Blockchain

Modelowanie zachowań oraz odkrywanie wzorców w grafie transakcji Ethereum

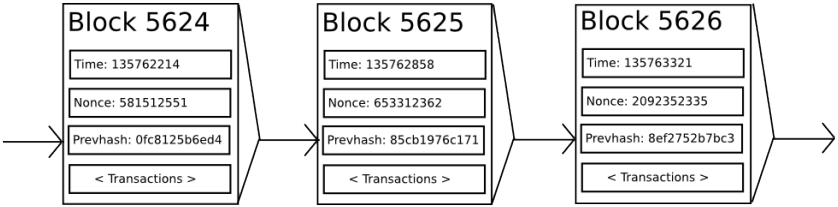
Mateusz Zaborski
M.Zaborski@mini.pw.edu.pl

Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska

Seminarium z Metod Inteligencji Obliczeniowej
23.05.2018

Spis treści

Blockchain - przypomnienie



Rysunek: Kopanie (źródło: https://github.com/ethereum/wiki/)

Cele
Współczesne zagadnienia

- Deanonimizacja
- Monitorowanie środków
- Monitorowanie powiązań
- Anti-Money Laundering

Cele
Cele badań

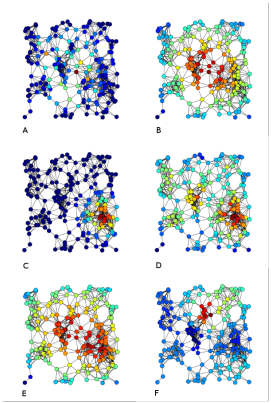
- Stworzenie modelu zachowań
- Wyodrębnienie kluczowych węzłów
- Analiza interakcji z innymi węzłami
- Predykcja ceny

Grafy
Transakcje

- Graf skierowany
- Węzły - adresy
- Krawędzie - transakcje
  - Waga - kwota
- Możliwe krawędzie równoległe

Grafy
Miary centralności

- Betwenness centrality (A)
- Closeness centrality (B)
- Eigenvector centrality (C)
- Degree centrality (D)
- Harmonic centrality (E)
- Katz centrality (F)



Przegląd literatury
Bitcoin Transaction Graph Analysis

Bitcoin Transaction Graph Analysis

Michael Fleder
mfleder@mit.edu

Michael S. Kester
kester@eecs.harvard.edu

Sudeep Pillai
spillai@csail.mit.edu

February 6, 2015

Rysunek: Artykuł 1

Bitcoin Transaction Graph Analysis

Idea

- System podzielony na dwie części
  - scrapowanie z publicznych forów
  - mechanizm przyporządkowywania użytkowników do transakcji
- Forum bitcointalk.org - 2322 użytkowników (2404 adresy)
- Uwzględniano wahania kwoty oraz czasu

Bitcoin Transaction Graph Analysis

Scrapowanie

help with Bitcoin development in php (variable parameters)
April 25, 2011, 02:17:14 AM

Hi all, I have run into some trouble using the bitcoin api with php. When I issue a command like:

$bitcoin->sendfrom($userid, $receiving_address, $amount);

I get an error like:

fopen(http://...@localhost:8332/): failed to open stream: HTTP request failed! HTTP/1.1 500 Internal Server Error

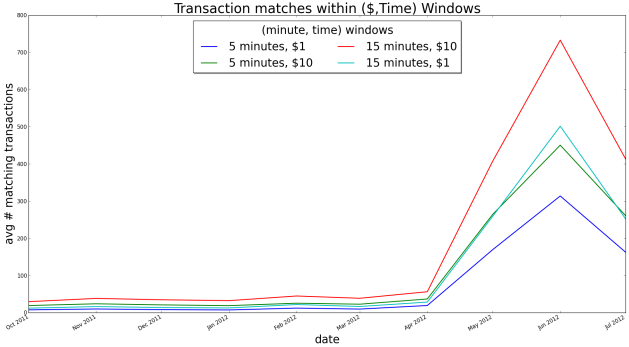
But when I hard code in the parameters:

$bitcoin->sendfrom("1", "1LDNLreKJ6GawBHPgB5yfVLBERi8g3SbQS", 10);

