The Predictor Impact of Web Search Media On Bitcoin Trading Volumes

Martina Matta, Ilaria Lunesu and Michele Marchesi
Universita' degli Studi di Cagliari
Piazza d'Armi, 09123 Cagliari, Italy
{*martina.matta, ilaria.lunesu, michele*}*@diee.unica.it*

Keywords: Bitcoin, Web Search Media, Google Trends, Cross Correlation Analysis

Abstract: In the last decade, Web 2.0 services have been widely used as communication media. Due to the huge amount of available information, searching has become dominant in the use of Internet. Millions of users daily interact with search engines, producing valuable sources of interesting data regarding several aspects of the world. Search queries prove to be a useful source of information in financial applications, where the frequency of searches of terms related to the digital currency can be a good measure of interest in it. Bitcoin, a decentralized electronic currency, represents a radical change in financial systems, attracting a large number of users and a lot of media attention. In this work we studied the existing relationship between Bitcoin's trading volumes and the queries volumes of Google search engine. We achieved significant cross correlation values, demonstrating search volumes power to anticipate trading volumes of Bitcoin currency.

1 INTRODUCTION

Internet has been one of the most revolutionary technologies in the last decades. The majority of daily activities radically changed, moving towards a "virtual sector", such as Web actions, credit card transactions, electronic currencies, navigators, games, etc. In recent years, web search and social media have emerged online. On one hand, services such as blogs, tweets, forums, chats, email have gained wide popularity. Social media data represent 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 (Kaplan and Haenlein, 2010).

Social media technologies have produced completely new ways of interacting (Hansen et al., 2010), bringing the creation of hundreds of different social media platforms (e.g., social networking, shared photos, podcasts, streaming videos, wikis, blogs). On the other hand, due to the huge amount of available information, searching has become dominant in the use of Internet. Millions of users daily interact with search engines, producing valuable sources of interesting data regarding several aspects of the world.

Recent studies demonstrated that web search streams could be used to analyze trends about several phenomena (Choi and Varian, 2012) (Rose and Levinson, 2004) (Bordino et al., 2012). In one of the most interesting works, Ginsberg et al. proved that search query volume is a sophisticated way to detect regional outbreaks of influenza in USA almost 7 days before CDC surveillance (Ginsberg et al., 2009). There are also studies that report another use in a search engine, namely as a possible predictor of market trends. Bollen et al. show that search volumes on financial search queries have a predictive power. They compared these volumes with market indexes such as Dow Jones Industrial Average, trading volumes and market volatility, demonstrating the possibility to anticipate financial performances (Bollen et al., 2011). 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. Bordino et al. prove that search volumes of stocks highly correlate with trading volumes of the corresponding stocks, with peaks of search volume anticipating peaks of trading volume by one day or more (Bordino et al., 2012).

Search queries prove to be a useful source of information in financial applications, where the frequency of searches of terms related to the digital currency can be a good measure of interest in the currency and it has a good explanatory power (Kristoufek, 2013). Mondria et al. proved that the number of clicks on search results stemming from a given country corre-

