UOC
Universitat Oberta de Catalunya
www.uoc.edu

UAB
Universitat Autònoma de Barcelona

Universitat Rovira i Virgili

UIB
Universitat de les Illes Balears

# The Bitcoin Transaction Graph: Anonymity

**Marc Santamaría Ortega**

**Master's Degree in Security of the Information and Communication Technologies (MISTIC)**
**Universitat Oberta de Catalunya (UOC)**
**Supervisor: PhD Student Cristina Pérez-Solà (UAB)**

**June 21, 2013**

# Introduction

First digital currency that relies on cryptography

- Most widely used
- Decentralized, no need of central banks
- Peer-to-peer, all users are equals
- No inflation
- Prevents double-spending

# Overview

- Bitcoin Description
  - Concepts: Blockchain, Blocks and Transactions
- Anonymity
  - Analysis
  - Previous works
- Aggregation of addresses
  - Suppositions
  - Linking methods
- Conclusions

# Bitcoin Description

UOC
Universitat Oberta de Catalunya
www.uoc.edu

# Blockchain

The Blockchain consists on a chain of blocks.

Modifying one block means recomputing all subsequent blocks.

Bitcoins are generated when creating a block.

The Blockchain is public.

# Blocks

A Bitcoin block contains the last Bitcoin transactions not recorded in the blockchain.

It has a reference to the previous block.

It contains a solution to a difficulty mathematical problem (proof-of-work).

The proof-of-work difficulty determines the generation rate of blocks.

# Transactions

Bitcoin transactions represent payments between users.

Three types of transactions: to an IP address, to a Bitcoin address and generation of new Bitcoins.

Transactions reference previous transactions and specify the amount transferred.

Inputs present the information needed to spend bitcoins

Outputs indicate the bitcoins spent and validation info.

# Anonymity

# Concepts

- Personal information from users is not required.
- Bitcoins are only linked with the Bitcoin address they belong and the public-private key related to an address.
- Users can generate all the Bitcoin addresses they want.
- Wallets store the link between an address and it’s user.
- Anonymity services like proxies or TOR can be used.
- All transactions are public, the origin can be traced.

# Previous works

Multi-input allows to relate Bitcoin addresses together.

Off-network information can be linked to addresses.

Transactions with two outputs, one is the payment and the other the change.

The origin of a transaction can be determined.

Zerocoin is a proposed anonymity extension.

# Analysis

Bitcoin clients do not allow to select the input address/es

A previous output received has to be fully spent, if not, change is produced.

eWallets, mixers or laundry services are discouraged.

A protocol which provides more anonymity is needed.

Zerocoin seems a good option but its computation is harder, and deployment could become an issue.

# Aggregation of addresses

# Aggregation

Suppositions will be established to determine the aggregation methods.

Objective: link addresses or obtain information of users.

Each method will be analyzed and results obtained from the different aggregation methods will be compared.

Anonymity of users and temporal correlations will also be studied.

# Linking of inputs - Suppositions

Linking of inputs mentioned in the Bitcoin definition.

Previously explored, with a shorter blockchain.

All the Bitcoin addresses appearing as an input in the same transaction can be related.

It is always true.

