Bitcoin Sentiment and Topic Analysis

Sentiment and Topic Composition as Tools for Crypto Investing

By: Michael Li

METIS®

Why Sentiment and Topic Analysis on Bitcoin?

Bitcoin is a Lucrative and Unique Investment

Bitcoin ROI:

- Best Performing Asset in History
  - 20X Return if you Invested in 2014

Digital Currency -- Price Depends On:

- Fixed Supply/Scarcity -- Blockchain and Code

- Social Acceptance/Confidence -- Legitimacy and Adoption
  - **Measured by Sentiment and Topic Composition**

Methodology

Methodology:

Data Gathering:

- Used Crypto News Websites and Aggregator APIs
  - BTC Price Data
  - 50,000 News Articles

Methodology:

Data Gathering:

- Used Crypto News Websites and Aggregator APIs
  - BTC Price Data
  - 50,000 News Articles

Sentiment Analysis:

- Utilized Textblob and Vader to generate an average composite score per doc
- Charted % Price Change and Sentiment Across Time

Methodology:

Data Gathering:

- Used Crypto News Websites and Aggregator APIs
  - BTC Price Data
  - 50,000 News Articles

Sentiment Analysis:

- Utilized Textblob and Vader to generate an average composite score per doc
- Charted % Price Change and Sentiment Across Time

Topic Modeling:

Data Preprocessing:

- Cleaned Data
- Lemmatized Text (SpaCy)
  - Tfidfvectorizer: Tokenize and Vectorize Words, placing more weight on less common words

Methodology:

Data Gathering:

- Used Crypto News Websites and Aggregator APIs
  - BTC Price Data
  - 50,000 News Articles

Sentiment Analysis:

- Utilized Textblob and Vader to generate an average composite score per doc
- Charted % Price Change and Sentiment Across Time

Topic Modeling:

Data Preprocessing:

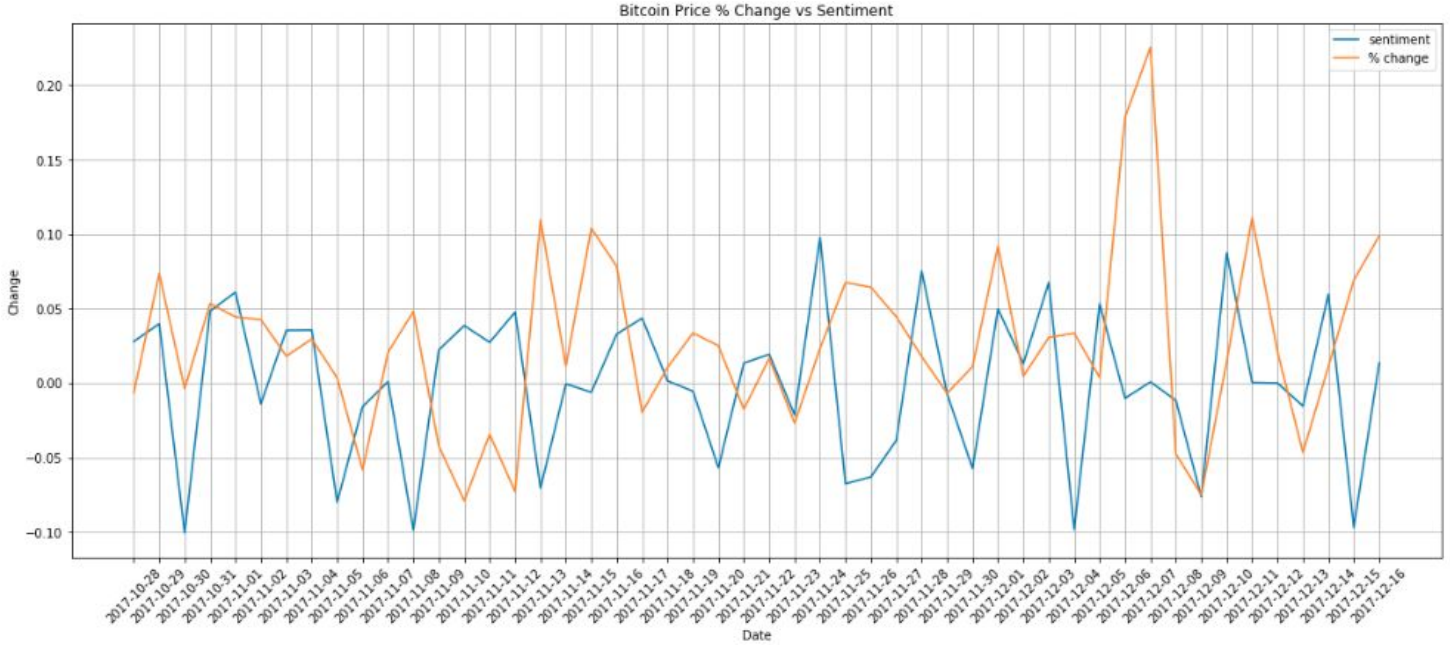
- Cleaned Data
- Lemmatized Text (SpaCy)
- Tfidfvectorizer: Tokenize and Vectorize Words, placing more weight on less common words

Dimensionality Reduction:

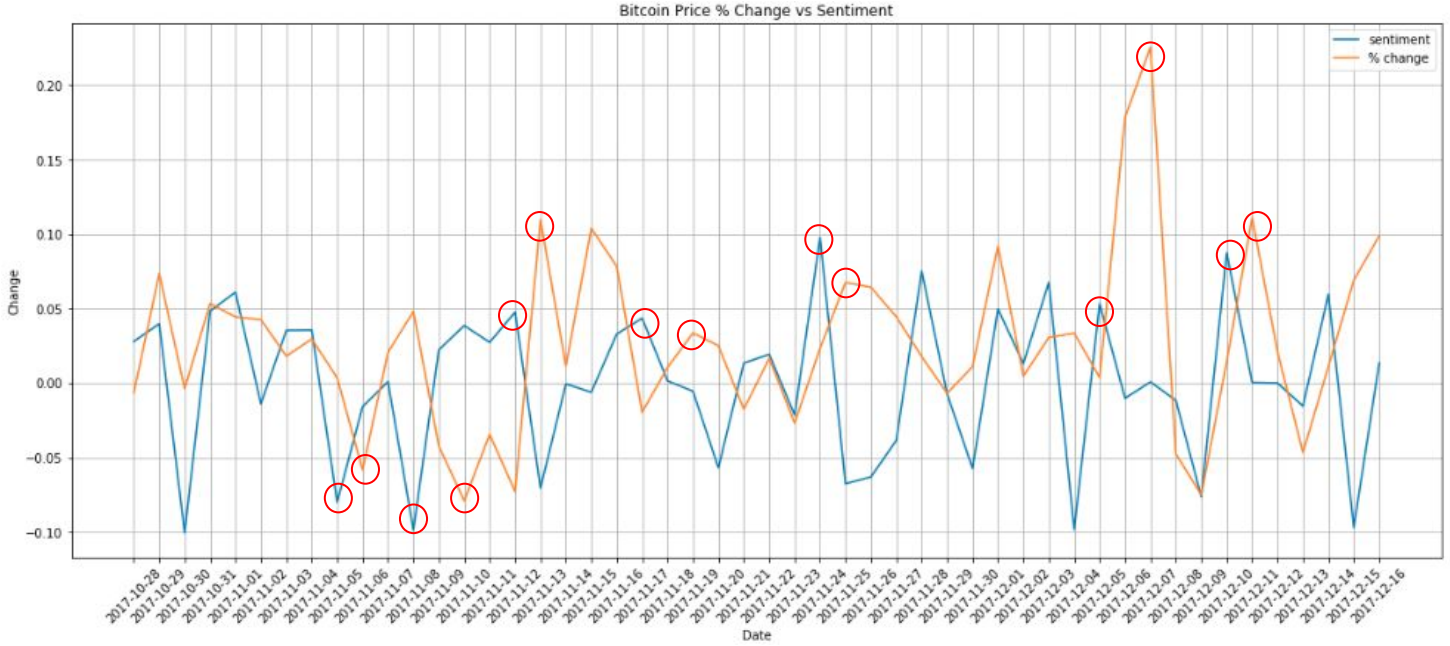
- Choose LDA: Articles can be Long
  - Adjusted Stopwords and # of Components
  - 5 Meaningful Topics Generated
  - Tracked Topic Changes and BTC Price over Time

Findings

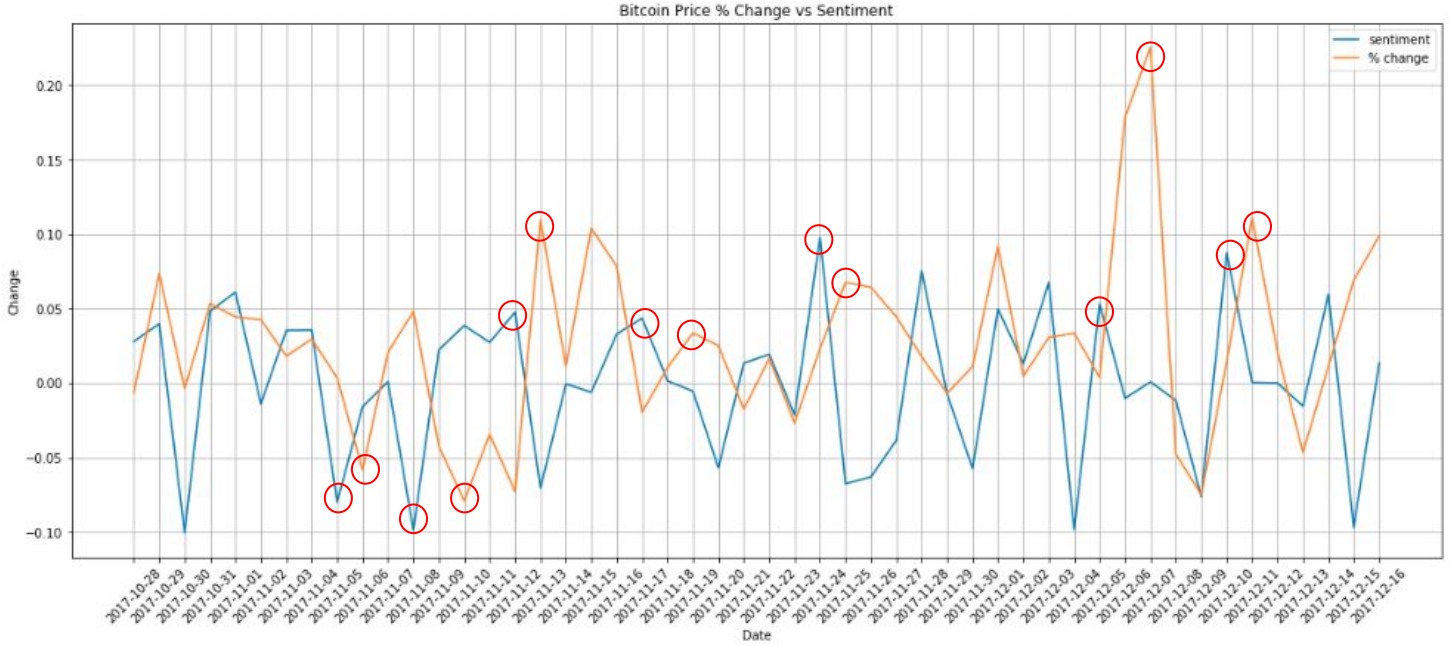
Sentiment: Leading Indicator for Short Term Trades



