

Bitcoin Spread Prediction Using Social And Web Search Media

Martina Matta, Ilaria Lunesu, Michele Marchesi

Università degli Studi di Cagliari
Piazza d'Armi, 09123 Cagliari, Italy
{martina.matta, ilaria.lunesu, michele}@diee.unica.it

Abstract. In the last decade, Web 2.0 services such as blogs, tweets, forums, chats, email etc. have been widely used as communication media, with very good results. Sharing knowledge is an important part of learning and enhancing skills. Furthermore, emotions may affect decision-making and individual behavior. Bitcoin, a decentralized electronic currency system, represents a radical change in financial systems, attracting a large number of users and a lot of media attention. In this work, we investigated if the spread of the Bitcoin's price is related to the volumes of tweets or Web Search media results. We compared trends of price with Google Trends data, volume of tweets and particularly with those that express a positive sentiment. We found significant cross correlation values, especially between Bitcoin price and Google Trends data, arguing our initial idea based on studies about trends in stock and goods market.

Keywords: Bitcoin, Twitter, Google Trends, Sentiment Analysis

1 Introduction

Bitcoin, a decentralized electronic currency system, represents a radical change in financial systems after its creation in 2008 by Satoshi Nakamoto [22]. Bitcoin stands for an IT innovation based on the advancement in peer-to-peer networks [21] and cryptographic protocols. Due to its properties, Bitcoin is not managed by any governments or bank. Like any other currency, a peculiarity of Bitcoin is to facilitate transactions of services and goods [20], attracting a large number of users and a lot of media attention.

Nowadays, Web 2.0 services such as blogs, tweets, forums, chats, email etc. are widely used as communication media, with satisfying results. Sharing knowledge is an important part of learning and enhancing skills. Through the use of social media services, team members have the opportunity to acquire more detailed information about their peers' expertise [7]. Social media data represents a collective indicator of thoughts and ideas regarding every aspect of the world. It has been possible to assist to deep changes in habits of people in the use of social media and social network. Twitter[10], an online social networking website and microblogging service, has become an important tool for businesses and individuals to communicate and share information with a rapid growth and

significant adoption. In addition, Twitter has rapidly grown as a mean to share ideas and thoughts on investing decisions.

In this work we analyze whether social media activity or information extracted by web search media could be helpful and used by investment professionals. There are several works that present predictive relationships between social media and bitcoin price where the relative effects of different social media platforms (Internet forum vs. microblogging) and the dynamics of the resulting relationships, are analyzed using cross-correlation such as [17] or linear regression analysis such as [6] or [5]. Social factors, that are composed of interactions among the actors of the market, may strongly drive dynamics of Bitcoin's economy [3].

We decided to apply automated Sentiment Analysis on shared short messages of users on Twitter in order to automatically analyze people's opinions, sentiments, evaluations and attitudes. We investigated whether public sentiment, as expressed in large-scale collections of daily Twitter posts, can be used to predict the Bitcoin market. We tried to discover if the chatter of the community can be used to make qualitative predictions about Bitcoin market, attempting to establish whether there is any correlation between tweet's sentiment and the Bitcoin's price ¹. The results suggest that a significant relationship with future Bitcoin's price and volume of tweets exists on a daily level. We also used Google Trends to analyze Bitcoin's popularity under the perspective of Web search, which provides a time series index of the volume of queries made by users in Google Search. We found a striking correlation between Bitcoin's price spread and changes in query volumes for the "*Bitcoin*" search term.

The body of this paper is organized in five major sections. Section 2, describes the background, section 3 presents the research steps of our study and section 4 summarizes and discusses our results. Finally, section 5 presents conclusions and suggestions for future work.

2 Background

In these decades, social web has been commercially exploited for goals such as automatically extracting customer opinions about products or brands, to find which aspects are liked and which are disliked [9]. In their work, Ye and Wu demonstrate how particularly interesting is the influence of Twitter users and the propagation of the information related to their tweets[18].