# Linking of inputs - Study

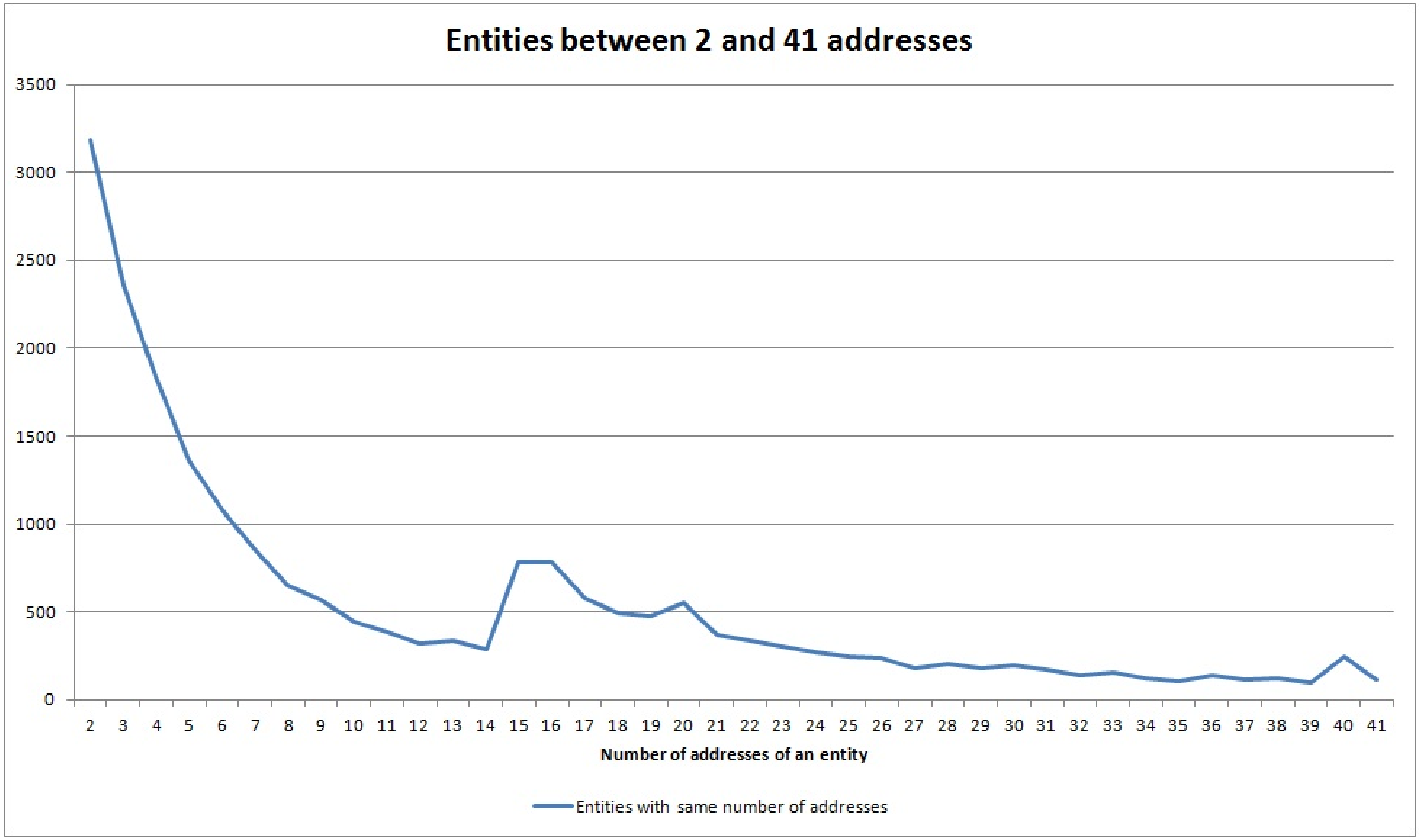
Simple in appearance but computationally difficult.

31,4% of all transactions (4520210) have several inputs.

