

EFFECT OF SOCIAL MEDIA SENTIMENT ON THE PRICE OF BITCOIN

by

Ngoc Tu

Submitted in partial fulfillment of the  
requirements for Departmental Honors in  
the Department of Information Systems and Supply Chain Management

Texas Christian University

Fort Worth, Texas

May 7, 2018

# EFFECT OF SOCIAL MEDIA SENTIMENT ON THE PRICE OF BITCOIN

Project Approved:

Supervising Professor: David Preston, Ph.D.

Department of Information Systems and Supply Chain Management

Antonio Sanchez, Ph.D.

Department of Computer Science

## ABSTRACT

Bitcoin started in 2009 as a peer-to-peer value transfer system. It quickly gained significant value against the dollar and became a traded asset. However, because of its age and its origin in the dark corners of the Internet, Bitcoin does not have formal communication channels where investors discuss its value. For example, a stock relies on its company for official announcement about the development of the company; investors could use this information to give valuation to and make decision to sell or buy the asset. Bitcoin does not have a formal channel because there is no entity behind it.

For that reasons, it is necessary to study the channels that people use to communicate about Bitcoin. I studied the conversations on Twitter and Reddit in the days leading up to and the days following the point of all-time-high price of Bitcoin. I preferred Twitter because its data are automatically labeled as either bullish or bearish, but a restriction on the amount of data I could get access turned to Reddit instead.

To study the sentiment hidden in the messages of Reddit, I wrote customary code to extract the most common words in the period of study. I planned on using them to construct a model that could automatically assign a sentiment to any Bitcoin conversation. Unfortunately, I discovered that conversations on Reddit are extremely specialized for Reddit users. Words on this website usually confer a different meaning from its regular dictionary meaning. To complete the model to truly extract the sentiment, there needs to be human effort put into manually labelling what sentiment the vocabulary on Reddit means. That next step would complete this thesis and give a more definitive answer to how informal and untraditional channels influence the price of an equally untraditional asset.

## Contents

Introduction.....	5
Literature Review.....	7
Supervised learning.....	16
Unsupervised learning.....	16
Research methodology.....	17
Period of time under study .....	18
Building code.....	19
Results.....	22
Discussion.....	25
Conclusion .....	26
Bibliography .....	27

## Introduction

On November 26, 2017, Bitcoin's price surpassed the \$9,000 mark for the first time. The breach prompted attention from established news sources like Fortune and Forbes, technology-focused outlets like TechCrunch and CNET, and general discussion forums like Reddit. Specifically, on Reddit, a large number of users flooded the subreddit /r/bitcoin, the primary channel for all discussion about Bitcoin, with memes relating to the price. Most memes make use of the character Vegeta saying "It's over 9000", referring to a moment in the anime Dragon Ball Z where Vegeta expressed disbelief about the power level of his nemesis.

Indeed, it is quite hard to believe the astronomical growth of Bitcoin and cryptocurrencies: the market cap of all cryptocurrencies rocketed from over \$13 billion in November 2016 to over \$300 billion in November 2017. This massive expansion followed the movement in Bitcoin price, whose market cap increased from \$11 billion to \$160 billion in the same time period.

How does Bitcoin, the very first cryptocurrency whose use case was confined to drug markets on the dark web and whose life has been less than ten years, achieve such substantial growth over such a short period of time? Stocks grow because the companies behind them grow. Bitcoin, by its very design, represents a decentralized network with no single entity taking the main role in representing it. It is more curious, then, how the exchange unit of a distributed network of machines running a common software could rise in value so quickly. Bitcoin's price is determined by the simple law of supply and demand. An investor would point to the rising demand of Bitcoin as a reason for its growth. A wise investor would try to capture when demand and subsequently price to buy and when the reverse situation happens to sell.

Financial analysts have been making forecasts about the stock market since the stock market inception. Now, with a wealth of data reflecting the market sentiment towards certain companies and a wide array of tools to extract, analyze, and aggregate readily available data, they can make increasingly accurate information about market sentiment. Market sentiment is useful for predicting how the market feels about certain companies and stocks. If the general sentiment is bullish, price will most likely rise. In contrast, bearish feeling could drive the price down.

Bitcoin is almost designed for market sentiment analysis. Having no central authority, it has no centralized or official channel of information about what will happen to it. Consequently, Twitter and Reddit became the de facto channels of announcement and communication for big players in the Bitcoin community. Most Bitcoin online news sources get their information from tweets generated by Bitcoin service companies, Bitcoin mining companies, and Bitcoin developers. I want to study the effect of these social media platforms as the de facto communication channels of Bitcoin on the cryptocurrency's price.

To achieve this, I categorized Bitcoin as a trading asset similar to stocks. I wrote customary code to extract the general sentiment from social media data for juxtaposition against the movement of the price of Bitcoin. I will then draw conclusion on the correlation between what people are talking about Bitcoin online and the fluctuation of its price.

## Literature Review

Predecessors to Bitcoin existed. Nick Szabo suggested bit gold and Wei Dai suggested b-money, both of which aimed to battle shortcomings of traditional fiat money (Szabo; Dai). Theoretical in nature, they inspired the design of Bitcoin.

