Adam Mickiewicz University Poznań

Poznań University of Economics and Business

Jacek Mizerka, Bartosz Biadasiewicz, Piotr Mizerka

Bitcoin properties – analysis of users graph

(still work in progress)

Presentation for INFINITI 2018 Asia-Pacific

Sydney 10-11 December 2018

The goals we want to achieve

General goal: Indicate the possibilities that are open to researchers thanks to access to blockchain database. With regard to bitcoins, the access allows the construction of a transaction hypergraph and users graph thanks to use of the original software based on heuristics.

Specific objectives

1. Identification of basic features of users graph (frequency of cluster sizes, number of edges, number of nodes versus turnover)

2. Investigating the wealth of users and distribution of users' richness (balances of users, Gini coefficient)

3. Checking the robustness of the users graph (clustering coefficient)

4. Examine the relationship between bitcoin price and user graph properties (number of buyers and sellers versus price of bitcoin, results of Principal Component Analysis)

Stages of the research

1. Data acquisition from the Bitcoins blockchain
2. Construction of the users graph on the basis of transactions hypergraph
3. Characteristics of the users graph properties
4. Inference about the features of the „cryptocurrency world" based on the characteristics of the users graph
5. Proposals for future research

Stage 1. Data acquisition

A. We have collected data from the period of 19.01.2009 – 09.02.2018

B. Data was obtained from the website of Hungarian researchers D. Kondor, M. Posfai, J. Szüle, I Csabai and G. Vattay

C. Other assumptions also have been made:
   - We limit ourselves to the users represented by at least 10 IP addresses.
   - We neglect transactions fees and taxes.

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 2 Construction of users' graph

Transactions Hypergraph:

The multigraph of transactions is the following triple: (A,T,f), where:

- A is the set of all addresses that are given in the blockchain.
- T is the set of all elementary transactions between particular addresses.
- $f : T \rightarrow A_{in} \times A_{out} \times \mathbb{R}_+ \times \mathbb{R}_+$

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 2 Construction of users' graph, cont.

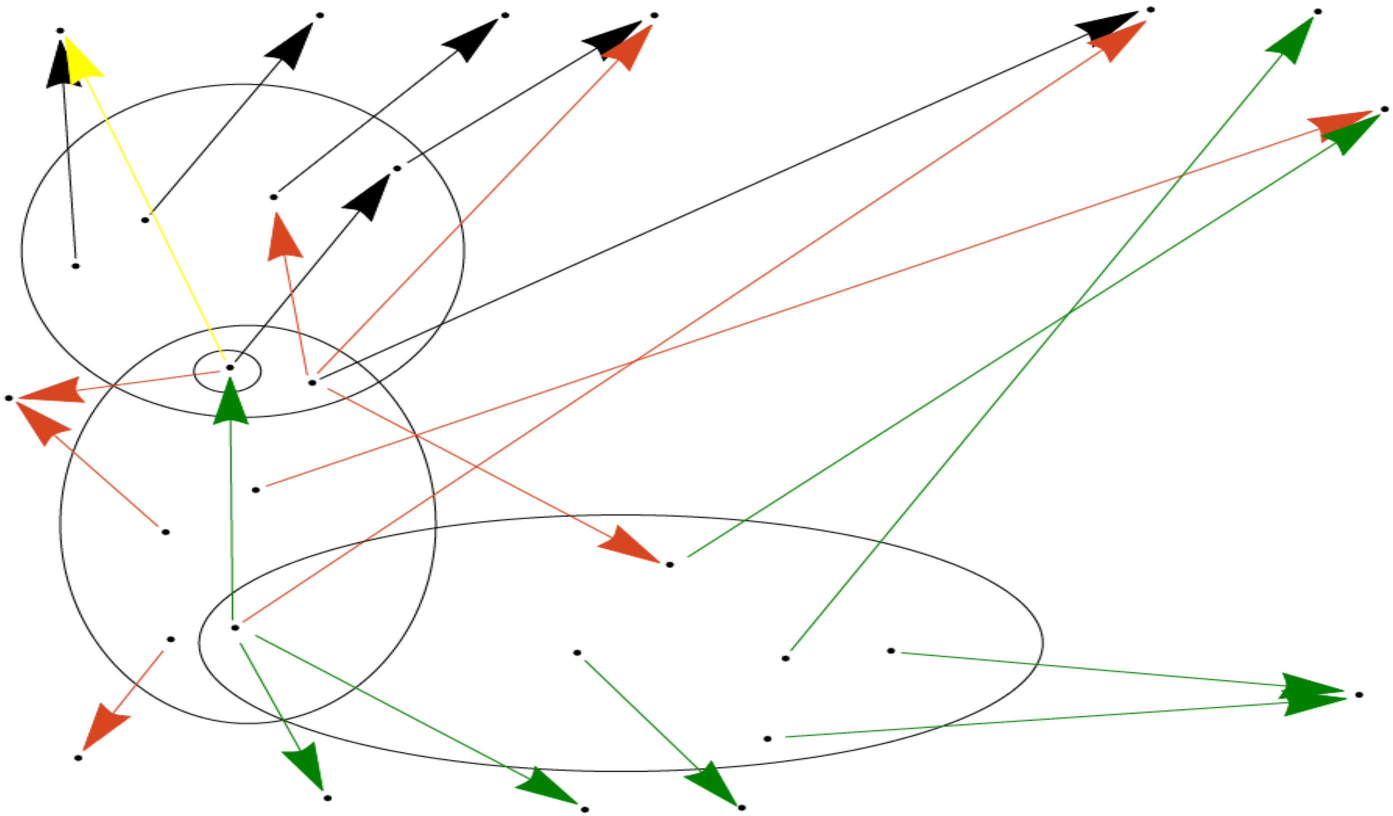
Transactions Hypergraph, cont.