Rysunek: Wyciek adresu z forum bitcointalk.org

Bitcoin Transaction Graph Analysis

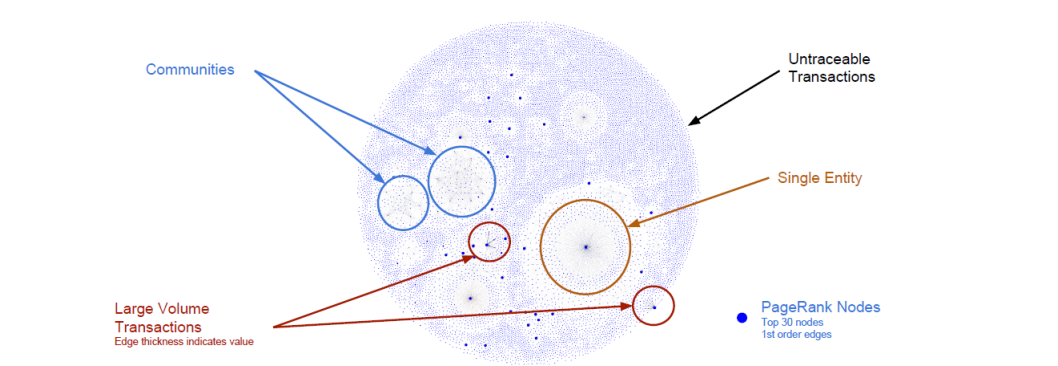
Przyporządkowanie transakcji



Rysunek: Niejednoznaczność transakcji wynikająca z wahań czasu oraz kwoty (źródło: Bitcoin Transaction Graph Analysis, M. Fleder et al.)

Bitcoin Transaction Graph Analysis

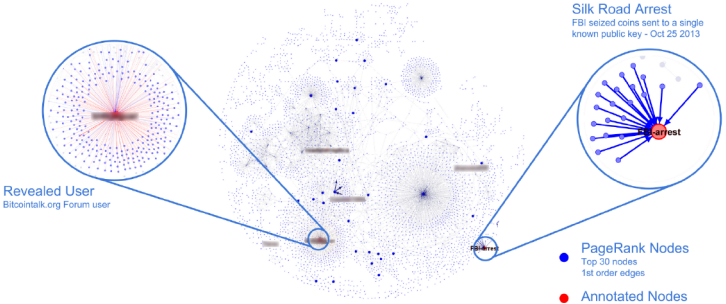
Graf transakcji



Rysunek: Graf transakcji z wybranego dnia (źródło: Bitcoin Transaction Graph Analysis, M. Fleder et al.)

Bitcoin Transaction Graph Analysis

Graf transakcji



Rysunek: Przejęcie środków Silk Road przez FBI (źródło: Bitcoin Transaction Graph Analysis, M. Fleder et al.)

Bitcoin Transaction Graph Analysis

Wnioski

- Zaobserwowano przejęcie aktywów Silk Road przez FBI dnia 25.10.2013
- Znaleziono bliskie powiązania miedzy użytkownikami Silk Road a rzeczywistymi użytkownikami

Przegląd literatury
Quantitative Analysis of the Full Bitcoin Transaction Graph

Quantitative Analysis of the Full Bitcoin Transaction Graph

Dorit Ron and Adi Shamir

Department of Computer Science and Applied Mathematics,
The Weizmann Institute of Science, Israel
{dorit.ron,adi.shamir}@weizmann.ac.il

Rysunek: Artykuł 2

Quantitative Analysis of the Full Bitcoin Transaction Graph

Idea

- Pobrano historię blockchaina
- Przeanalizowano wiele właściwości statystycznych
- Pierwszy raz odpowiedziano na pytania o zachowanie użytkowników
- Skupiono się na grupowaniu adresów
  - wariant algorytmu Union-Find
- Wyizolowano wszystkie duże transakcje i zbadano powiązania
- Sprawdzano cyrkulację BTC

Quantitative Analysis of the Full Bitcoin Transaction Graph

Analiza grafu

- Początkowo 3.73 mln adresów
- 609 tys. wyłącznie jako odbiorcy
- 3.12 mln adresów przyporządkowane do 1.85 mln podmiotów ("entities")
- Jeden podmiot posiada ponad 156 tys. adresów - Mt. Gox

Quantitative Analysis of the Full Bitcoin Transaction Graph

Grupowanie adresów

Table 1. The distribution of the number of addresses per entity

Larger or equal to	Smaller than	Number of entities
1	2	2,214,186
2	10	234,015
10	100	12,026
100	500	499
500	1,000	35
1,000	5,000	41
5,000	10,000	5
10,000	50,000	5
50,000	100,000	1
100,000		1

