A Bayesian Approach to Identify Bitcoin Users

Péter László Juhász

Department of Physics of Complex Systems
Eötvös Loránd University

March 24, 2017

1. Introduction
2. Data Collection
3. Main Steps of Solution
4. Results

Introduction

Basic Properties of Bitcoin

Basic Properties of Bitcoin

Digital

Basic Properties of Bitcoin

Digital

Cryptography

Basic Properties of Bitcoin

Digital

Cryptography

Network

Basic Properties of Bitcoin

Digital

Cryptography

Network

Anonymity

Basic Properties of Bitcoin

Digital

Cryptography

Network

Anonymity

Public transactions

Goal of Research

- Goal: identification of Bitcoin users; determine geographical distribution and flow of Bitcoin
- Challenge:
  - Anonymous users
  - Parameters of interest are hidden
- Solution: analysis of the time delay and content of the messages propagating in the network, develop and apply a probabilistic model

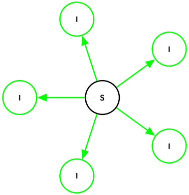
Propagation of a Transaction

Propagation of a Transaction

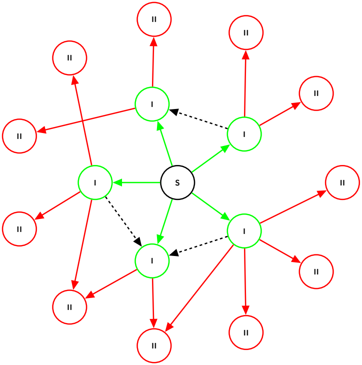
I am sending some Bitcoins!!!
s

Propagation of a Transaction

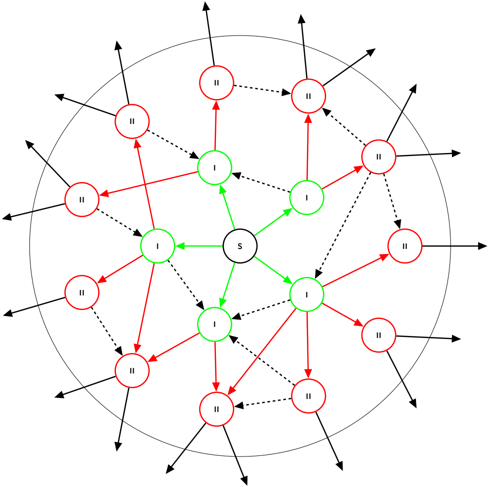
Propagation of a Transaction



Propagation of a Transaction

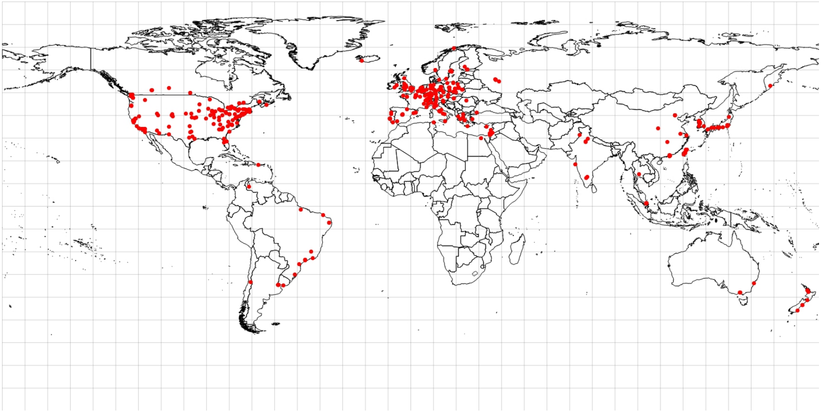


Propagation of a Transaction



Data Collection

Data Collection



$\sim$ 2 billion messages $\rightarrow$ database server

Main Steps of Solution

Main Steps of the Solution

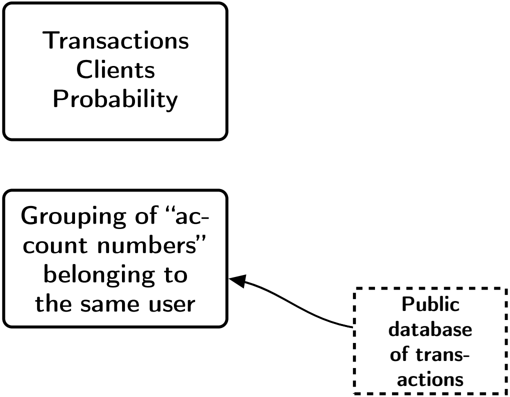
Main Steps of the Solution

Identification of Bitcoin clients creating the transactions

Main Steps of the Solution

Transactions
Clients
Probability

Main Steps of the Solution

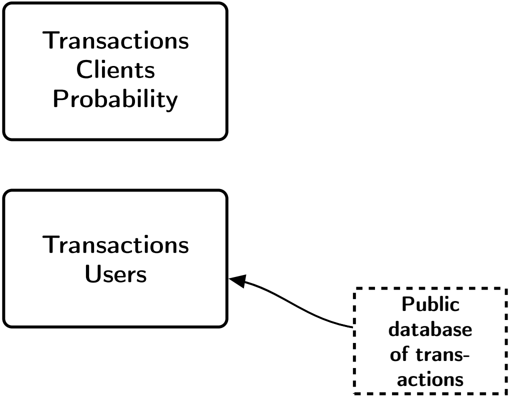


Transcriptions
Clients
Probability

Grouping of "account numbers" belonging to the same user

Public database of transactions

Main Steps of the Solution

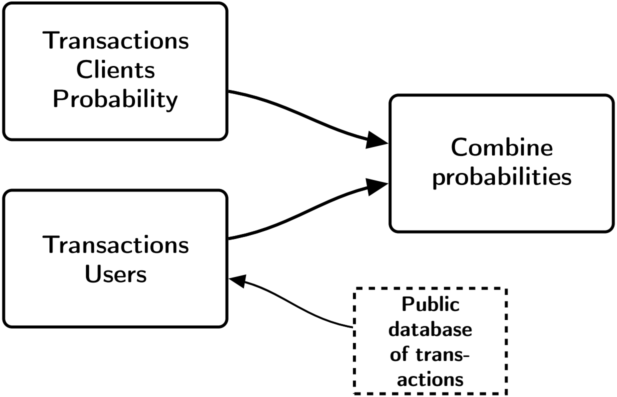


Transcriptions
Clients
Probability

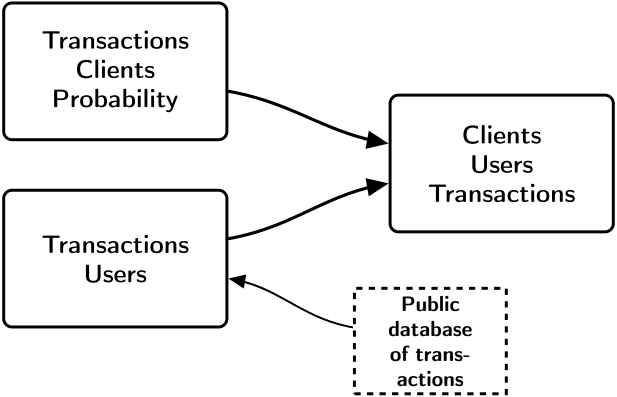
