Case Study 1: Twitter Analysis for Cryptocurrency Investment

Group 1: Claire Danaher
Jonathan Friedman
Janvi Kothari
Renee Sweeney
Erin Teeple

DS 501 September 9, 2017

Tables and Figures

Tables

Figures

Motivation

Blockchain is a technology that is an “incorruptible digital ledger of economic transactions” (1). Digital cryptocurrencies are built off of this blockchain technology and are used for recording peer-to-peer network transactions without verification of transfers by a central authority such as a bank (2, 3). This allows trading to potentially be conducted at any time (2, 3). These currencies fluctuate in value relative to other currency forms, thus they can be traded for profit. In order to realize a profit, trades must occur when currency has increased in value and before value is lost. For this project, we explore Twitter mentions and sentiments for several commonly traded cryptocurrencies: Bitcoin, Litecoin, XRP, Etherium, and Namecoin. Twitter data has the potential to capture temporal trends in cryptocurrency name recognition and sentiments towards cryptocurrency investments. The business application explored in this project is the use of Twitter data to inform and time cryptocurrency investments. Understanding which currencies dominate the market, how sentiments towards cryptocurrency fluctuate, and how such information from Twitter relates to price changes provides information to guide investment choices and to possibly identify times of opportunity for profitable trades.

Scientific Approach and Analysis Plan

With these considerations in mind, for this case study, we sampled Twitter data over five days at three times each day: morning (approximately 7AM), midday (approximately noon) and evening (approximately 7PM), Eastern Standard Time. Twitter data was processed according to the case study instructions and the results of these procedures were then interpreted with respect to our identified business interest. These analyses included focused word frequency assessment, sentiment fluctuation analysis, and an examination of the relationships between Twitter use patterns related to cryptocurrency, tweet sentiments, and cryptocurrency prices. Methods used in this study for collecting and processing Twitter data are based on those previously published by M.A. Russell (4). Our Twitter search terms were 'bitcoin', 'XRP', 'etherium', 'litecoin', and 'namecoin'.

We also identified a popular user that tweets about the topic of cryptocurrency (@cryptocointalk). We obtained lists of this user’s friends and followers and applied setwise operations to identify “mutual friends”. Such mutual friends are individuals in the friends list who are also being followed by the popular user. This information can provide some insight into the community interested in receiving information from the user @cryptocointalk. A high number of mutual friends suggests that the information in the user’s tweets is being shared within the context of a more connected community. In contrast, a finding of few intersections suggests a less personally connected audience following this popular user for information (4).

Word frequency comparisons allowed us to assess the most frequent content and variety of language used in tweets mentioning cryptocurrency. Sentiment analysis using Textblob (5) was then further applied to explore volatility of tweet-expressed sentiments toward cryptocurrency topics. As the final component of our analysis, fluctuations in relative cryptocurrency prices were then compared to variations in Twitter sentiment toward cryptocurrency by day of data collection, in order to assess whether there might exist a relationship between Twitter sentiment and cryptocurrency prices.

Results

Problem 1 - Sampling Twitter Data with Streaming API about Our Topic

The total number of tweets collected was 21,044. To investigate the share of Twitter activity related to our cryptocurrencies of interest, we obtained Tweet counts at varied time points over several days (Table 1). Comparing morning, midday, and evening samples, mean number of tweets returned in our samples (mean ∓ standard deviation) were the following: Morning 2094.5 ∓ 719.5 tweets/sample; Midday 2308.3 ∓ 528.3 tweets/sample; and Evening 1435.3 ∓ 395.1 tweets/sample. The samples containing the fewest tweets were obtained on a Saturday evening (848 tweets/sample) and Saturday morning (1727 tweets/sample).

Table 1. Tweet counts by date and time of day of collection

	Time		
Date	Morning	Midday	Evening
09/13/17 Wed	-	1968	1645
09/14/17 Thur	3162	2040	1688
09/15/17 Fri	1874	2917	1560
09/16/17 Sat	1615	-	848
09/17/17 Sun	1727	-	-
Total			21,044

An interesting consideration regarding these tweet counts is that since the tweets about our topic of interest obtained for each time point are only a small percentage of tweets being sent out during that interval, these numbers are relative to the total number of tweets being circulated. It therefore cannot be definitively stated that the absolute number of tweets about cryptocurrency is lower on weekends, only that the relative fraction of tweets discussing the cryptocurrencies of interest was observed to be somewhat lower in some of our samples, in particular those obtained on Saturday, September 16, 2017.

