Assessing Holistic Impacts of Major Events on the Bitcoin Blockchain Network
1

Introduction
Bitcoin

- The pioneer of blockchain technology
  - Decentralized public ledger
  - Solves double spending
- Deflationary cryptocurrency
  - 2040 = no more Bitcoins created
  - Current Market Cap: ~70 billion
- Proof of Work Consensus Algorithm
### Bitcoin Transactions

- Transactions are a transfer of BTC from one Bitcoin address(es) to others
  - Broadcast to nodes (unconfirmed transactions)
- Unconfirmed transactions are added to the memory pool (Mempool)
Bitcoin Blocks

- Set of validated transactions
- New block is mined every ~10 minutes -> added to blockchain (chronological)
  - Miner rewarded with Bitcoins
  - Transactions in that block are removed from mempool
    - Unconfirmed -> Confirmed transactions
    - Miners typically add transactions with a higher fee first
1. **Bitcoin user** sends a transaction (Transaction 86).

2. **Transactions** are stored in the memory pool.

3. **Miner** validates all transactions and creates a block.

   - **Block header** contains:
     - Block hash (Block ID)
     - Previous block id
     - Transactions hash (Merkle root)
     - Number of transactions

4. **Block** gets added to the node's Blockchain.

**BLOCK**

**Transactions**

**BLOCK IN PROGRESS**
Purpose

Events and Bitcoin
We define this as the aggregate of Bitcoin addresses, transactions, blockchain data, and memory pool.

Abbreviated BBN
Price of Bitcoin

From Blockchain.com Charts
**Bitcoin’s Price Has Soared. What Comes Next?**

SAN FRANCISCO — **Bitcoin** has been in a bull market like few the world has ever seen.

At the beginning of the year, the price of a Bitcoin was below $1,000. It hit $8,000 in October, then doubled by late November. And on Thursday, less than two weeks later, the price of a single Bitcoin rose above $20,000 on some exchanges, according to Coinmarketcap.

The latest price spike has been credited to signs that Wall Street companies plan on bringing their financial might into the market.

At the current cost, the value of all Bitcoin in circulation is about $300 billion. To get a sense of how big that is, all the shares of Goldman Sachs are worth about $80 billion.

The gains have been driven by several other factors — perhaps the most important being the irrational mentality that can take over in speculative bubbles.
Is there a quantifiable change in the Bitcoin ecosystem (aside from price) in response to worldwide events?

How can we measure this change? Do certain types of events cause a particular type of change?
Challenges

Holistic State Capture
How can we quantitively capture holistic states of a cryptocurrency?

Change vs. Variation
How can we distinguish events from fluctuations?
Existing Research

- Bitcoin transaction & users network [5][6]
  - Overall topology
  - Anonymity
- Bitcoin exchange price and factors [7][8][10]
  - User activity
  - Trends, wavelet coherence analysis
- Visualization [9]
  - Inspiration: DDOS attack and money laundering
3

Framework

Bitcoin Blockchain Network Analysis
Impact-Score
Initial Exploration

- Bitcoin transaction & users graph
  - Node = Bitcoin address
  - Edge = flow of Bitcoin
  - Edge color/weight = value
- Visualized before/after
Ideal Approach

- Graph similarity/matching
- Network property comparison
  - Clustering coefficient, weighted degree, Eigenvector centrality, Betweenness centrality, PageRank, etc.
**Problem:** Graph Size!
- ~1 million edges, ~500k nodes
- K-Core Filtering and Giant Component
  - Graph is still too big!
- Not computationally feasible

8 hour frame with Giant Component Filtering
1. “Features” of key Bitcoin attributes
2. BNS (Blockchain Network Structure) Vectors
   a. State comparisons without graph matching
      i. Vector Similarity
3. Impact-Score
   a. Standardize event-associated change against fluctuations for given configuration
Attribute Selection

Framework/Methods
Bitcoin Attribute Types

Transaction
Tx Value, Tx Fees, Tx Size, etc.

Address
Unique Ads, Clustering Coefficient, etc.

Mempool
Mempool Growth, Mempool Size, etc.

Block
Block Size, Tx/Block, etc.
Feature Processing
Feature Processing

- Used to reduce multidimensional size-variant raw data into BNS (Blockchain Network Structure) vector with consistent specs
- EX: 96 hour time frame
  - ~600 blocks
  - ~1 million transactions
Feature Processing

1122 Features

Derived from 27 attributes

Feature Selection

99 Features

Derived from 13 attributes
Why Feature Processing?

