

“Bitcoin Bundle”

Seminar in Distributed Computing

Papers

“An Analysis of Anonymity in the Bitcoin System”

Fergal Reid and Martin Harrigan
Univeristy College Dublin, Ireland

“Quantitative Analysis of the Full Bitcoin Transaction Graph”

Dorit Ron and Adi Shamir
The Weizmann Institute of Science, Israel

Agenda

- 1. What is “Bitcoin”?**
2. Anonymity Analysis
3. Quantitative Analysis
4. Discussion

What is “Bitcoin”?

What is “Bitcoin”?

“A digital currency in which transactions can be performed without the need for a central bank.”

‘bitcoins can be used for online transactions between individuals’

source: oxforddictionaries.com

What is “Bitcoin”?

“A digital currency in which transactions can be performed without the need for a central bank.”

‘bitcoins can be used for online transactions between individuals’

source: oxforddictionaries.com

#p2p public-keys reassignments signatures public consensus

What is “Bitcoin”?

Example of a transaction on the blackboard.

Scenario: Alice sends 20 BTC to Bob.

What is “Bitcoin”?

For more details, please read:

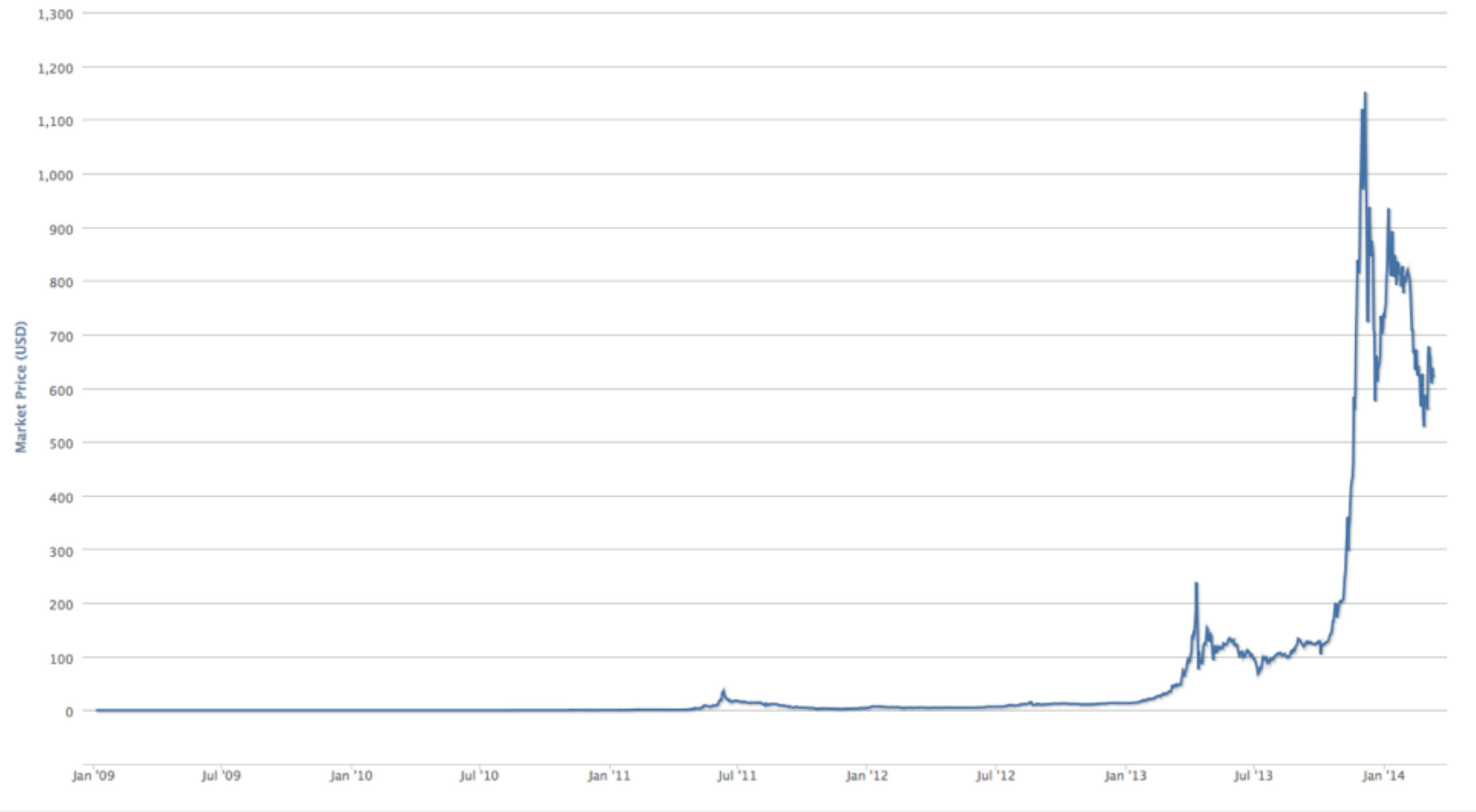
“Bitcoin: A Peer-to-Peer Electronic Cash System”,
Satoshi Nakamoto, 2008

Participants

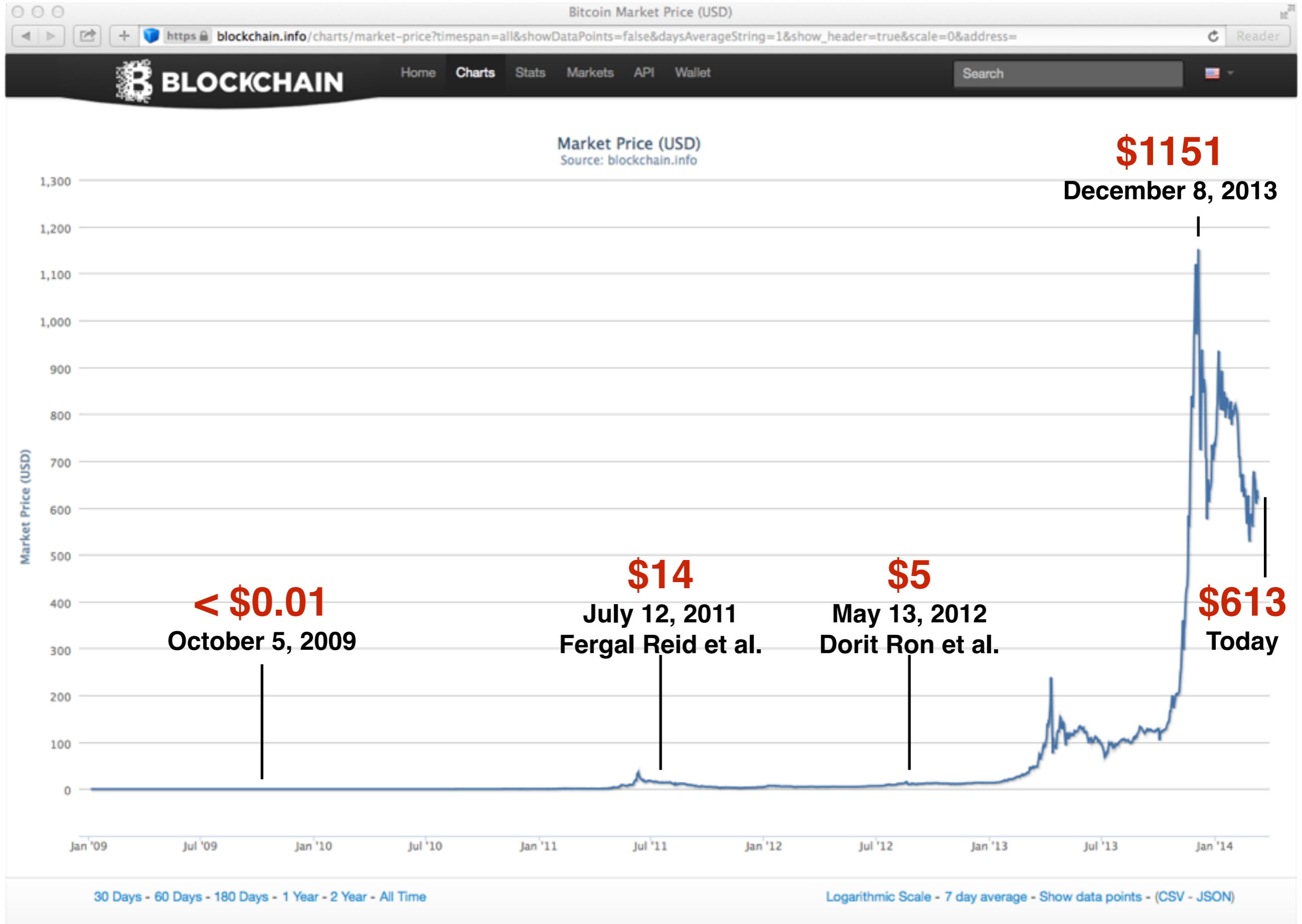
- People (Alice, ...)
- Companies (WordPress, ...)
- E-Wallets & Exchanges (CoinBase, *Mt.Gox*, ...)
- Mining Pools (Deepbit, ...)
- Bitcoin Faucets
- Bitcoin Browser (Blockchain.info, ...)