For a transaction t such that $f(t) = (a_{in}(t), a_{out}(t), v(t), d(t))$ this function can be interpreted in the following way:

- $a_{in}(t)$ is the address of a person who sends bitcoins,
- $a_{out}(t)$ is the address of the person who receives bitcoins,
- $v(t)$ is the amount of money transferred,
- $d(t)$ is the exact time, it is measured since the first bitcoin transaction occured

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 2. Construction of users' graph



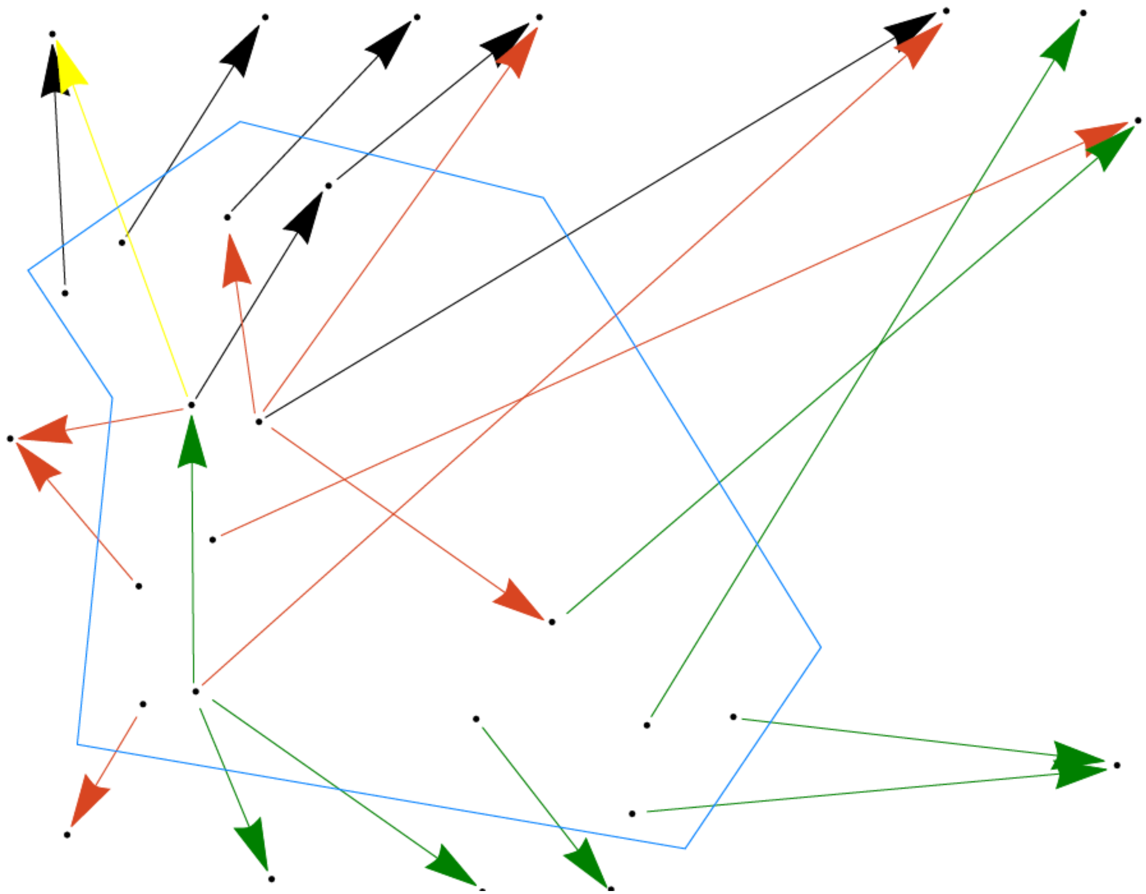
Arrows denote transactions; the same colour of arrows means the same transaction.

Dots (·) denote IP addresses

Poznań University
of Economics
and Business

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 2. Construction of users' graph, cont



The area outlined by the blue line includes all IP addresses which belong to one user

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Characteristics of the user's graph properties