The creation of Bitcoin was largely informal. The domain name bitcoin.org was registered on 18 August 2008 (Kelleher; Whois Lookup). The identity of who registered the domain name is kept secret. On October 31, 2008, Satoshi Nakamoto made an announcement on the cryptography mailing list at metzdowd.com about his work - "a new electronic cash system that's fully peer-to-peer, with no trusted third party". No one knew who Satoshi was, and no one does today. All we know is that Satoshi mined the first Bitcoins, released the first version of Bitcoin, and created the first transaction between himself and Hal Finney, a developer and cryptographic activist (History of Bitcoin). On October 12, 2009, the first Bitcoin communication channel was #bitcoin-dev channel on freenode Internet Relay Chat, a discussion network for free and open-source community (History of Bitcoin). On November 22, 2009, Satoshi wrote a welcome message in bitcointalk.org, a Bitcoin discussion forum intended to replace the discussion on sourceforge.net where the Bitcoin project started (satoshi). From this point, communication about Bitcoin exploded.

Designed to be a payment system, Bitcoin needed a value estimation. New Liberty Standard is a website that took the liberty to calculate the worth of Bitcoin by estimating the cost of electricity used in Bitcoin mining (2009 Exchange Rate). At its first point of estimation, Bitcoin was worth roughly \$0.00076 (or \$1 was worth 1,309 Bitcoin). However, Bitcoin did not have real life application. In the early days, Bitcoin enthusiasts made various attempts to bring Bitcoin to life.

The legendary story of Bitcoin's first real world application where someone paid 10,000 BTC for two pizzas still widely circulates Internet discussion today (Price).

Bitcoin did not really pick up until the online drug exchange Silk Road was created (Chen). The creator of Silk Road, Ross Ulbricht, wanted an unregulated and unsupervised marketplace where illegal drugs could freely exchange between anonymous users (O'Neill). Bitcoin provided the solution: it offered an adequate level of anonymity and could be converted to fiat currencies. The practical use case increased the value of Bitcoin, raising its price to be on parity with the US Dollar (CoinDesk). Bitcoin price continued to fluctuate, eventually hitting and surpassing the \$100 market in 2013, three years after it came into existence.

To understand why and how Bitcoin got its value, we need to look at traditional asset and currency and how and why they got their values.

As civilization grew and started producing surplus, humans began trading. Barter was the first system in used where goods were directly exchanged for other goods. As surplus increased, however, so did the need for a standardized system for trading and a means of measurement. Governments issued currencies to impose a default system for merchants to trade commodities against. Some civilizations used rare metals while others used shells or cocoa beans (Burn-Callander). China issued paper currency, which quickly spread to Europe. Initially, currencies could be redeemed for gold or silver under the gold standard, which allowed anyone to exchange paper money for gold (Amadeo). The Bretton Woods system replaced gold with the US Dollar because the US held three-fourths of the world's supply of gold while still allowed other countries to covert the US Dollar to gold (Amadeo). President Nixon suspended the ability to convert in 1971, giving currencies value when comparing against one another (Encyclopædia

Britannica). As such, fiat currency's price complies with the supply and demand in the market: the more currency there is, the lower its price.

Gold, once a currency, now most often trades as an asset. Early civilizations found gold fascinating and useful, adopting it as both ornaments and currency. Through the gold standard and the Bretton Woods system, gold backed the value of currency by allowing people to exchange currency for gold. Specifically, gold was traded at a fixed rate of \$35 per ounce. After the Nixon Shock in 1971, the US government refused to redeem dollars for gold because money printing had raised the price of gold in the free market beyond the fixed rate (Hur). Gold continued to be a valuable asset, a safe-haven store of value through time of turmoil as it is not reliant on a centralized entity like a government (Money Metals Exchange). The price of gold fluctuates but is generally high because of a few reasons, as listed by bebusinessed.com (Hur).

- Scarcity: gold is hard to find and to extract in the real world.
- Physical characteristics: gold is an excellent conductor, highly malleable and highly ductile
- Aesthetic attributes: early civilizations used gold extensively for their rulers' construction. Gold remains a desirable material for jewelry today.
- Wealth storage: gold's increase in value has always coincided with poor economic situation. As people lose faith in their country's currencies, they buy gold to retain the value of their wealth.

The last point is but a reason that influences the price of gold. A trading asset, gold relies on other factors to determine its value. The Motley Fool identifies the following as common factors that have an impact on the price of gold:

- Monetary policy: interest rates can greatly influence gold. A higher interest rate on bonds compel investors to buy bonds instead of gold because they can get a higher return. If the interest rate goes so low that the inflation rate beats the interest rate, investors would switch to gold to protect the value of their asset.
- Economic data: in general, a strong economy drives down the price of gold as investors can spend their money to make profit out of a growing economy.
- Supply and demand: if demand grows at a faster rate than supply does, the price of gold will increase.
- Inflation: inflation is the rise of price of goods and services or the loss of value of the currency. Inflation usually correlates to a booming economy where the money supply expands. The higher number of currency, the lower its value. Generally, high level of inflation tends to push the price of gold higher.
- Currency movement: the US Dollar and gold tends to have an inverse relationship: a weak dollar is synonymous with rising gold price.
- Electronic-traded funds (ETFs): Motley Fool defined ETFs as "basket funds investors can purchase that allow for increased liquidity and the potential ability to spread their risks over a large number of assets for a minimal cost." As demand for gold changes, ETFs reflect the movement and thus affect the price of gold (Williams).