Transactions
Users

Public database of transactions

Main Steps of the Solution



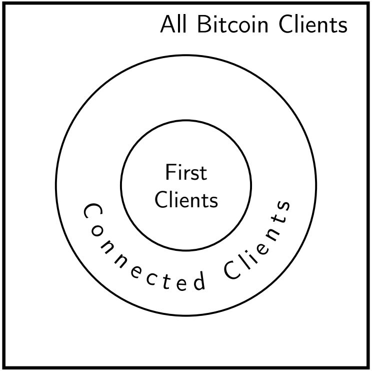
Main Steps of the Solution



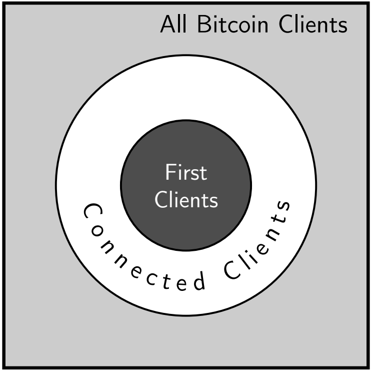
- Combination of Probabilities with naive Bayes classification
- Bitcoin clients can be localized through their IP address

Continue

Creators of Transactions



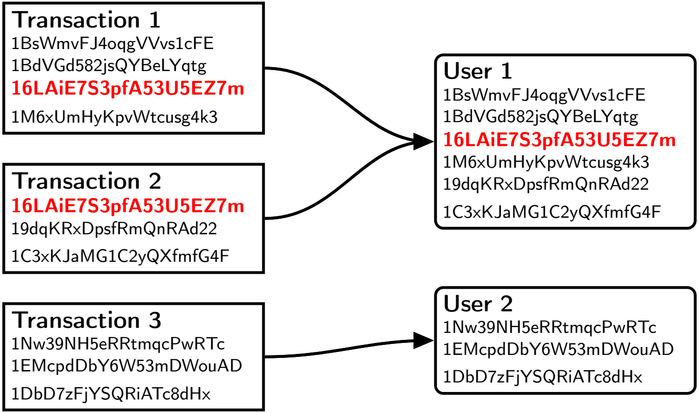
Creators of Transactions



Back

Grouping of Bitcoin Users

Input addresses of a transaction belong to the same user. (Usually, a user uses several addresses in one transaction as inputs.)



Transaction 1
1BsWmvFJ4oqgVVvs1cFE
1BdVGd582jsQYBeLYqtg
16LAiE7S3pfA53U5EZ7m
1M6xUmHyKpvWtcusg4k3

Transaction 2
16LAiE7S3pfA53U5EZ7m
19dqKRxDpsfRmQnRAd22
1C3xKJaMG1C2yQXfmfG4F

Transaction 3
1Nw39NH5eRRtmqcPwRTc
1EMcpdDbY6W53mDWouAD
1DbD7zFjYSQRiATc8dHx

User 1
1BsWmvFJ4oqgVVvs1cFE
1BdVGd582jsQYBeLYqtg
16LAiE7S3pfA53U5EZ7m
1M6xUmHyKpvWtcusg4k3
19dqKRxDpsfRmQnRAd22
1C3xKJaMG1C2yQXfmfG4F