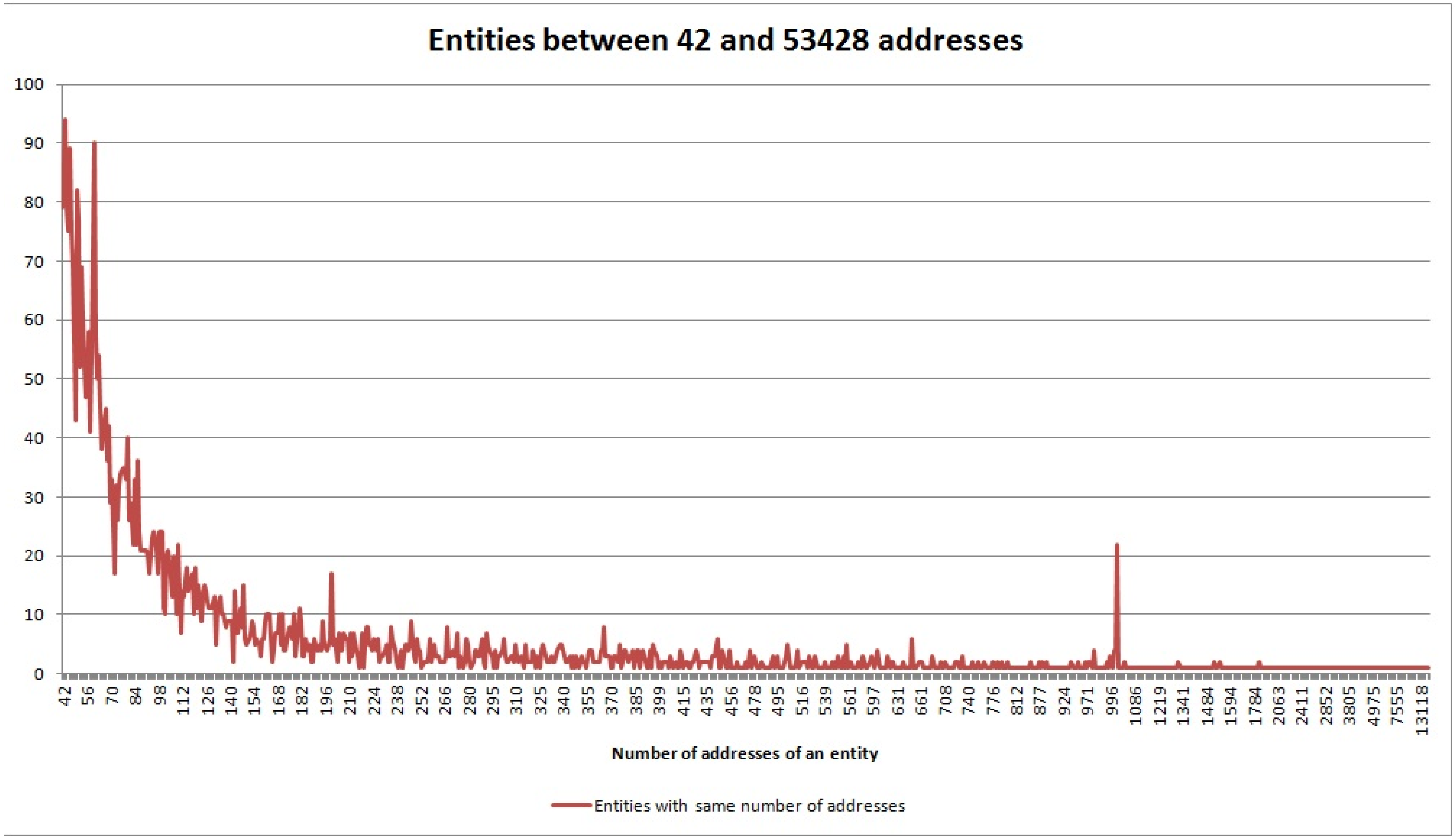
Bash script developed for doing the aggregation.

Out of 6865143 inputs analyzed (34,17%), 1985379 Bitcoin addresses have been obtained, resulting in 47366 identities. Around 42 addresses per user.

# Linking of inputs - Charts



# Linking of inputs - Charts



# Linking of outputs - Suppositions

In most transactions with two outputs one of them will be the change.

It probably is the output with more decimals in its value:

- Prizes are in general rounded
- The smallest unit from most currencies is several magnitudes bigger than a Satoshi unit
- The lowest units are seldom used in most currencies

# Linking of outputs - Study

Best to center in transactions with two outputs.