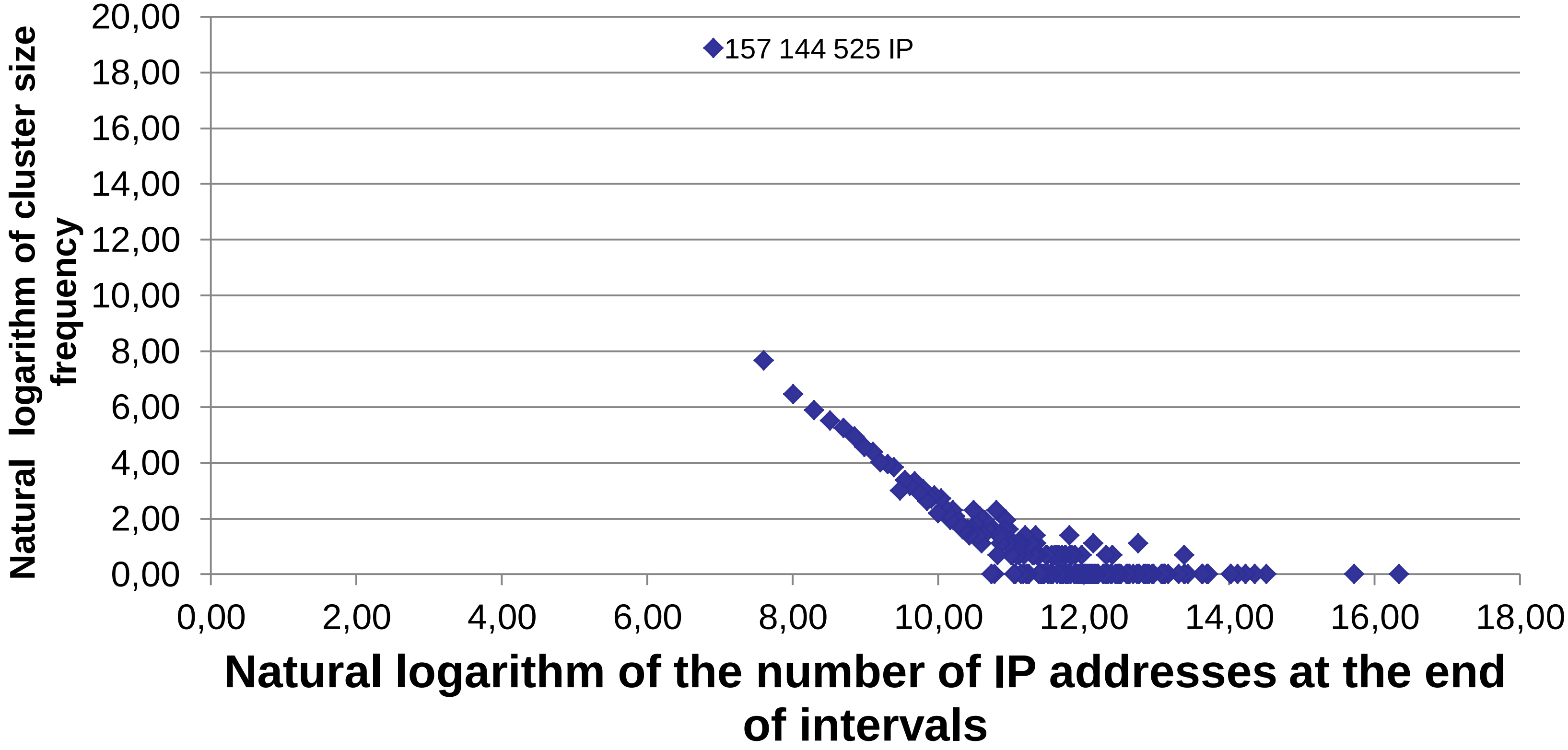
Number of Addresses, i.e. A – *340 000 0000 - 350 000 000*
Number of Transactions, i.e. T – *298 000 000*
Number of Clusters, i.e. nodes of G – *ca 160 000 000*
Number of Arcs of G – *850 000 000*

The 10 biggest clusters	**Size in no. of IP addresses**
1	12 465 626
2	6 748 815
3	2 016 103
4	1 718 254
5	1 513 284
6	1 356 687
7	1 234 814
8	1 230 950
9	902 665
10	896 231

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3. Objective 1 Characteristics of the user's graph properties, cont.