lates with the amount of investment in that country (Mondria et al., 2010). Further studies showed that changes in query volumes for selected search terms mirror changes in current volumes of stock market transactions (Preis et al., 2010).
Technology always had a strong impact on financial markets and it has favored the emergence of Bitcoin, a digital currency created in 2008 by Satoshi Nakamoto (Nakamoto, 2008). It has been created for the purpose to replace cash, credit cards and bank wire transactions. It is based on advancements in peer-to-peer networks (Ron and Shamir, 2013) and cryptographic protocols for security. Due to its properties, Bitcoin is completely decentralized and not managed by any governments or bank, ensuring anonymity. It is based on a distributed register known as ”block-chain” to save transactions carried out by users. Like any other currency, a peculiarity of Bitcoin is to facilitate transactions of services and goods with vendors that accept Bitcoins as payment(Grinberg, 2012), attracting a large number of users and a lot of media attention.
The Bitcoin represents an important new phenomenon in financial markets. Mai et al. 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 auto-regressive vector and error correction vector models (Mai et al., 2015).
Matta et al. examined the striking similarity between Bitcoin price and the number of queries regarding Bitcoin recovered on Google search engine (Matta et al., 2015). In their work, Garcia et al. (Garcia et al., 2014) proved 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 provided evidence that Bitcoins growing popularity causes an increasing search volumes, which in turn result a higher social media activity about Bitcoin. A growing interest inspires the purchase of Bitcoins by users, driving the prices up, which eventually feeds back on the search volumes.
There are several works that present predictive relationships between social media and bitcoin volume[1] 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 (Constantinides et al., 2009) or linear regression analysis (Bollen et al., 2011) (Mittal and Goel, 2012). Social factors, that are composed of interactions among market actors, may strongly drive the dynamics of Bitcoin's economy (Garcia et al., 2014).
In this work we study the relationship that exists between trading volumes of Bitcoin currency and the queries volumes of search engine. The frequency of searches of terms about Bitcoin could be a good explanatory power, so we decided to examine Google, one of the most important search engine. We studied whether web search media activity could be helpful and used by investment professionals, analyzing the search volumes power of anticipate trading volumes of the Bitcoin currency.
We compared USD trade volumes about Bitcoin with those in a media, namely, Google Trends. This is a feature of Google search engine that illustrates how frequently a fixed search term was looked for. Following this kind of approach, we evaluated how much *bitcoin* term, for the specific time interval, is looked for using Google's search engine.

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

2 METHODOLOGY

2.1 Google Trends

Google Trends [2] is a feature of Google Search engine that illustrates how frequently a fixed term is looked for. Through this, you can compare up to five topics at one time to view their relative popularity, allowing you to gain an understanding of the hottest search trends of the moment, along with those developing in popularity over time. The system provides a time series index of the volume of queries inserted by users into Google.
Query index is based on the number of web searches performed with a specific term compared to the total amount of searches done over time. Absolute search volumes are not illustrated, because the data are normalized on a scale from 0 to 100.
Google classifies search queries into 27 categories at the top level and 241 categories at the second level through an automatic classification engine. Indeed, queries are given out to fixed categories due to natural language processing methods.
The query index data are available as a CSV file in order to facilitate research purposes. Figure 1 depicts

[1] https://markets.blockchain.info/

[2] http://trends.google.com

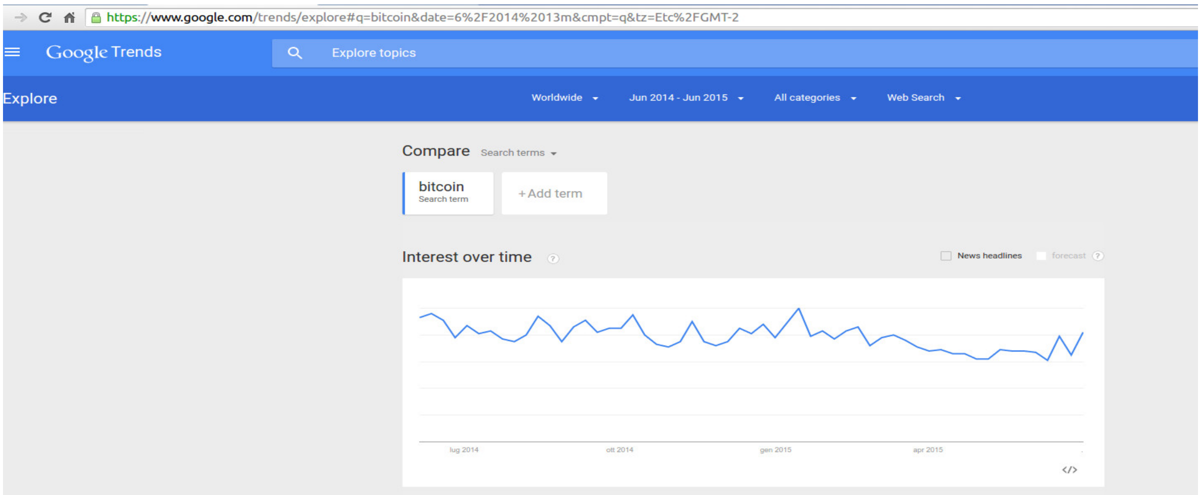

Figure 1: Example of Google Trends usage for the query "Bitcoin".

an example from Google Trends for the query "Bitcoin". We downloaded data about how much the term "Bitcoin" was referred to last year.