Gold relies largely on governmental regulation and the overall state of the economy for its value. Equities, or stocks, however, relies on more than just general information. Stock is defined as a share in the ownership of a company. Owning stocks means a person actually owns a part of the company. Stocks give owners the right to dividend, which is a portion of profit the company pays to its shareholders. Private companies issue stocks for the public to buy when they need to

raise money. Afterwards, the public could vote on company decisions to determine which direction to steer the company forward (HowTheMarketWorks). There are many factors that contribute to the price of stock. Investopedia defines two fundamental factors:

- Earning base: A prime example is the earnings per share (EPS) figure. Common stock gives a person a claim on earnings. A stock purchase is a purchase of a share of future earnings.
- Valuation multiple: A prime example is the Price-Earnings (P/E) ratio. The P/E ratio measures its current share price relative to its per-share earnings. The valuation multiple expresses expectations about the future.

Beyond fundamental factors, there are technical factors that make impacts on stock prices as well. Investopedia defined the following factors:

- Inflation: low inflation has an inverse correlation with valuations. Low inflation drives high multiples, which give stock higher value.
- Economic strength of market and peers: stocks rise and fall in sync with the market and with their sector and peers. Some argue that the combination of overall market and sector movements decide the majority of a stock's movement instead of the stock's internal factors.
- Substitutes: stocks compete with other asset classes. Investors might choose to buy bonds or foreign currencies if they believe they could earn better profits that way.
- Incidental transaction: incidental transactions are purchases of stock motivated by something other than belief in the intrinsic value of the stock. These transactions affect supply and demand and thus have an impact on the price.

- Demographics: age demographics has a specific implication on the stock market. Middle-aged investors who are peak earners tend to invest in stocks. Older investors, in contrast, are more likely to pull out to meet demands of retirement. The larger the former group, the larger the demand for equity, the higher the stock price.
- Trends: a stock can gain momentum and ride on its previous success to higher valuation. However, the opposite situation can also happen leading to an undervaluation of a stock.
- Liquidity: liquidity refers to how much investor interest and attention a specific stock has. Highly liquid stocks are highly responsive to material news – news released by a company that can strongly influence the perception of its stocks or investors' decisions.

Another factor that recently came into the picture as a big influencer of stock price is market sentiment. The Oxford Reference defines market sentiment as "An optimistic feeling that will push up the general level of prices on a financial market, commodity market, etc., or a pessimistic view that will depress the general level of prices. Financial market sentiment can be triggered by bad results or profit warnings from leading companies, apprehension regarding government policy, interest rates and general economic data, or market rises or falls on other exchanges" (Law). Market sentiment is being explored in the new field of study – behavioral finance. Oxford Reference defines it: "The study of the role played by psychological factors in financial decision making and hence their effect on overall market outcomes. In particular, behavioral finance studies the ways in which individual and group behavior deviates from the rational pursuit of self-interest posited by classical economic theory" (Law).

Market sentiment is useful in forecasting where the market is going next. The Wall Street Journal considered the extremes of market sentiment turning points of the market: if everyone has sold off their stocks then stock price can only go up (Constable). Market sentiment might be so

powerful that short-term investments might rely almost exclusively on whether people feel good or bad about investing instead of looking at fundamental criteria like earnings, sales, inflation and interest rates (Andrews). Research has shown relationship between sentiment and the price of the stock market. Some researcher even used sentiment to predict Dow Jones performance (Liu).

With the advent of a wealth of information readily generated by the users on social media platforms, sentiment analysis evolved from text mining and natural language processing to determine sentiment on a specific topic, especially sentiments from unstructured human-authored documents (Yu, Duan, and Cao). Sentiment analysis aims to compute polarity, which is the degree to which a document is negative, neutral, or positive. Polarity of opinion about certain asset is useful because it narrows down the overall feeling on social media about stock performance to two comprehensive extremes: bull or bear. A study showed that sentiment on both social media and conventional media have a strong impact on stock performance; however, news on social media tend to have stronger influence on the movement of stocks (Yu, Duan, and Cao).

Bitcoin does not have an account on social media that announces news about its activities. Instead, due to its decentralized design, many have power over Bitcoin. Miners, developers, governments, exchanges, individual users with large amount of Bitcoin, and media coverage can all influence the market (Callahan). Each plays a different role in determining the next step Bitcoin will take and each reaches a different group of the Bitcoin community. The amount of influence each exerts on the market is a measurement still to be decided.

This paper aims to locate the news surrounding a sharp price drop or increase. From the news, I want to identify the actors that the news focused on and the relative sentiment associated with

such news. The correlation or lack of correlation between market sentiment and price could support a relationship between how much influence an actor has on Bitcoin. For example, if a government decided to shut down Bitcoin and the bearish news surrounding the decision dominated the conversation about Bitcoin, the price is expected to tumble. If it does, there is a correlation. If it does not, that government's decision fails to change how the market trades or other minor influencers have a much greater effect.