Cluster size frequency



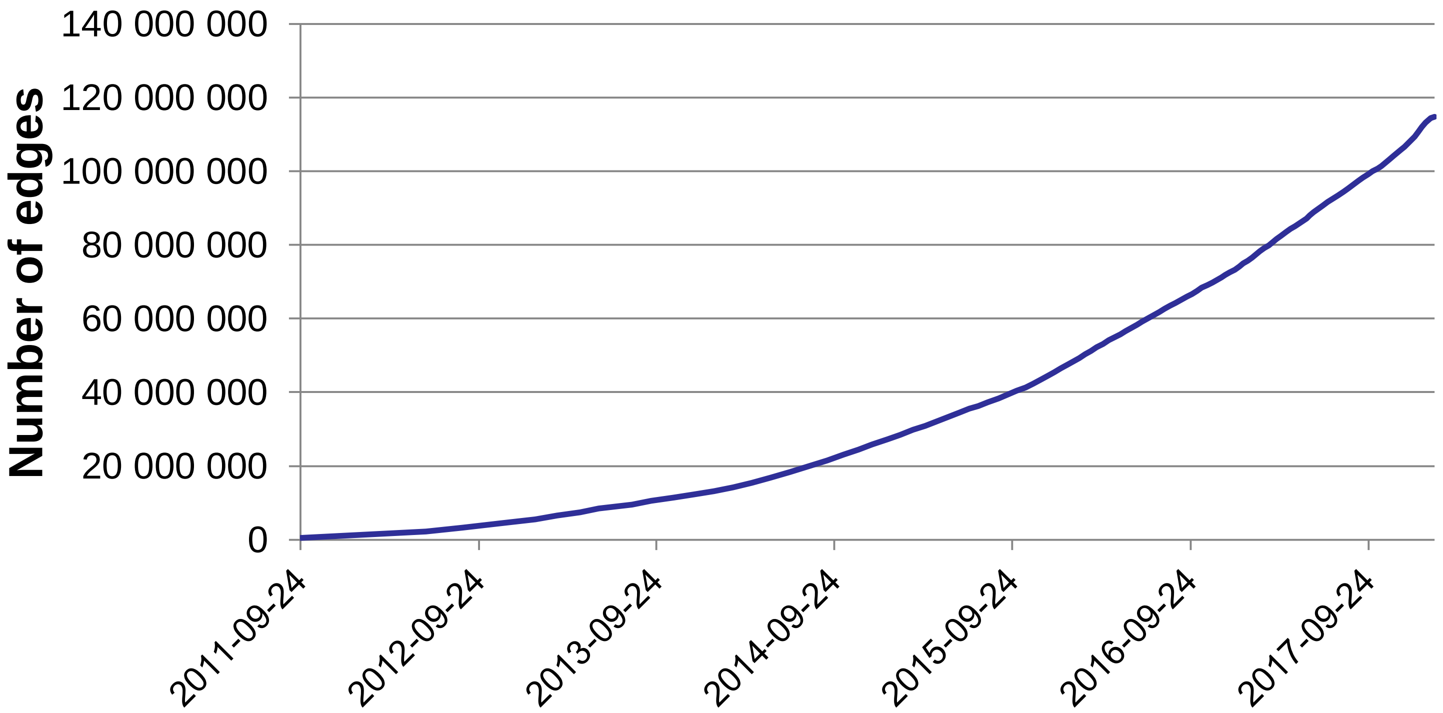
Comment 1

Among clusters containing at least 10 IP addresses, almost 100% clusters (99.997%) are clusters containing at most 1000 IP addresses

However, several clusters containing more than 1 million IP addresses have been found; the largest cluster contains about 12.47 million addresses

Poznań University of Economics and Business
Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3. Objective 1. Characteristics of the user's graph properties, cont.
Evolution of the number of edges in the period: 24.09.2011- 6.02.2018



Comment 2

During the study period, a sharp increase in the number of edges connecting nodes containing at least 10 IP addresses was observed