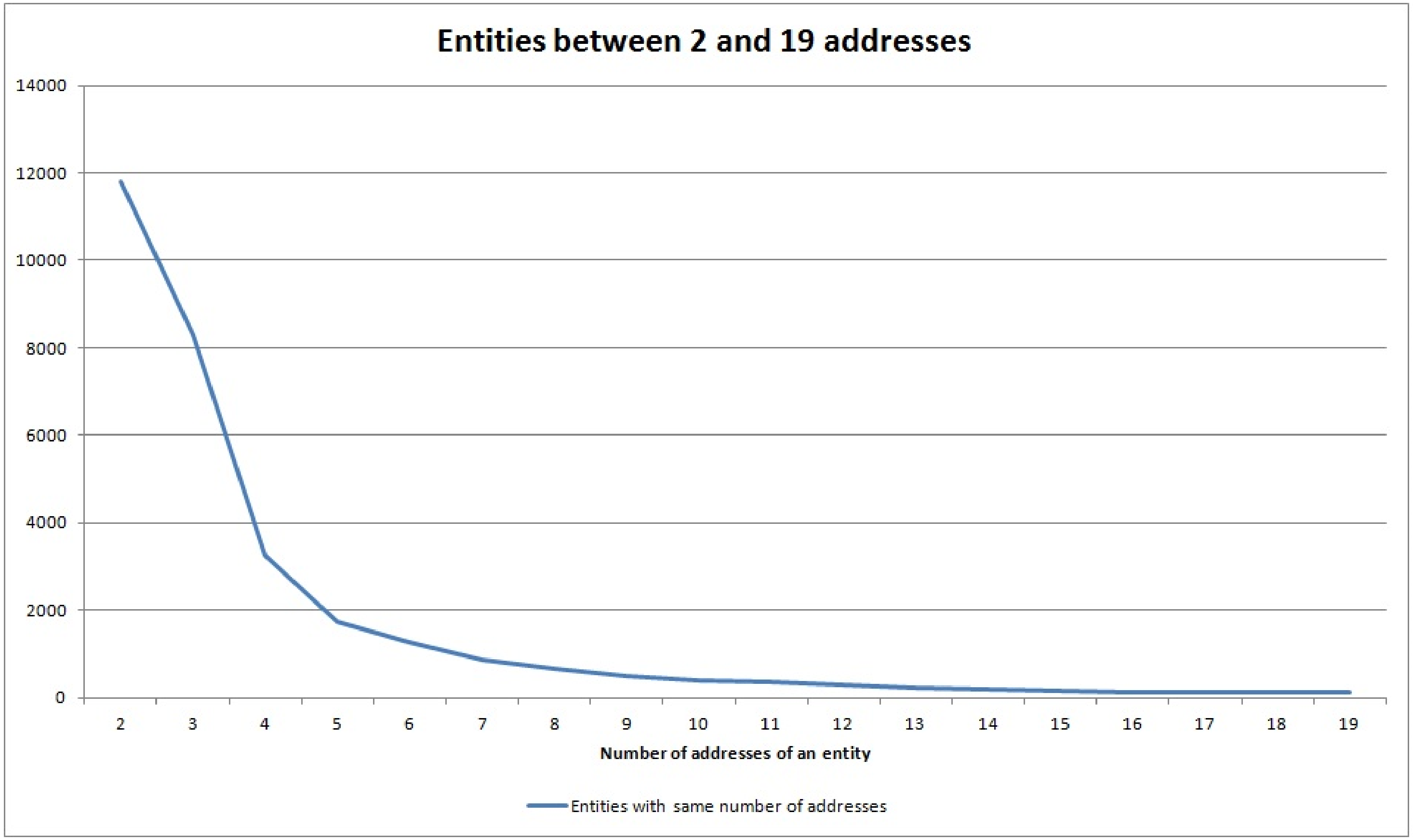
It is also computationally difficult.

89,98% of all transactions (13061821) have two outputs.