Conversation around Bitcoin does not gravitate towards any single social media platform. As mentioned, the number and diversity of players in the Bitcoin market require a wide range of platforms for different needs. The first online forum for Bitcoin on the code sharing site SourceForge is now lost. BitcoinTalk forum replaced it in November 2009 and remains today as an active online discussion platform for the Bitcoin community (Bitcoin Wiki). However, BitcoinTalk does not have an extensive reach to audience outside of developers and early adopters and offers limited ability to query data from all conversations surrounding the cryptocurrency. For this paper which focuses on market sentiment, I needed a different platform that 1) could be easily queried and have data available for extraction and 2) cover a diverse user base with different opinions and experiences.

I started my research by looking at Twitter data. Many large players in the Bitcoin community have Twitter accounts with which they provide updates on their activities. For example, GDAX/Coinbase, a top ten exchange by trading volume, has more than 875,000 followers on Twitter (@coinbase; CoinMarketCap). Twitter provides an Application Programming Interface (API) for developers to request data from the platform (Twitter). The wealth of information and the ease to retrieve it made Twitter an attractive option to begin research.

The Twitter API has serious limitations, however. A developer can only request data seven days from the date of request. I did not realize this limit exists until after determining that I would choose Twitter as my platform of choice. I had to choose a different platform that reflects market sentiment regard Bitcoin. One of the most notable is a subreddit on Reddit intuitively called "Bitcoin" in 2012 (Reddit Metrics).

Reddit is a social networking site with numerous sub-communities called subreddits that functions on members' content and curation (Oxford University Press). Most recently, it received \$200 million in funding and a valuation of nearly \$1.8 billion - the result of strong and continuous user base growth (Huddleston). The Bitcoin subreddit grew with the site. In February 2018, it has over 700,000 subscribers (Reddit). As such, it aggregates much information about Bitcoin. Updates on Bitcoin include subreddit member generated content such as personal anecdotes, investment guides, opinions, memes and links to news about Bitcoin from other sites. Because of the wealth of the information available on Reddit, I choose the social networking site to gather data and perform sentiment analysis on the data.

To decide whether a piece of social media is bullish or bearish, Sarkar suggests giving the piece a score based on its polarized level followed by a corresponding label (Sarkar). To understand how this is achieved, we need to understand that sentiment analysis is a usage of text classification. Text classification categorizes any textual data in various media into common groups such as content-based groups (same topic) or request-based groups (targeted towards users' request of how data should be grouped together) (Sarkar). Sentiment analysis of social media on asset categorizes data into two groups: bull or bear. There are two major methods of text classification.

### Supervised learning

Supervised learning, according to *Text Analytics with Python*, "refers to specific machine learning techniques or algorithms that are trained on pre-labelled data samples known as training data" (Sarker). There are two outcomes for supervised learning: classification, a process which provides outcome of categories, and regression, the result of which is a prediction on continuous numeric variable (Sarker).

Within the scope of this paper, supervised learning could be applied to predict the category of social media data. From a pre-labelled training set of data which consists of data assigned to either bull or bear category, a model is built to predict the bullish and bearish level of the entire data set.

### Unsupervised learning

*Text Analytics with Python* defines unsupervised learning as "specific machine learning techniques or algorithms that do not require any pre-labelled training data samples to build a model" (Sarker). Within the scope of this paper, unsupervised learning could be used to group together social media data with a similar level of sentiment.

With the two mentioned social media platforms and a sentiment analysis technique, I can begin extracting sentiment from social media and comparing it against the price of Bitcoin. I want to write code customized to this purpose because 1) no readily available tools serve the scope of this project and 2) it is easier to add features to my code to accommodate the final goal of the project

## Research methodology

I choose Python as my language of choice to customize coding programs. There is a diverse community of developers who actively create various Python libraries that serve every conceivable purpose. To scrape data from Twitter/Reddit and to get the pricing information, I would be using my own Python codes and would import external libraries to extract the most out of available online data.

Because I used Twitter API extensively before discovering their rate limit, I would credit the Twitter tools that I used. I utilized these libraries to aid my research:

- TwitterSearch library from GitHub: the author to the library describes the function: "This library allows you easily create a search through the Twitter API without having to know too much about the API details. Based on such a search you can even iterate throughout all tweets reachable via the Twitter Search API" (TwitterSearch). After issuing a call to the Twitter API, the user would receive a JSON file that stores all the tweets in the JSON format (see fig. 1).

```
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': '<a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>',
'text': '@joshdallas @ginnygoodwin home during wintertime@ThatsEarth: Hohenzollern Castle f',
'truncated': False,
'user': {'contributors_enabled': False,
'created_at': 'Fri Aug 14 09:15:27 +0000 2009',
'default_profile': False,
'default_profile_image': False,
'description': 'Scorpio. 23. MBA Graduate.',
'entities': {'description': {'urls': []},
'url': {'urls': [{'display_url': 'fanfiction.net/u/4764512/',
'expanded_url': 'http://www.fanfiction.net/u/4764512',
'indices': [0,
22],
'url': 'http://t.co/sEKQ1M85H2'}]}}
```

Fig. 1. TwitterSearch JSON result

- Python Reddit API Wrapper (PRAW): the GitHub page of PRAW describes it as a python package that allows for simple access to Reddit's API. PRAW aims to be easy to

