Sentiment Analysis of Blockchain Companies Based on Twitter

Abstract

This project aims to analyze the influence blockchain companies have and challenges they are facing with the assistance of Twitter-based sentiment analysis. After an introduction about blockchain technology, a series of problems including the general attitudes, popularity and temporal/spatial distribution of these companies are explored through visualization. Conclusions are made in the end.

Key words: *Blockchain, Cryptocurrency, Twitter, Sentiment Analysis*

1. Introduction

1.1 Blockchain technology

Blockchain technology is an incorruptible digital ledger that can record transactions between two parties efficiently in a verifiable and permanent way [1]. A typical process in blockchain is: one party firstly requests a transaction which is funneled into the network, represented by a block and broadcast to each individual node(computer) with the highest level of security. Then individual nodes compete to validate the transaction for getting paid part of the crypto currency as a reward. After the validation, the block will be added to the existing blockchain and the transaction is nailed. Blockchain has many advantages. It provides greater transparency and traceability in transactions, enhances security and ensures everyone's access to the global financial system. There are many companies in this technology. While some of them have great influence on blockchain ecosystems, some are facing the challenges of trust and reputation.

1.2 Cryptocurrency

Cryptography is the technique of protecting information by encrypting it into a format that can only be decrypted by someone who possesses a secret key[2]. Cryptocurrencies such as Bitcoin, are secured via this technique using an ingenious system of public and private digital keys.

2. Problem statement

2.1 Goal of project

The main goal of this project is to analyze the influence blockchain companies have and challenges they are facing. The main method used is sentiment analysis of Twitter messages. After analysis, better understood information will become knowledge that may help people make wiser decisions on selecting Bitcoin trading platform and timing, or even predict the trend of cryptocurrencies.

2.2 Problems of interest

- What are people's general attitudes towards blockchain technology, companies and cryptocurrencies?
- What is the general popularity of blockchain, companies and cryptocurrencies?
- Does the price fluctuation of cryptocurrencies influence people's attitudes towards these companies?
- Geographically, which areas are promising for blockchain companies?

3. Dataset description

3.1 Data collection

This project uses Tweepy package in python to scrape information from Twitter. Since the topic is blockchain company, the keywords for searching tweets are names of typical blockchain companies, cryptocurrencies (for popularity and price trend analysis only) and the word "blockchain". The nine companies selected in this study are: Binance, Bitcoin.com, Bitmain, Coinbase, Bitpay, Bitstamp, Kraken, Xapo and ShapeShift; Two selected typical cryptocurrencies are: Bitcoin and BCH. Tweets about each keyword are scraped and manipulated separately which allows independent analysis and comparison. For each keyword, Tweepy captures a series of attributes from every tweet, including tweet id number, user id number, text, time of creation, location of creation, the number of followers and follows of the user account, whether the user is verified and repost count.

3.2 Quantity and quality evaluation

The time span of data collection is four weeks, from 10/27 to 11/26, 2018. The raw data

amount is about 8000 texts for each keyword except Xapo and Shapeshift which have only limited data about 5000. After repetition and null value cleaning which eliminate 1000 texts on average for each. Therefore, the temporal distribution and magnitude of quantity is ideal for sentiment analysis. When it comes to data quality, since it's based on keyword searching method, the accuracy of estimation is relatively low. Also, while most data are in English, there are a few texts which are written in other languages or even emoji. These data can be invalid in terms of English-based sentiment analysis. Though the overall quality of results won't be influenced much, solving the language problem will still be desirable for the improvement of this study in the future.

3.3 Data operation

All the data, after acquired from Twitter, are stored into csv files (one file per keyword) for operation. The first step is basic data cleaning. The main content of this step is getting rid of repetition, noises and null values within Excel. Then, to adjust to analysis of different problems, the same datasets are manipulated accordingly. For sentiment analysis, the sentiment score of each tweet text is calculated using TextBlob package in python. Then, to better illustrate the sentiment, three groups of sentiment score are defined: scores less than -0.00001 are defined as negative, scores greater than 0.00001 as positive, and the rest as "0" or neutral. For Geographical analysis, the location of each tweet is first translated to latitude and longitude with the assistance of Google Geocoding API for further analysis.