Market Price (USD)

Source: blockchain.info



30 Days - 60 Days - 180 Days - 1 Year - 2 Year - All Time Logarithmic Scale - 7 day average - Show data points - (CSV - JSON)



Agenda

1. What is “Bitcoin”?
- 2. Anonymity Analysis**
3. Quantitative Analysis
4. Discussion

Anonymity Analysis

Idea

3 features of Bitcoin are of importance:

1. Public availability of Bitcoin transactions
2. input-output relationship between transactions
3. re-use and co-use of public-keys

These 3 features provide a basis for two distinct network

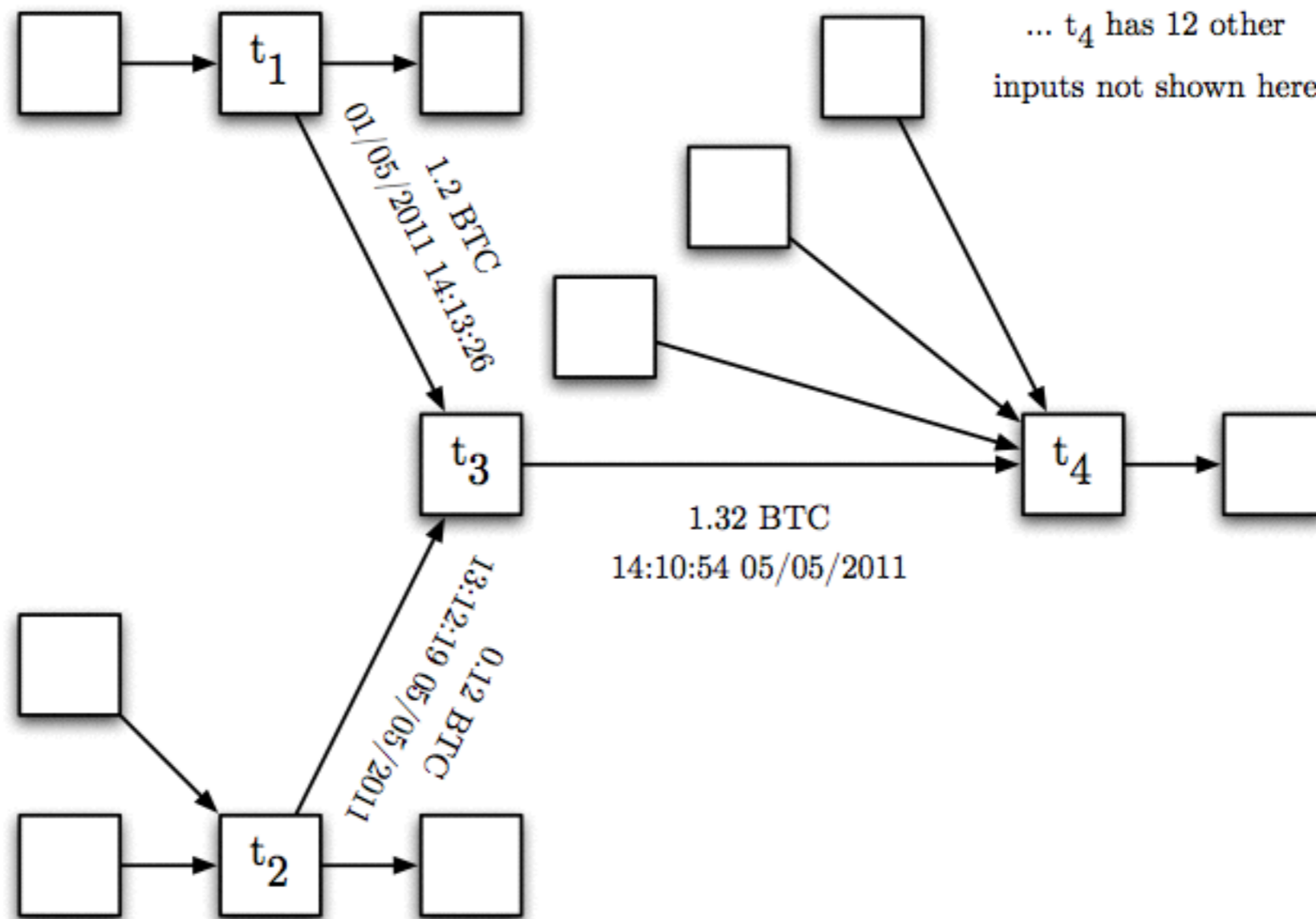
structures: a **Transaction Network** \mathcal{T} and a **User Network** \mathcal{U}

Transaction Network \mathcal{T}

\mathcal{T} represents the flow of Bitcoins between **transactions** over time

- Vertices represent transactions
- Directed edges represent input-output transaction pairs
- Directed Acyclic Graph

Transaction Network \mathcal{T}



User Network \mathcal{U}

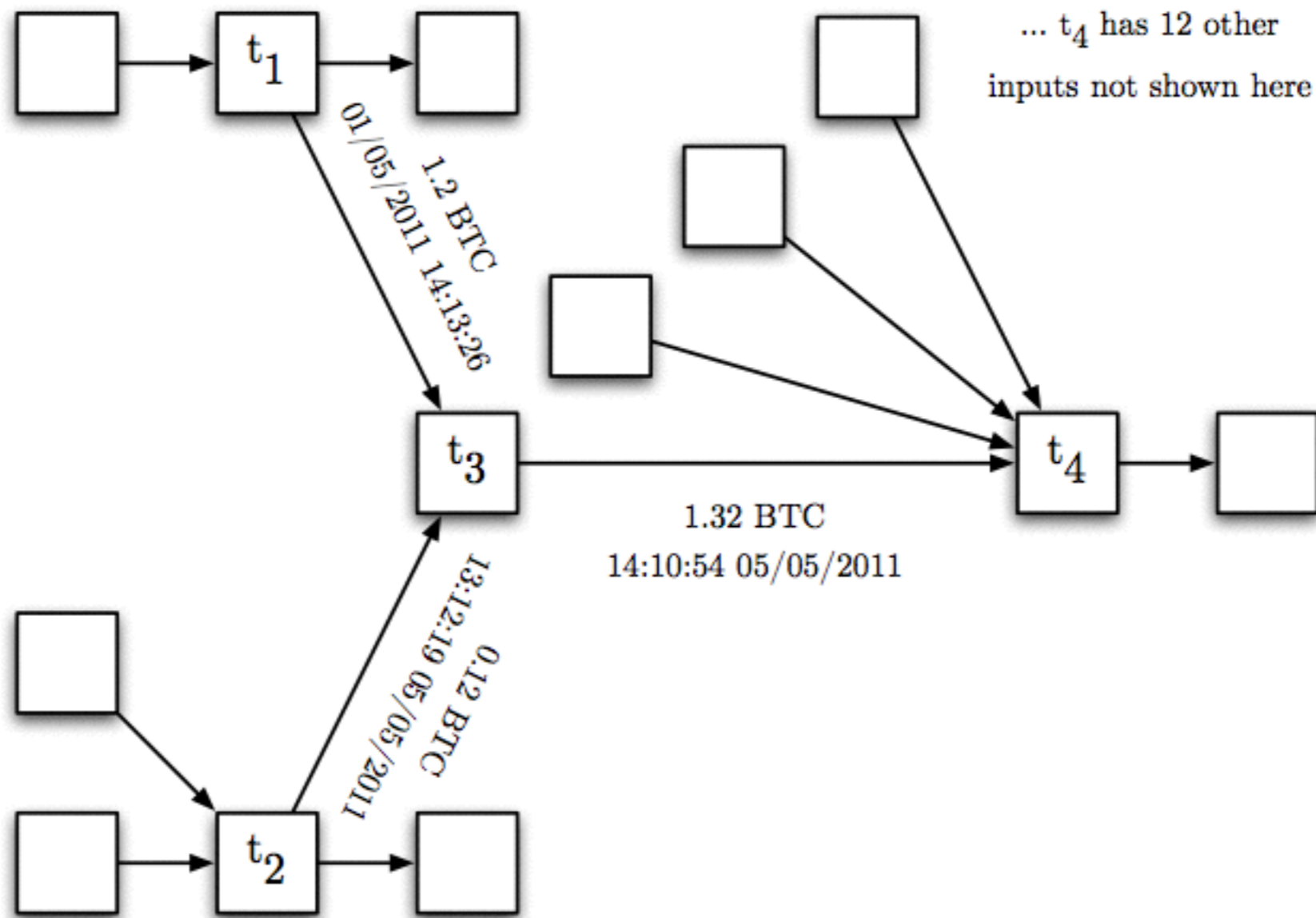
\mathcal{U} represents the flow of Bitcoins between **users** over time