Problem 2 - Frequency Analysis: Tweets and Tweet entities

Word frequencies were computed for the top 30 words used in tweets about our topic of interest. These words and and their frequencies are presented in Table 2. We see that bitcoin, the cryptocurrency with the largest share of this market, appears earlier than the other searched cryptocurrency names. Thematically, these top words can be qualitatively described as focusing on coin technology, trading, and pricing, and they suggest relatively focused interests among Twitter users tweeting about cryptocurrencies.

Table 2. Top words used and word counts

Top Words	Count
https	5173
co	5017
bitcoin	3375
xrp	2690
namecoin	2372
btc	2278
rt	2183
1	2101
usd	1943
0	1851
litecoin	1816
eth	1645
ltc	1442
cryptocurrency	1412
ripple	1147
etherium	1051
30	880
mins	809
nmc	787
ethereum	735
etc	700
price	699
dash	634
neo	629
crypto	572
blockchain	553
market	491
changed	443
omg	434
bch	412

The most popular 10 tweets, defined as those with the largest number of retweet counts, are shown in Table 3. Based on examination of tweets written in English, a similar focus on cryptocurrency technology, trading, and pricing also appears. We likewise identified the most popular Tweet entities in our collection of tweets. Table 4 presents the top 10 hashtags and top 10 user mentions from our collection of Tweets. Once again, the themes of technology, trading, and pricing appear in this collection.

Table 3. Most popular 10 tweets and retweet counts

Count	Top Tweets
34169	"RT @lorde: VOGUE BABY AHHH! ✨🌙☺ https://t.co/pxHDeyeFhd"
28713	"RT @stem910: 先日頭のいい友人に「喋り方がバカっぽいのをなんとかしたい」と相談したら「バカっぽいというのは本当はバカではないのにバカと思われてしまうということだが、お前は事実バカだろ」と指摘されたので「なる程。ではバカを強調する喋り方をやめたい」と相談し直したら“あ…"
14256	"RT @tsumland_jp: ✨最大100万円分の海外ディズニー旅行が当たるかも✨ ディズニー ツムツムランドの事前登録キャンペーンを実施中！ このアカウントをフォローするだけで豪華賞品が当たるチャンス🎵 この機会をお見逃しなく☆ ＃ツムツムランド 詳細↓ https://…"
7079	"RT @aco220: 友達が「彼氏と別れそう」って言うから、何事だ何があったのかどうしたのかって聞いたら、「Excelで下にずらして数字が並ぶの知らない人とは付き合えない」って、オートフィルわかんないとフラれるのやばいな"
5978	"RT @emartineeez: RT THIS TWEET IF YOU WANT TWO VIDEOS TODAY😁"
5651	"RT @at_raku: 東京ドームシティ アトラクションズでは、9/30（土）-「活撃 刀剣乱舞」とのコラボイベント『出陣！活撃 刀剣乱舞 in 東京ドームシティ』を開催！限定クリアファイルが貰えるスタンプラリー、コラボフード&グッズも販売！https://t.co/KDW2…"
5604	"RT @ErikVoorhees: My memory is failing, was it Bitcoin or was it JP Morgan that was bailed out by the government? https://t.co/DHqFzr5UJN"
3828	"RT @huga731: #btc #bitcoin #xem #comsa #nem 100 000 уже зарегистрировалось в Японском ico COMSA .Это будет самое масштабное айсио💪https://t…"
3244	"RT @imartinezp_: RETWEET MY LAST TWEET FOR A FOLLOW"
2864	"RT @RivetzCorp: Rivetz Enables Provable Cybersecurity #cryptocurrency #blockchain #ethereum #bitcoin #btc #bitcoins #ICO #token #crowdsale…"

Table 4. Top 10 hashtags and users mentioned

Top Hashtags	Count
#bitcoin	1188
#cryptocurrency	1180
#namecoin	737
#nmc	708
#Bitcoin	687
#litecoin	567
#etherium	475
#Litecoin	421
#XRP	376
#xrp	352