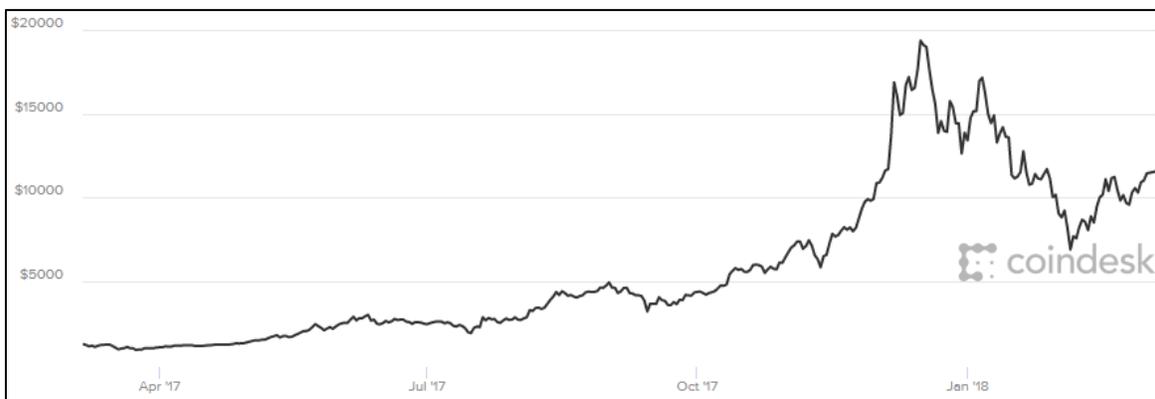
use and internally follows all of Reddit's API rules" (praw-dev). With PRAW, a developer can post content to Reddit or request data about certain subreddits. Unlike Twitter API, Reddit does not limit the time period of the requested content.

- Coindesk API: Coindesk API aids financial data gathering. Their website describes how they calculate the price of Bitcoin: "CoinDesk Bitcoin Price Index (XBP) represents an average of bitcoin prices across leading global exchanges that meet criteria specified by the XBP" (CoinDesk). A simple call to their API could return data about Bitcoin price in JSON format. It is also possible to sort data by date or by price.

### Period of time under study

Ideally, I want to compare social media sentiment and Bitcoin price over a long period of time.

However, I encountered three limitations: 1) the social media APIs have limits on the amount of data that can be called, 2) the large amount of data returned by social media and my inability to store them and query them all, and 3) my lack of experience in machine learning. Because of these reasons, I decided to choose a specific time of significant price change and the movement of social media sentiment leading up and following that change. Specifically, I want to examine the price peak of 17 December 2017 and the social media conversation around it.



The price chart is provided by CoinDesk's Bitcoin Price Index (CoinDesk). According the price chart, on December 17<sup>th</sup>, Bitcoin hit an all-time high (ATH) price. The peak signifies a twenty time increase in value from the beginning of 2017. However, Bitcoin did not quite reach the \$20,000 mark when it began to lose its value on the same day of trading. The downward trend continued to the point of one third the all-time high. I want to juxtapose the sentiment on the subreddit Bitcoin on the way to its peak and on the way down.

### Building code

My code needs to accomplish two goals: 1) to retrieve data confined by the period of interest and by relevance to Bitcoin and 2) to extract useful information from this data.

My interaction with social media platforms takes the form of receiving JSON files. Json.org defines JSON as "a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate" (JSON). Twitter API and PRAW returns JSON files with data limited by the parameters of the query passed to them (see fig. 1).

Twitter returns useful information about the content of the Tweet, about the user, and other related information regarding time, location, and relationship to other entities on Twitter.

Twitter API allows users to access all tweets that mention a certain phrase. For example, a call to the API requesting all tweets with the word "bitcoin" would return all conversations that has the word "bitcoin" from all accounts from the most recent to the least topping off at the API limit call which is 1500 tweets. I initially set the parameter to capture all conversations with the word "bitcoin". Another restriction I put on my queries was time. The code needed to capture information from only the mentioned timeframe. Initial results only returned recent data. I discovered that Twitter API only allowed data query of up to seven days from the point of query.

An alternative solution is to capture the conversation from leading accounts that discuss Bitcoin and cryptocurrencies. The reason follows that their large number of followers would be enough to steer market sentiment. For example, Coinbase, a popular US exchange, has more a large customer base in the US which is reflected in their 800,000 followers on Twitter. I tweaked my code to capture the timeline of major cryptocurrency service companies, the top crypto news sources, and the top cryptocurrency evangelists. The problem with this approach arose when I found out that these top accounts did not produce much content at all. Cryptocurrency exchanges and cryptocurrency service providers only went on Twitter to announce the release of a new feature or a fix of a bug, which is not inherently bullish or bearish. The conversation that followed, in the form of comments on the original Tweets, provided better indication of sentiment, but extracting these comments was impossible through Twitter API. Cryptocurrency news sources on Twitter produced more content but had a strong bullish skew on the market with no healthy skepticism of Bitcoin.

These factors meant that I could not use Twitter data for my research purpose.

At this point, I turned to other sources for market sentiment. As mentioned, Reddit is a social network with diverse audience who discuss Bitcoin regularly. Reddit has a much less convenient API, but it offered data in the right time range. In addition, there is a special community dedicated to conversation about Bitcoin. I limit my data query to extracting information specifically from this community.