- Vertices represent users
- Directed edge represent input-output pair of a transaction with corresponding user public-keys
- Unlike \mathcal{T} , \mathcal{U} has multi-edges, loops and directed cycles

“Some linking is unavoidable with multi-input transactions, which necessarily reveal that their inputs were owned by the same owner. The risk is that if the owner of a key is revealed, linking could reveal other transactions that belonged to the same owner.” *

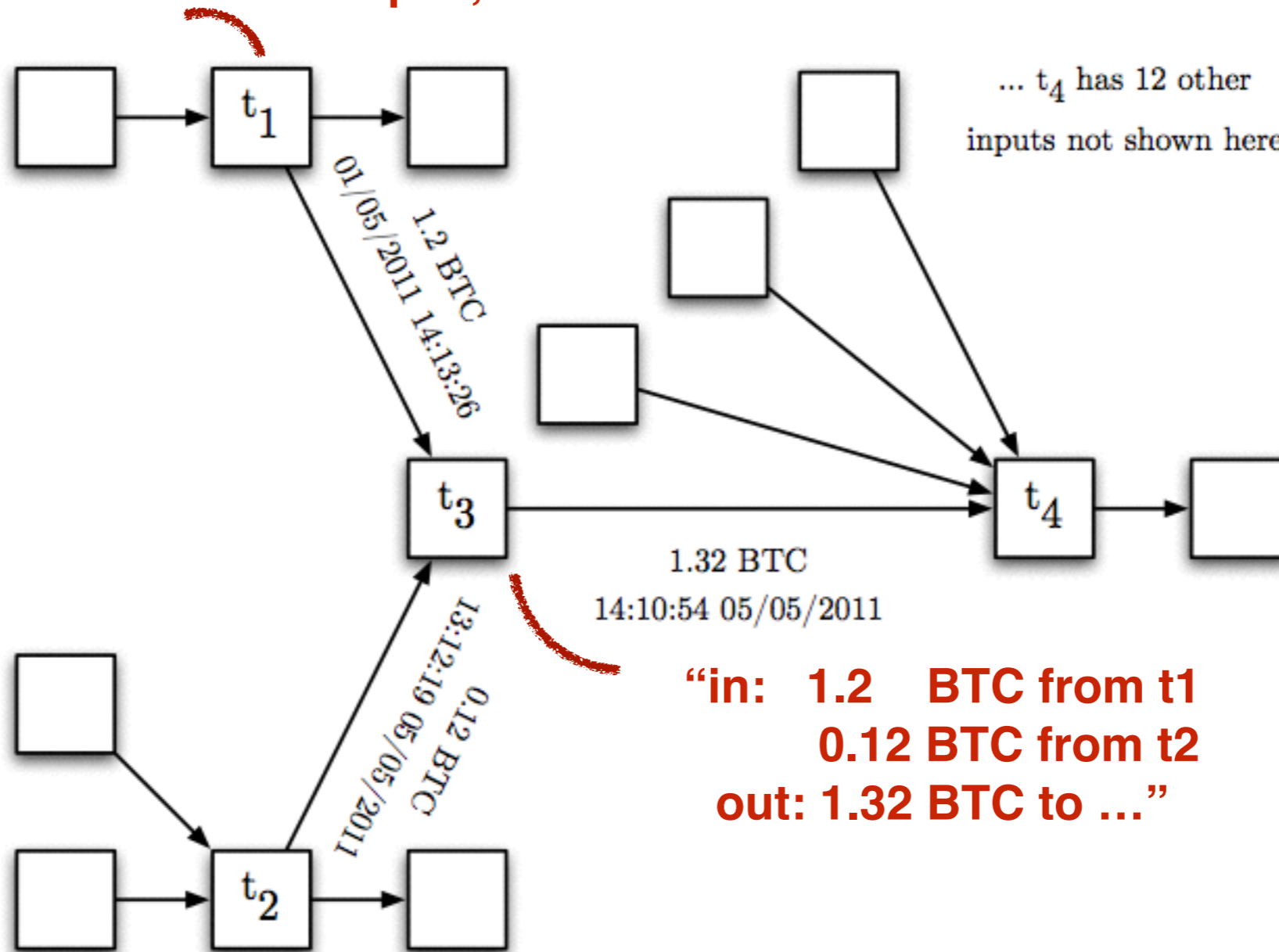
* “Bitcoin: A Peer-to-Peer Electronic Cash System”, Satoshi Nakamoto

User Network \mathcal{U}



User Network \mathcal{U}

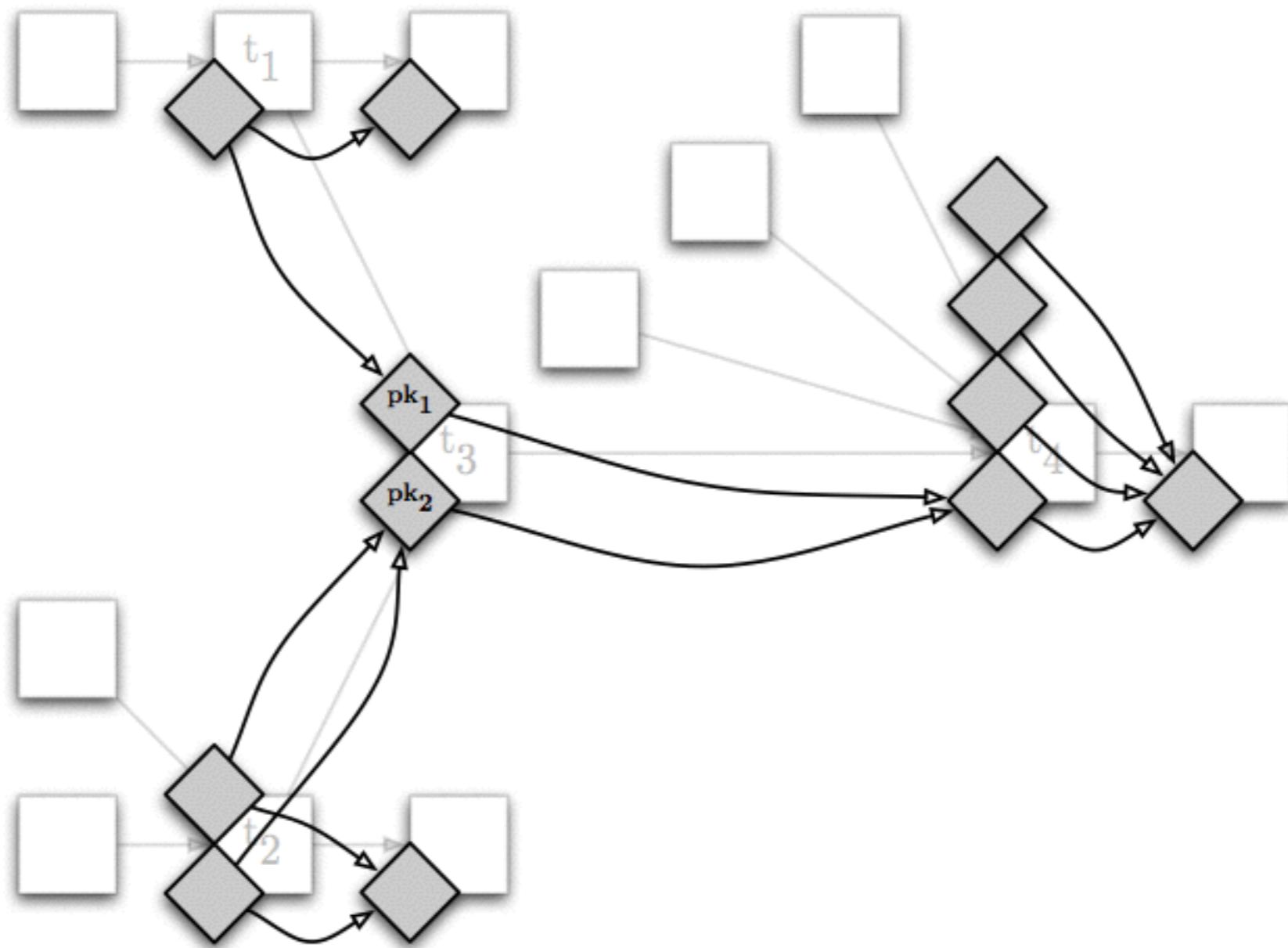
“in: ... out: 1.2 BTC to pk1, ...”