Top Users Mentioned	Count
@Ripple	125
@Keiki_XRP	116
@cryptopayments2	71
@erishiiiii	71
@SatoshiLite	69
@TO30447473	68
@toyokichimaru	66
@cannavinothc	59
@MarketNmc	51
@cryptographicc	49

Problem 3 - Getting all friends and all followers of a particular user

Examining intersections among the friends and followers of popular user whose tweets focus on cryptocurrency topics allows us to further consider the social dynamics of these discussions. We chose the popular user (@cryptocointalk) and obtained the list of friends (n = 998) and followers (n = 29,344) for this user. A plot listing 20 followers of this user by identification (ID) number and screen name is presented in Table 5. Table 6 similarly presents a plot of 20 friends of this user.

Table 5. Screen names and ID numbers of followers of our popular user

	Followers ID	Followers Screen name
0	865297070768259077	crypto_Pickle
1	2233501680	luishut
2	2460243694	mistertapps
3	890289232484466688	YohannesGee
4	905865817501429761	coolix1254
5	886367528	vrdci
6	899895351641088000	trader_altcoin
7	907853605855744000	Mikey105uk
8	142529167	elpablo090
9	3314023754	rakibnazir4
10	147753974	xarihanTW
11	339623336	sadtol
12	908493566476820480	AlphaCvk
13	785477330651217920	AmirZokan
14	732148431695818752	SedatDin13
15	591770023	skliask
16	2427122690	Station12vc
17	896782003856695297	TheCryptoCorn
18	864987913846575104	_CryptoBeggar
19	4871612482	VRMobile_Com

Table 6. Screen names and ID numbers of friends of our popular user

	Friends ID	Friends Screen name
0	812491802955644928	EmbermineDrake
1	797930343277989888	TheEmbermine
2	104259801	Squidoogeek
3	262755165	SkinnerLiber8ed
4	788886561731538944	YVerif
5	902772556670914560	VUnioninfo
6	1479248557	msjemmagreen
7	4341007829	uquidcard
8	897380780510502912	CryptoTickets
9	2309343738	OfficialTitcoin
10	888615399281102848	GoldenFleece_co
11	2600194316	NewKoreCoin
12	1664012648	chimaera_tech
13	2313671966	NEMofficial
14	221071894	drbitq
15	867574261195522048	matryx_ai
16	873598891723239424	BlockDevCo
17	769176690908069888	godzillion_io
18	775658150179508224	iEx_ec
19	857130360903274500	Bitcore_BTX

Mutual friends are those users in the followers group who are also in the friends group. Table 7 presents the ID numbers and screen names of users in this mutual friends group. The total number of mutual friends and followers was found to be 995. Consistent with our observations about the frequently used words, users, and popular hashtags, many of these mutual friends reference technology, trading, or coins in their screen names. The relatively larger number of followers compared to friends (29,344 followers versus 998 friends) suggests that the information tweeted out by this popular user is delivered to an interested audience, rather than being shared within a community of individuals mostly known to each other outside of their Twitter connection.

Table 7. ID numbers and screen names of users in the mutual friends group

	Mutual ID	Mutual Screen name
0	788886561731538944	YVerif
1	902772556670914560	VUnioninfo
2	897380780510502912	CryptoTickets
3	888615399281102848	GoldenFleece_co
4	867574261195522048	matryx_ai
5	873598891723239424	BlockDevCo
6	769176690908069888	godzillion_io
7	775658150179508224	iEx_ec
8	857130360903274500	Bitcore_BTX
9	895895550494728192	MonacoCoin
10	890901570312957953	hearthyco
11	874241762075832322	mindpass2050
12	848547538957328387	bonzocorleonee
13	867498616730025985	HonestisN
14	867456902627721218	QchainPlatform
15	824342681912561678	MARXCOLLECTIVE1
16	862041575081246723	CONTENT_COIN
17	274456588	cryptocoinfan
18	2335207442	coinshost
19	3662979086	customminer
20	43192340	JackPhan
21	2150123534	newsbtc
22	17186834	djspang
23	42584086	siavash
24	240633881	BenBschor
25	1950640170	joebitcoinorg