According to Alexa [12], Twitter had become the world's seventh most popular website by March 2015. Twitter [10] is an online social networking website and microblogging service that allows users to post and read text-based messages of up to 140 characters, known as "tweets". Launched in July of 2006 by Jack Dorsey, Twitter is now in the top 10 most visited internet sites with a total amount of 645,750,000 registered users. Java et al. affirm that it seems to be used to share information and to describe minor daily activities [14]. The short format of a tweet is a defined characteristic of the service, allowing informal collaboration and quick information sharing. For business, Twitter can be used to

¹ <https://markets.blockchain.info/>

broadcast company's latest news, posts, read comments of the customers or interact with them. A communicative feature of Twitter is the hashtag: a metatag beginning with the character #, designed to help others find a post.

Twitter is a rich source of real-time information regarding current societal trends and opinions. There are also studies that report another use of Twitter, namely as a possible predictor of market trends. Indeed, in 2010, a publication of the professor Johan Bollen showed that combining information on Wall Street with the millions of Tweets and posts, makes possible to anticipate financial performance [6]. In this work, Granger causality analysis and a Self-Organizing Fuzzy Neural Network are used to investigate the hypothesis that public mood states, as measured by the OpinionFinder and GPOMS mood time series, are predictive of changes in DJIA closing values. The analysis of Tweets made by Bollen would have had 87% of chance to successfully predict prices of the stock, 3 or 4 days in advance. This study and analysis of millions of posts on Twitter represents a thermometer of emotions, on a large scale, which reflects the whole of society.

Earlier studies had found that blogs can be used to evaluate public mood, and that tweets about movies can predict box office sales. Investigating the literature related to different uses of social media, and Twitter in particular, we collected information about the use of Twitter for seeking real world emotions that could predict real financial markets trend [1]. In their paper, Rao and Srivastava investigate the complex relationship between tweet board literature (like bullishness, volume, agreement etc) with the financial market instruments (like volatility, trading volume and stock price) [2].

The Bitcoin represents an important new phenomenon in financial markets. Mai et al. [4] examine predictive relationships between social media and Bitcoin returns by considering the relative effect of different social media platforms (Internet forum vs. microblogging) and the dynamics of the resulting relationships using vector autoregressive and vector error correction models.

In their work, Garcia et al. [3] show the interdependence between social signals and price in the Bitcoin economy, namely a social feedback cycle based on word-of-mouth effect and a user-driven adoption cycle. They provide evidence that Bitcoin's growing popularity causes an increasing search volumes, which in turn result a higher social media activity about Bitcoin. More interest inspire the purchase of bitcoins by users, driving the prices up, which eventually feeds back on the search volumes.

We compared Twitter's trending topic about Bitcoin with those in other media, namely, Google Trends [8]. This is a feature of Google search engine that illustrates how frequently a fixed search term was looked for. Through this, you can compare up to five topics at one time to view relative popularity, allowing you to gain an understanding of the hottest search trends of the moment, along with those developing in popularity over time. Following this kind of approach, we evaluated how much "*bitcoin*" term, for the analyzed time interval, is looked for using Google's search engine.

3 Methodology

3.1 Sentiment Analysis

Tweets sometimes express opinions about different topics, and for this reason we decided to evaluate user's opinion about Bitcoin. We also investigated its power at predicting real-world outcomes. In order to evaluate if a user really appreciates the Bitcoin spread, we tried to predict sentiments analyzing tweets collection. In recent years, there is a wide collection of research surrounding machine learning techniques, in order to extract and identify subjective information in texts. This area is known as sentiment analysis or opinion mining [15]. Sentiment techniques are able to extract indicators of public mood directly from social media content [24].

Pang et al. argue that the research field of sentiment analysis has developed many algorithms to identify if the opinion expressed is positive or negative. In fact, algorithms to recognize sentiment are required to understand the role of emotions in informal communications [15]. Go et al. affirmed the strength of the sentiment analysis applied to the Twitter domain by using similar machine learning techniques to classifying the sentiment of tweets [19].