Sentiment: Leading Indicator for Short Term Trades



Sentiment: Leading Indicator for Short Term Trades



Not Always Reliable, but could be a part of a Robust Trading System with Technical Analysis

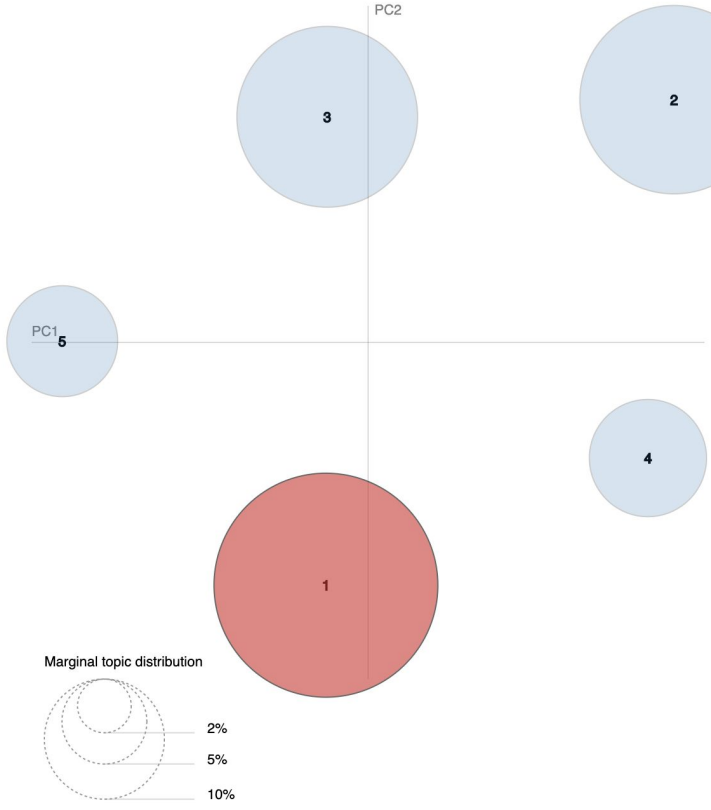

Topics could be used to Analyze **trends** for **long term** movements

LDA: Generated 5 Meaningful Topics

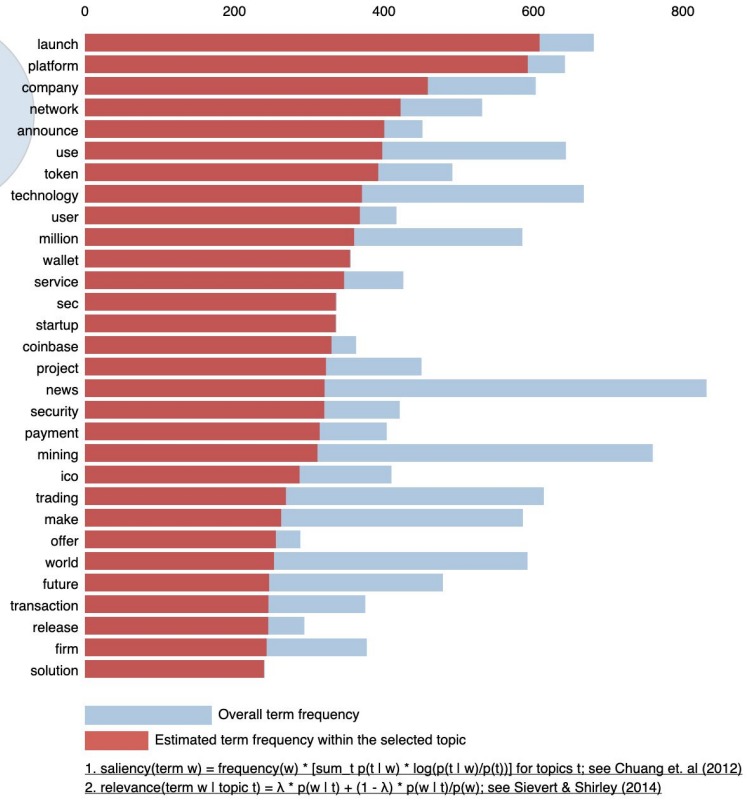
1. **FOMO**(Positive Tech Driven News)
2. **Financial** News and **Regulation**
3. **Market** Analysis
4. **Altcoins** News
5. **FUD** (National Bans, Scams, 51% Attacks)