4 Analysis results

Visualization with text is the main form of illustration in this study. Plots of general sentiments, popularity and fluctuations are rendered with R. Geographical plots are drawn using folium package in python.

4.1 General attitudes

People's general attitudes are presented by two polar bar plots and a bar plot in Fig 1. The first polar bar plot shows that more than half of the sentiment scores are "0". The second polar bar plot, in which the tweets in"0" group are eliminated, clearly shows the

portion of positive tweets for each keyword. Then the average sentiment scores of each keyword are revealed in the bar plot on the right. It's obvious that sentiments are overall positive. The keywords which have the highest and lowest sentiment score are Binance and Bitmain respectively. While the success of Bitmain can be explained by its innovation on wallet application and mining, Bitmain's waterloo is probably due to its involvement in the fork of BCH.

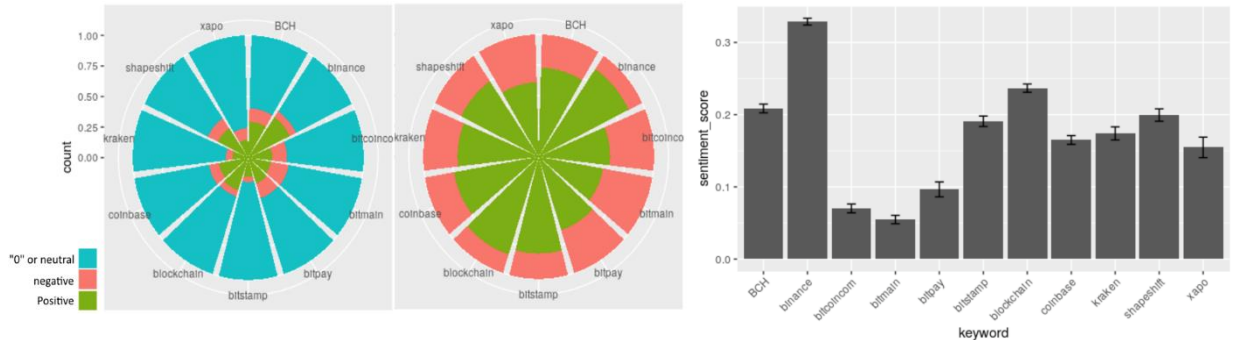

Fig 1. General attitudes and average sentiment score

4.2 Popularity

Popularity can be explored by focusing on two attributes: repost count and number of followers. The key point is to figure out the relationship between them and sentiment score. Firstly, repost count is often considered as a factor indicating the popularity of a