**Constant Specifications**
N blocks, transactions, etc. fluctuates. Features allow us to maintain constant BNS vector dimensionality & size.

**Remove Temporal Component**
Did not want the time series of blocks, transactions, or mempool values.

**Computational Efficiency**
> 1 million BNS vectors used in our analysis.

**Con: Decreased Data Resolution**
Feature Sets & Subnetworks

Activity
40 features
5 attributes

Transaction
37 features
3 attributes

Fee
27 features
3 attributes
Impact Score

- Change associated with event must be compared against natural variations
  - (natural, “non-event”) variations
- Distance Measure between BNS vector = change
- Universal measure across cryptocurrencies
\[ I = \frac{e_d - MD(B_d)}{STD(B_d)} \]

**Background Fluctuation Calculations**

Vector pairs are offset by 1 hour \((s = 1)\)

**Event**
August 1, 2017

**Background Distances**

\[ B_{\Delta} \]

Summary Statistics
Mean, Median, SD

**Data Frame Length**

\( l_d \)

**Gap Length**

\( l_s/2 \)

**Not to scale**

**I Score**

**Vector Pair Distance**

\[ \Delta([v_c, v_{rep}] \]

**BNS Vector**

\( v_c \)

\( v_{rep} \)
Impact Score Advantage

- Direct cross-comparisons within ecosystem across all configurations
  - Network/subnetwork
    - Fee vs. Overall
  - Gap lengths
    - 0 vs. 200 hour (natural variations)
  - Data frame length
- I score value is directly cross-comparable

Compared to using vector distances
Impact Score Advantage

Compared to using vector distances

- Event Impact Significance
  - Separating event-induced change from natural variations
- Natural variation dependent on configuration
- Approximates Z-score cube-root normalized
Impact Score Weakness

- Assumes relatively event-free background time period
- Underestimates event-impact during high-event time period
  - Late-2017 to early 2018
4

Data Management
API key from Blockchain.com (formerly Blockchain.info)
  ▶ Block #300,000-534,000
Mempool & Transaction Rate data from Blockchain Charts
Analysis Workflow

1. Blockchain to MongoDB server
2. BNS Vector Generation
3. BNS Vector Distance Calculations
4. Temporal Impact Scores & Graphs

Java

Python
Time Performance

For all 16 events, run on a server with 24 cores, 128 GB RAM

Feature Processing & BNS Vector Creation

Parallelized: ~100 hours (data frame length = 96 hours)

Impact Score Analysis

Single Thread: ~2 hours
Events Selected

16 events selected from 2016-2018
- No mempool data before 2016

3 Types

Global
Security, political, or governmental events that impact the Bitcoin ecosystem either directly or indirectly.

Financial
Events that impact or are the result of a change in the state or value of Bitcoin as a cryptocurrency.

Regulatory
Events that are the result of either positive or negative regulation directly or indirectly affecting the Bitcoin ecosystem.
Global
- Brexit
- Bitfenix Hacked
- Election of Trump
- BTC-BCH Hard Fork
- Coincheck NEM Hack

Regulatory
- SEC denies Winklevoss ETF
- Japan legalizes BTC
- South Korea announces BTC ban
- Facebook bans Crypto ads
- SEC requires BTC exchange registration
- Google bans Crypto ads

Financial
- 1 BTC=$1000
- CME Announces Bitcoin Futures
- 1 BTC = $10,000
- CBOE BTC Futures Launch
- BTC Price Peak
Impact Score Threshold

- **Substantial Impact: $I > 2.0$ (1.9-2.1)**
  - Eventful background period
  - $\pm 0.1$ if event time adjusted $\pm 3$ hours

- **Significant Impact: $I > 3.0$ (2.9-3.1)**
Distance Measure Selection

Squared Euclidean Distance

\[ d(p, q) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2 \]

5 Distance Measures Tested:

- Euclidean, Cosine, Squared Euclidean, Chebyshev, and Canberra.
Analysis Process

- Event analysis configurations
  - 4 featuresets
  - Gap lengths: 0-480 hours
  - Data frame length: 96 hours

- Scikit-Learn
  - Vector standardized (0,1) Min-Max Scaler
  - Squared Euclidean Distance
  - Temporal Graphs: Matplotlib
  - Cubic spline interpolation
Analysis & Results
15/16 associated with substantial change in one or more of the four networks defined (overall, activity, fee, transaction)