2.2 Blockchain.info

Blockchain.info[3] is an online system that provides detailed information about Bitcoin market. Launched in August 2011, this system shows data on recent transactions, plots on the Bitcoin economy and several statistics. It allows users to analyze different Bitcoin aspects:

- Total Bitcoins in circulation
- Number of Transactions
- Total output volume
- USD Exchange Trade volume
- Market price (USD)

We decided to study a time series regarding the USD trade volume from top exchanges, analyzing its trends.

2.3 Data Collection

Search query volumes regarding Bitcoin were collected from *Google Trends* website, capturing all searches, inserted from June 2014 to July 2015, with "Bitcoin" word as keyword .
Trading volume data were acquired from *blockchain.info* website, in order to evaluate daily trends of Bitcoin currency. We assessed the relationship over time between number of daily queries related to the trading volume of Bitcoin.

[3]http://www.blockchain.info

To better understand whether search engine can be seen as a good predictor of trading volumes, we applied an analysis of correlation between these data expressed in time series, a time-lagged cross-correlation study, concluding with a Granger-causality test.

3 RESULTS

In order to decide the correct strategy of analysis for studying the relationship among Bitcoins trading volume and others meaningful parameters, the available related literature has been examined in depth. Most of articles (Bollen et al., 2011) (Kaminski and Gloor, 2014) (Rao and Srivastava, 2012) reports analysis about the existent relationship between volume of media and market evolution. In general, Bollen et al. proved that tweets can predict market trend 3-4 days in advance, with a good chance of success. We extract from both data sources time series composed by daily values in the time interval ranging from June 2014 to July 2015 in order to evaluate their relationship and the capability of prediction. We run statistical analysis and the computation of correlation, cross-correlation and Granger causality test yielded interesting results.

3.1 Pearson Correlation

Pearson's correlation r is a statistical measure that evaluate the strength of a linear association between two time series G and T. We assumed G as query data and T as trading volumes.

$$r = \frac{\sum_i (G_i - \overline{G})(T_i - \overline{T})}{\sqrt{\sum_i (G_i - \overline{G})^2}\sqrt{\sum_i (T_i - \overline{T})^2}} \quad (1)$$

The correlations have values between -1 and +1, the bounds indicate maximum correlation and 0 indicating no correlation. A high negative correlation indicates a high correlation but of the inverse of one of the series. We calculated the Pearson correlation between queries search data and trading volume and we found a result equal to 0.60. This similarity is also clearly visible in the figure 2.

Following this kind of analysis, we demonstrated the striking similarity existing between the time series. This result means that the trading volumes follows the same direction pace of queries volumes. Figure 3 reveals an obvious correlation due to peaks in one time series that occur close to peaks in the other. In this Figure it is possible to see that solid line, correspondent to search volumes, very often anticipated the dotted line correspondent to trading volumes. The most significant peaks occurred in the interval between August and September 2014, between September and October 2014, between November and December 2014 and between January and February 2015. During other periods the same phenomenon is less evident but anyway present.

Radical changes in peaks are due to several factors. One of the most evident peak is visible in Figure 3 corresponding to the interval between end of June and beginning of July. This is the period of the greek crisis acme, that causes changes also in the Bitcoin market. Indeed, a lot of people already started to invest in Bitcoin business. When people try to move money out of the country the government blocks this process, thus Bitcoin are the only way to transfer their wealth. In fact Greeks would use bitcoin to protect the value of their money at home. Ten times more Greek than usual are being recorded at the company *'German Bitcoin.de'*[4] to buy electronic currency. This situation is clearly visible in the right part of Figure 3, where curve correspondent to queries index volumes regarding Bitcoin considerably grew up, followed by an increase of curve correspondent to trading volumes after some days. In these mentioned cases it is clear how search volumes predict trading volumes preceding it, as confirmed by correlation values.