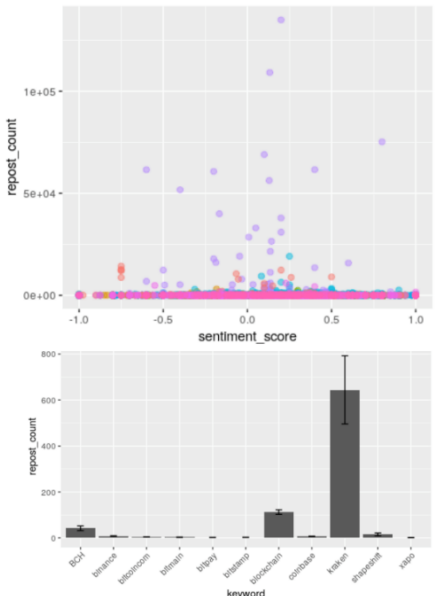

Fig 2. Repost count of all keywords

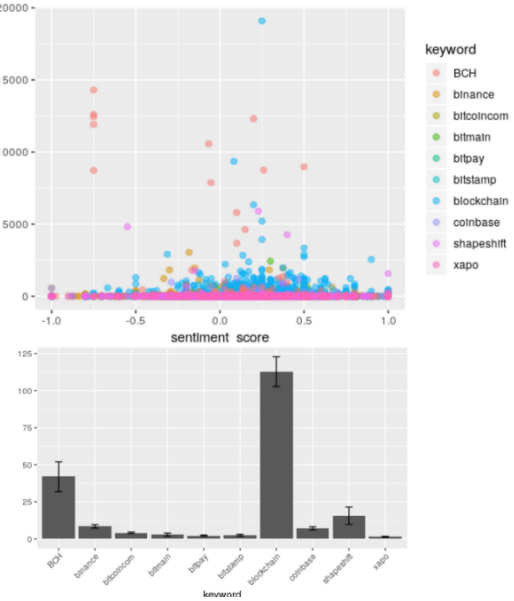

Fig 3. Repost count of all keywords except "kraken"

tweet. Fig 2 shows the count of reposts for each keyword with a bar plot and the distribution along sentiment score with a scatter point plot in which each point

represents a tweet. Here the keyword “kraken” shows a significantly higher repost rate than others, which is unusual because as a relatively small company, it has greater repost count than the keyword “blockchain”. When these tweets are traced back to Twitter, they are found to come from an identical user. So, this can be seen as a disturbance of the analysis. To eliminate this noise, the bar and scatter plot are replotted without the keyword “kraken”, as is shown in Fig 3. The new result in which “blockchain” is the keyword of highest repost count and "BCH" the second is more reasonable because general keyword should be more popular than specific company keywords.

Another factor that can reflect popularity is the number of followers. Scatter plots indicating relationship between number of followers and sentiment score of each keyword are shown in Fig 4. In some of the plots, a horizontal line pattern (marked by

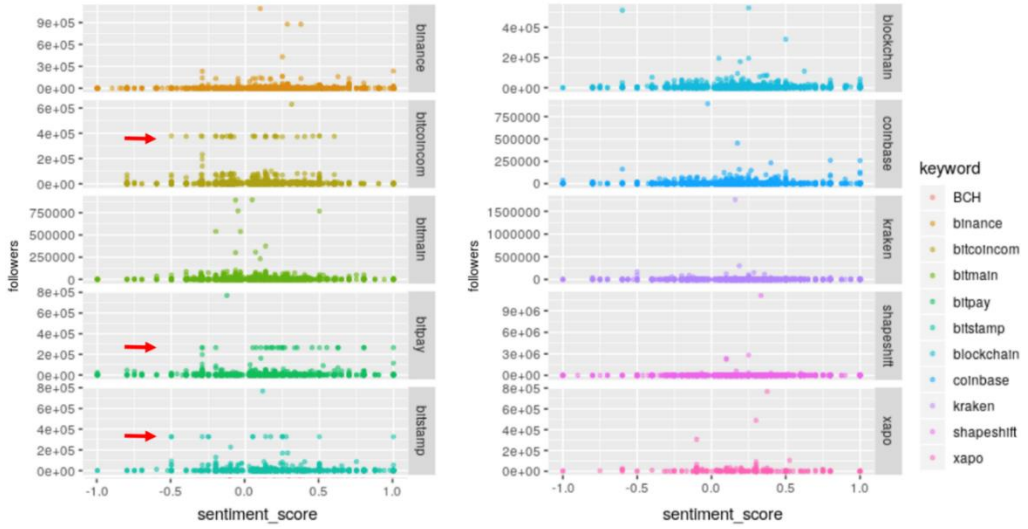

Fig 4. Followers count of all keywords

a red arrow in the plot) formed by data points is very interesting. The posters of these tweets have very similar follower numbers and they come from an identical user. This indicates that the user could be the promotion account of the company. Then, following this path, companies like Bitcoincom, Bitpay and Bitstamp are suspects of promoting themselves on Twitter purposefully. For Chinese companies like Bitmain and Binance, the pattern is not that strong. This makes sense because it’s relatively less cost-effective to manage a promotion Twitter account in China. But these two companies are still popular in Twitter. They show great potential and if they make more promotion on Twitter, their business could be boosted.