On Reddit, a new discussion topic is called a thread. A thread has a title and a body. Users participate in the conversation by making comments on these threads. Initially, I extracted only the threads of the subreddit "bitcoin". Popular threads can have thousand comments while less influential ones might have none. For that reason, a popular comment on a popular thread might

have more influence on the market than a non-popular thread. Initially, I started with extracting only the titles of threads and the number of threads started each day to identify a general sense of the conversation about Bitcoin. To run the data through Natural Language Processing (NLP) code, I cleaned data for processing by excluding punctuation, English stop words like personal pronouns and prepositions, and numbers.

## Results

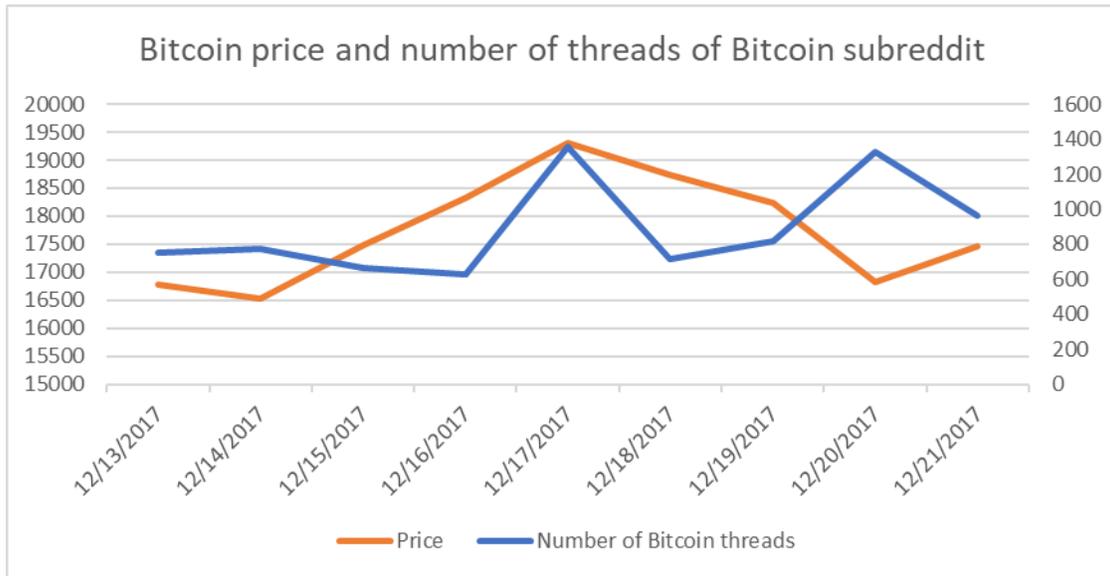


Fig 2: Bitcoin price and number of threads of Bitcoin subreddit

There's a spike in the number of conversation when the price hit all-time-high on the 17<sup>th</sup> and when price dipped on the 20<sup>th</sup>.

The top 20 words for each of the dates in the study period are as following. Column F stands for Frequency.

12/13/2017	F	12/14/2017	F	12/15/2017	F	12/16/2017	F
bitcoin	221	bitcoin	196	bitcoin	185	bitcoin	158
btc	45	btc	44	segwit	29	btc	44
wallet	24	wallet	36	coinbase	26	buy	23
crypto	22	buy	29	wallet	23	price	22
like	22	coinbase	28	transaction	23	new	21
cryptocurrency	21	get	22	btc	20	wallet	20
coinbase	20	think	21	get	19	get	20
buy	20	bitcoins	20	buy	19	mining	17
network	19	transaction	20	crypto	17	bitcoins	16
people	19	people	19	futures	16	segwit	16
futures	19	price	19	good	16	network	15
get	19	way	18	way	15	one	15
sell	19	sell	16	bitcoins	15	crypto	14
new	17	would	15	money	15	money	13
still	17	still	15	need	14	coinbase	13

bitcoins	16	money	15	trading	13	sell	13
price	16	value	15	anyone	13	like	13
lightning	16	one	14	help	13	cryptocurrency	13
exchange	15	cryptocurrency	14	network	12	bubble	12
worth	14	help	14	still	12	exchange	12

12/17/2017	F	12/18/2017	F	12/19/2017	F	12/20/2017	F
bitcoin	340	bitcoin	205	bitcoin	195	bitcoin	298
btc	68	btc	45	btc	38	coinbase	128
get	45	futures	39	transaction	32	btc	85
futures	45	wallet	29	wallet	28	bch	62
wallet	44	transaction	24	use	23	bcash	51
buy	41	buy	24	buy	21	trading	46
transaction	35	crypto	20	exchange	20	buy	38
price	33	people	19	like	19	cash	37
help	31	bitcoins	18	crypto	18	wallet	33
like	29	get	17	fee	18	one	31
money	28	money	17	would	17	price	31
coinbase	27	new	16	money	16	exchange	31
fee	27	trading	15	coinbase	16	like	31
would	26	coinbase	14	know	16	time	30
cryptocurrency	26	fees	14	coins	15	going	30
bitcoins	24	would	13	get	15	gdax	29
time	24	way	13	lightning	15	transaction	29
one	24	tax	13	network	14	network	27
new	23	like	13	still	14	new	25
anyone	22	time	12	think	14	people	25

12/21/2017	F
bitcoin	221
coinbase	55
btc	53
transaction	39
fees	30
trading	25
wallet	25
fee	24
lightning	24
network	22

new	22
price	22
help	21
buy	20
one	19
going	19
get	19
segwit	18
transactions	17
money	17

Table 1: Most common words in Bitcoin threads' title in period 12/13/2017 – 12/21/2017

## Discussion

The number of threads started has a small correlation to the price. For example, when price peaked on the 17<sup>th</sup>, so did the number of threads started. When price dipped on the 20<sup>th</sup>, the number of threads came close to the level of the peak. The studied period is not long enough to draw any meaningful conclusion about the movement of price and the amount of conversation surrounding Bitcoin. Future projects should look at longer periods of price and number of threads on Reddit to gain more meaningful insights.

