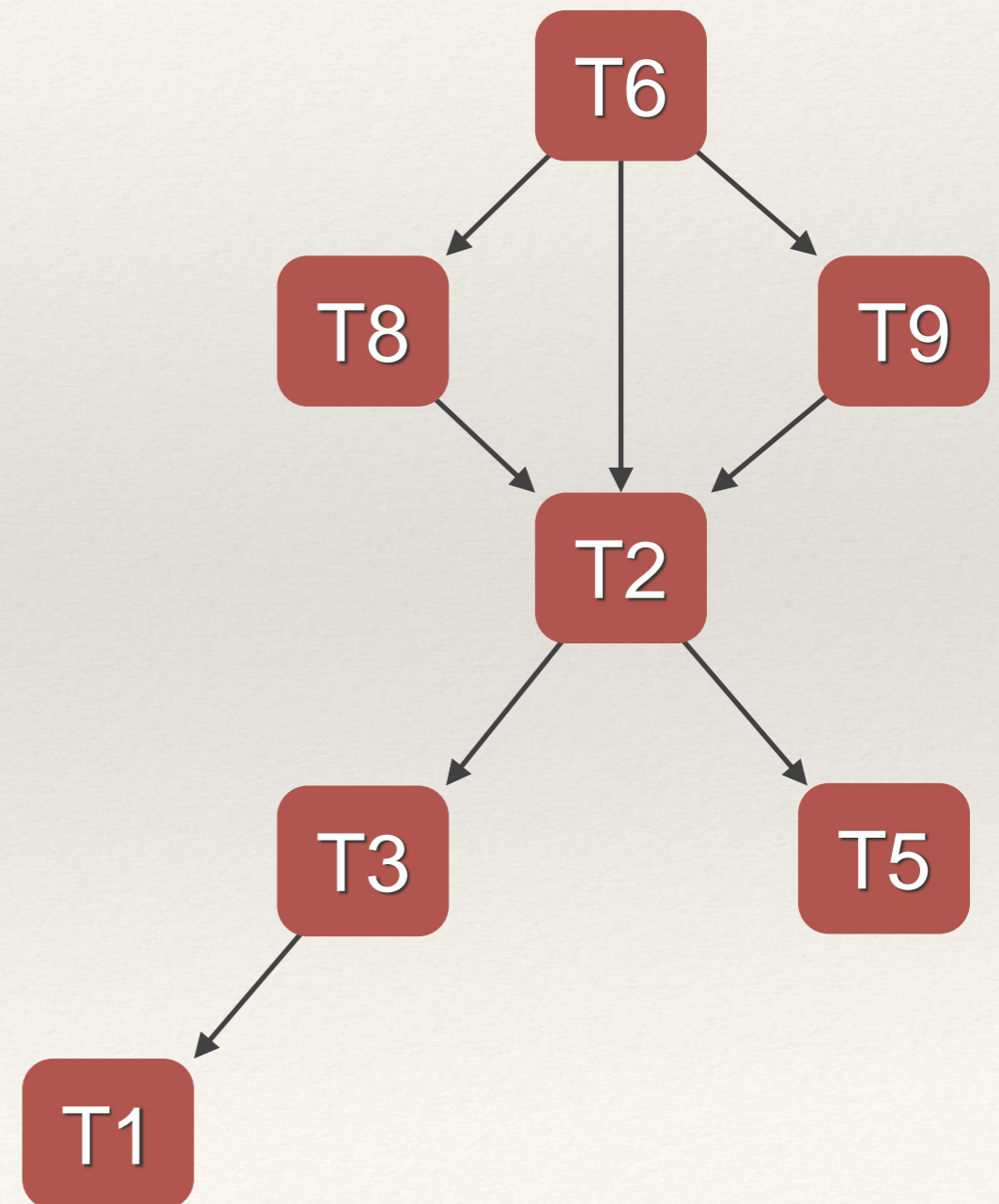
# Algorithm to Label Transactions with LLC

LLC (tx) = Length of Longest Chain tx lays on

Input: A subgraph of the transaction graph induced by two block heights:  $G(b', b'') = (V, E)$

Output: a label for each node  
**LLC: length of the longest chain**

Intermediate step:  
Label each node with two numbers  
**b: length of the longest chain backward**  
**f: length of the longest chain forward**



# Algorithm to Label Transactions with LLC

LLC (tx) = Length of Longest Chain tx lays on

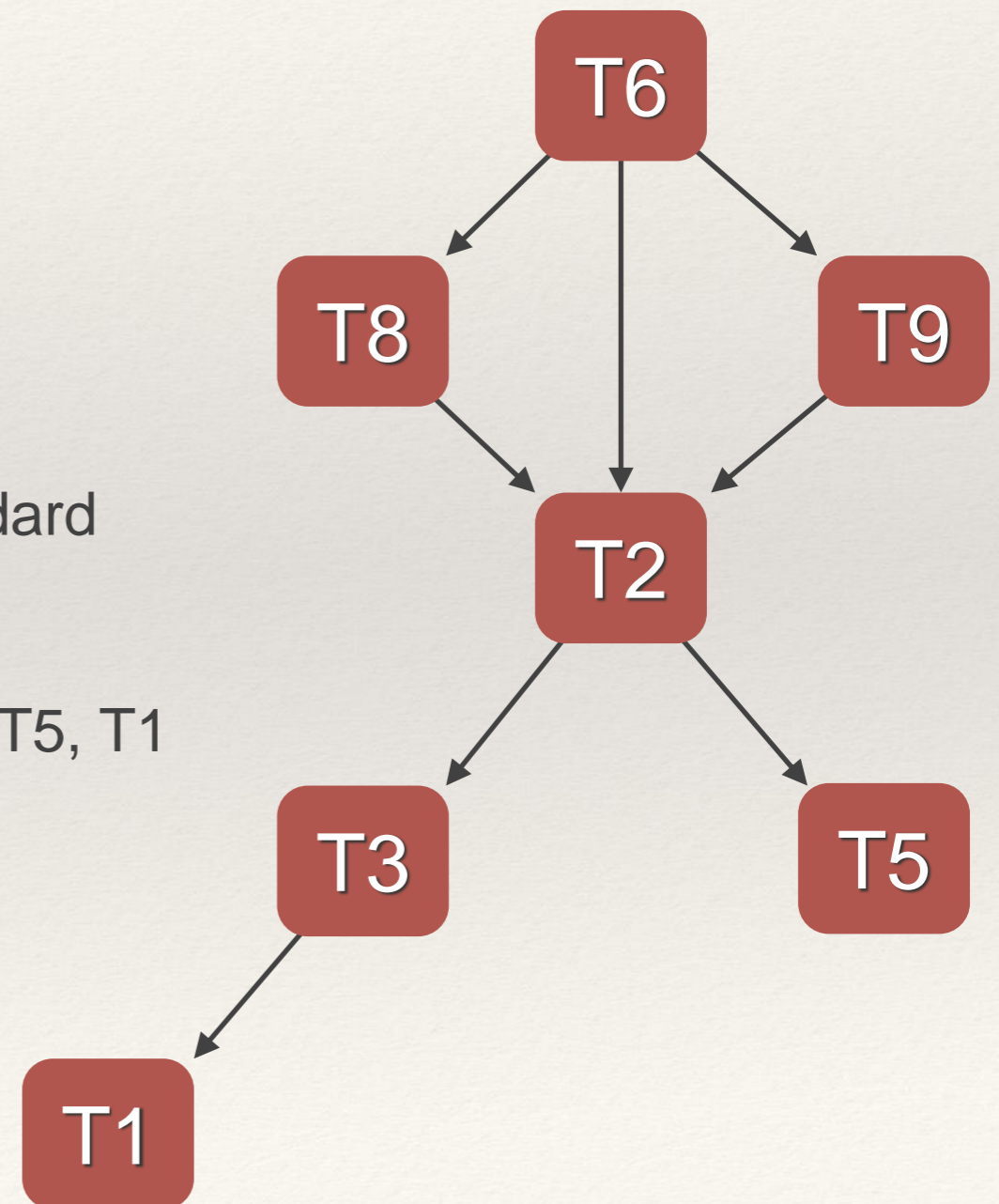
Step 1:

Compute a *topological ordering* of the nodes of  $G$

A *topological ordering* of the vertices of a directed acyclic graph is an ordering such that for each edge  $(i, j) \in E$ , vertex  $i$  precedes vertex  $j$

This can be done in linear time with a dfs and standard data structures such as adjacency lists and queues

TOPOLOGICAL ORDERING = T6, T8, T9, T2, T3, T5, T1



# Algorithm to Label Transactions with LLC

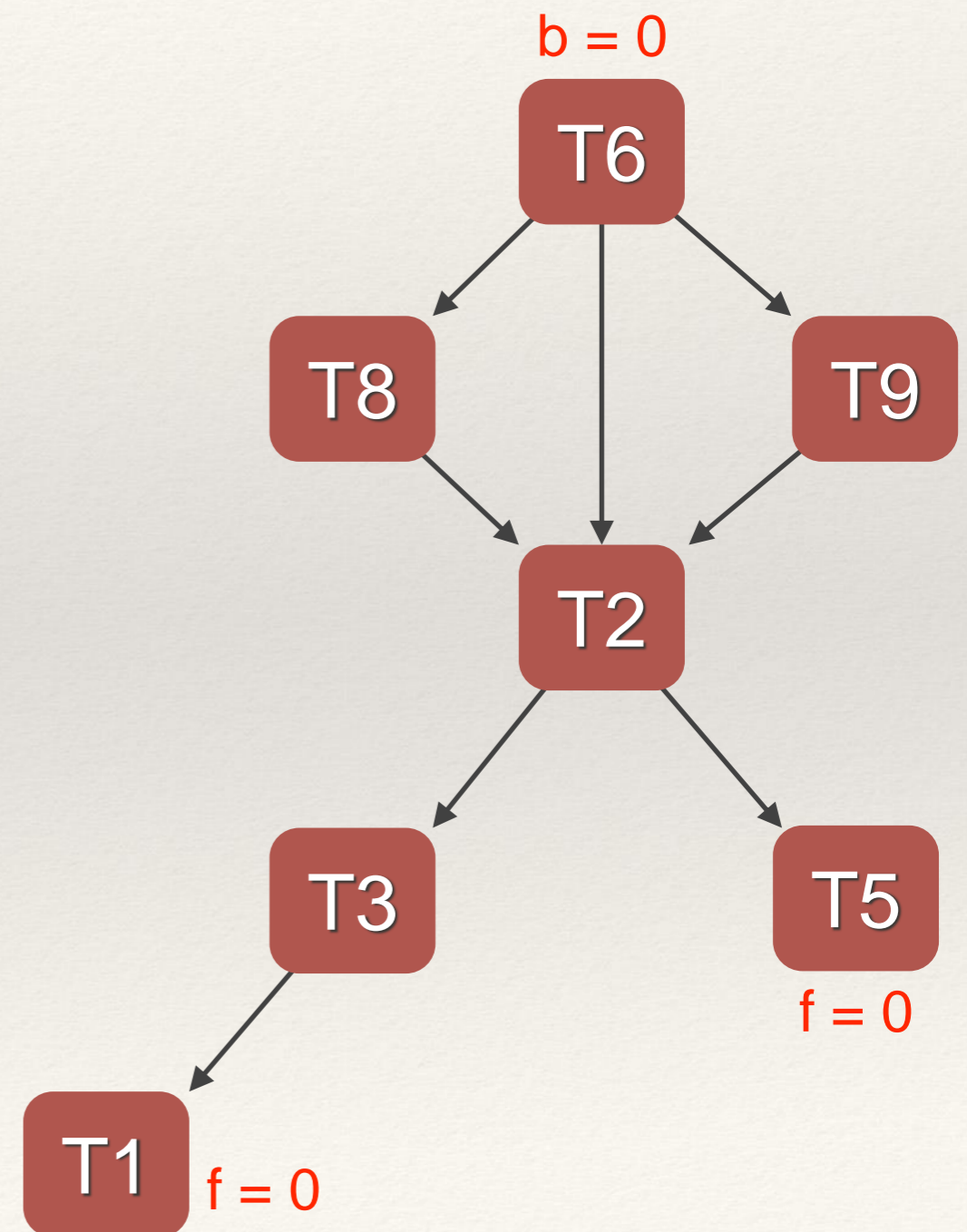
LLC (tx) = Length of Longest Chain tx lays on

Step 2:

Label each node T with in-degree = 0 with  $b = 0$

Label each node T with out-degree = 0 with  $f = 0$

This can be done in linear time



# Algorithm to Label Transactions with LLC

LLC (tx) = Length of Longest Chain tx lays on

Step 3:

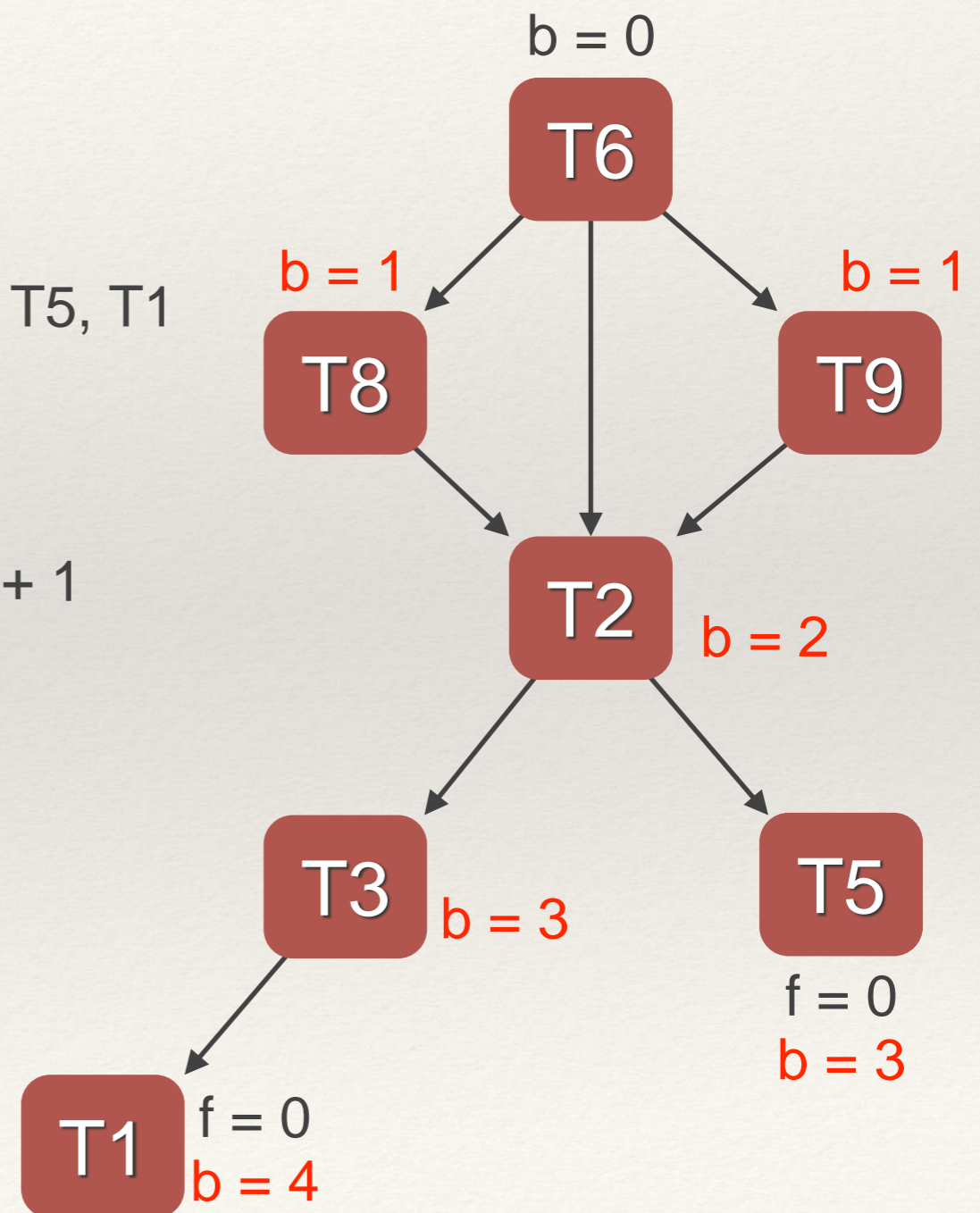
Following the TOPOLOGICAL ORDERING = T6, T8, T9, T2, T3, T5, T1 do the following:

if the number of predecessors of tx T is  $\neq 0$

$$b(T) = \max(\text{over all } b \text{ attributes of predecessors}) + 1$$

Note that  $b(T2) = \max(1, 0, 1) + 1 = 2$

This can be done in linear time



# Algorithm to Label Transactions with LLC

LLC (tx) = Length of Longest Chain tx lays on

Step 4:

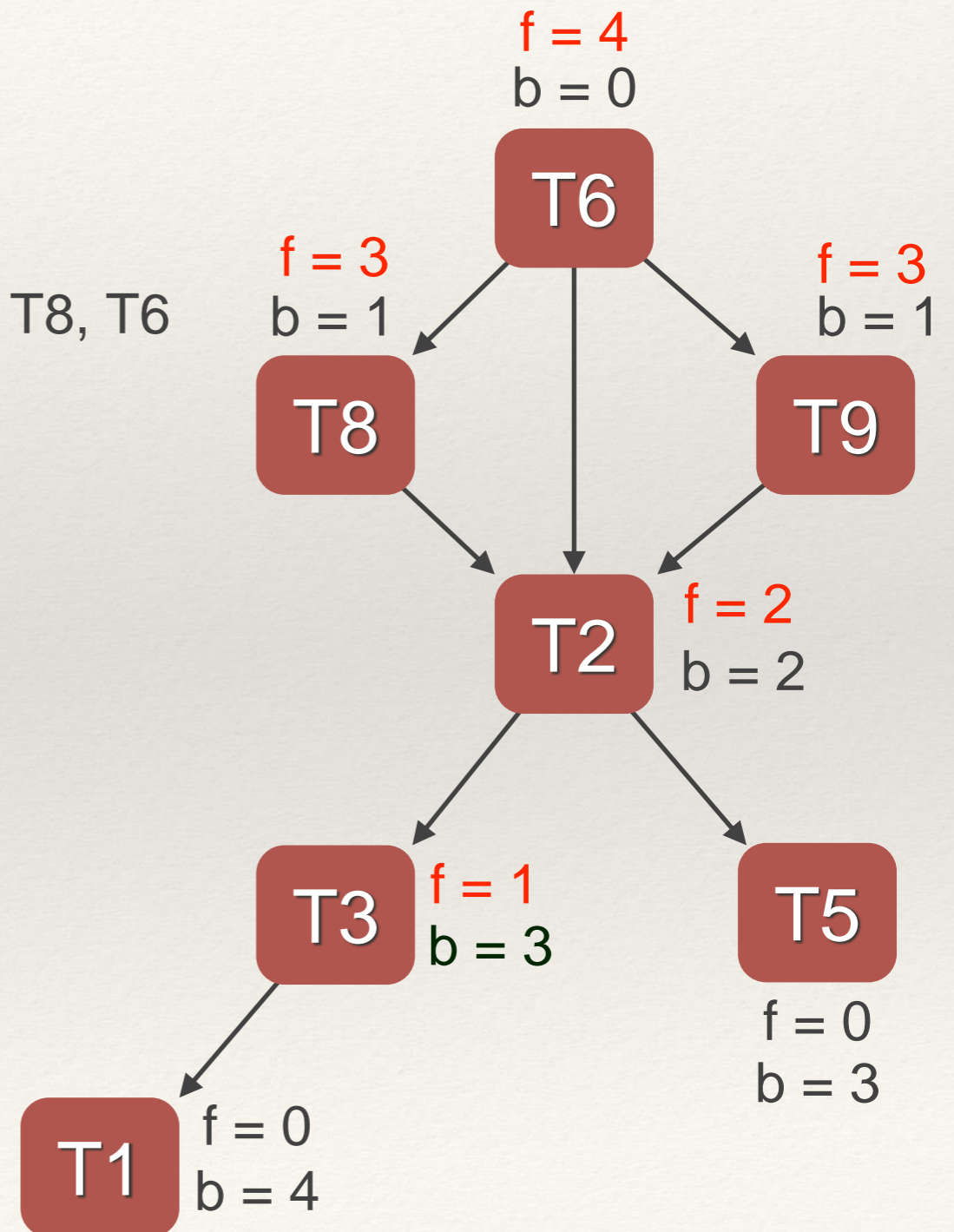
Following the **reversed**  
TOPOLOGICAL ORDERING = T1, T5, T3, T2, T9, T8, T6  
do the following:

if the number of successors of tx T is  $\neq 0$

$$f(T) = \max(\text{over all } f \text{ attributes of successors}) + 1$$

Note that  $f(T6) = \max(3, 2, 3) + 1 = 4$

This can be done in linear time



# Algorithm to Label Transactions with LLC

LLC (tx) = Length of Longest Chain tx lays on

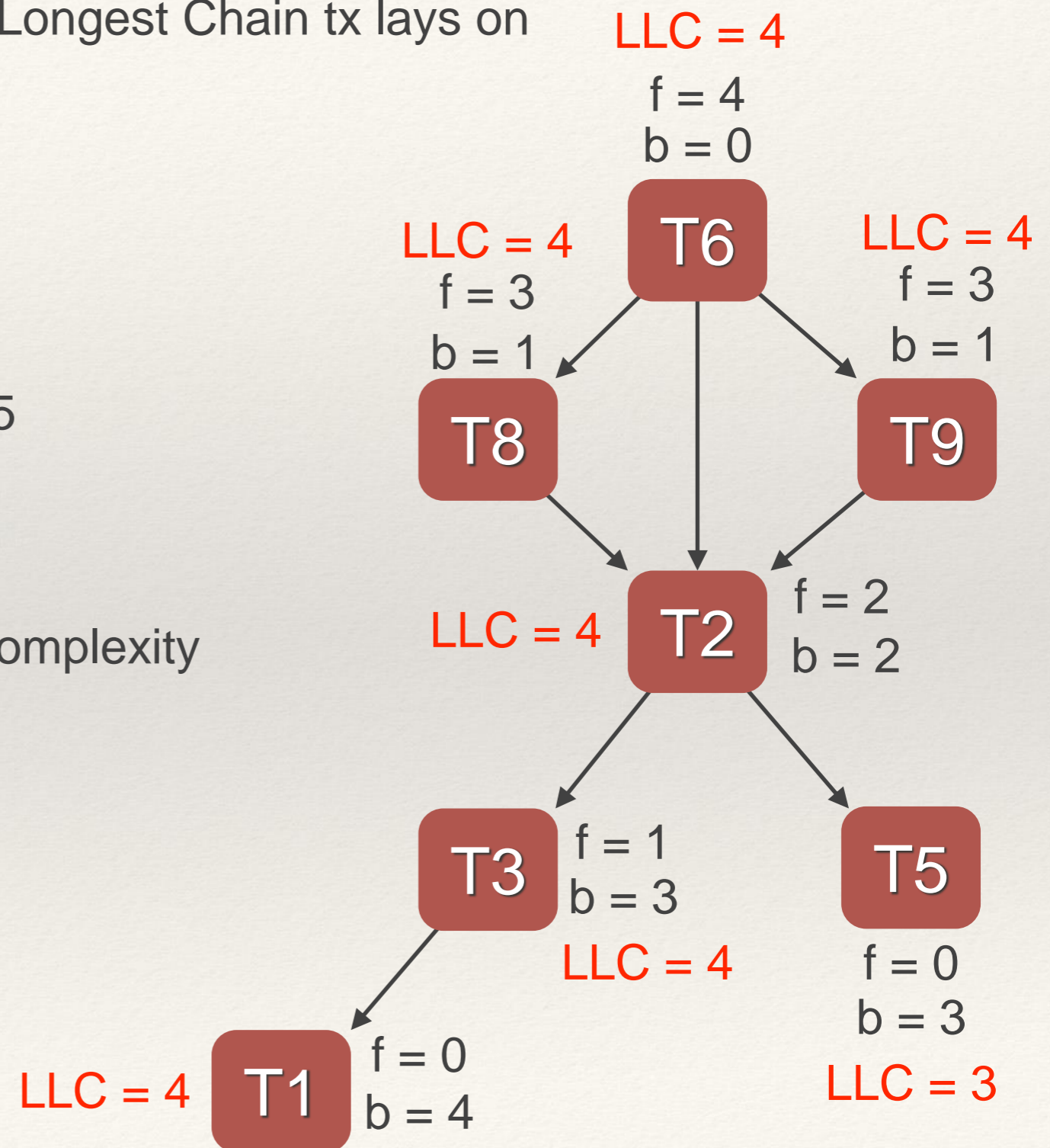
Step 5:

For each tx T:  $LLC(T) = b(T) + f(T)$

All transactions  $T_i$  have  $LLC = 4$  except  $T5$

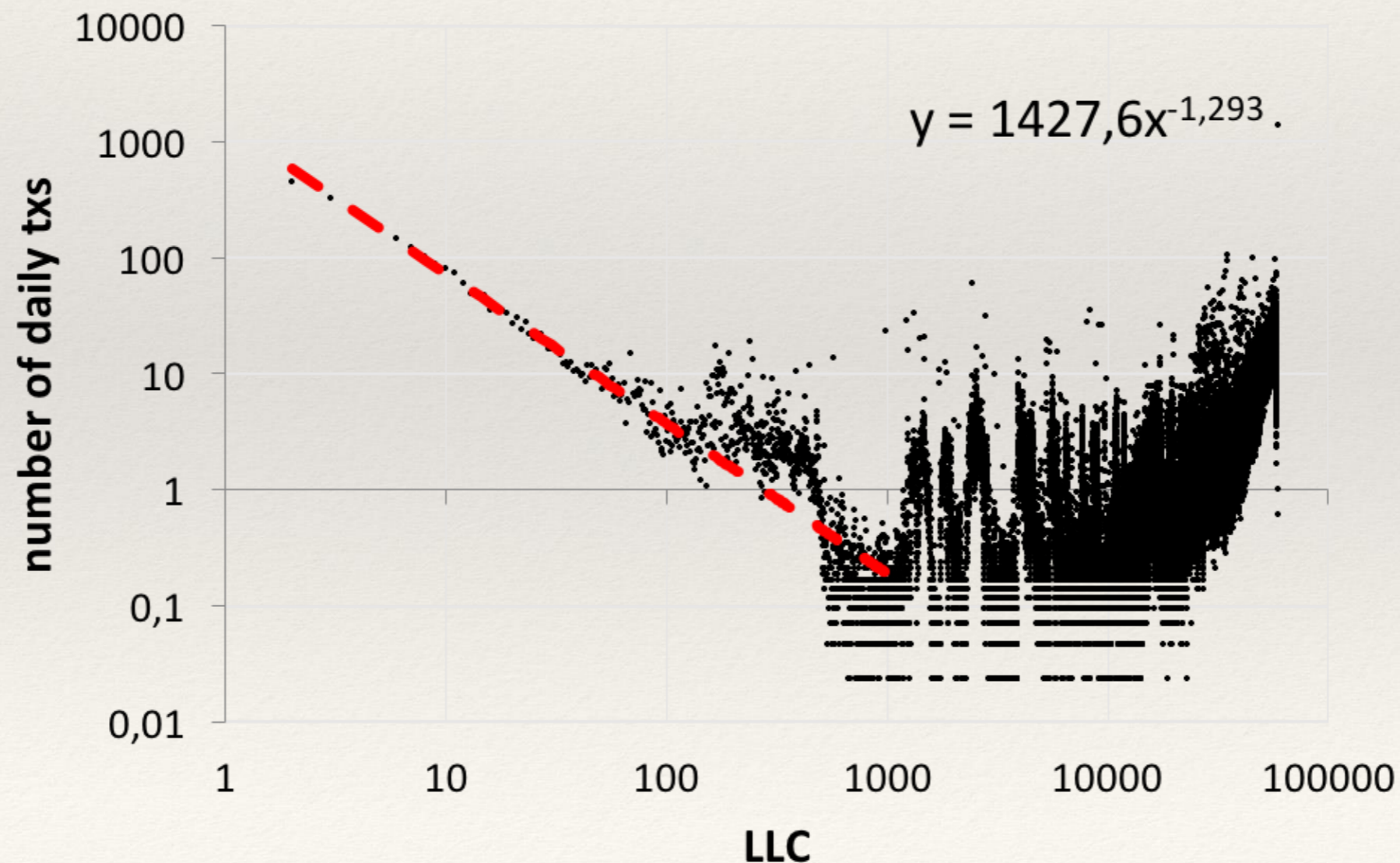
This also can be done in linear time

The algorithm has an overall linear time complexity



# Distribution of LLC over 42d

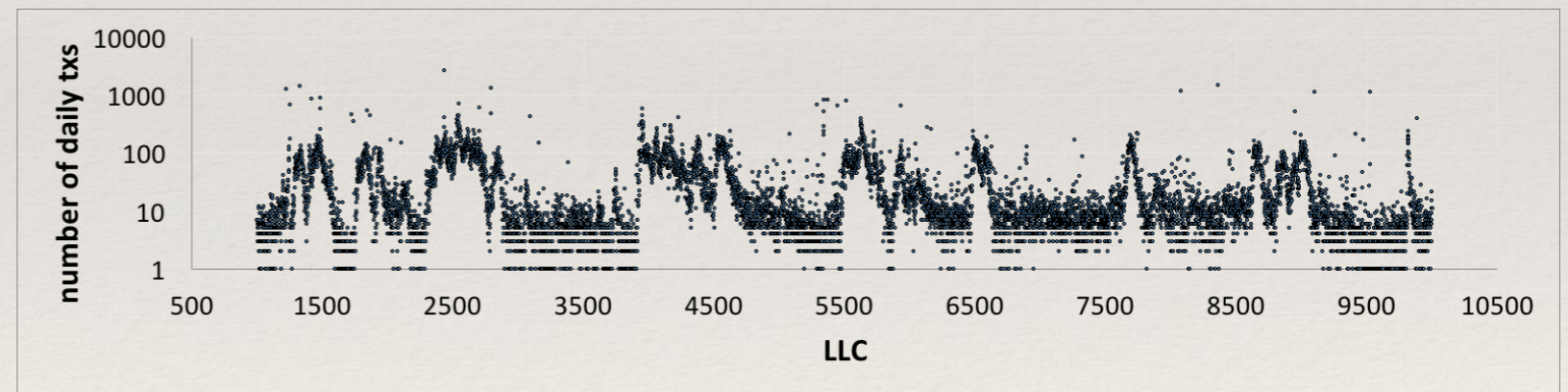
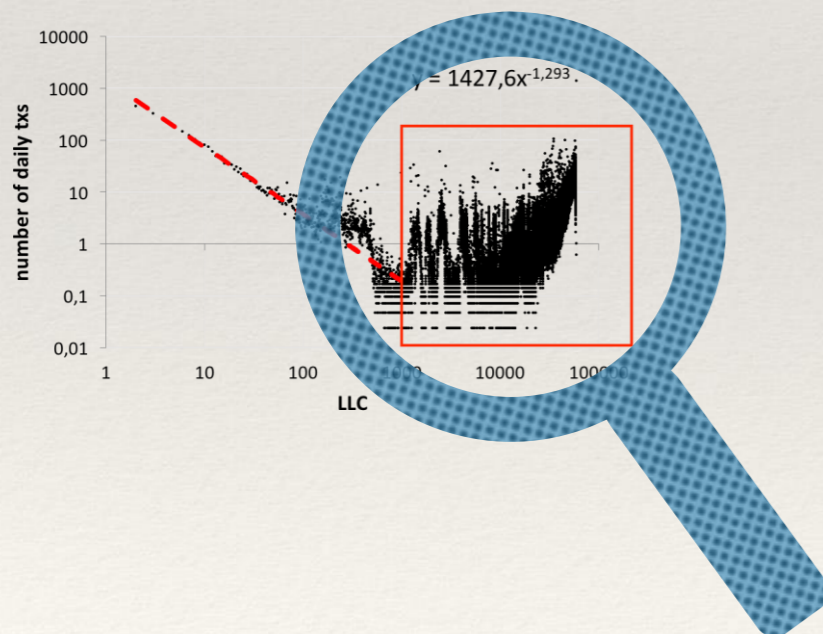
- Build the tx-graph corresponding to 42d of activity (the most we could do with our machines...)
- Look at the PDF and CDF of LLC of all nodes normalizing values to 24h