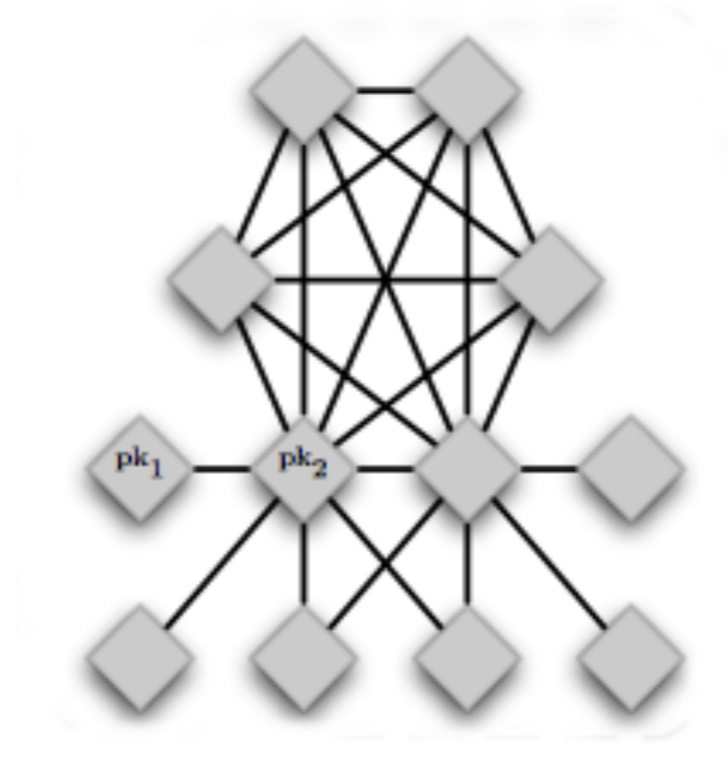
“in: 1.2 BTC from t_1
0.12 BTC from t_2
out: 1.32 BTC to ...”

“in: ... out: 0.12 BTC to pk2, ...”

User Network \mathcal{U}

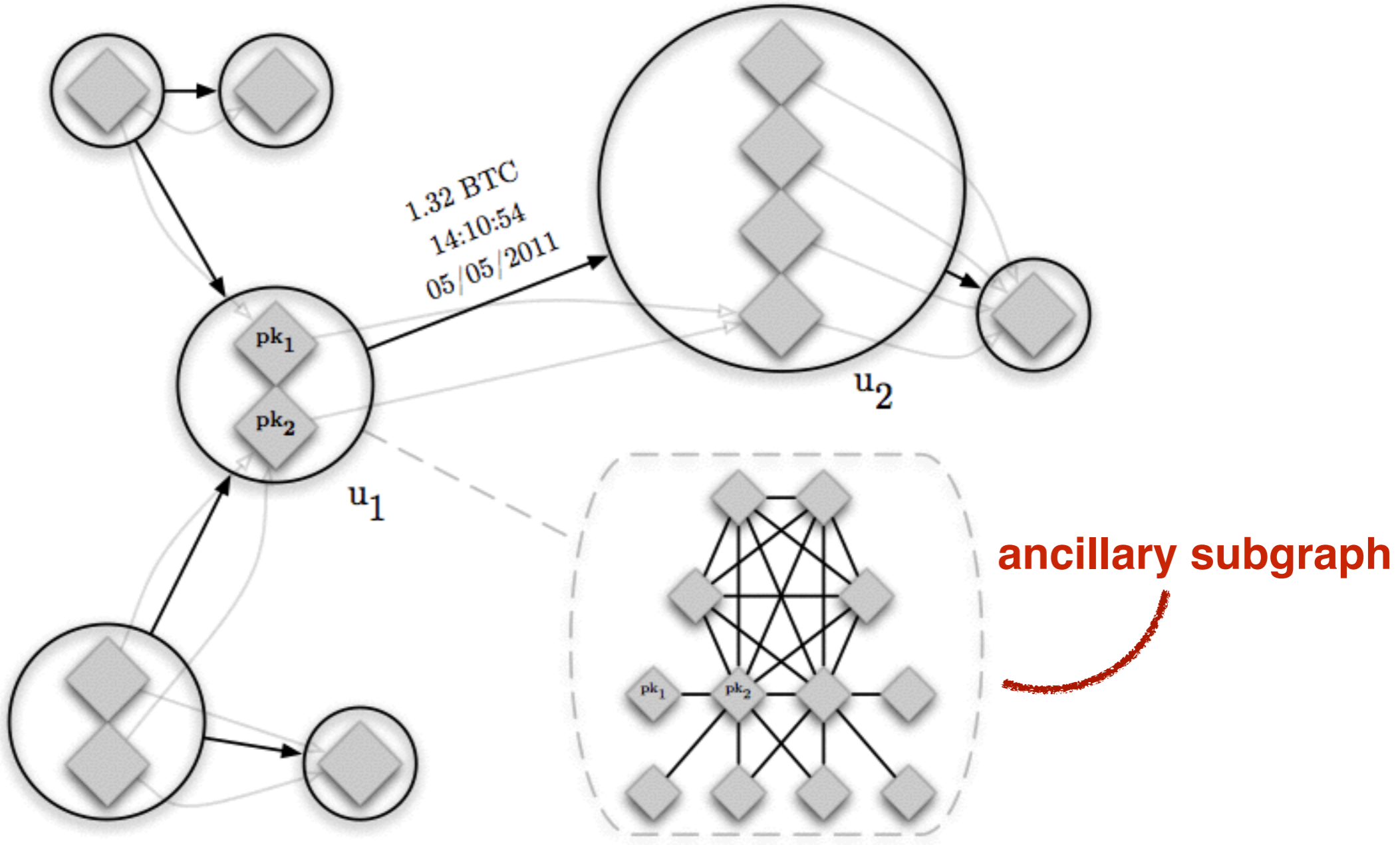


User Network \mathcal{U}



ancillary subgraph

User Network \mathcal{U}



Data & Numbers

Cut-off date: July 12, 2011

- Transaction Network size: 974'520 vertices, 1'558'854 direct edges
- User Network size: **881'678 vertices** (86'641 non-trivial mcc), 1'961'636 edges

Deducing Information

How can the user network \mathcal{U} be used to deduce information about Bitcoin users?

- A. Integrating Off-Network Information
- B. Egocentric Analysis and Visualization of \mathcal{U}
- C. Context Discovery
- D. Flow and Temporal Analysis




4 Methods

A. Integrating Off-Network Information

Organizations that accept Bitcoins have access to **identifying information** (email, shipping addresses, CC, bank account details, IP addresses, ...)