Problem 4 - Using Twitter data to answer a business question

Suppose a company was debating whether to diversify their investment portfolio to include cryptocurrencies. Cryptocurrencies are highly unregulated and as such, information-gathering approaches differ from those used for more conventional regulated markets. Therefore, a potential business concern would be formulating a strategy for monitoring this unique market. Twitter, similar to the cryptocurrency environment, is also highly unregulated and attracts a segment of the population actively engaged with technology. Given these commonalities, our team theorized that participants in the cryptocurrency market might be relatively more likely to use Twitter to obtain and share information about this market. Such self-selected participation by cryptocurrency traders and the ability to aggregate substantial volumes of Twitter data could then possibly be used to track, and perhaps even predict, cryptocurrency markets. Thus, we designed and implemented an experiment to test this theory.

In addition to the analyses requested in the project instructions, the text from the collected tweets was further analyzed to assess the sentiment of the users generating tweets about cryptocurrency. Sentiment analysis was performed using the Python add-on TextBlob (5). Textblob is a natural language processing program which scores text strings based on the domains of polarity and subjectivity. Scores range from highly negative [-1] to highly positive [1] and very objective [0] to highly subjective [1].

To determine whether Twitter sentiment might be related to cryptocurrency price fluctuations, we looked at sentiment versus various different price metrics. After studying the data, we chose to focus on Twitter sentiment positivity over the data collection period as other data sets such as negative sentiment seemed to be completely independent of the collected data. We also obtained data on price changes and directionality, rate of price change, for each of the cryptocurrencies included in our search during the data collection period (6, 7). Figure 1 shows the relative price changes among our cryptocurrencies of interest during the study period. Figure 2 presents price directionality together with positive cryptocurrency sentiment trend. We see clearly in Figure 2 that an increase in price directionality across all cryptocurrency types is preceded during our study period by an uptick in positive sentiment (Figure 2, data column segments 5-6). Individual graphs of overall cryptocurrency sentiment versus individual currency values are presented in Figure 3 and also show most cryptocurrency values appearing to track somewhat together with cryptocurrency sentiment. Supplementary trading views for each of the cryptocurrencies of interest are also included in the Appendix. Among the cryptocurrencies studied, Namecoin (NMC) is notable for having some differences in its sentiment-value relationships. NMC is both outside the tight band of price fluctuations in Figure 1, and it also demonstrates different trends in comparison to the other studied cryptocurrencies with respect to price fluctuations versu positive sentiment. We theorize that this observed variant behavior may be due to NMC being a much less mainstream, less frequently traded, and differently valued cryptocurrency. These characteristic may account for differences in its price response to the social trends that appear to drive the price movements of the other main cryptocurrencies we studied.

Figure 1. Relative price change of selected cryptocurrencies over five-day data collection period