# Overlapping phenomena?

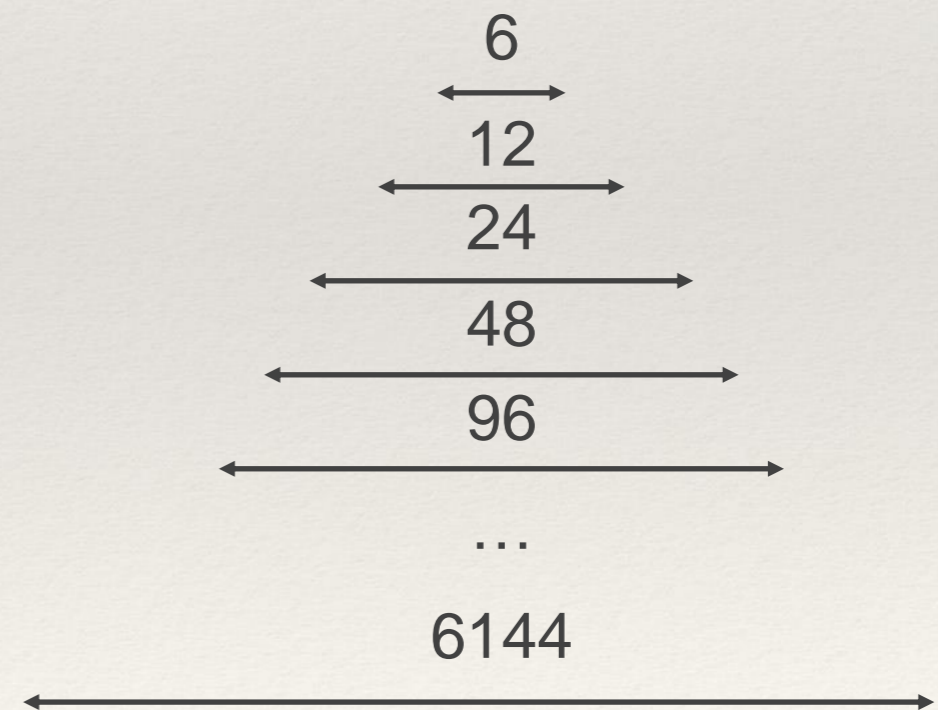
- Zoom of the tail of the previous figure
- A series of consecutive peaks that might be interpreted as a sequence of automatic phenomena, each of which introduces at its own frequency new “artificial” transactions in the Blockchain



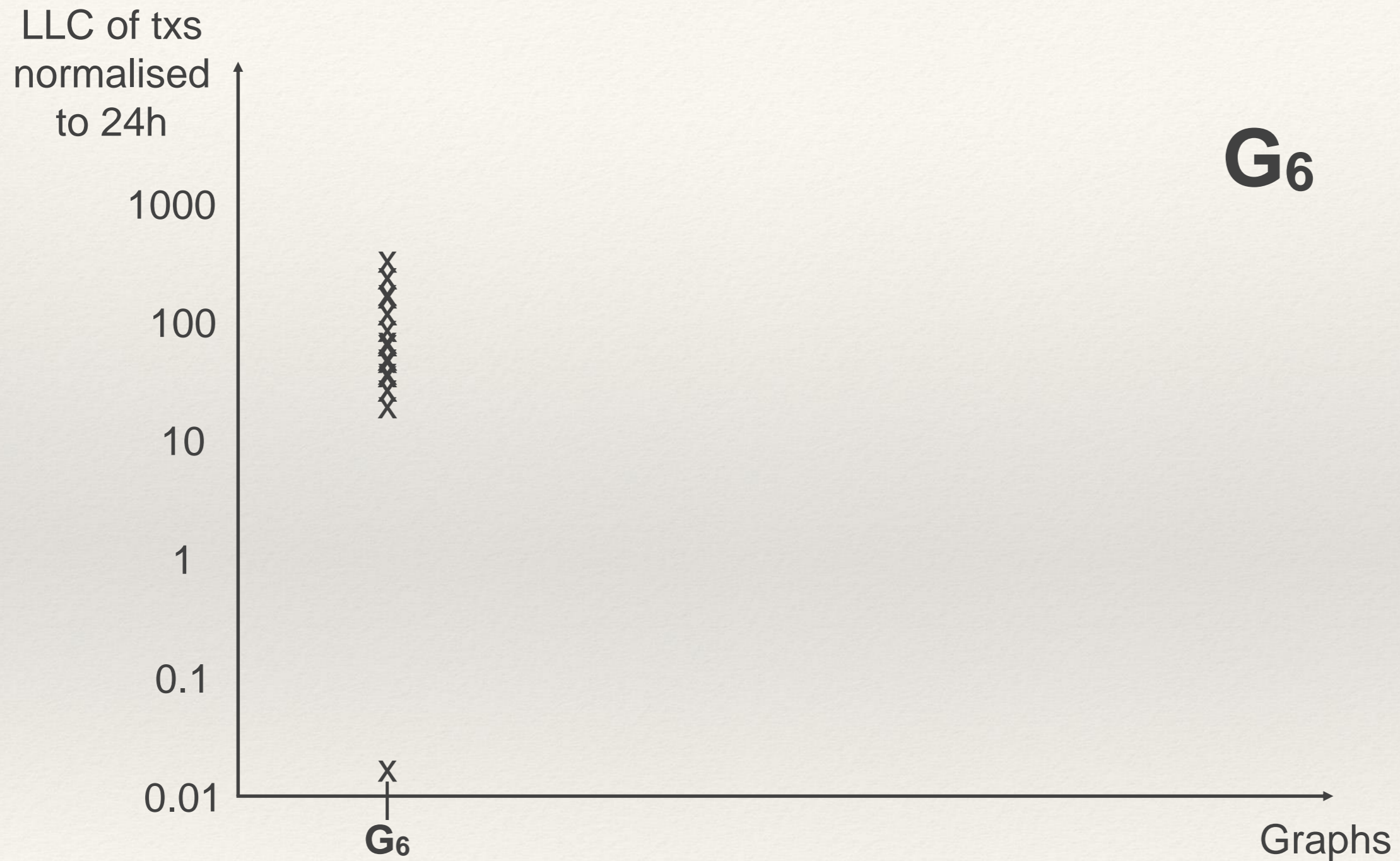
linear scale for x-axis and log-scale for y axis