Rysunek: Grupowanie adresów (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Stan konta

Table 4. The distribution of the maximal balance of BTC's ever seen per entity and per address

Larger or equal to	Smaller than	Number of entities	Number of addresses
0	0.1	547,763	1,063,876
0.1	10	668,247	1,160,170
10	100	945,083	1,188,596
100	1,000	259,142	276,613
1,000	10,000	36,769	37,087
10,000	50,000	3,513	3,521
50,000	100,000	163	159
100,000	200,000	40	41
200,000	400,000	26	26
400,000	500,000	68	129
500,000		2	0

Rysunek: Stan konta (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Liczba transakcji

Table 5. The 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

Rysunek: Liczba transakcji (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Największe podmioty

Entity ID	Number of Addresses	Accumulated Incoming BTC's	Number of Transactions
A	78,251	2,886,650	246,012
B (Mt.Gox)	156,722	2,206,170	477,526
C	13,289	941,013	77,525
D	12,520	867,996	48,347
E	191	692,864	1,353
F	12	660,000	23
G (Instawallet)	23,649	633,606	92,593
H	9	580,000	59
I	10,561	514,066	49,550
J	4	500,021	6
K	134	479,254	1,039
L (Deepbit)	2	452,929	814,044
M	9	442,000	10
N	128	432,161	137
O	10	432,286	14
P	1	432,078	3
Q	14	430,490	23
R	2,124	321,866	300,486
S	1,037	20,308	197,334

Rysunek: Największe podmioty (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Wielkość transakcji

Table 6. The 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

Rysunek: Wielkość transakcji (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

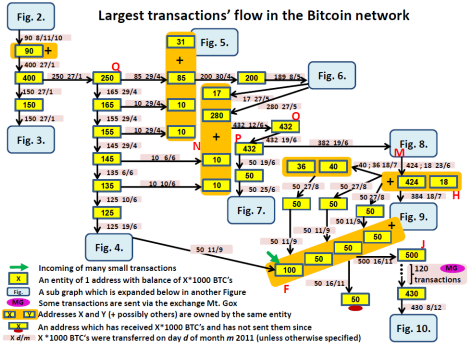
Quantitative Analysis of the Full Bitcoin Transaction Graph

Największe transakcje

- Wzięto pod uwagę pierwszą największą transakcję - 90'000 BTC, 8.10.2010
- 348 (wszystkich - 368) było następnikami pierwszej transakcji
- Wykryto ciekawe wzorce
  - Długie łańcuchy (do 350 transakcji)
  - Wzorce "fork-merge" i pętle własne
    - Często różne kwoty
  - Trzymanie BTC w "bezpiecznych kontach"
  - Drzewa binarne

Quantitative Analysis of the Full Bitcoin Transaction Graph

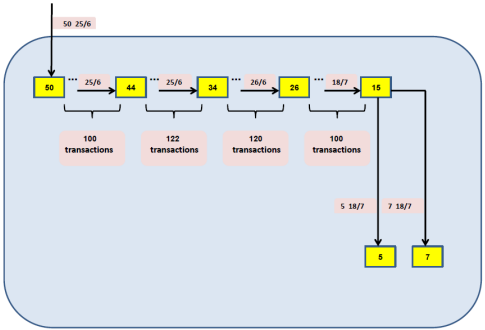
Duża transakcja



Rysunek: Duża transakcja (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

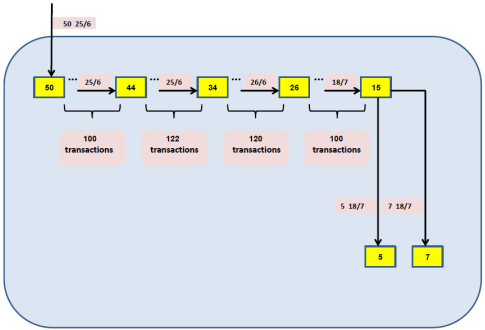
Długie łańcuchy



Rysunek: Długie łańcuchy (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

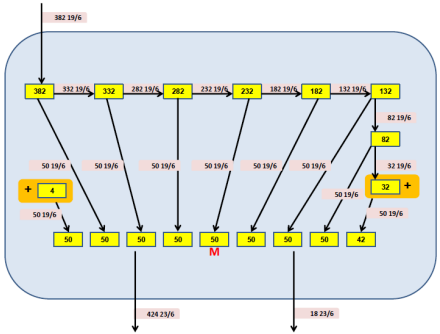
Długie łańcuchy



Rysunek: Długie łańcuchy (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

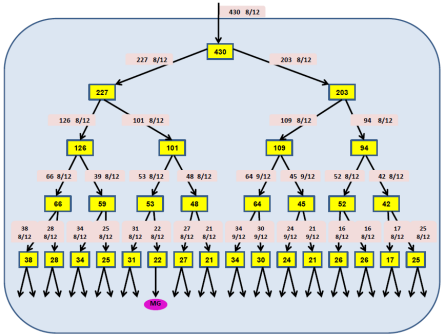
Wzorzec "fork-merge"