Voluntary disclosure of public-keys by users (Bitcoin forums, Twitter, ...), often also indexed by search engines

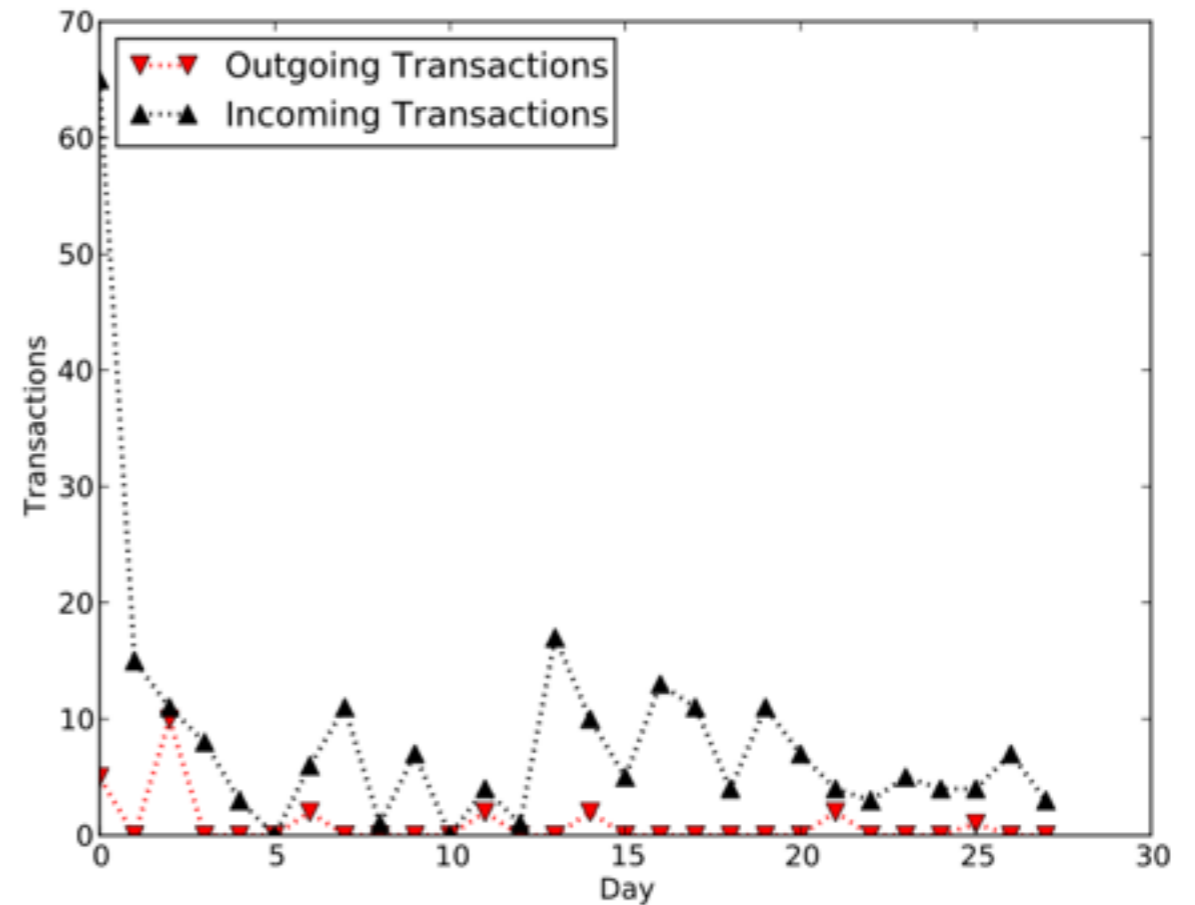
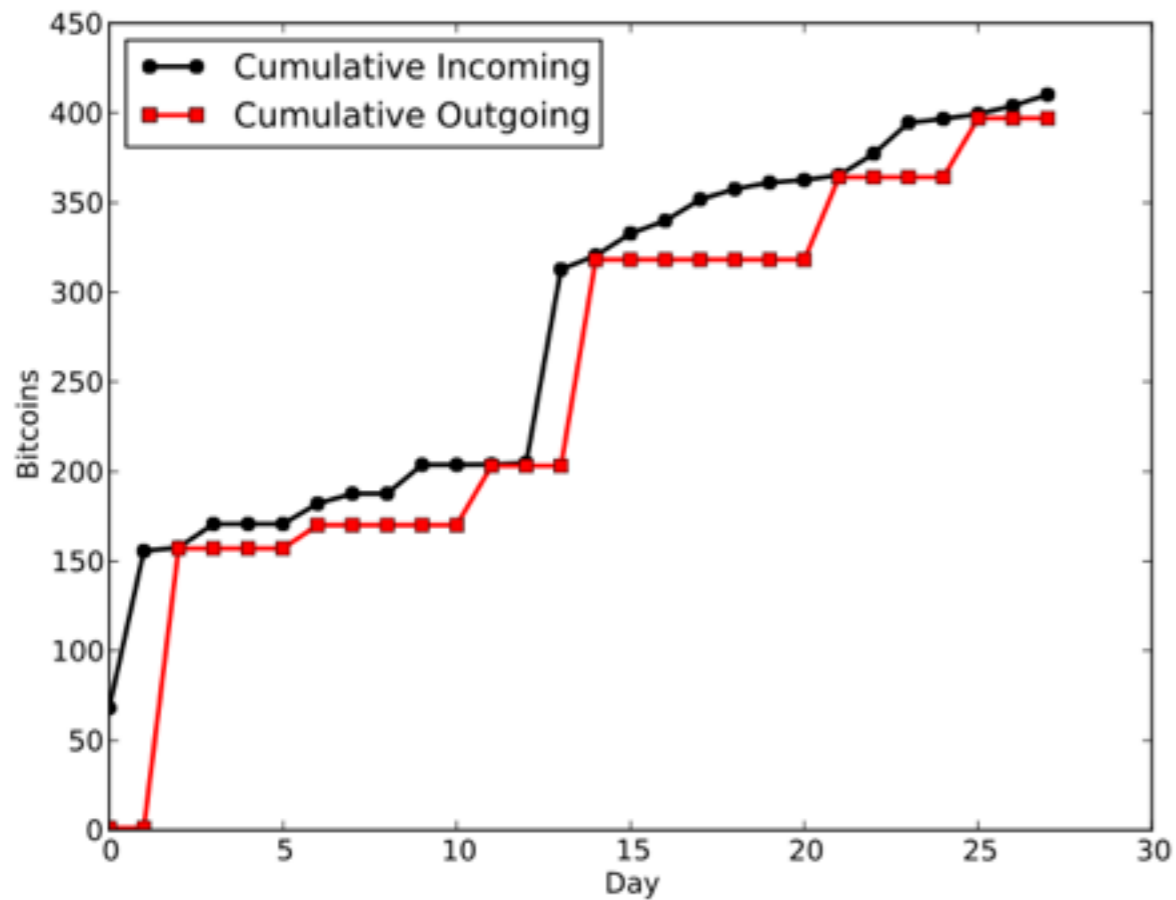
A. Integrating Off-Network Information

ranimes Full Member 	 Re: Seminar in Distributed Computing March 19, 2014, 03:33:30 PM	#2
Activity: 1337	Hope you're enjoying the presentation.	
		