We chose to use automated sentiment analysis techniques to identify the sentiments of tweets in the matter of Bitcoin. Since the goal of this research is neither to develop a new sentiment analysis nor to improve an existing one, we used "SentiStrenght", a tool developed by a team of researchers in the UK that demonstrated good outcomes [11]. SentiStrength estimates the strength of positive and negative sentiments in short texts. It is based on a dictionary of sentiment words, each one associated with a weight, which is its sentiment strength. In addition, this method uses some rules for non-standard grammar.

Based on the formal evaluation of this system on a large sample of comments from MySpace.com, the accuracy of predicting positive and negative emotions was something similar to that of other systems (72.8% for negative emotions and 60.6% for positive emotions, based on a scale of 1-5). Compared to other methods, SentiStrenght showed the highest correlation with human coders [13]. The tool is able to assess each message separately and, at the end, it returns one singular value: a positive, a negative or a neutral sentiment.

3.2 Data Collection

Tweets are available and are easily retrieved making use of Twitter Application Programming Interface (API) [16]. Composing the hashtag #Bitcoin or @bitcoin, we are able to gather all tweets that mentioned the analyzed subject. We briefly describe the different components of our system. An overview of this architecture is shown in Figure 1. The system consists of four components:

- *Twitter Streaming API*: it provides access to Twitter data, both public and protected, on a nearly real-time basis. A persistent connection is created between our system and Twitter. As soon as tweets come in, Twitter notifies our system in real time, allowing us to store them into our database.

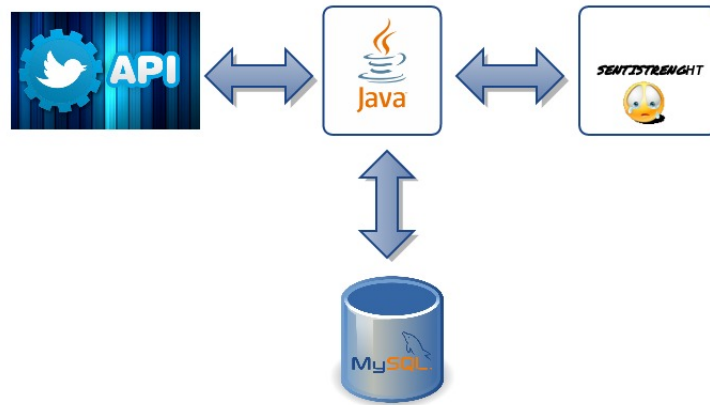


Fig. 1. System Architecture

- *DataStore*: our datastore consists of a back-end database engine, using MySQL as RDBMS, that repeatedly saves the incoming tweets from the Twitter Streaming API.
- *SentiStrenght tool*
- *Java Module*: this component allows us to send automated requests to Twitter Streaming API, to recover new tweets about Bitcoin, to parse the data gathered and to store them into our datastore. In a later stage, these data are sent to SentiStrenght tool in order to automatically evaluate the users' opinion.

We analyzed a collection of tweets, regarding Bitcoin, posted on Twitter between January 2015 and March 2015 (60 days). During this time 1,924,891 tweets were collected. The tweets were analyzed to determine its identifier, the date-time of the submission, its type, and its text content, which is limited to 140 characters. Comparing the timeline of tweets and the fluctuations in the Bitcoin market, we determined the specific day that provide a better correlation value. We then used SentiStrenght to evaluate comments extracted from Twitter. Given as input all tweets, the system assigned a score for each comment:

- 1 if the comment is positive
- -1 if the comment is negative
- 0 if the comment is neutral

4 Results

In order to decide the correct strategy of analysis for studying the relationship among Bitcoin's price and others meaningful parameters, the available related literature has been examined in depth. Most of articles [6] [1] [2] reports analysis about the existent relationship between the volume of tweets and the market evolution. In general, Bollen et al. demonstrated that tweets can predict the market trend 3-4 days in advance, with a good chance of success. We analyzed the Bitcoin price's behavior comparing its variations with the number of tweets, with the number of tweets with positive mood, and with Google Trends results. The computation of cross-correlation yielded interesting results.