Understandably, the most common word in a subreddit dedicated to Bitcoin is “bitcoin”. However, just like other words in the top 20, “bitcoin” does not signify a bullish or bearish sentiment. I tried to use a dictionary with pre-labelled dictionary of words and their usual associated sentiment in the English language. This approach is bad, however, because many of the most common words on Reddit received their meaning from the Reddit community. What might mean positive in the English language might mean the complete opposite on Reddit. Therefore, to determine the sentiment of the conversation on Reddit about Bitcoin. There needs to be a human who categorizes whether a term reflects bullish or bearish sentiment.

The scope of this project does entail going through a long list of the most common words on the subreddit bitcoin to categorize their respective sentiment. The next logical step would be to employ human effort, specifically those with knowledge of how people discuss Bitcoin and Bitcoin trading, to decide the sentiment of individual words. Only then can a concrete conclusion about the sentiment exposed on social media be reached.

## Conclusion

Studies have shown that market sentiment has a great influence on the price of stock and gold. Bitcoin, a recently popular traded asset, has not existed long enough to have attract the quantity of research that traditional assets enjoy. In addition, conversation about Bitcoin trade circulates in untraditional channels that lack formal languages used to express opinion about its price.

My project aims to understand the influence of these untraditional channels' influence on Bitcoin price. Instead, I discovered that they could not currently be formally studied because of the lack of information on how to interpret their meaning, not mentioning their effects on the cryptocurrency. As Bitcoin grows, understanding of how people talk about it and what influences it would grow as well, which will not only help us understand what affects the price of Bitcoin but also help us extract data from and understand fledgling communication channels.

## Bibliography

@coinbase (Coinbase). *Twitter*, 15 Mar. 2018.

“/r/Bitcoin Metrics (Bitcoin - The Currency of the Internet).” *Reddit Metrics*,  
redditmetrics.com/r/bitcoin.

“24 Hour Volume Rankings (Exchange).” *CoinMarketCap*,  
coinmarketcap.com/exchanges/volume/24-hour/.

CoinDesk. “About the Bitcoin Price Index.” *CoinDesk*, CoinDesk,  
www.coindesk.com/price/bitcoin-price-index/.

Amadeo, Kimberly. “The Pros and Cons of the Gold Standard, and Why We Can’t Go Back.” *The Balance*, The Balance, 17 Apr. 2018, www.thebalance.com/what-is-the-gold-standard-3306137.

Amadeo, Kimberly. “How the Bretton Woods Agreement Introduced a New World Order.” *The Balance*, The Balance, 10 Jan. 2018, www.thebalance.com/bretton-woods-system-and-1944-agreement-3306133.

Andrews, Elyse. “The Importance of Investor Sentiment.” *NASDAQ.com*, Nasdaq, 17 Apr. 2010, www.nasdaq.com/article/the-importance-of-investor-sentiment-cm18961.

“Bitcoin History: The Complete History of Bitcoin [Timeline].” *Bitcoin History: The Complete History of Bitcoin [Timeline]*, historyofbitcoin.org/.

“BitcoinTalk.” *Bitcoin Wiki*, 28 May 2015, en.bitcoin.it/wiki/BitcoinTalk.

Britannica, The Editors of Encyclopaedia. “Fiat Money.” *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 3 Feb. 2017, www.britannica.com/topic/ fiat-money.

Burn-Callander, Rebecca. "The History of Money: from Barter to Bitcoin." *The Telegraph*, Telegraph Media Group, 20 Oct. 2014, [www.telegraph.co.uk/finance/businessclub/money/11174013/The-history-of-money-from-barter-to-bitcoin.html](http://www.telegraph.co.uk/finance/businessclub/money/11174013/The-history-of-money-from-barter-to-bitcoin.html).

Callahan, Ann. "Factors Influencing Bitcoin Price – Hacker Noon." *Hacker Noon*, Hacker Noon, 11 Jan. 2018, [hackernoon.com/factors-influencing-bitcoin-price-cfabdf634894](http://hackernoon.com/factors-influencing-bitcoin-price-cfabdf634894).

Chen, Adrian. "The Underground Website Where You Can Buy Any Drug Imaginable." *Gawker*, 1 June 2011, [gawker.com/the-underground-website-where-you-can-buy-any-drug-imag-30818160](http://gawker.com/the-underground-website-where-you-can-buy-any-drug-imag-30818160).

ckoepp. "TwitterSearch." *GitHub*, 1 Jan. 2018, [github.com/ckoepp/TwitterSearch](https://github.com/ckoepp/TwitterSearch).

CoinDesk. "Bitcoin Price Index - Real-Time Bitcoin Price Charts." *CoinDesk*, CoinDesk, [www.coindesk.com/price/](http://www.coindesk.com/price/).