Ignore	Will sing for bitcoins: 1A1337b9e3b1a0bca72e9bbe7cc705baca5	

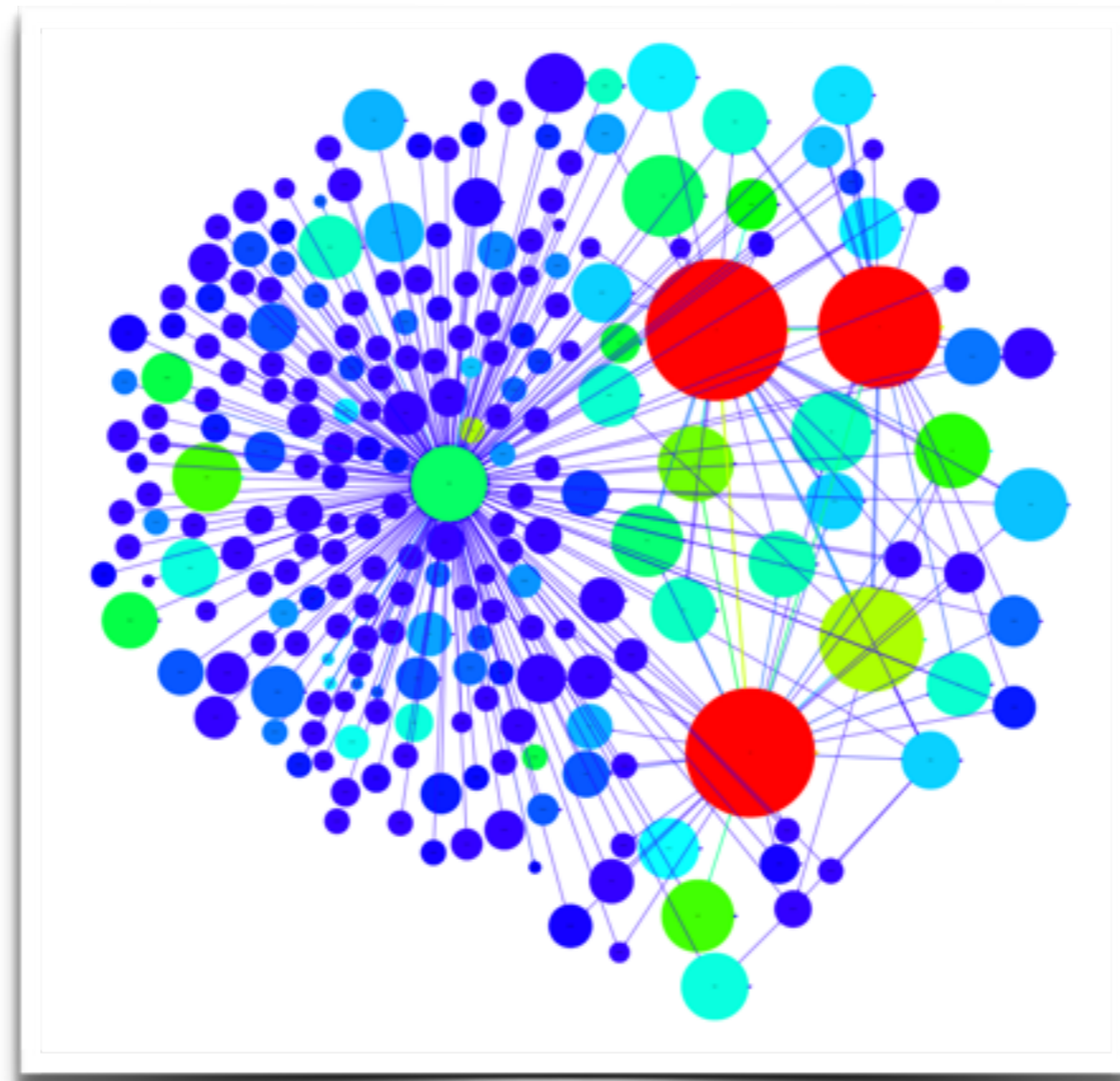
B. Egocentric Analysis and visualization of the user network

Examples: **balance** held by a single public-key, **aggregation of balances** belonging to public-keys that are controlled by a particular user

B. Egocentric Analysis and visualization of the user network



B. Egocentric Analysis and visualization of the user network



Egocentric visualization of the vertex in the imperfect user network.
Color = Volume, Size = Degree

C. Context Discovery

Examples: **Shortest paths** between a set of vertices, **maximum number of Bitcoins that can flow** from a source to a destination given the transactions and their 'capacities' in an interesting time-window

D. Flow and Temporal Analyses

Example: user **receives a large volume of Bitcoins** relative to their estimated balance, and, **shortly after, transfers a significant proportion** of those Bitcoins to another user: interesting?

Conclusions

Using an appropriate network representation, it is possible to associate many public-keys with each other, and with external identifying information.

Activity of known users can be observed in detail.

Large centralized services are capable of identifying and tracking considerable portions of user activity.

Users have to be aware that **strong anonymity is not a prominent design goal** of the Bitcoin system.

Agenda

1. What is “Bitcoin”?
2. Anonymity Analysis
- 3. Quantitative Analysis**
4. Discussion

Quantitative Analysis

Idea

The types of statistics already available tend to describe global properties of the network over time.

With help of graph \mathcal{U} , one can go much further than that and answer questions such as:

Idea

- What is the typical behavior of account owners?
- How do they acquire and spend their Bitcoins?
- What is the balance of Bitcoins kept in accounts?
- How do they move Bitcoins between their various accounts?
- What are the structures of the largest transactions?

Data & Numbers

Cut-off date: May 12, 2012

- Input: 7'134'836 single sender and single receiver transactions
- Result size: **2'460'814**
(1'851'544 different owners, 609'270 public keys only received BTC)

5 Discoveries

Discovery #1

- The 609'270 receiver-only addresses *put aside* 7'019'100 BTC (~78% of the existing Bitcoins)
- The total number of Bitcoins participating in all the transactions since the establishment of the currency is more than 420'000'000

Implication: Each coin which is in circulation had to be moved a much larger than expected number of times!

Discovery #2

Distribution of the accumulated incoming Bitcoins per entity and per address

Larger or equal to	Smaller than	Number of entities	Number of addresses
0	1	893,763	1,497,451
1	10	389,302	698,132
10	100	881,273	1,206,209
100	1,000	255,826	285,820
1,000	10,000	36,713	38,484
10,000	50,000	3,593	3,723
50,000	100,000	181	190
100,000	200,000	55	50
200,000	400,000	30	29
400,000	800,000	76	129
800,000		4	1

Discovery #2

Distribution of the accumulated incoming Bitcoins per entity and per address

Larger or equal to	Smaller than	Number of entities	
0	1	893,763	36%
1	10	389,302	52%
10	100	881,273	88%
100	1,000	255,826	
1,000	10,000	36,713	
10,000	50,000	3,593	
50,000	100,000	181	
100,000	200,000	55	
200,000	400,000	30	
400,000	800,000	76	
800,000		4	

Discovery #2

The total number of Bitcoins received by most owners is negligible.

Discovery #3

Distribution of the current balance of Bitcoins per entity and per address

Larger or equal to	Smaller than	Number of entities	Number of addresses
0	0.01	2,097,245	3,399,539
0.01	0.1	192,931	152,890
0.1	10	95,396	101,186
10	100	67,579	68,907
100	1,000	6,746	6,778
1,000	10,000	841	848
10,000	50,000	71	65
50,000	100,000	5	3
100,000	200,000	1	1
200,000	400,000	1	1
400,000		0	0

Discovery #3

The current balance of almost 97% of owners as of May 13, 2012 was less than 10 Bitcoins.

There are only 78 owners with current balance larger than 10'000 Bitcoins.

Discovery #4

Distribution of the number of transactions per entity and per address

Larger or equal to	Smaller than	Number of entities	Number of addresses
1	2	557,783	495,773
2	4	1,615,899	2,197,836
4	10	222,433	780,433
10	100	55,875	228,275
100	1,000	8,464	26,789
1,000	5,000	287	1,032
5,000	10,000	35	51
10,000	100,000	32	24
100,000	500,000	7	3
500,000		1	2

Discovery #4

97% of all owners had fewer than 10 transactions each, while 75 owners use the network very often and are affiliated with at least 5000 transactions.

Discovery #5

Distribution of the size of the transactions in the Bitcoin scheme

Larger or equal to	Smaller than	Number of transactions in the graph of entities	Number of transactions in the graph of addresses
0	0.001	381,846	2,315,582
0.001	0.1	1,647,087	4,127,192
0.1	1	1,553,766	2,930,867
1	10	1,628,485	2,230,077
10	50	1,071,199	1,219,401
50	100	490,392	574,003
100	500	283,152	262,251
500	5,000	70,427	67,338
5,000	20,000	6,309	6,000
20,000	50,000	1,809	1,796
50,000		364	340

Discovery #5

73% of the transactions involve fewer than 10 BTC.