Topic 1: FOMO (Positive Tech Driven News) Adds to Legitimacy

Intertopic Distance Map (via multidimensional scaling)

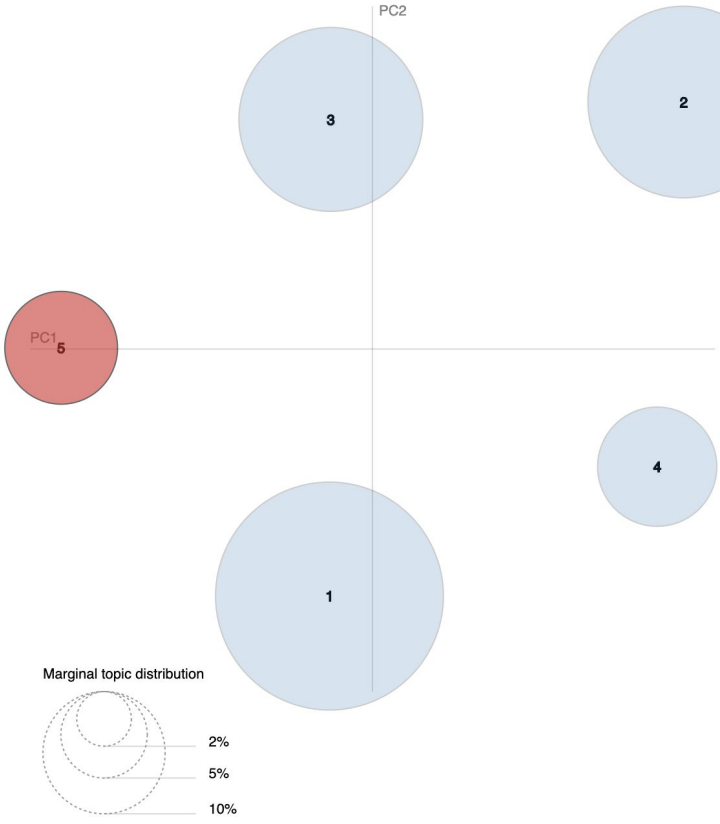


Top-30 Most Relevant Terms for Topic 1 (34.8% of tokens)

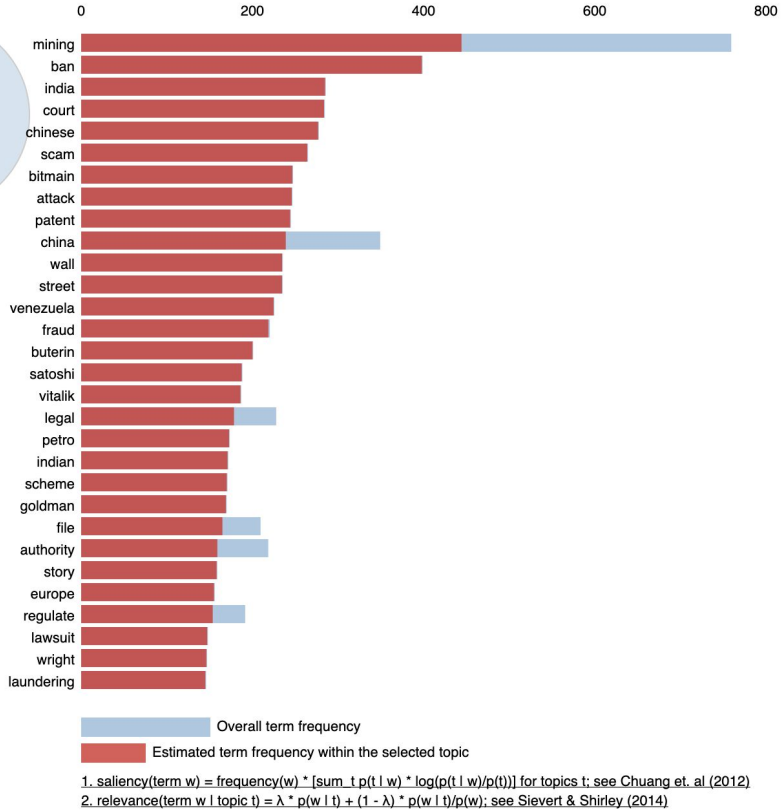


Topic 5: FUD (National Bans, Scams, 51% Attacks)

Intertopic Distance Map (via multidimensional scaling)