3.2 Cross Correlation

We investigated whether query volumes can anticipate trading volume of Bitcoin. We calculated the cross correlation values between query data G and trading volumes T as the time lagged Pearson cross correlation between two time series G and T for all delays d=0,1,2,..5.

$$r(d) = \frac{\sum_i (G_i - \overline{G})(T_{i-d} - \overline{T})}{\sqrt{\sum_i (G_i - \overline{G})^2}\sqrt{\sum_i (T_{i-d} - \overline{T})^2}} \quad (2)$$

We chose to evaluate a maximum lag of five days and, also in this case, the correlation ranges from -1 to 1. In Table 1, the results obtained from these experiments are reported. Each column shows the cross correlation result corresponding to different time-lag. We can observe that cross correlation results for positive delays are always higher than the ones with negative time lag. Indeed, the results with positive delays achieve values always higher than 0.64 and with negative delays report values always lower than 0.55. It means that query volumes is able to anticipate trading volumes in almost 3 days.

Figure 4 shows the cross correlation results with a maximum lag of 30 days, just to highlight that the best result is given by a lag of almost 3.

3.3 Granger Causality

We performed a Granger causality test in order to verify whether web search queries regarding Bitcoin are able to anticipate particular trends in some days. The Granger-causality test is used to determine whether a time series G(t) is a good predictor of another time series T(t) (Granger, 1969). If G Granger-causes T, then G^{past} should significantly help predicting T^{future} via T^{past} alone. We compared query volumes G with trading volume T with the null hypothesis being that T is not caused by G. An F-test is then used to determine if the null hypothesis can be rejected.
We performed two auto-regression vectors as follows in the formula 3 and 4, where L represents the maximum time lag.

$$T(t) = \sum_{l=1}^{L} a_l T(t-l) + \varepsilon_1 \quad (3)$$

$$T(t) = \sum_{l=1}^{L} a'_l T(t-l) + \sum_{l=1}^{L} b'_l G(t-l) + \varepsilon_2 \quad (4)$$

We can affirm that G causes T if eq(4) is statistically better significant than eq(3). We applied the test in both directions, as an instance G $\rightarrow$ T means that the null hypothesis is “G doesn’t Granger-cause T”.
Table 2 shows the results of the Granger causality test, where the first column represents the direction of the applied test, the second one the delay, and then the F-test result with its p-value. This parameter represents the probability that statistic test would be at least

[4] https://www.bitcoin.de/

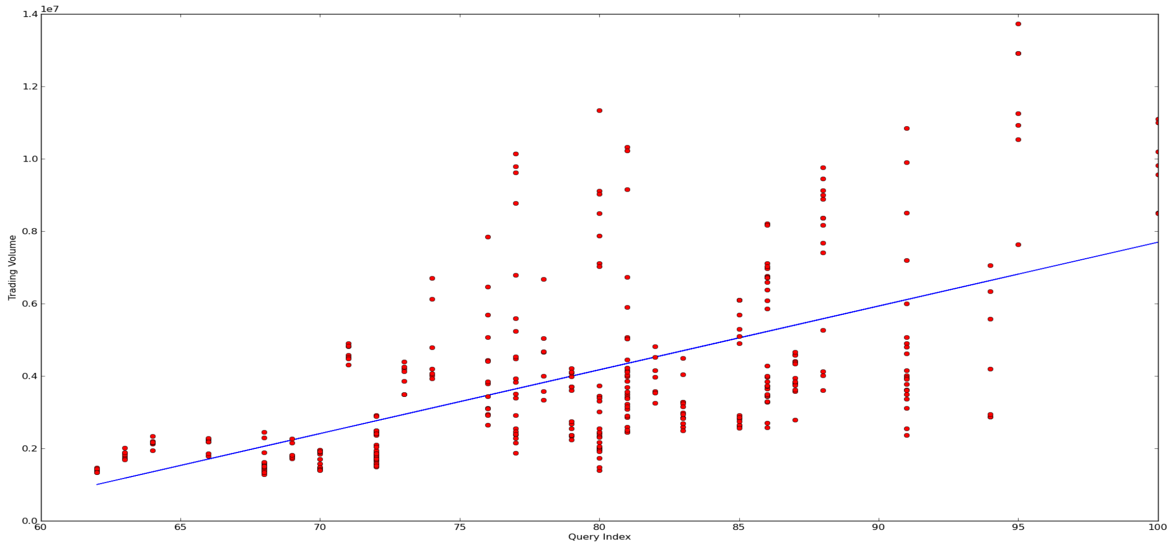

Figure 2: Correlation between Trading Volume and Queries Volume about Bitcoin.