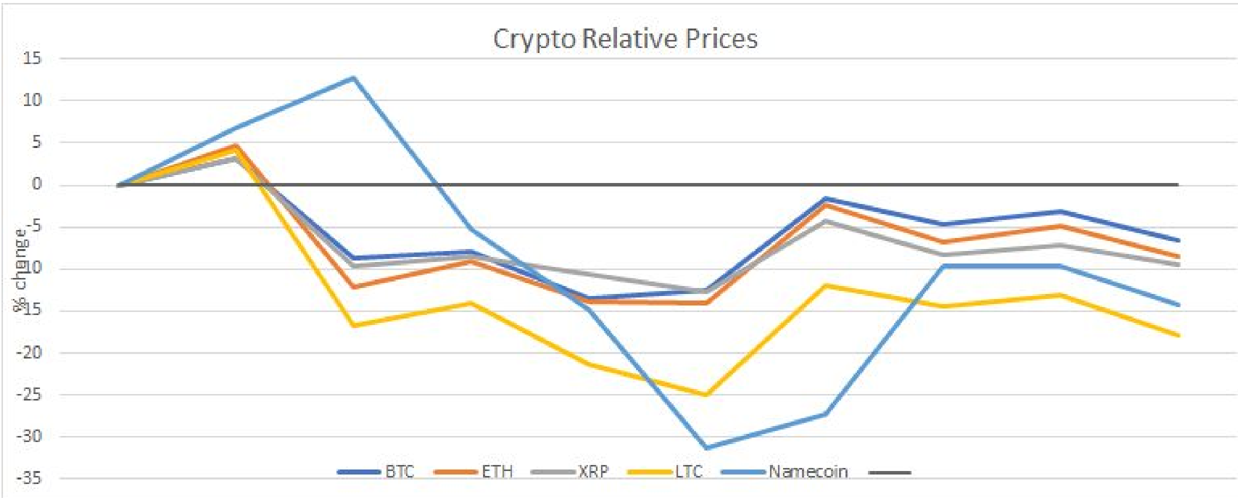


Figure 2. Cryptocurrency sentiment and price directionality

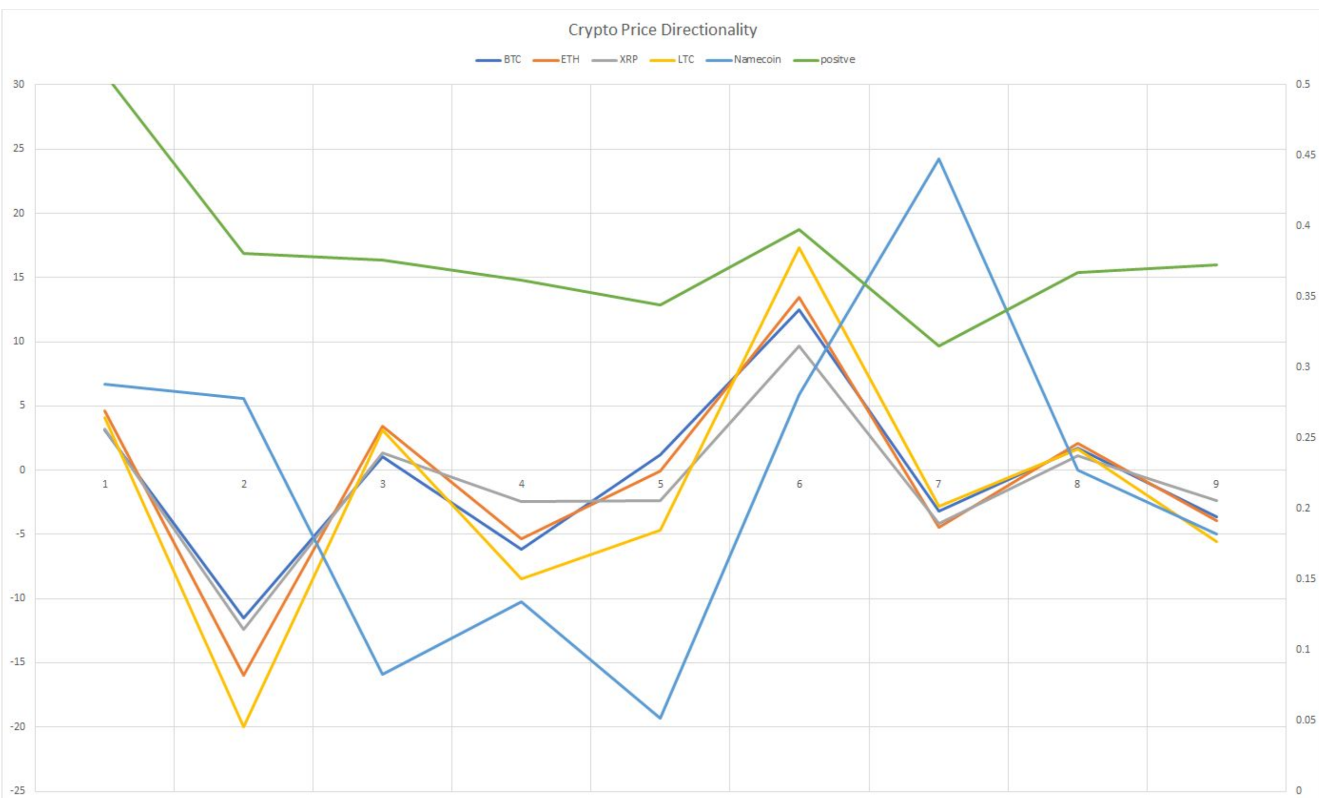


Figure 3. Individual cryptocurrency values versus overall cryptocurrency sentiment