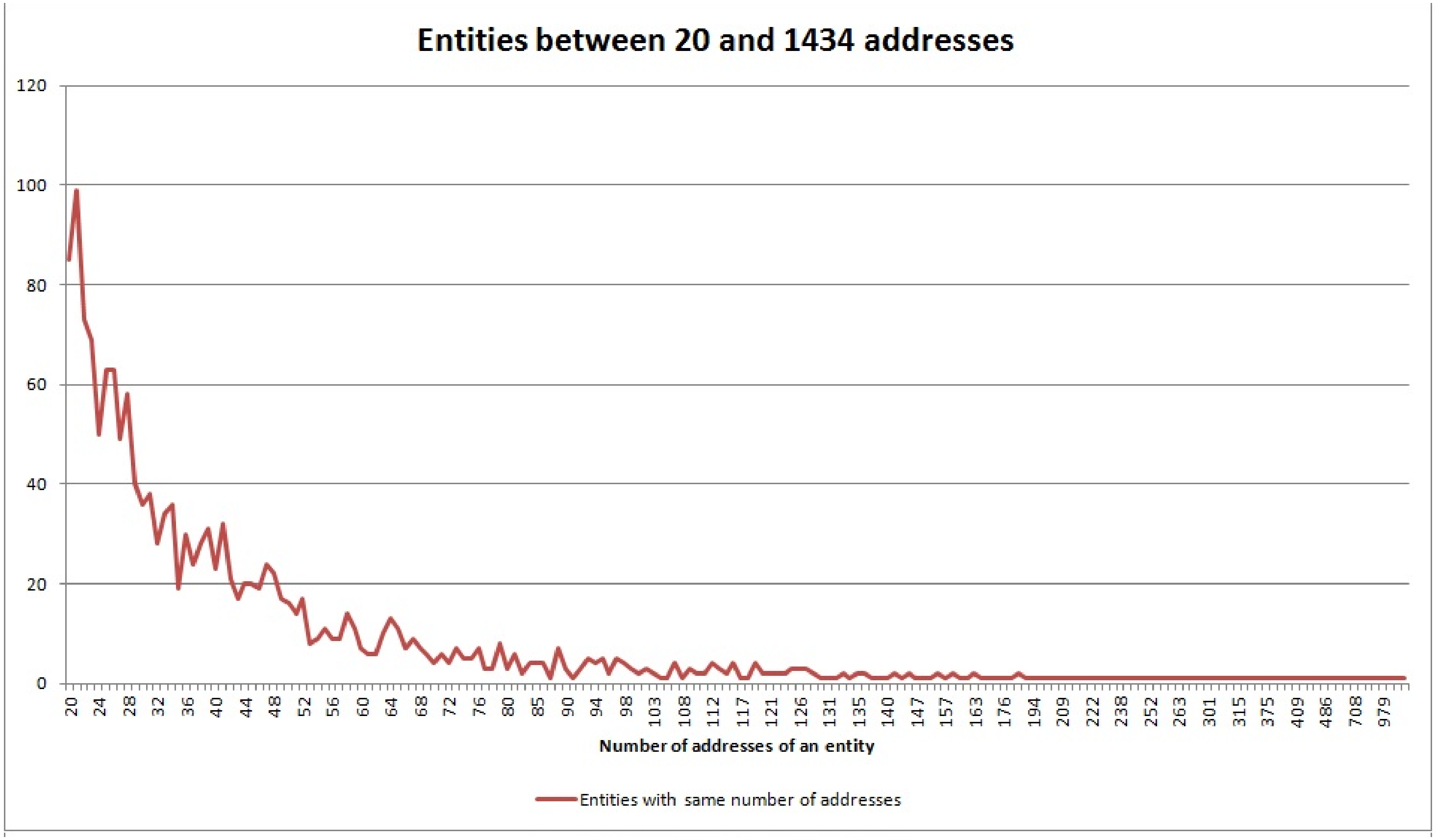
Bash script developed for doing the aggregation.

Out of 451639 transactions analyzed (3,45%), 626604 Bitcoin addresses have been obtained, resulting in 45083 identities. Around 14 addresses per user.

# Linking of outputs - Charts



# Linking of outputs - Charts



# Linking of IPs - Suppositions

Possible to obtain IP address of a transaction.

First IP broadcasting a transaction is the responsible of the transaction.

Most IP addresses are not static, but several transactions close in time could be related.

Using anonymizing services difficults aggregation.

# Linking of IPs - Study

DB of transactions is not useful in this case.

Download daily all the transactions made from a source like Blockchain which include the IP source.

Between 40000 and 69000 transactions per day, with around 4000-5000 different IPs.

IPs most used in June: 5.9.24.81(230450), 127.0.0.1(174559), 85.17.239.32(32167) and 199.48.164.36(17154).

Around 1% of transactions are made using TOR or proxies.

# Comparison of aggregation methods

Aggregation of inputs is the most reliable.

The rest of methods are based on assumptions.

The aggregation of outputs defined also includes the linking of inputs.

Aggregation of inputs more efficient due to previous ordering, aggregation of outputs more stable.