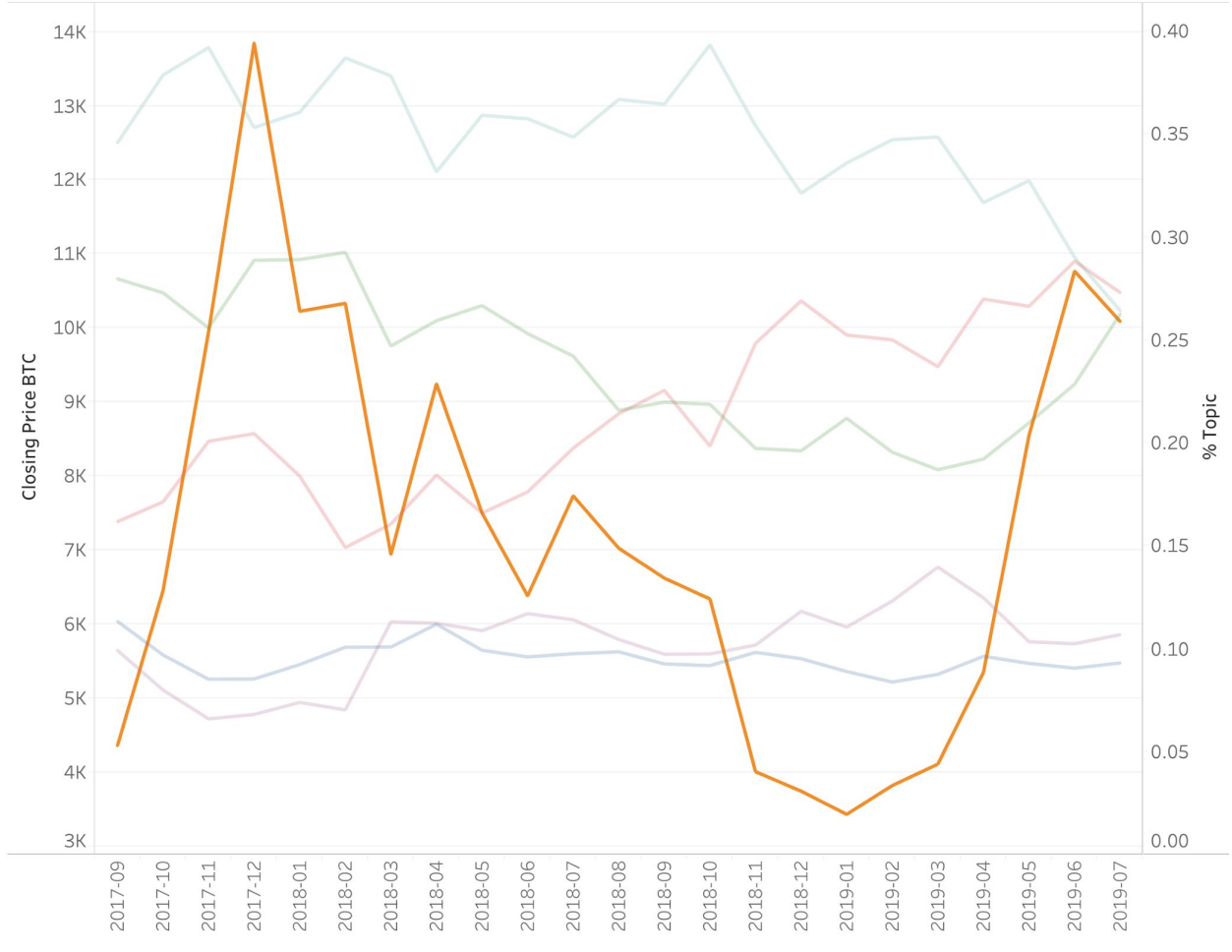


Top-30 Most Relevant Terms for Topic 5 (8.5% of tokens)