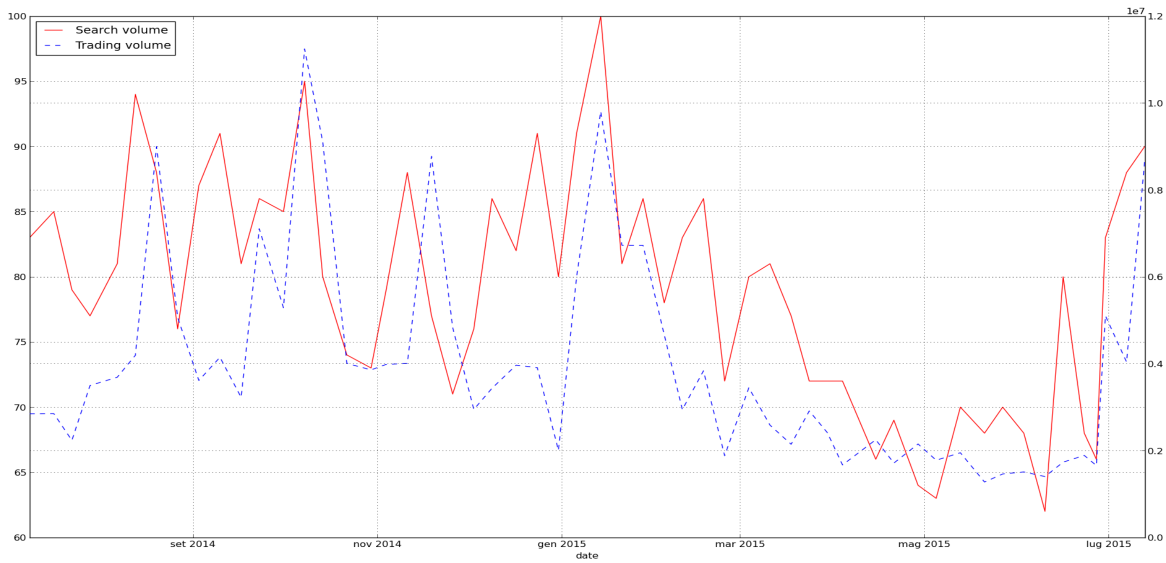

Figure 3: Correlation between Trading Volume and Queries Volume about Bitcoin.

Table 1: *Cross-correlation results*

Delay	-5	-4	-3	-2	-1	0	1	2	**3**	4	5
Cross-Corr Value	0.36	0.40	0.44	0.50	0.55	0.60	0.64	0.67	**0.68**	0.67	0.64

as extreme as observed, if the null hypothesis were true. So, we reject the null hypothesis if p-value is inferior to a certain threshold ($p<0.05$). Our analysis demonstrated that trading volumes can be considered Granger-caused by the query volumes. It is clearly shown that time-series G influences T, given by the p-value <0.001 for lags ranging from 1 to 5. So, the null hypothesis is completely rejected. On the other hand, the F-value test applied to the direction T$\rightarrow$G reported a p-value always greater than 0.1. Trading volume T doesn't have significant casual relations with changes in queries volumes on Google search engine G. So, null hypothesis cannot be rejected.