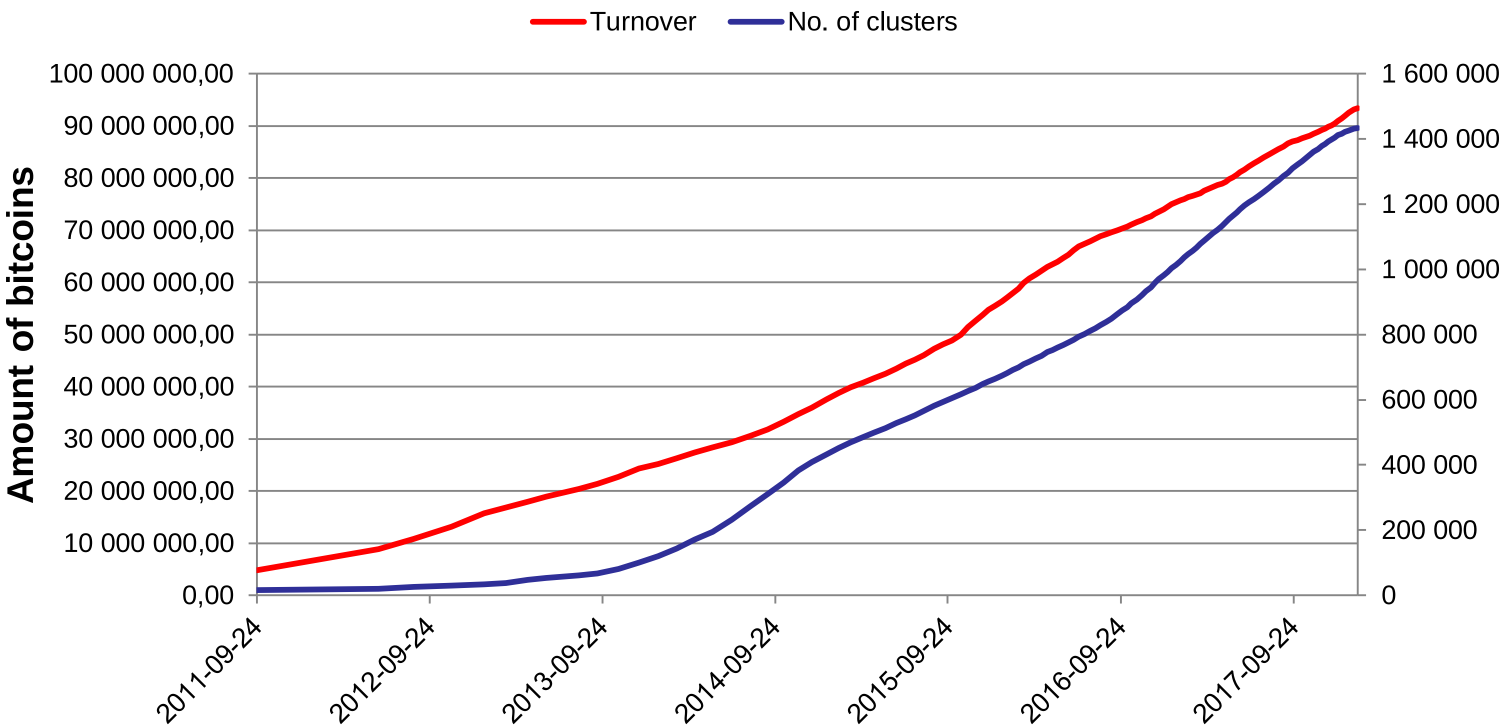
Conjecture 1

The distribution of number of edges follows a power law.

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 1 Characteristics of the user's graph properties, cont.

Evolution of the number of nodes and bitcoins turnover in the period: 24.09.2011-6.02.2018



Comment 3

In the period under consideration, an increase in the number of clusters containing at least 10 IP addresses as well as bitcoins turnover were observed. The curve showing the increase in turnover is similar to the curve reflecting the increase in the number of users

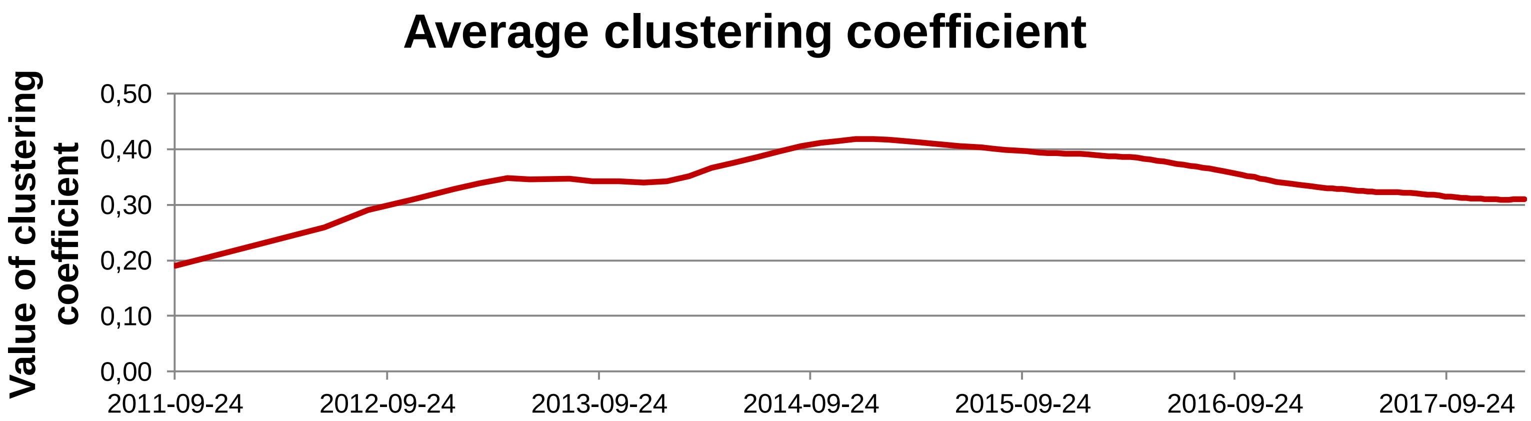
POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 3 Average clustering coefficient

Clustering coefficient – the coefficient defined for each node; it is a ratio of number of edges between neihgbours of that node to the number of all possible edges between these neighbours;