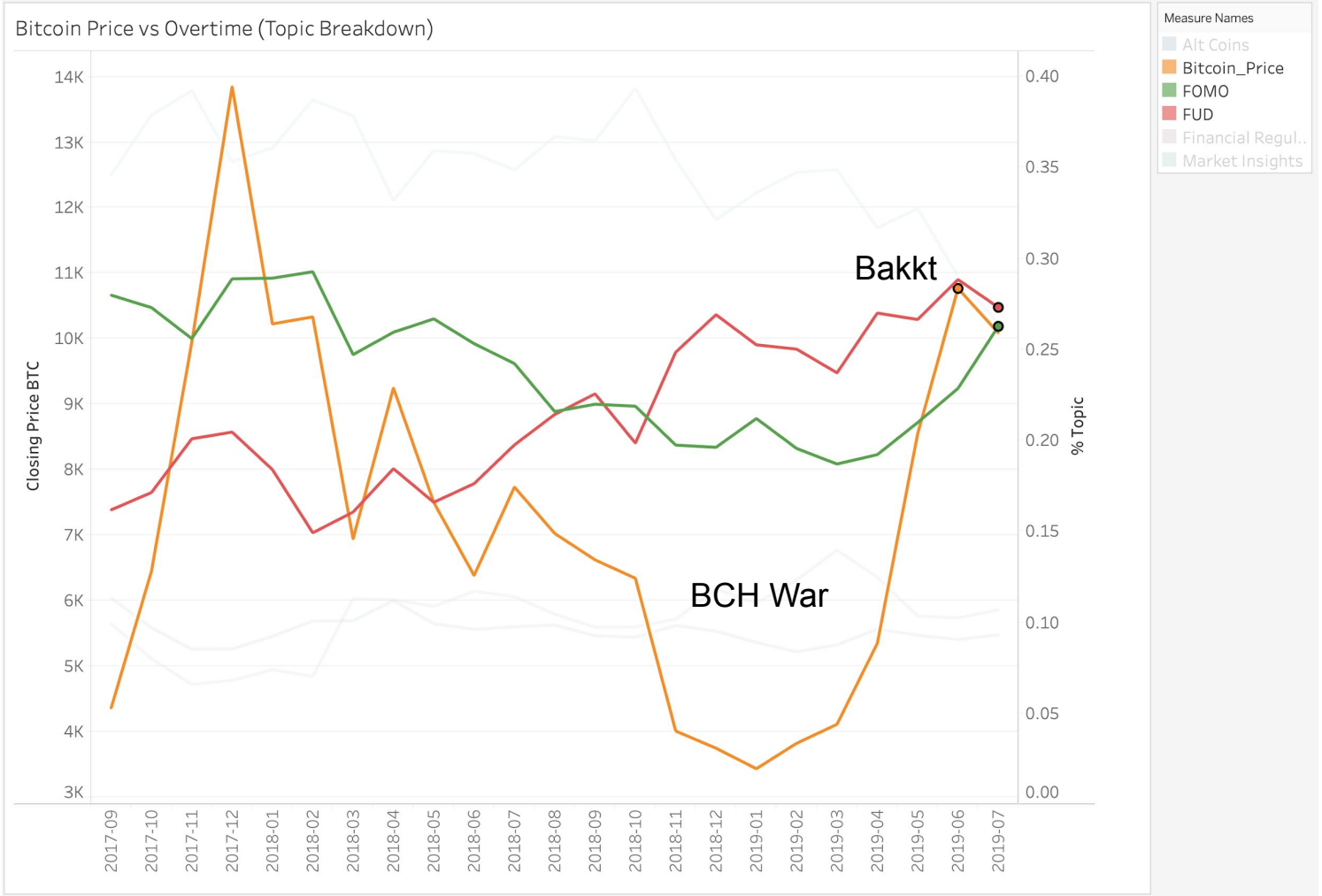


1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Bitcoin Price vs Overtime (Topic Breakdown)



Measure Names
- Alt Coins
- Bitcoin_Price
- FOMO
- FUD
- Financial Regul..
- Market Insights


Bitcoin Price vs Overtime (Topic Breakdown)
Measure Names
Alt Coins
Bitcoin_Price
FOMO
FUD
Financial Regul..
Market Insights
Closing Price BTC
% Topic
14K
13K
12K
11K
10K
9K
8K
7K
6K
5K
4K
3K
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
2017-09
2017-10
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07
Bakkt
BCH War

Bitcoin Price vs Overtime (Topic Breakdown)

Measure Names
- Alt Coins
- Bitcoin_Price
- FOMO
- FUD
- Financial Regul..
- Market Insights

Closing Price BTC

14K
13K
12K
11K
10K
9K
8K
7K
6K
5K
4K
3K

% Topic

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Bakkt

BCH War

2017-09
2017-10
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

Bitcoin Price vs Overtime (Topic Breakdown)

Measure Names
- Alt Coins
- Bitcoin_Price
- FOMO
- FUD
- Financial Regul..
- Market Insights

Closing Price BTC

14K
13K
12K
11K
10K
9K
8K
7K
6K
5K
4K
3K

% Topic

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Bakkt

BCH War

2017-09
2017-10
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

Bitcoin Price vs Overtime (Topic Breakdown)

Measure Names
- Alt Coins
- Bitcoin_Price
- FOMO
- FUD
- Financial Regul..
- Market Insights

Closing Price BTC

14K
13K
12K
11K
10K
9K
8K
7K
6K
5K
4K
3K

% Topic

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Bakkt

BCH War

2017-09
2017-10
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

Bitcoin Price vs Overtime (Topic Breakdown)

Measure Names
- Alt Coins
- Bitcoin_Price
- FOMO
- FUD
- Financial Regul..
- Market Insights

Closing Price BTC

14K
13K
12K
11K
10K
9K
8K
7K
6K
5K
4K
3K

% Topic

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Bakkt

BCH War

2017-09
2017-10
2017-11
2017-12
2018-01
2018-02
2018-03
2018-04
2018-05
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07

Takeaways:

- Sentiment can be used to predict short-term movements
- Proportion of Topics of News Articles can indicate overall climate of the market - **Spread of FUD and FOMO**
- Currently BTC is at a pivotal point where FOMO and FUD converge
- Future Work: Incorporate Sentiment and Topics over time into a Trading Algo with TA indicators