# Analyzing the Evolution of a Specific Set of txs

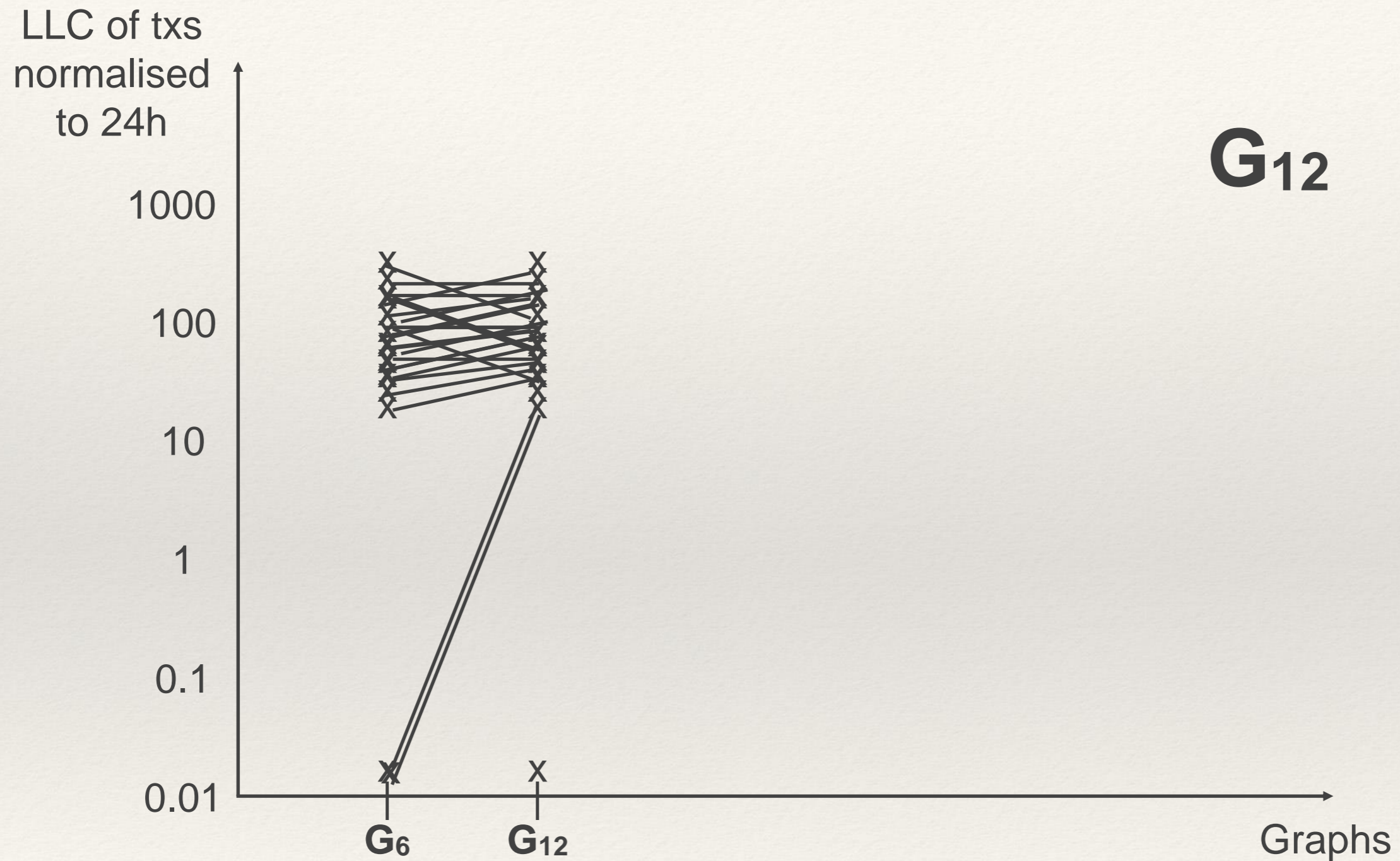
- Question: how does LLC evolve over time?
- Experiment 3
- Randomly picked one recent block B and consider all of its txs  $T_i$
- Considered growing intervals of blocks centered in B and built the corresponding graphs  $G_k$
- At each iteration tracked the values of LLC for all txs  $T_i$
- Normalize values to 24h