Our result seems to confirm that volumes of exchanged tweets may predict the fluctuations of Bitcoin's price. Furthermore, the comparison between tweets with a positive mood and trend of Bitcoin's price seems to prove this behavior. The

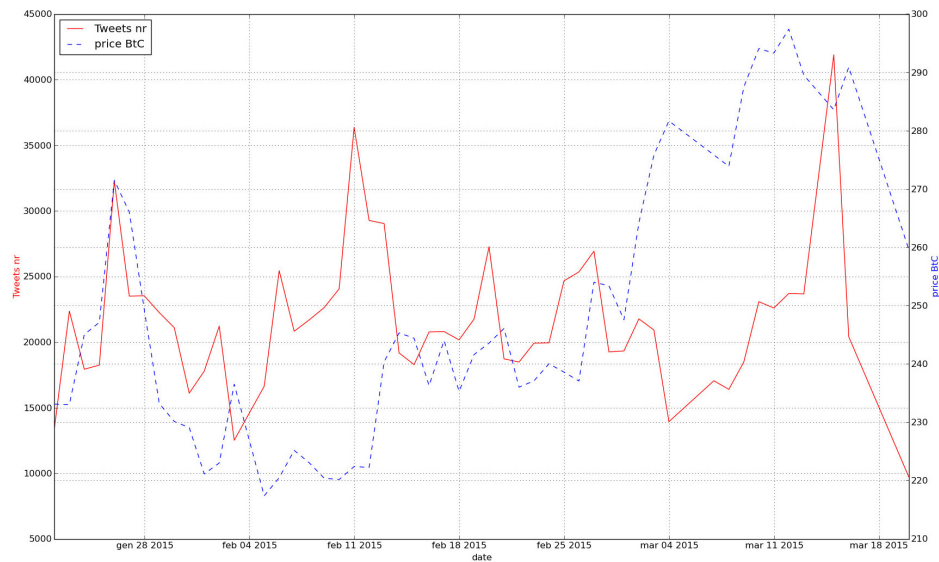


Fig. 2. Similarity between Bitcoin's price and number of Tweets

examined literature shows different ways to highlight the existent relationship between big volumes of exchanged tweets and meaningful variations in the Bitcoin's price. Some papers show studies using regression methodology [4] or causality analysis [6]. Rao et al.[2] and Mittal et al.[5] showed how goods and stocks markets may be influenced by a big exchanged of tweet's volume. Inspired by these works, we tried to demonstrate how chatter of tweets might predict the price's variations of Bitcoin. Figure 2 illustrates the curve trend of Bitcoin prices, expressed in dollars, and Twitter volume. We calculated the cross-correlation and,

analyzing the results, we found that, in minimal degree, tweets volume is related to price with a maximum cross correlation value of 0.15 at a lag of 1 day (this is not very significant). Nevertheless, if we observe Figure 2, we can notice how, also at a glance, there are peaks in tweets trend that precede peaks in price, suggesting a relationship between the two time series. A patent peak of tweets on 11 February, is followed by a growth of Bitcoin's price. The same circumstance is visible in the following days: January 23, February 3, February 25 and so on. We also analyzed tweets with positive mood and we noticed a two-fold increase in cross-correlation value. Figure 3 shows this result and it's well rendered that positive tweets can predict the fluctuations of the Bitcoin's price. It is proven by a maximum cross correlation value of -0.35 with a positive delay of almost 4 days. We can confirm that positive mood could predict the Bitcoin's price almost 3-4 days in advance. All patent peaks in the positive tweets plot precede a significant change in the Bitcoin's price after some days.