# Anonymity of users

To study anonymity, first, clients using anonymizing services like TOR, proxies, ... should be found.

Services used by Bitcoin clients could be monitored and external information obtained.

For example using crawlers for forums like Bitcointalk, pastes like Bitbin or Pastebin, ...

Around 5400 Bitcoin addresses were obtained, only 2245 were valid,  and assigned to 1248 forum users.

# Anonymity of users

Results from the crawlers should be revised to avoid addresses from online services.

List of addresses from these services to mark matches.

The scripts should be executed periodically with an automatic process, to have up-to-date information.

This process could be extended to other services, correlating different sources will improve the final result.

# Temporal correlations

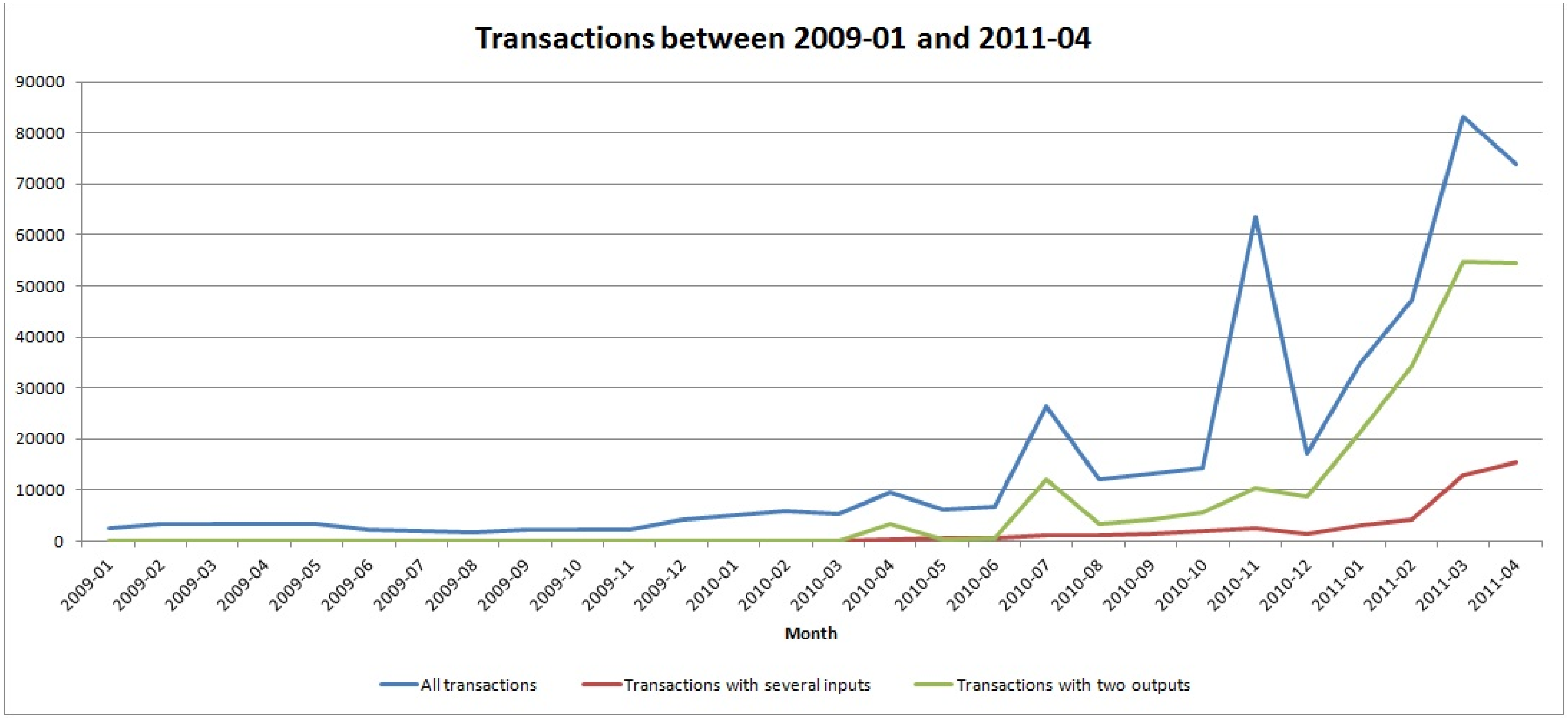
Determine the reliability of aggregation methods of inputs and outputs comparing them overtime.