Clustering coefficient is a measure of graph's robustness



Comment

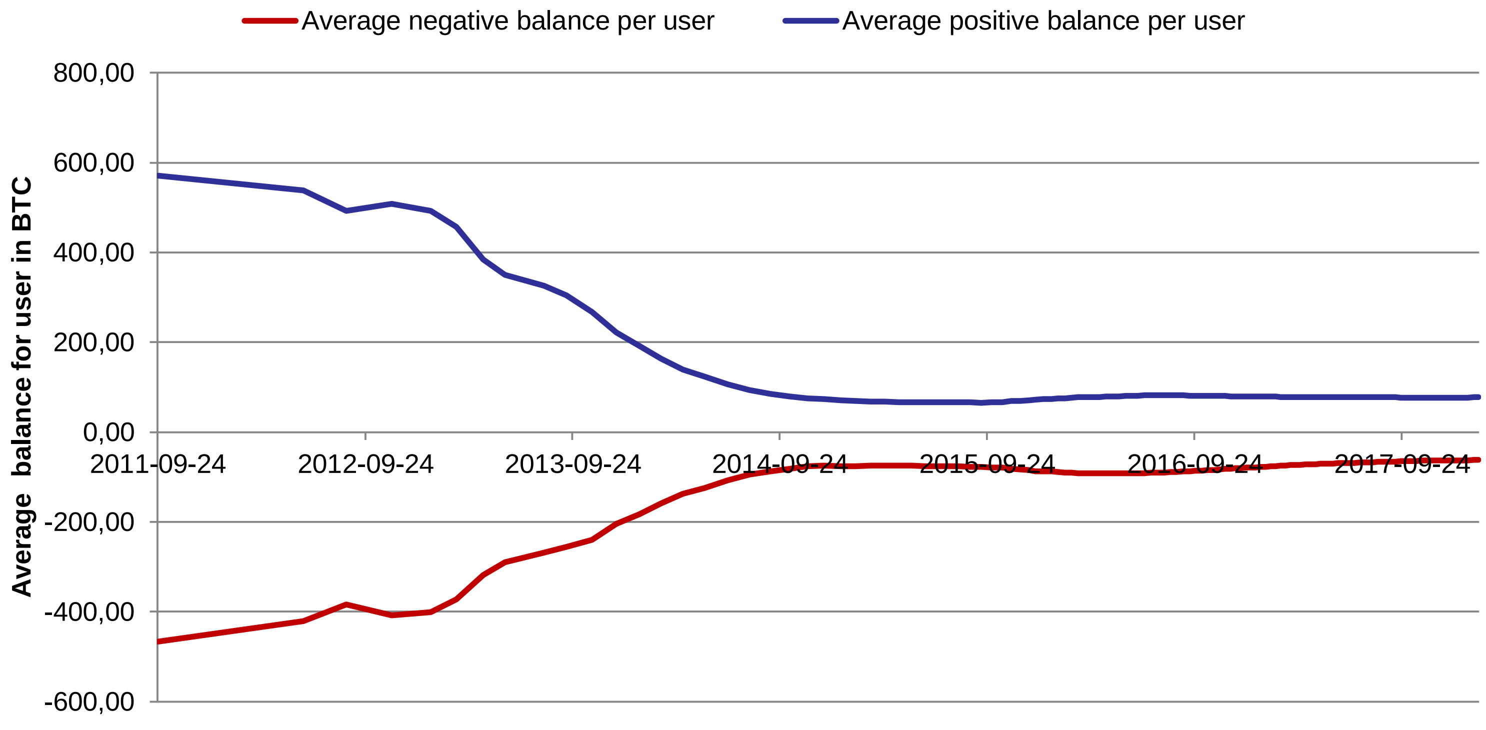
From the beginning of 2017, the value of average clustering coefficient stabilizes at around 0.32

Conjecture (based on changes of edges and vertices number in time)

The increasing number of edges is concentrated among the minority of the most active users.

POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS
Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 2 Characteristics of the user's graph properties, cont. Average balance (negative or positive) per one user in the period: 24.09.2011- 6.02.2018



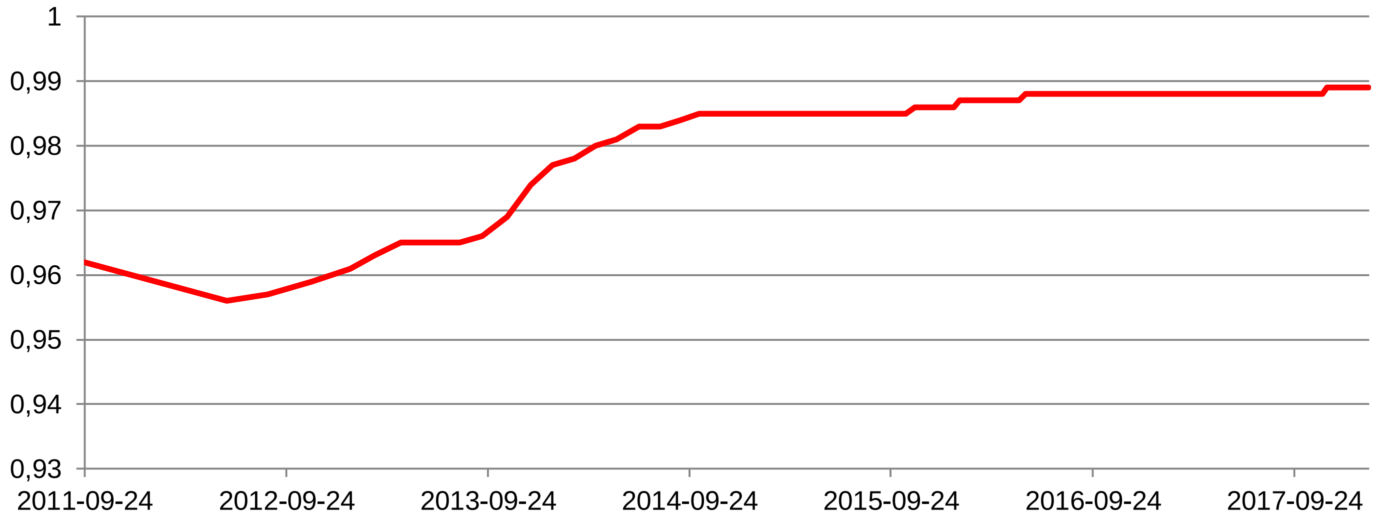
Comment 4

The tendency to stabilize both the positive and negative balance of transactions per one user is observed

Poznań University of Economics and Business

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 2 Characteristics of the user's graph properties, cont.
Value of Gini coefficient in the period: 24.09.2011- 6.02.2018



Gini coefficient - $G = \frac{2\sum_{i=1}^{n} i x_i}{n \sum_{i}^{n} x_i} - \frac{n+1}{n}$

where:
n denotes total number of nodes
x_i is the richness (measured for users with positive balance of transactions) of *i*th node (x_i is indexed in non-decreasing order); *i*= 1…n