# Analyzing the Evolution of a Specific Set of txs

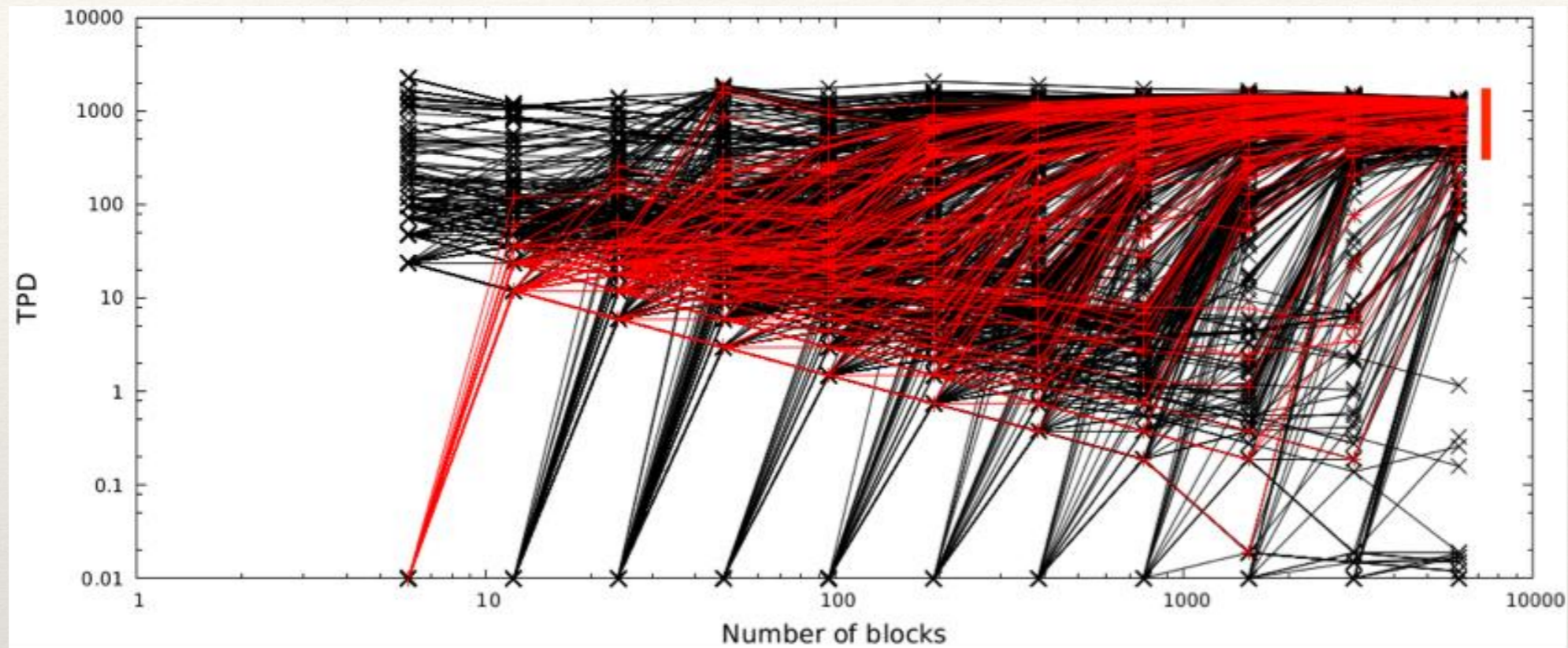


# Analyzing the Evolution of a Specific Set of txs



...and so on...

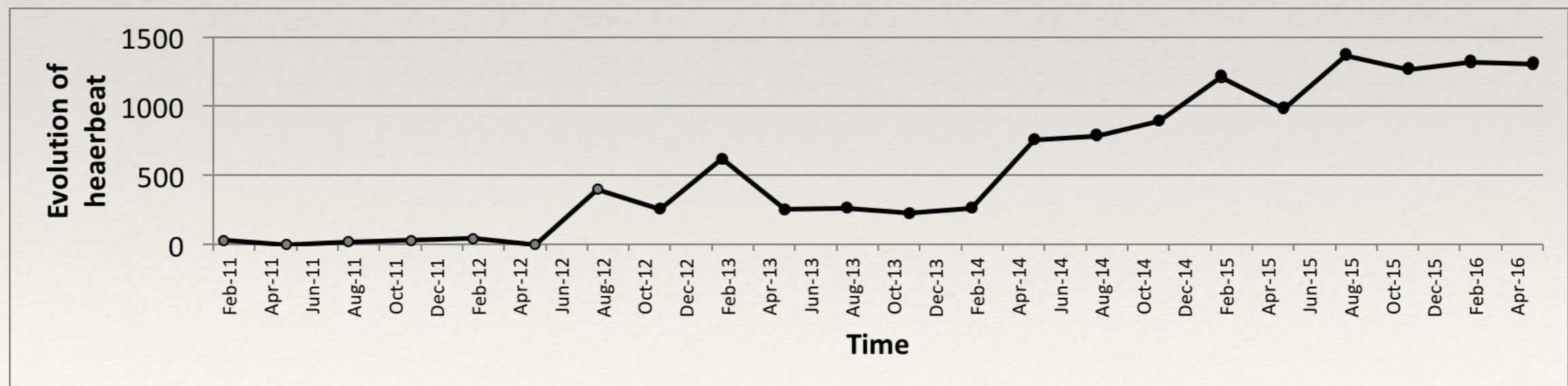
# Analyzing the Evolution of a Specific Set of txs



- X-Axis: Number of blocks considered for graph  $G_k$ : [6, 12, 24, ..., 6144]
- Y-Axis: Values of LLC of txs in  $G_k$  normalized to 24h (which is the Number of TxS Per Day)
- Each tx is represented by a set of points linked by a curve, each showing its TPD in a graph  $G_k$
- Red curves refer to txs whose LLC changes simultaneously when going from  $G_6$  to  $G_{12}$
- Interestingly, TPD for almost all txs, in the long run, converges to a value included in [300, 1300]
- This suggests that after some time, most txs in B will be connected to chains that evolve at the pace of  $h$  TPDs, with  $h \in [300, 1300]$  (see red bar)

# A Bitcoin Heartbeat

- Question: How did  $h$  change over time?
- Experiment 4
- Build 22 families of graphs such that each of them refers to 6144 consecutive blocks
- The 22 families of graphs correspond to intervals of blocks centred in a random block of the first day of the months Feb., May, Aug. and Nov. of years 2011 - 2016
- For each family, build a graph similar to the previous one and consider red txs only
- Compute the  $h$ -interval for each family of graphs
- Since the  $h$ -interval is the set of frequency values where txs tend to converge over time, we call its average value the *Bitcoin Heartbeat*



---

# Conclusions

---

- The distribution of the lengths of the longest chains passing through txs exhibit a shape that is hard to believe to be produced by explicit human activities (low frequency portion that resembles a power-law distribution + high frequency portion that contains several peaks)
- In the long term, txs surprisingly tend to lay on chains with frequencies distributed in a somehow small interval. We call the average of such interval the *Bitcoin Heartbeat*
- The *Bitcoin Heartbeat* has a rather stable value that has slowly grown over time.

---

Thank you

---

Questions?