Rysunek: Wzorzec "fork-merge" (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Wzorzec - drzewo binarne



Rysunek: Wzorzec - drzewo binarne (źródło: Quantitative Analysis of the Full Bitcoin Transaction Graph, D. Ron and A. Shamir)

Quantitative Analysis of the Full Bitcoin Transaction Graph

Wnioski

- Oszacowano liczbę adresów WikiLeaks (>83) i sumę dotacji (2605.25 BTC)
- 73% BTC jest nieużywanych (jest w adresach, które nic nie wysyłają)
- 51% BTC jest nieużywanych od ponad 3 miesięcy
- Odkryto wzorce rozchodzenia się dużej transakcji w sieci

Przegląd literatury
Trading Bitcoin and Online Time Series Prediction

Trading Bitcoin and Online Time Series Prediction

Muhammad J Amjad MAMJAD@MIT.EDU
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Devavrat Shah DEVAVRAT@MIT.EDU
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technolog
Cambridge, MA 02139, USA

Editor: Oren Anava, Marco Cuturi, Azadeh Khaleghi, Vitaly Kuznetsov, Alexander Rakhlin

Rysunek: Artykuł 3

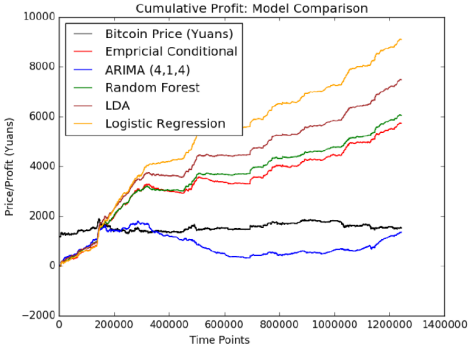
Trading Bitcoin and Online Time Series Prediction

Idea

- Celem jest predykcja ceny BTC
- Klasyczne podejście w oparciu o szeregi czasowe
- Drugie kluczowe zagadnienie - zwieranie transakcji
- Podział na 3 okresy
  - Uczący
  - Walidujący
  - Testowy

Trading Bitcoin and Online Time Series Prediction

Wyniki



Rysunek: Porównanie wybranych modeli (źródło: Trading Bitcoin and Online Time Series Prediction, M. J. Amjad, D. Shah)

Przegląd literatury

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Alex Greaves, Benjamin Au

December 8, 2015

Rysunek: Artykuł 4

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Idea

- Stosowanie uczenia maszynowego do predykcji ceny BTC
- Stworzenie grafu transakcji
  - Kopanie jest przedstawione jako pętla własna
- Pobrano historyczne ceny BTC
  - Przyporządkowano ceny do transakcji (krawędzi)
- Celem jest predykcja ceny na godzinę do przodu

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

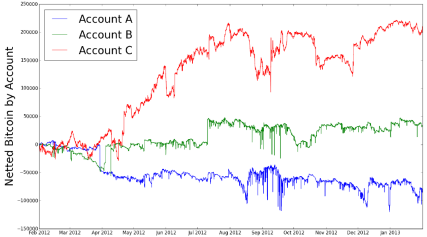
Cechy użyte w modelu

- Aktualna cena BTC
- Liczba transakcji na godzinę
- Średnia wartość transakcji
- Mediana stopni wierzchołków
- Liczba nowych węzłów w godzinie
  - Kłopotliwa kwestia
- Średnia wartość depozytu w nowych węzłach
- Liczba transakcji wykonanych przez nowe węzły
- Miara centralności (*closeness*)
- etc.