Comment 5

After initial some fluctuations, the value of Gini coefficient stabilized at high level

Conjecture 2

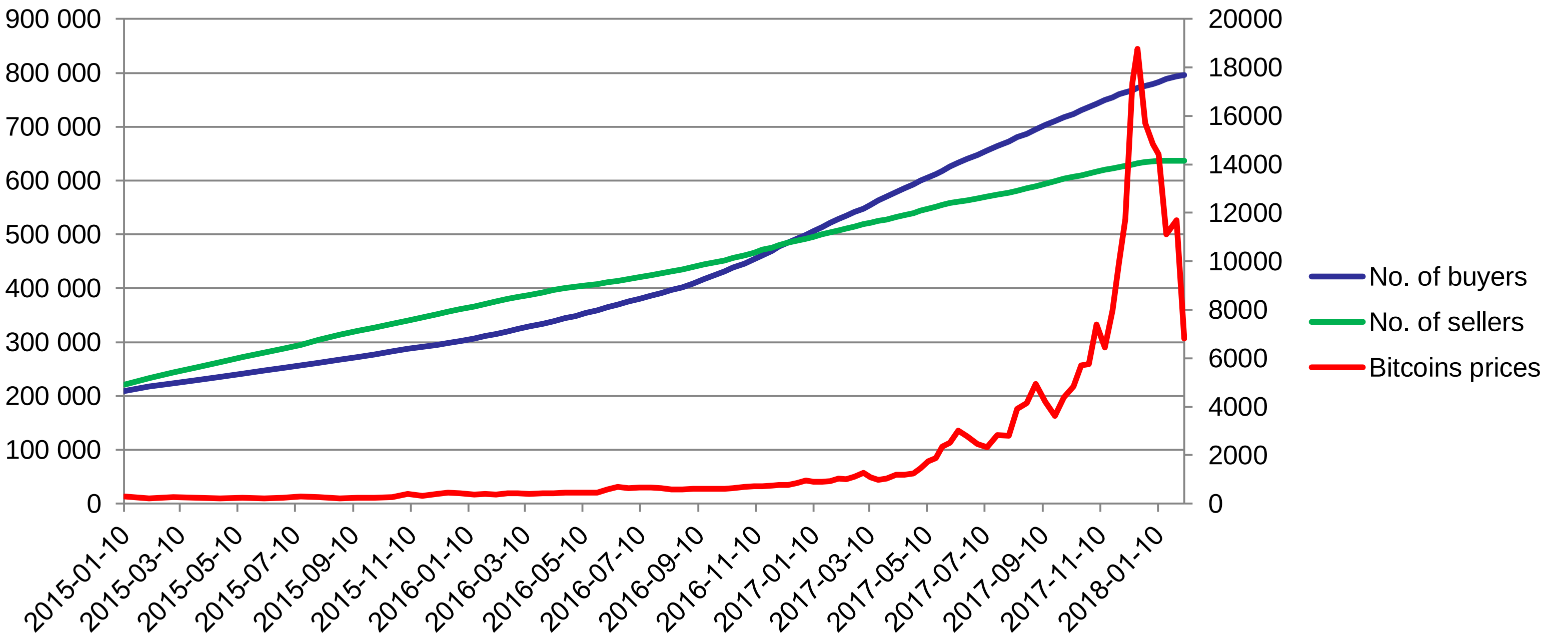
The bitcoin system is characterized by high stratification of users wealth; the value of Gini coefficient stabilized at a level close to 1.

POZNAN UNIVERSITY
OF ECONOMICS
AND BUSINESS

Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 4 Characteristics of the user's graph properties, cont.

Number of buyers and sellers versus price of bitcoin in the period: 10.01.2015 - 6.02.2018