Transactions grouped monthly.

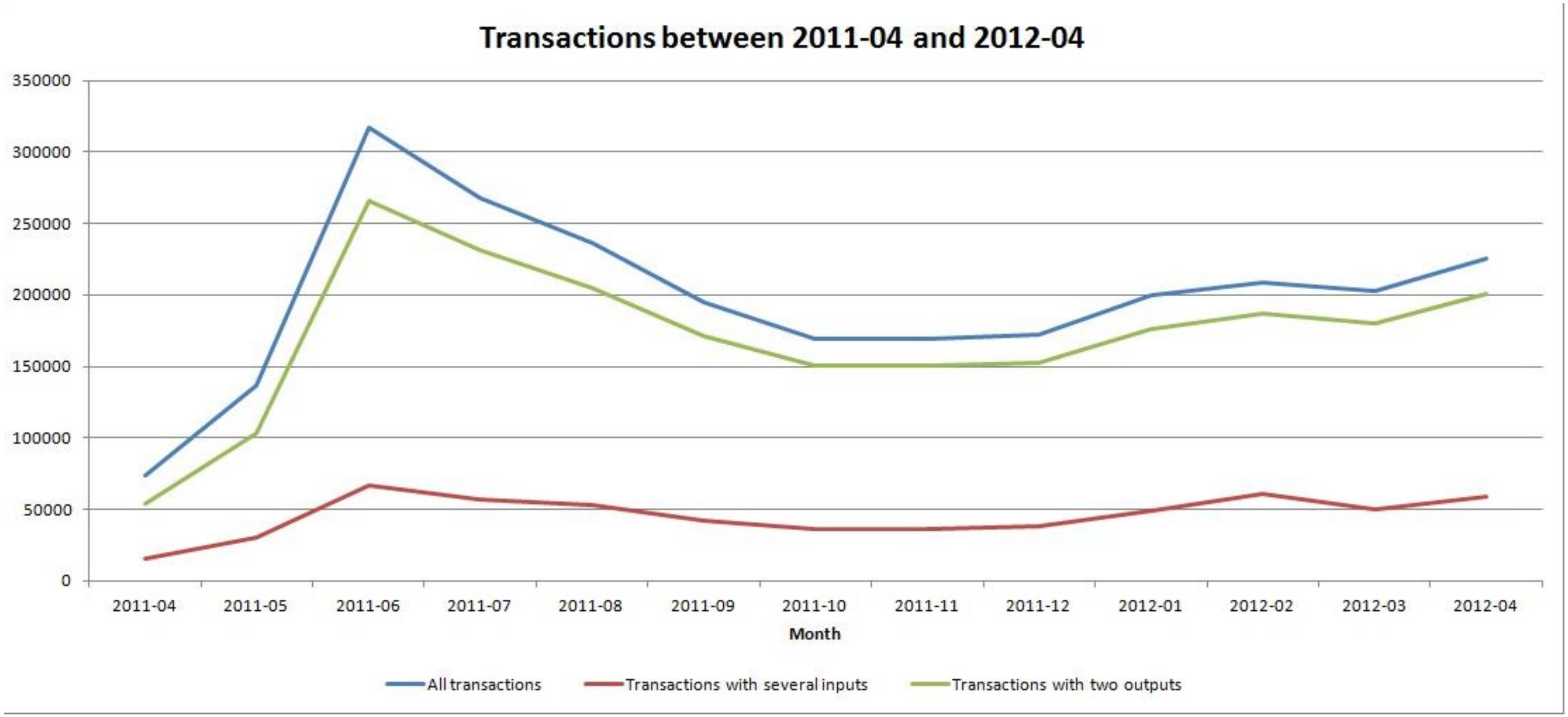
Three groups: all transactions, transactions with several inputs and transactions with two outputs.

Results obtained presented in three charts.