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Obserwacje

- 3 węzły odpowiadają za ponad 10% obrotu BTC w sieci
  - Prawdopodobnie jest to Mt. Gox



Rysunek: Trzy kluczowe węzły w sieci (źródło: Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin, A.Greaves, B. Au)

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Algorytmy

- Algorytmy sieciowe
  - Union Find
- Algorytmy uczenia maszynowego
  - Linear Regression
  - Logistic Regression
  - Support Vector Machine
  - Neural Network (2-hidden-layer)
  - K-Nearest Neighbors
- Model referencyjny
  - $P_{t+1h} = P_t * 1.01$

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Wyniki

Model	MSE
Model referencyjny	2.02
Linear Regression	1.94
SVM Regression	1.98

Tabela: Wyniki modeli regresji

Model	MSE
Model referencyjny	53.4%
Logistic Regression	54.3%
SVM	53.7%
Neural Network	55.1%

Tabela: Wyniki modeli klasyfikacji

Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin

Wnioski

- Najbardziej istotną cechą (*informative*) okazała się aktualna cena
- Na wybranym okresie testowym udało się nieznacznie poprawić wynik względem modelu referencyjnego

Środowisko i technologia

Przetwarzanie

- Rozmiar blockchaina - ponad 140GB
- AWS Amazon
  1. Węzeł (geth + web3.js API)
  2. Skrypt korzystający z web3.js API
  3. Skrypt do analiz grafu
     - Julia + LightGraphs
     - Python + networkX

Środowisko i technologia

Analiza blockchaina Ethereum

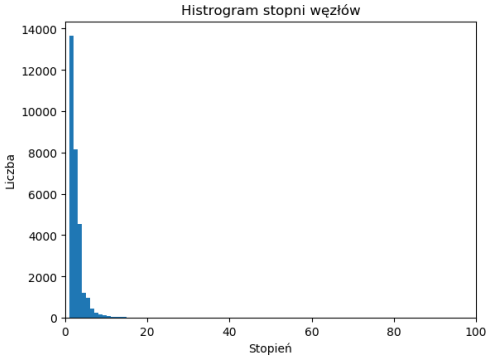

```
{
  blockHash: "0x8e38b4dbf6b11fcc3b9dee84fb7986e29ca0a02cecd8977c161ff7333329681e",
  blockNumber: 1000000,
  from: "0x32be343b94f860124dc4fee278fdcbd38c102d88",
  gas: 50000,
  gasPrice: 60000000000,
  hash: "0xe9e91f1ee4b56c0df2e9f06c2b8c27c6076195a88a7b8537ba8313d80e6f124e",
  input: "0x",
  nonce: 17387,
  r: "0x3b08715b4403c792b8c7567edea634088bedcd7f60d9352b1f16c69830f3afd5",
  s: "0x10b9afb67d2ec8b956f0e1dbc07eb79152904f3a7bf789fc869db56320adfe09",
  to: "0xdf190dc7190dfba737d7777a163445b7fff16133",
  transactionIndex: 1,
  v: "0x1c",
  value: 437194980000000000
}
```


Rysunek: Transakcja pozyskana przez web3 API

Badany podgraf

- Okres - ... - 31.12.2015
- Liczba węzłów - 29917
- Minimalny stopień - 1
- Średni stopień - 3.49
- Mediana stopni - 2
- Maksymalny stopień - 6260
  - 2. największy - 2899
  - 3. największy - 2560
  - 4. największy - 2368

Badany podgraf



Rysunek: Histogram stopni węzłów

Badany podgraf

Adres	Degree centrality
0x2910543Af39abA0Cd09dBb...	0.0969
0x52bc44d5378309EE2abF15...	0.0856
0xC47Aaa860008be6f65B58c...	0.0792
0x120A270bbC009644e35F0b...	0.0738
0x793AE8C1b1a160BFc07BFB...	0.0670

Tabela: Miary centralności

Badany podgraf

Adres	Degree Centrality
0x2910543Af39abA0Cd09dBb...	Kraken
0x52bc44d5378309EE2abF15...	Nanopool
0xC47Aaa860008be6f65B58c...	Cryptsy
0x120A270bbC009644e35F0b...	Shapeshift1
0x793AE8C1b1a160BFc07BFB...	Contract

Tabela: Miary centralności

Podsumowanie

- Blockchain (Ethereum) nadaje się do modelowania jako graf
- Badania przedstawiają obiecujące wyniki
- Wbrew przekonaniom technologia przeczy anonimowości
- Ethereum słabo zbadane
- Interesujący obszar behawioralno - ekonomiczny
- Interesujące nowe obszary i modele
  - Transakcje oczekujące
  - Aktualna prowizja