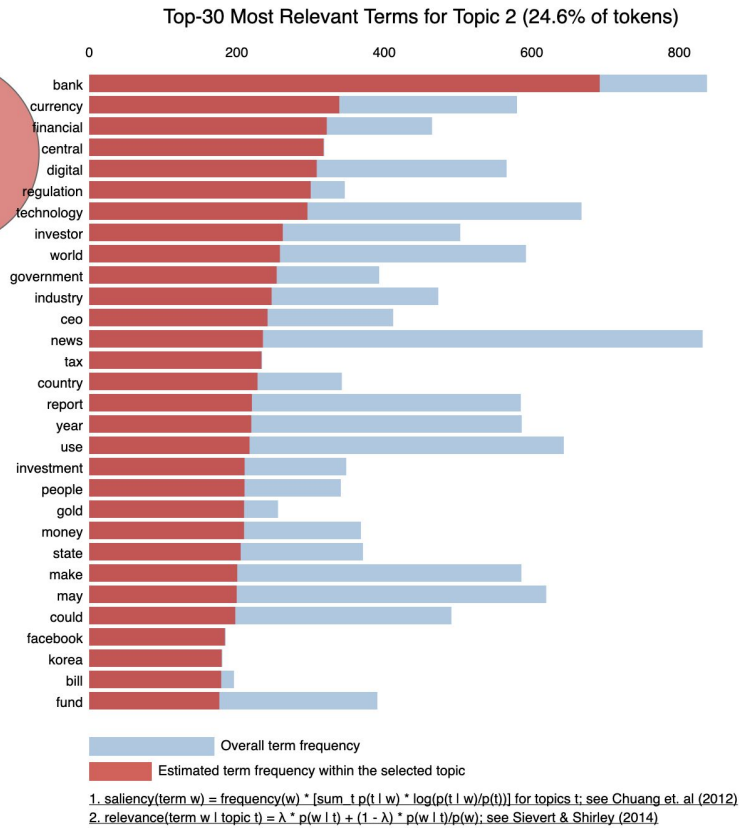
Questions?

Appendix

What is Bitcoin?

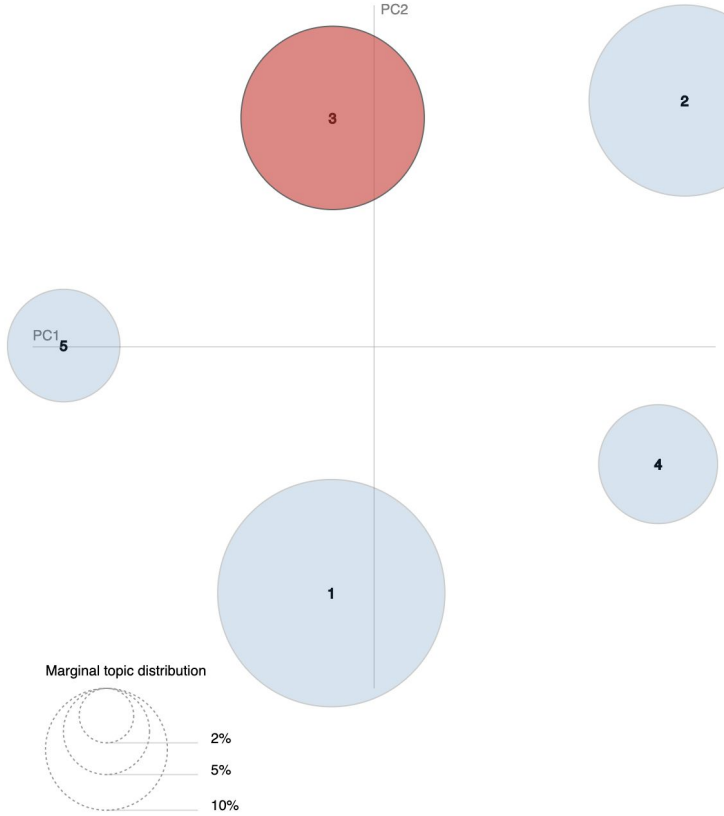
- Decentralized Digital Currency
- Created by Satoshi Nakamoto in 2008
  - Financial Crisis
  - Alternative to Bank Controlled Currencies
- Reward from a process called mining
  - Energy and Computing Power intensive process
- Transactions
  - Recorded on a public distributed ledger (Blockchain)
  - Verified by peer to peer network of computers
    - Rather than a centralized entity