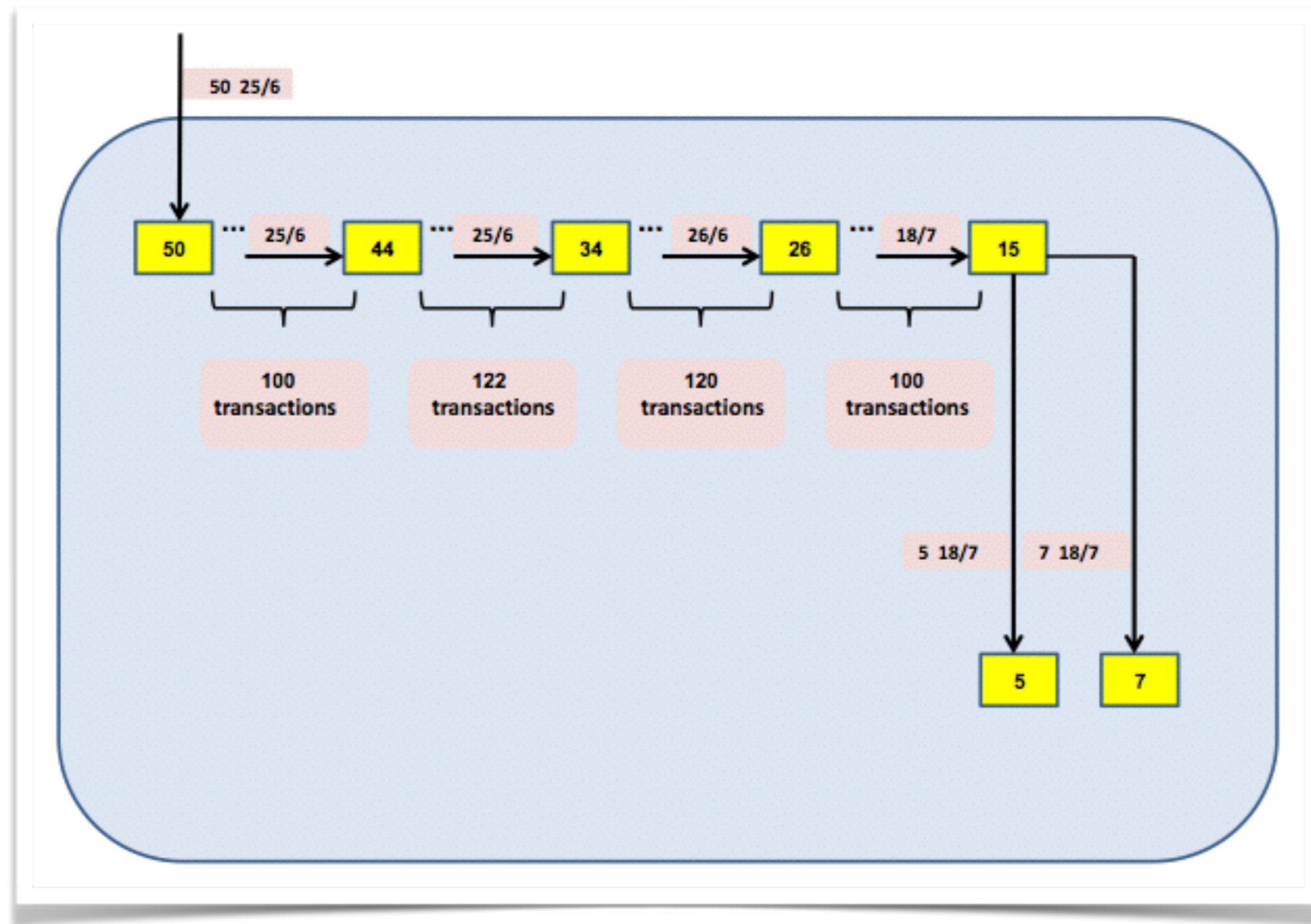
Large transactions are rare at Bitcoin; there are only 364 transactions larger than 50'000 BTC.

Structures of the largest transactions

Tracing the 364 largest transactions reveals that 348 were actual successors of the earliest of those transactions of 90'000 BTC made on November 8, 2010