4.3 Cryptocurrency price fluctuation

The Bitcoin Cash split of November 2018 was an incident initiated by companies influential among the community, Nchain and its ally Coingeek. Both of them are also miners who had significant share in terms of mining Bitcoin Cash. This incident had great influence on trend of cryptocurrency price. It's of great interest to figure out whether the influence has been spread to the blockchain companies. Line plots about the trend of sentiments around the data of the incident are shown in Fig 5.

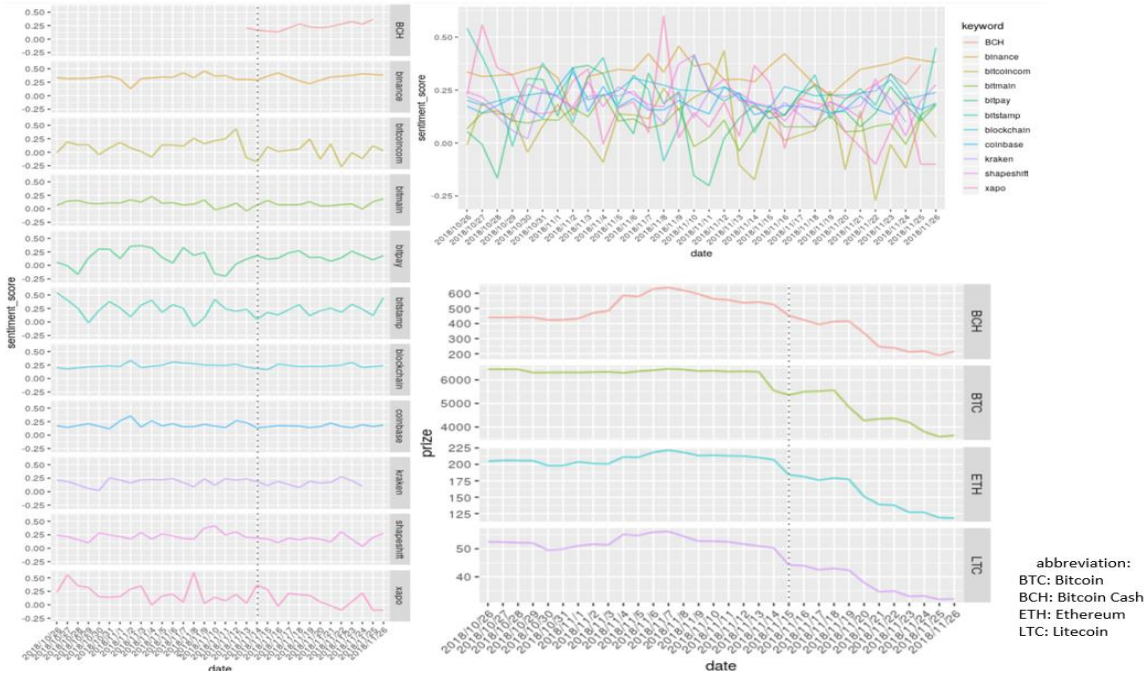

Fig 5. Fluctuation of cryptocurrency price and sentiments

Unlike the vigorously busted prices of cryptocurrencies shown in the down-right plot, sentiments about companies is relatively stable (left plot). This is because as a growing technique, blockchain is extending to various fields and the ecosystem of blockchain is becoming more and more immune to the mutation of cryptocurrency price. However, it's still too early to assert that those companies are completely independent to the price fluctuation of cryptocurrencies.

4.4 Geographical analysis

A good way to understand the sentiment distribution of a keyword is to view it geographically. A series of bubble maps are generated for each keyword. The bubbles are plotted based on the geographical information of corresponding tweets. The

sentiment score is reflected on the color and diameter of the bubble. A green one represents a positive sentiment, while a red one means a negative sentiment. The diameter is directly proportional to how positive/negative the sentiment score is depending on the color. Shown below are some selected examples of the bubble maps this study delves into and draws conclusions on. Fig 6 is the bubble map for keyword "blockchain". As is seen in the figure, the density of bubble is very high in the US and Europe region. The majority of bubbles are green and a few are red. This can be interpreted as the popularity of blockchain technology is highest in Europe and US. This distribution pattern is reasonable since technology hubs are located mainly in Europe, east and west coast of the US, like London, New York, Bay area and Seattle. What's more, in Asia, blockchain technology is emerging and developing with support in India, Japan and Korea. However, since Twitter is officially banned in China, very few data there are available.