Topic 2: Financial News and Regulation

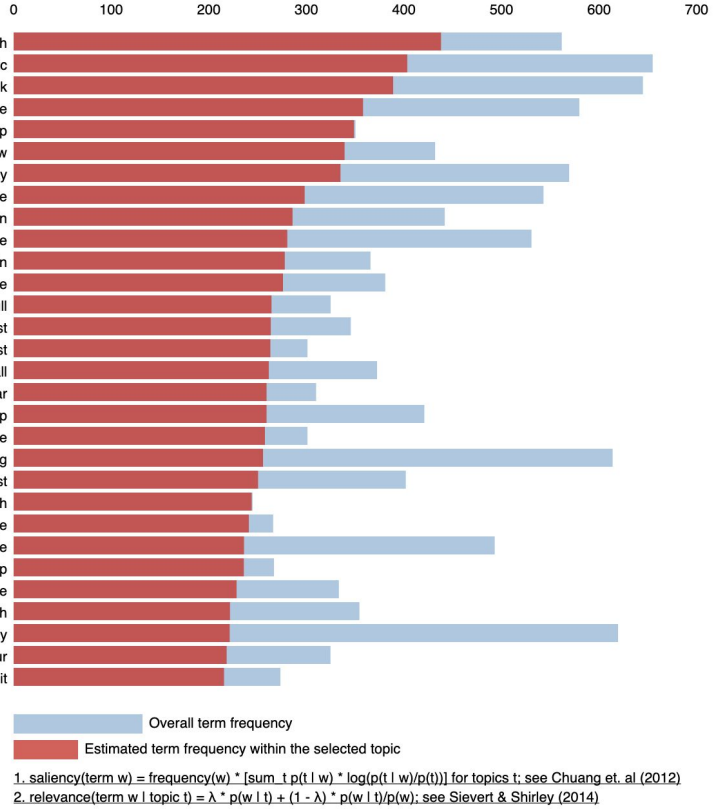


Topic 3: Market Analysis

Intertopic Distance Map (via multidimensional scaling)

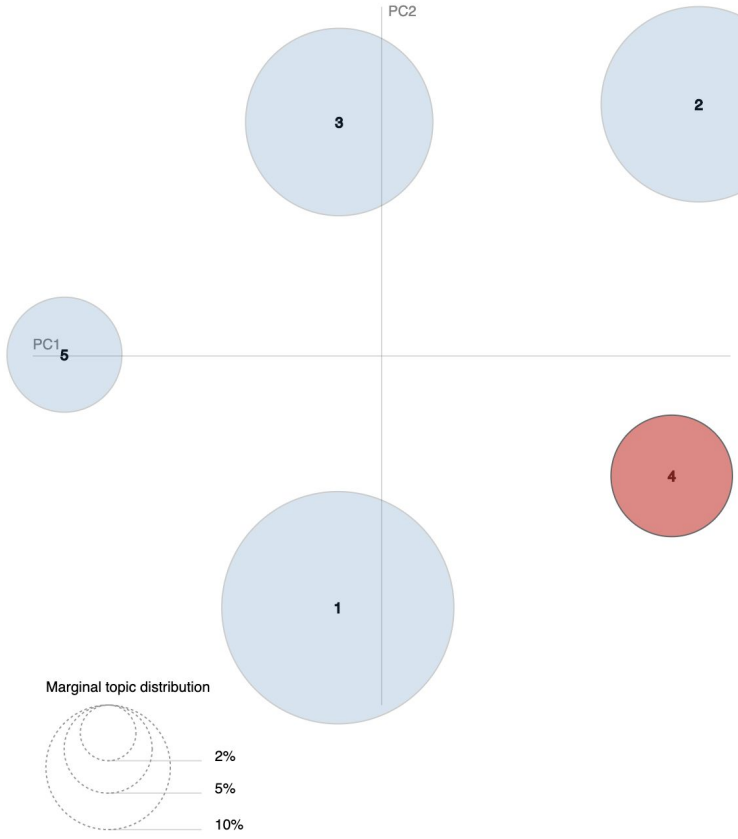


Top-30 Most Relevant Terms for Topic 3 (22.6% of tokens)

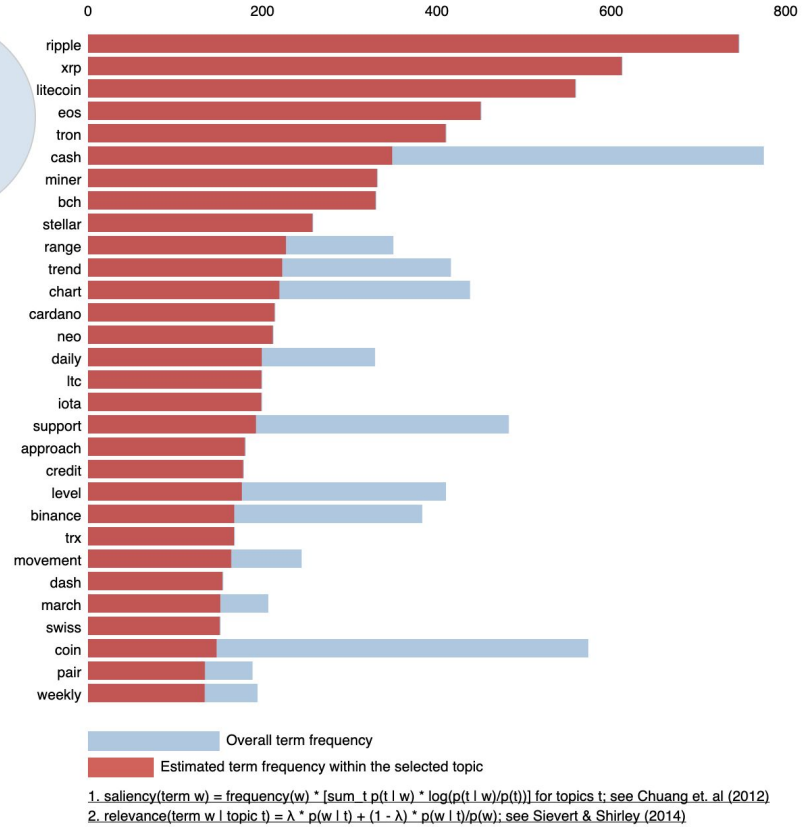


Topic 4: Altcoins News

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (9.5% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)