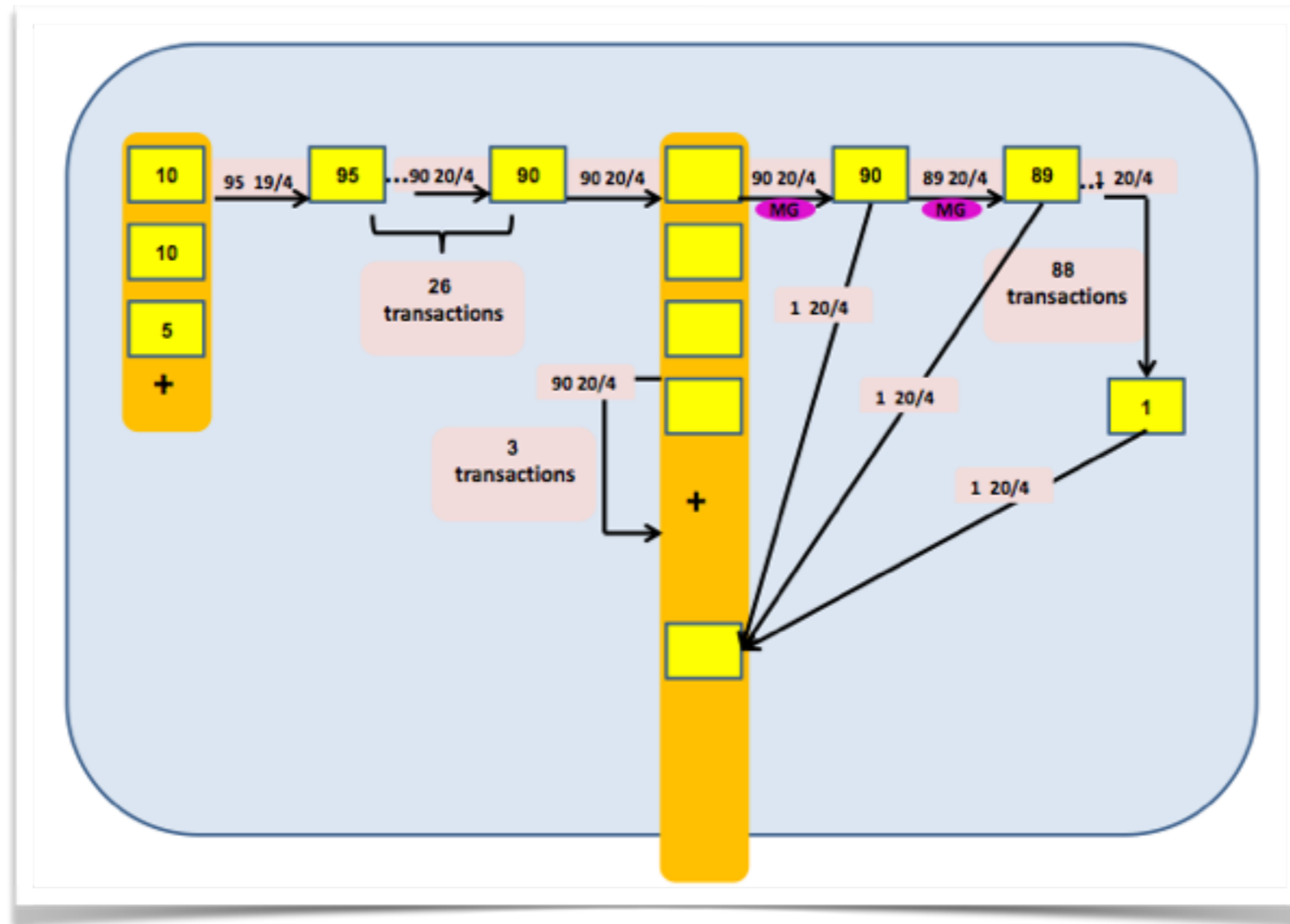
4 Structures

1. Long Chains



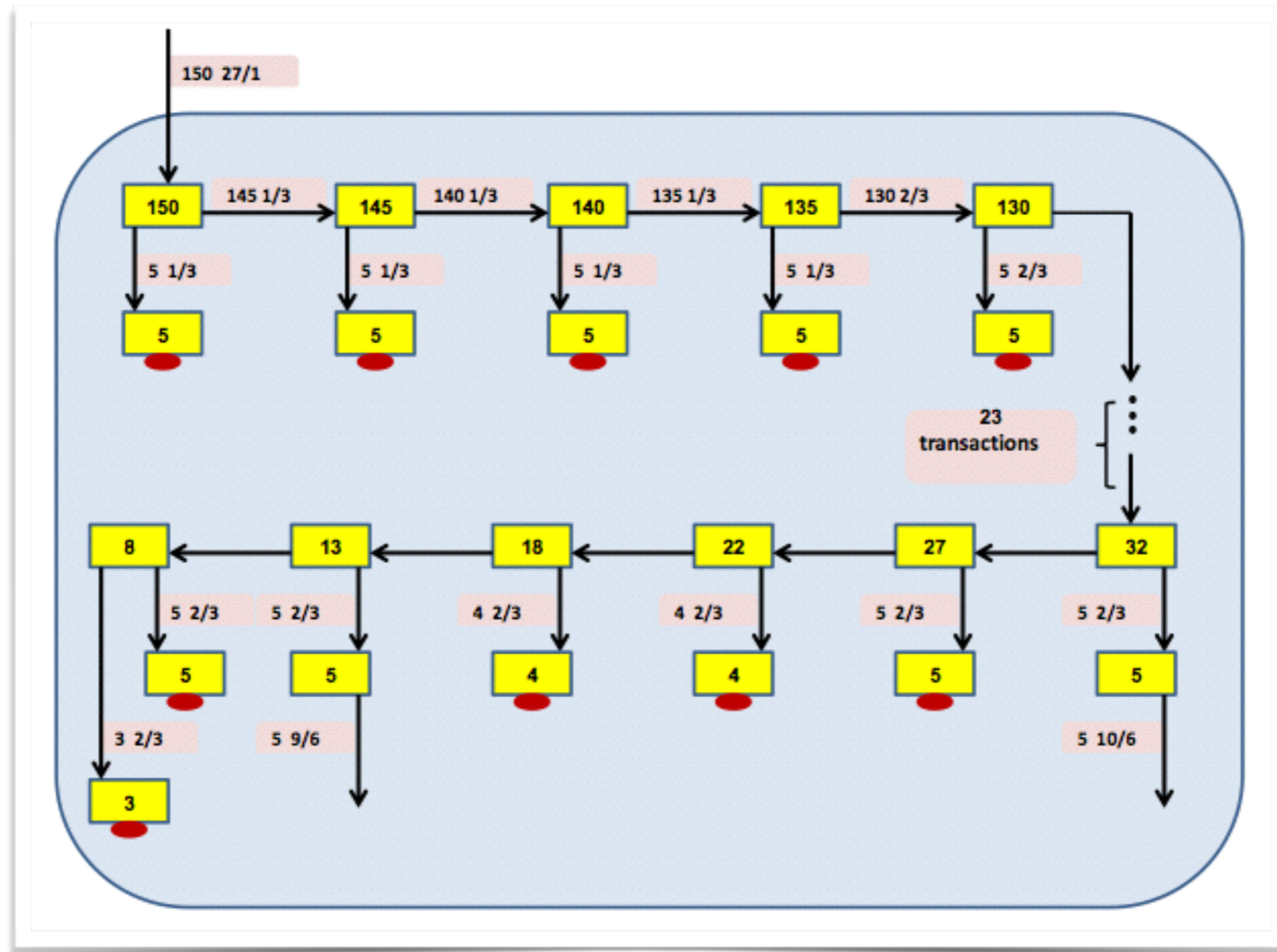
Large amounts of Bitcoins are rapidly transferred in a very long chain of hundreds of transactions in a very short period of time.

2. Fork-Merge Patterns and Self Loops



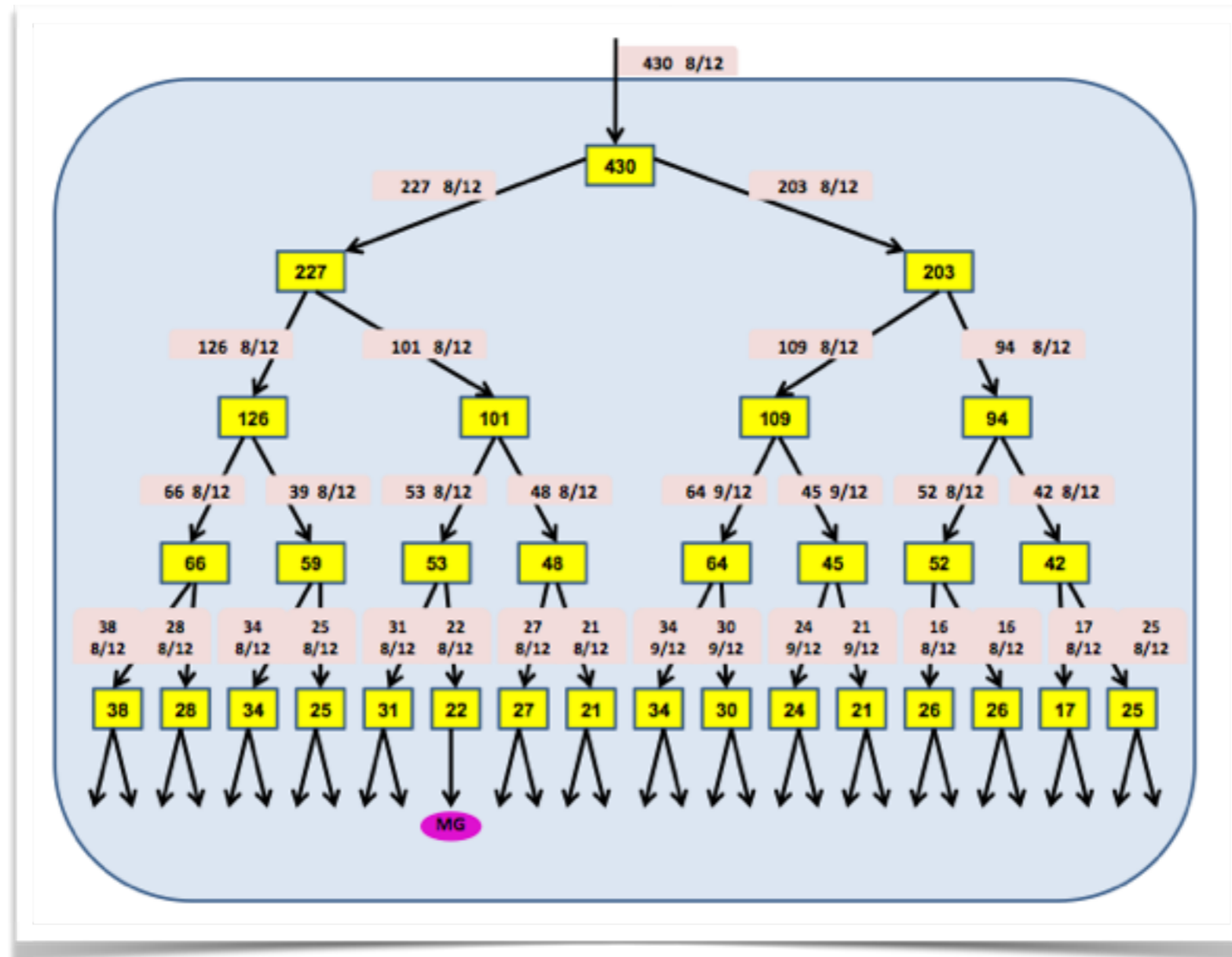
An owner sends 90'000 BTC to himself in a self loop, then transfers it forward but gets it back via 90 transfers of 1'000 BTC each.

3. Keeping Bitcoins in “saving accounts”



This chain puts aside 5'000 BTC at 28 of 30 steps. The accumulated sum of 140'000 BTC has never been spent since.

4. Binary Tree-like Distributions



A large amount of Bitcoins is distributed among many addresses via a binary tree-like structure.

Conclusions

Most of the minted Bitcoins remain dormant.

Very large number of tiny transactions

The number of large transactions are in the hundreds.

Almost all those large transactions were descendants of one single transaction.

Subgraph containing those large transactions contains peculiar structures which could be an attempt to conceal the existence and relationship between those transactions.

Such an attempt can be foiled by following the money trail in a sufficiently persistent way.

Agenda

1. What is “Bitcoin”?
2. Anonymity Analysis
3. Quantitative Analysis
- 4. Discussion**

Discussion