Constable, Simon. "The Market's Bullish, the Market's Bearish: What Does It Matter?" *The Wall Street Journal*, Dow Jones & Company, 7 Sept. 2014, [www.wsj.com/articles/market-sentiment-why-it-matters-1410120108](http://www.wsj.com/articles/market-sentiment-why-it-matters-1410120108).

Dai, Wei. "Bmoney." *Wei Dai's Home Page*, [www.weidai.com/bmoney.txt](http://www.weidai.com/bmoney.txt).

"Gold Spot Price Per Ounce Today, Live & Historical Charts in USD." *Money Metals Exchange*, Money Metals Exchange, 7 Mar. 2018, [www.moneymetals.com/precious-metals-charts/gold-price](http://www.moneymetals.com/precious-metals-charts/gold-price).

HTMW Team. "What Is a Stock?" *HowTheMarketWorks Education Center*, 1 Nov. 2015, [education.howthemarketworks.com/beginners/types-of-investments/stocks/intro-to-stocks/](http://education.howthemarketworks.com/beginners/types-of-investments/stocks/intro-to-stocks/).

Huddleston, Tom. "Reddit Is Now Valued at Nearly \$2 Billion." *Fortune*, Fortune, 31 July 2017, fortune.com/2017/07/31/reddit-200-million-funding/.

Hur, Johnson. "The History of Gold." *History of The US Minimum Wage - From The Very First Minimum Wage*, 17 Nov. 2017, bebusinessed.com/history/the-history-of-gold/.

"Introducing JSON." *JSON*, www.json.org/.

Kelleher, John. "Bitcoin." *Investopedia*, Investopedia, 7 Mar. 2018, www.investopedia.com/terms/b/bitcoin.asp.

Law, Jonathan. "Market Sentiment - Oxford Reference." *Social Class and Sport - Oxford Reference*, 12 Mar. 2018, www.oxfordreference.com/view/10.1093/acref/9780199664931.001.0001/acref-9780199664931-e-2262.

Law, Jonathan. "Behavioural Finance - Oxford Reference." *Social Class and Sport - Oxford Reference*, 12 Mar. 2018, [www.oxfordreference.com/view/10.1093/acref/9780199664931.001.0001/acref-9780199664931-e-5281](http://www.oxfordreference.com/view/10.1093/acref/9780199664931.001.0001/acref-9780199664931-e-5281).

Liu, B. (2015). Introduction. In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (pp. 1-15). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139084789.002

"New Liberty Standard." *2009 Exchange Rate - New Liberty Standard*, WikiFoundry, 5 Feb. 2010, newlibertystandard.wikifoundry.com/page/2009 Exchange Rate.

O'Neill, Patrick Howell. "Meet Dread Pirate Roberts' Secret Mentor." *The Daily Dot*, 11 Dec. 2015, [www.dailydot.com/crime/dpr-silk-road-variety-jones/](http://www.dailydot.com/crime/dpr-silk-road-variety-jones/).

praw-dev. "PRAW: The Python Reddit API Wrapper." *GitHub*, [github.com/praw-dev/praw](https://github.com/praw-dev/praw).

Price, Rob. "Someone in 2010 Bought 2 Pizzas with 10,000 Bitcoins - Which Today Would Be Worth \$20 Million." *Business Insider*, Business Insider, 22 May 2017, [www.businessinsider.com/bitcoin-pizza-day-passes-2000-20-million-2017-5](http://www.businessinsider.com/bitcoin-pizza-day-passes-2000-20-million-2017-5).

"r/Bitcoin." *Reddit*, Reddit, [www.reddit.com/r/Bitcoin/](http://www.reddit.com/r/Bitcoin/).

reddit. Oxford University Press, 2016.

Sarkar, Dipanjan. *Text Analytics with Python: a Practical Real-World Approach to Gaining Actionable Insights from Your Data*. Apress, 2016.

Szabo, Nick. "Bit Gold." *Wet Code and Dry*, Blogspot, 27 Dec. 2008, [unenumerated.blogspot.com/2005/12/bit-gold.html](http://unenumerated.blogspot.com/2005/12/bit-gold.html).

"Twitter Developer Platform - Twitter Developers." *Twitter*, Twitter, [developer.twitter.com/](https://developer.twitter.com/).

satoshi. "Welcome to the New Bitcoin Forum!" *Bitcointalk*, 22 Nov. 2009, [bitcointalk.org/index.php?topic=5.0](http://bitcointalk.org/index.php?topic=5.0).

"Whois Lookup." *Whois Domain Lookup Search - Whois Lookup: Find Who Registered a Website, Check Domain Name*, [www.register.com/whois.rcmx;jsessionid=97FF461163A7C262CFE616D581F64E7F.janus-production](http://www.register.com/whois.rcmx;jsessionid=97FF461163A7C262CFE616D581F64E7F.janus-production).

Williams, Sean. "7 Common Factors That Influence Gold Prices." *The Motley Fool*, The Motley Fool, 13 Oct. 2016, [www.fool.com/investing/2016/10/13/7-common-factors-that-influence-gold-prices.aspx](http://www.fool.com/investing/2016/10/13/7-common-factors-that-influence-gold-prices.aspx).

Yu, Y., WJ Duan, and Q. Cao. "The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach." *Decision Support Systems*, vol. 55, no. 4, 2013, pp. 919-926.