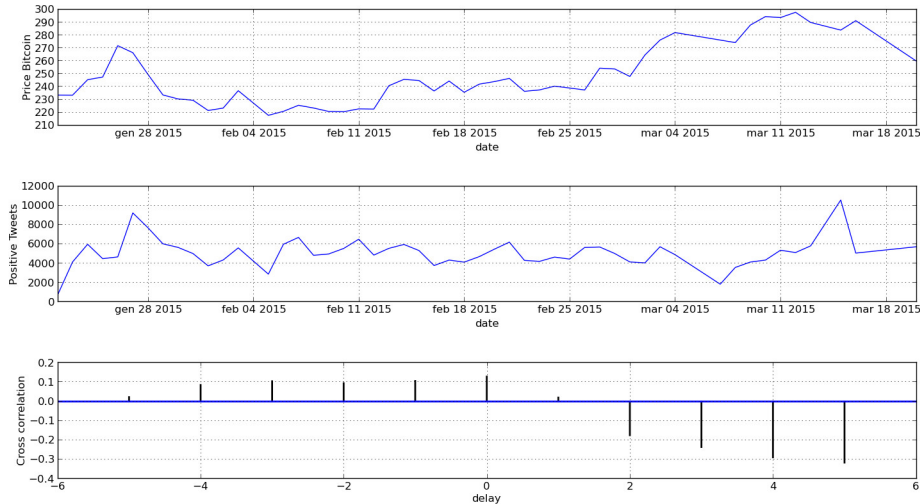


Fig. 3. Cross-correlation between positive Tweets and Bitcoin's price

The cross-correlation result between Google Trends data and Bitcoin's price also looks significant. The cross-correlation value increase up to a value of 0.64, that is quite substantial. This result is shown also by a little significant relationship that exists between positive tweets and Google Trends data. Figure 4 shows how Google Trends proceeds in the same direction of Bitcoin's price and highlighting a striking similarity between them. Table 1 summarizes the cross-correlation results, obtained comparing the spread among Bitcoin price and different volumes of data.

Table 1. Cross-correlation results

Compared Systems	Cross-correlation value	delay
Bitcoin price-Tweets volume	0.15	1
Bitcoin price-Positive tweets	-0.35	3-4
Bitcoin price-Google Trends data	0.64	0

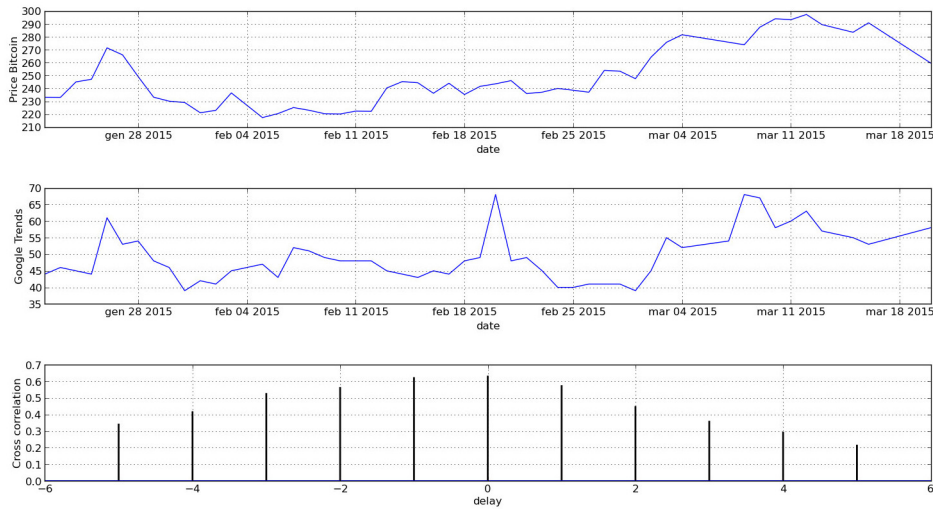


Fig. 4. Cross correlation between Google Trends and Bitcoin’s price, expressed in dollars

5 Conclusions

In this paper, we studied whether social media activity or information extracted by web search media could be helpful and used by investment professionals in Bitcoins. Since the use of Bitcoins is increasingly widespread, we decided to analyze the market, in order to predict the evolution of its price.