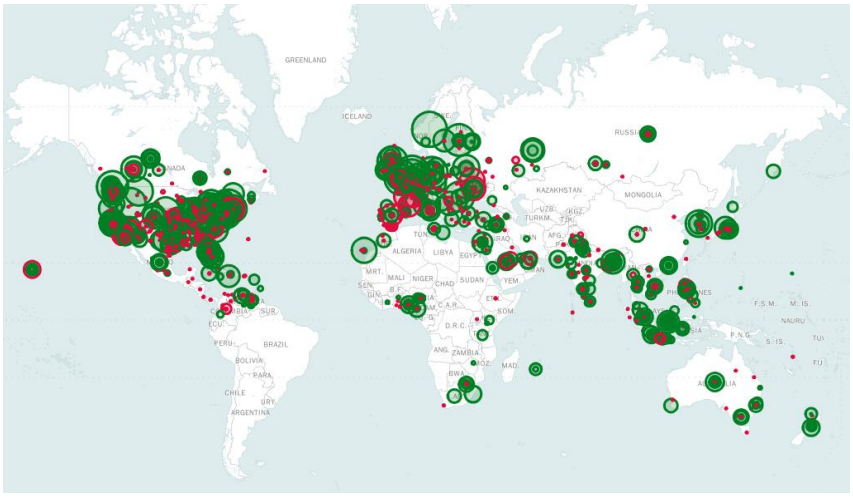

Fig 6. bubble map of "blockchain"

Fig 7 and Fig 8 are meant to show the sentiment distribution of two popular companies: Coinbase and Binance. Clearly, Coinbase has more negative sentiments coming from the mid-west of the US compared to the east and west coast, whereas Binance is quite consistent across the entire US in terms of color and density. This discrepancy might be caused by the difference in their user base. While Coinbase's mobile-based exchange application allows easy access to crypto wallets due to its simple user-friendly design, Binance uses a web-based exchange site which is more complicated to use and targeted towards people with more advanced knowledge in crypto and traditional stock

Fig 7. bubble map of Coinbase in the US

Fig 8. bubble map of Binance in the US

exchange. Thus, the consistent negative sentiment of Binance can be imputed to its user-unfriendly design of website. This could potentially be a guidance for Binance: to be more competitive, optimize the website and exchanging mechanism first. From another point of view, though Coinbase has less reds (negative sentiments) than Binance in the middle region, there are some reds with very large diameter which means extremely negative attitude. This is probably because the user-friendly application of Coinbase sets very low threshold of entrance. Then, people who have little experience with trading and no expectations about the risks were brought into the crypto space due to the hype. They are more susceptible to the fluctuation of the crypto market and tends to have a negative connotation towards cryptocurrency after losing their investment.

5. Conclusions

- Binance is the most welcomed company due to its innovation on wallet application while Bitmain's waterloo might be due to its involvement in the fork of BCH.
- Bitcoincom, Bitpay and Bitstamp, though popular, are suspects of using promotion Twitter accounts. Bitmain and Binance show great potential in terms of popularity.
- Blockchain companies are not very sensitive to mutation of cryptocurrency price.
- Geographically, blockchain technology is most popular in Europe and US because technology hubs are located mainly in these regions.

Reference

[1] Iansiti, Marco and Karim R. Lakhani."The truth about blockchain." Harvard Business Review 95.1 (2017): 118-127.

[2] https://en.wikipedia.org/wiki/Cryptocurrency

Link of data: https://github.com/vikkki/twitter-analysis-blockchain-companies sentiment.zip