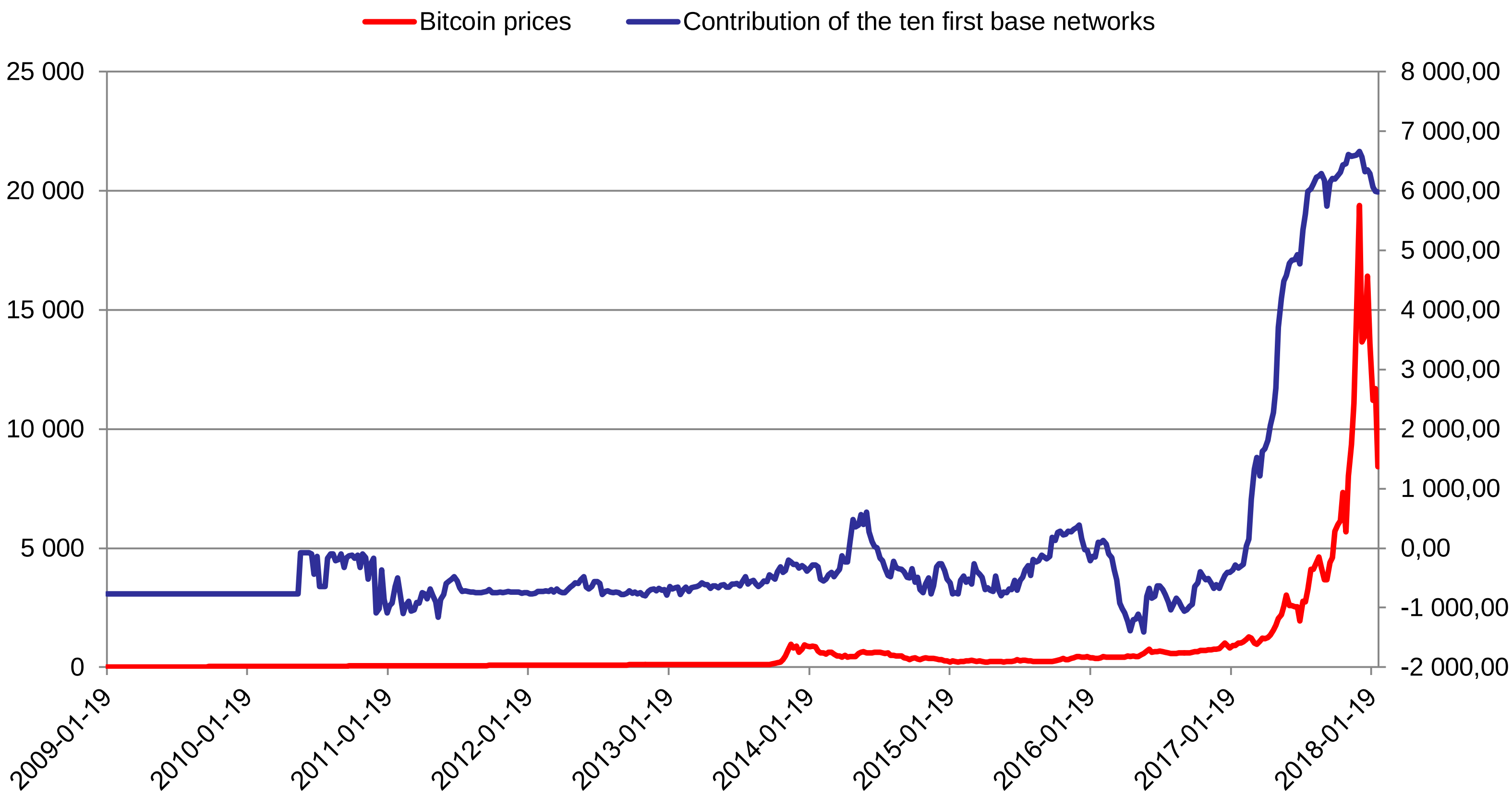


Comment 7

The number of buyers begins to outweigh the number of sellers since the rise of the Bitcoin prices (July 2017)

POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS
Al. Niepodległości 10
61-875 Poznań, Poland

Stage 3 Objective 4 Characteristics of the user's graph properties, cont.
Principal Component Analysis – ten base components



Comment 9

Ten base components are better correlated with the bitcoin price

Poznań University of Economics and Business
Al. Niepodległości 10
61-875 Poznań, Poland

Conclusions

We have identified basic features of users graph:

- Most users are represented by clusters containing at most 1000 IP addresses.
- We suppose that the distribution of number of edges follows a power law.
- Number of nodes and bitcoin turnover have similar distribution in time.
- Users' balances (positive and negative) tend to stabilization.
- The bitcoin system is characterized by high stratification of users wealth (value of Gini coefficient is close to one).
- The number of buyers begins to outweigh the number of sellers since the rise of the Bitcoin prices (July 2017).
- The results of Prinipal Component Analysis indicate that the linear combination of the first ten base network components is strongly correlated with bitcoin price.

Al. Niepodległości 10
61-875 Poznań, Poland

Proposals for future research

- Establishing causal relationship between the observed features, and possibly predicting price changes based on structural changes in the users' graph.

- From mathematical point of view: one can try to construct higher-dimensional objects from the users' graph – focus on construcion of such objects from the „base networks" computed with PCA. Further, investigate their structure by topological methods. There are interesting invariants allowing to capture imortant features of topological spaces (e.g. homology). In many cases, there is a possibility to compute these invariants.

POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS
Al. Niepodległości 10
61-875 Poznań, Poland

References

A.M. Antonopoulos, *Mastering Bitcoin*, O'Reilly, 2010

S. Corbet, B. Lucey, A. Urquhart, L. Yarovaya, *Cryptocurrencies as a Financial Asset: A systematic analysis,* Preprint submitted to Elsevier, April 26, 2018

D. Kondor, M. Posfai, J. Szüle, I Csabai and G. Vattay, *Inferring the interplay between network structure and market effects in Bitcoin*, New Journal of Physics (The open access journal), (16), 2014

Maesa, D. D. F., A. Marino, and L. Ricci (2017), *Data-driven analysis of Bitcoin properties: exploiting the users graph*. Int J Data Sci Anal.

Maesa, D. D. F., A. Marino, and L. Ricci (2017). *Detecting artificial behaviours in the bitcoin users graph*. Online Social Networks and Media 3, 63–74.

Dataset was obtained from the website:

https://senseable2015-6.mit.edu/bitcoin/

POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS

Al. Niepodległości 10
61-875 Poznań, Poland

Thank you for your attention