To this purpose, we presented an analysis of a corpus of tweets about Bitcoin, considering a total amount of 1,924,891 tweets. The corpus covers a period of 60 days between January 2015 and March 2015. We applied automated Sentiment Analysis on these tweets in order to evaluate whether public sentiment could be used to predict Bitcoin’s market. We also used Google Trends media to analyze Bitcoin’s popularity under the perspective of Web search. In this preliminary study, we examined the Bitcoin price’s behavior comparing its variations with these of tweets volume, tweets with positive mood volume and Google Trends data. From results of a cross correlation analysis between these time series, we can affirm that positive tweets may contribute to predict the movement of Bitcoin’s price in a few days. Google Trends could be seen as a kind of predictor, because of its high cross correlation value with a zero lag. Our results confirm those found in the previous works, based on a different corpus of tweets and

referred to a different Bitcoin market trend.

While the current data is only 60 days already looks promising, a consecutive analysis of more than 6 months might provide a better result quality. In further studies, we also plan to take into account the number of retweets and favorites for the tweet's corpus analyzed. Along these lines, we could check whether results stay unchanged with the addition of this variable.

Acknowledgments. This research is supported by Regione Autonoma della Sardegna (RAS), Regional Law No. 7-2007, project CRP-17938 LEAN 2.0.

References

1. Kaminski, Jermain, and Peter Gloor. "Nowcasting the Bitcoin Market with Twitter Signals." arXiv preprint arXiv:1406.7577 (2014).
2. Rao, Tushar, and Saket Srivastava. "Analyzing stock market movements using twitter sentiment analysis." Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012). IEEE Computer Society, 2012.
3. Garcia D, Tessone CJ, Mavrodiev P, Perony N. 2014 The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. J. R. Soc. Interface 11: 20140623. <http://dx.doi.org/10.1098/rsif.2014.0623>
4. Mai, Feng and Bai, Qing and Shan, Zhe and Wang, Xin (Shane) and Chiang, Roger H.L., From Bitcoin to Big Coin: The Impacts of Social Media on Bitcoin Performance (January 6, 2015). Available at SSRN: <http://ssrn.com/abstract=2545957>
5. A. Mittal, A. Goel, "Stock Prediction Using Twitter Sentiment Analysis" in proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2013
6. Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of Computational Science 2.1 (2011): 1-8.
7. K. Dessai and M. Kamat, Application of social media for tracking knowledge in agile software projects, Available at SSRN 2018845, 2012.
8. Google Trends. <http://www.google.it/trends/> . Accessed November 6,2014.
9. Thelwall, M., Buckley, K., Paltoglou, G. : Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406-418 (2011).
10. Twitter [online] <https://twitter.com/> .Accessed November 6,2014.
11. SentiStrenght. [Online] <http://sentistrength.wlv.ac.uk/> .Accessed November 6,2014.
12. Alexa. Top sites: The top 500 sites on the Web. Accessed September 30, 2014, from <http://www.alexa.com/topsites/global>
13. Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., Kappas, A. : Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558 (2010).
14. Java, A., Song, X., Finin, T., Tseng, B. Why we twitter: Understanding microblogging usage and communities. Proceedings of the Ninth WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (pp. 56–65). New York: ACM Press (2007).

15. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 1(1–2), 1–135. (2008).
16. Twitter API. <https://dev.twitter.com/rest/tools/console> . Accessed November 25, 2014.
17. SOCIAL MEDIA AND MARKETS: The New Frontier
18. S. Ye and F. Wu. Estimating the size of online social networks. In *Proc. of the IEEE 2nd Intl. Conf. on Social Computing (SocialCom)*, pages 169–176, Aug. 2010
19. A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report*, Stanford, 2009.
20. Grinberg, Reuben, Bitcoin: An Innovative Alternative Digital Currency, *Hastings Science & Technology Law Journal*, Vol. 4, p.160, Dec 2011
21. Dorit. Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph.
22. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system.
23. F. Reid and M. Harrigan. An analysis of anonymity in the bitcoin system. In *Proc. of the Conference on Social Computing (socialcom)*, 2011. Reid F. Harrigan M.
24. Pak, A and Paroubek, P. (2010) *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. (European Language Resources Association (ELRA), Valletta, Malta).