User 2
1Nw39NH5eRRtmqcPwRTc
1EMcpdDbY6W53mDWouAD
1DbD7zFjYSQRiATc8dHx

Back

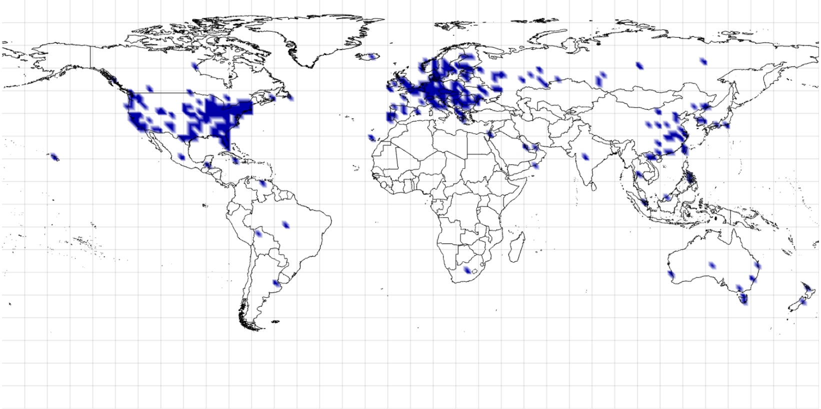
Combination of Probabilities

$$\mathbb{P}\left(Y|\mathbf{tx}\right) = \frac{\prod_{i=1}^{m} \mathbb{P}\left(Y|tx_i\right)}{\mathbb{P}\left(Y\right)^{m-1}\left[\frac{\prod_{i=1}^{m} \mathbb{P}(Y|tx_i)}{\mathbb{P}(Y)^{m-1}} + \frac{\prod_{i=1}^{m} \mathbb{P}(N|tx_i)}{\mathbb{P}(N)^{m-1}}\right]}$$

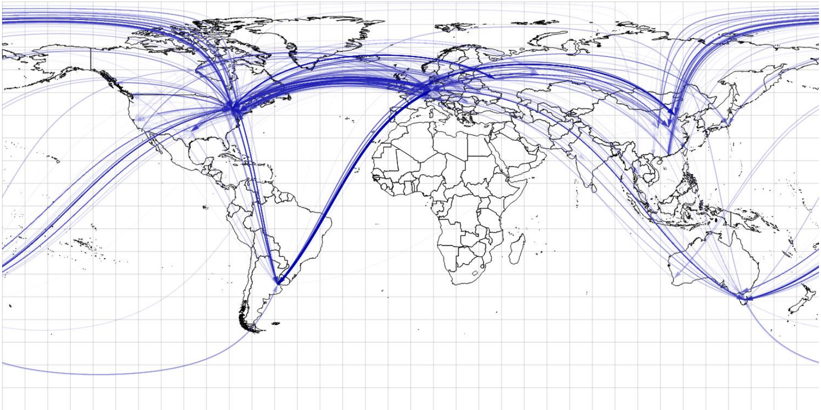
Back

Results

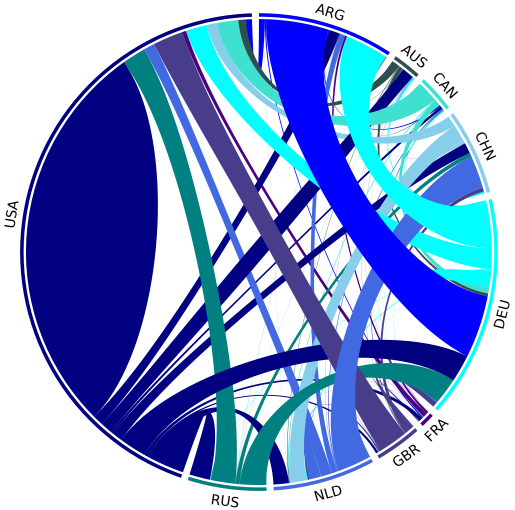
Geographical Distribution of Bitcoin



Flow of Bitcoin



Flow of Bitcoin



Summary

- Goal: determine the distribution and flow of Bitcoin

Summary

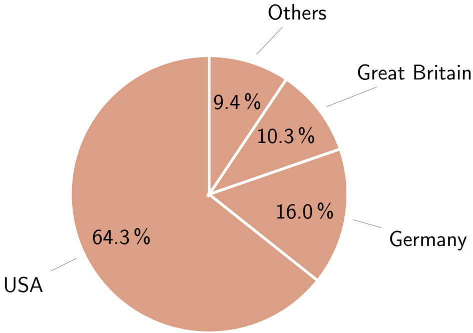
- Goal: determine the distribution and flow of Bitcoin
- Mathematical model → localization of users and transactions

Summary

- Goal: determine the distribution and flow of Bitcoin
- Mathematical model → localization of users and transactions
- Results: deanonymization of Bitcoin; the distribution of Bitcoin correlates with the economic development of the geographic regions

Summary

- Goal: determine the distribution and flow of Bitcoin
- Mathematical model → localization of users and transactions
- Results: deanonymization of Bitcoin; the distribution of Bitcoin correlates with the economic development of the geographic regions



Thanks for your attention!