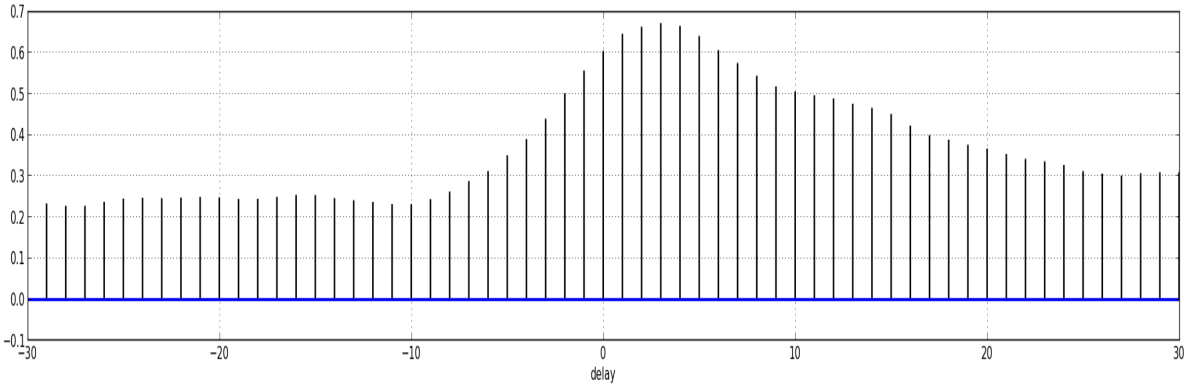

Figure 4: Cross Correlation results between Trading Volume and Queries Volume about Bitcoin with a maximum lag of 30 days.

Table 2: *Granger-causality tests*

Direction	Delay	F-value Test	P-value
G→T	1	41.8135	p<0.001
	2	15.1435	p<0.001
	3	12.9332	p<0.001
	4	15.1546	p<0.001
	5	12.9279	p<0.001
T→G	1	0.5450	p=0.46
	2	2.3006	p=0.10
	3	1.4878	p=0.21
	4	1.5336	p=0.19
	5	1.2297	p=0.29

4 CONCLUSIONS

In this paper, we evaluated whether the 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 trading volume.
To this purpose, we presented an analysis of a corpus of queries index about Bitcoin compared to its trading volume. We selected a corpus that covers a period of almost one year, between June 2014 and July 2015. We chose Google Trends media to analyze Bitcoins popularity under the perspective of Web search. We examined the Bitcoin tradings behavior comparing its variations with Google Trends data. From results of a cross correlation and Granger causality analysis between these time series, we can affirm that Google Trends is a good predictor, because of its high cross correlation value. Our results confirm those found in previous works, based on a different corpus and referred to a different Bitcoin market trend.
As future advancement, we are thinking about the possibility to apply this kind of approach to different contexts in order to better understand the predictive power of web search media. An other likelihood could be to consider not only search media but also social media like Twitter, Facebook and Google+.

Acknowledgments.

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

REFERENCES

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., and Weber, I. (2012). Web search queries can predict stock market volumes. *PloS one*, 7(7):e40014.

Choi, H. and Varian, H. (2012). Predicting the present with google trends. *Economic Record*, 88(s1):2–9.

Constantinides, E., Romero, C. L., and Boria, M. A. G. (2009). Social media: a new frontier for retailers? In *European Retail Research*, pages 1–28. Springer.

Garcia, D., Tessone, C. J., Mavrodiev, P., and Perony, N. (2014). The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *Journal of the Royal Society Interface*, 11(99):20140623.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Grinberg, R. (2012). Bitcoin: an innovative alternative digital currency. *Hastings Sci. & Tech. LJ*, 4:159.

Hansen, D., Shneiderman, B., and Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann.

Kaminski, J. and Gloor, P. (2014). Nowcasting the bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577*.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68.

Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3.

Mai, F., Bai, Q., Shan, Z., Wang, X. S., and Chiang, R. H. (2015). From bitcoin to big coin: The impacts of social media on bitcoin performance.

Matta, M., Lunesu, I., and Marchesi, M. (2015). Bitcoin spread prediction using social and web search media. *Proceedings of DeCAT*.

Mittal, A. and Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University, CS229*.

Mondria, J., Wu, T., and Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82(1):85–95.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28.

Preis, T., Reith, D., and Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1933):5707–5719.

Rao, T. and Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 119–123. IEEE Computer Society.

Ron, D. and Shamir, A. (2013). Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security*, pages 6–24. Springer.

Rose, D. E. and Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, pages 13–19. ACM.