- Exception: Japan Legalizes Bitcoin (April 1, 2017)
  - Prior announcement

Greatest Impact: Election of Trump [I Score = 5.459] @ 60 hrs on Overall
Overall BBN
### Overall BBN

<table>
<thead>
<tr>
<th>N &amp; Type</th>
<th>Event</th>
<th>Max I-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, G</td>
<td>Trump Elected</td>
<td>5.5</td>
</tr>
<tr>
<td>16, R</td>
<td>Google bans Crypto Ads</td>
<td>3.9</td>
</tr>
<tr>
<td>7, G</td>
<td>BTC-BCH Hard Fork</td>
<td>3.6</td>
</tr>
<tr>
<td>1, G</td>
<td>Brexit</td>
<td>3.4</td>
</tr>
<tr>
<td>11, F</td>
<td>BTC Price Peak</td>
<td>3.3</td>
</tr>
<tr>
<td>5, R</td>
<td>SEC denies ETF</td>
<td>2.3</td>
</tr>
<tr>
<td>10, F</td>
<td>CBOE BTC Futures Launch</td>
<td>2.3</td>
</tr>
<tr>
<td>8, F</td>
<td>CME Announces BTC Futures</td>
<td>2.1</td>
</tr>
<tr>
<td>14, R</td>
<td>Facebook bans Crypto Ads</td>
<td>2.0</td>
</tr>
</tbody>
</table>

- Only event with no measurable impact
  - Bitfenix Hacked (I Score = 0.02)
5 Sig. Events

- Significant (I Score > 3)
- Top 4 associated with immediate substantial change
  - Exception: BTC Price Peak
  - Hindsight nature
No strong correlations between event type and temporal impact on overall

- Red = Regulatory
- Blue = Global
- Green = Financial
BTC = $1000 and BTC = $10k

Two BTC Exchange Price Milestone Events
Event Type Impact
Global Events

Non-Temporal Max I Score

Trump Elected  BTC-BCH Hard Fork  Brexit  Coincheck NEM Hack  Bitfenix Hacked
Security Events

Non-Temporal Max I Score

Fee Subnetwork
- Bitfenix Hacked
- Coincheck NEM Hack

Activity Subnetwork
- Bitfenix Hacked
- Coincheck NEM Hack

Transaction Subnetwork
- Bitfenix Hacked
- Coincheck NEM Hack

Subgroup of Global Events
Political Events

Non-Temporal Max I Score

- Fee Subnetwork
- Activity Subnetwork
- Transaction Subnetwork

Brexit vs. Trump Elected

Subgroup of Global Events
Regulatory Events

Non-Temporal Max I Score

- Google bans Crypto Ads
- SEC Exchange Register
- SEC denies ETF
- Japan legalizes BTC
- Facebook bans Crypto Ads
- South Korea BTC Ban

Transaction Subnetwork

Fee Subnetwork

Activity Subnetwork
Financial Events

Non-Temporal Max I Score

- BTC Price Peak
- 1 BTC = $1000 USD
- 1 BTC = $10k USD
- CBOE BTC Futures Launch
- CME Announces BTC Futures

Transaction Subnetwork

Fee Subnetwork

Activity Subnetwork
Event Specific Analysis
Election of Trump

- Political event associated with largest I Score on Overall BBN
  - FUD -> Bitcoin “safe haven” nature
- Traditional Markets [12]
  - Gold spiked 4% but fell
  - US dollar index dropped but recovered quickly
- Event Time: Nov 9 2016 08:00:00 GMT
Conclusion
Framework for holistic analysis of event-induced change
- Applicable to other cryptocurrencies

I-Score facilitates separation of event-induced change from background variations
- Cross-comparisons across all configurations (temporal possible)

Applied to simple Bitcoin case study (simple fset)
Future Work

- I-Score weakness
  - More complex anomaly detection (unsupervised ML)
- Bitcoin transaction/user network topology
Acknowledgements

I would like to thank Dr. Dianxiang Xu, my research advisor for the past two years, for his guidance and support. I would also like to thank Boise High school for creating a flexible schedule to allow me to conduct research.

This research was supported in part by the National Science Foundation under grant CNS 1461133.

We would like to acknowledge high-performance computing support of the R2 compute cluster (DOI: 10.18122/B2S41H) provided by Boise State University’s Research Computing Department.

We would like to thank Blockchain.com for granting an API key for analysis and Dr. Edoardo Serra for providing his Deep Learning servers for our use.
References