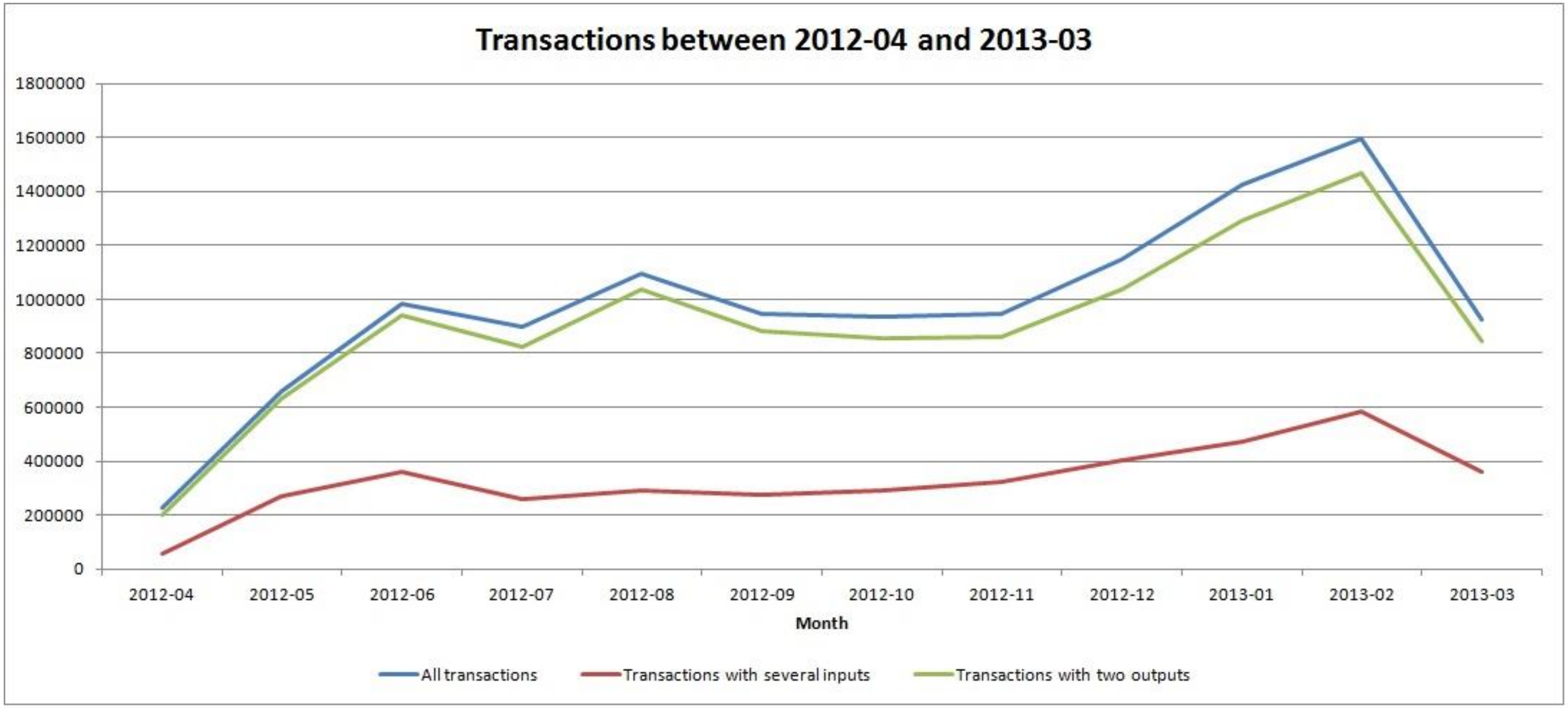
# Temporal correlations - Charts



# Temporal correlations - Charts



# Temporal correlations - Charts



# Conclusions

# Conclusions

- Bitcoin is secure against malicious attacks or double spending but weak against anonymity analysis.
- Anonymity depends mainly on users:
  - Proxies or TOR can be used to anonymize connections
  - Mixers or laundry services for blurring the source of a transaction
  - Using as less inputs as possible avoids input aggregation
  - Using more than two outputs prevents output aggregation
  - Not posting personal information in forums
- Bitcoin still has several challenges to overcome

# Questions?

msantamaria@uoc.edu
https://github.com/Asgarath/