Conclusions/Future Directions

In this case study, we found that Twitter data analysis provides an intriguing window into cryptocurrency communications. During our relatively brief period of data collection, positive cryptocurrency tweet sentiment was observed to track with cryptocurrency price trends. These initial findings suggest that Twitter data could potentially provide interesting and useful insights to guide cryptocurrency investments, although further data collection and model refinement are necessary before firm conclusions can be established. The unpredictability of Twitter usage, the interests of cryptocurrency buyers and sellers, and external factors impacting value and sentiments towards cryptocurrency may all limit the generalizability of our findings outside the period of data collection. Nonetheless, our observation that increased positive Twitter sentiment tracked with cryptocurrency value increases is intriguing and deserving of further study.

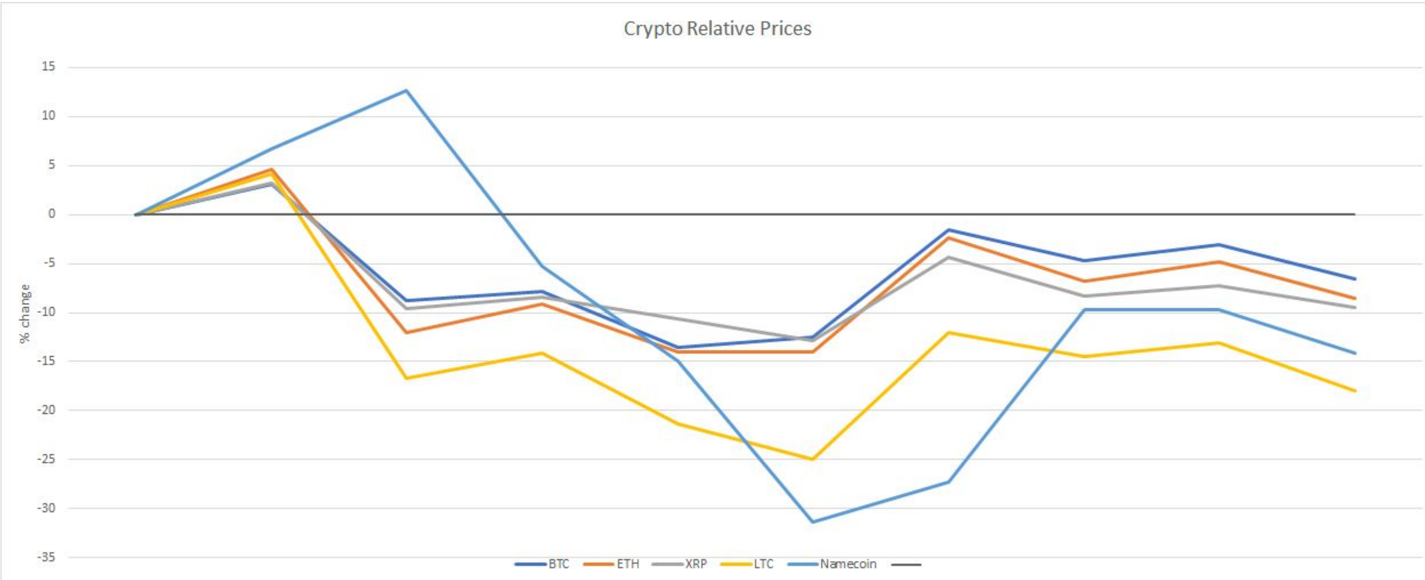
While cryptocurrency technology does have legitimate uses in permitting secure peer-to-peer transactions, it is not entirely clear to what degree legitimate uses drive the value of cryptocurrencies versus social dynamics. The focused content of Twitter communications related to this topic and the disconnection among individuals following our popular user in this domain do raise interesting questions about social communication trends and cryptocurrency appreciation. Another important issue to note relative to our findings is that our data collection period coincided with several events that would reasonably be expected to adversely impact cryptocurrency values. These events included recent government action to close bitcoin exchanges in China (8) and the publication of a JP Morgan report critical of cryptocurrency investments that described Bitcoin as a “fraud” (9). It is not clear if decreases in cryptocurrency value occur regularly in response to media attention, regulatory activity, or other world events, only to rebound as a result of continued investment attention by cryptocurrency enthusiasts. In order to answer these questions, additional data collection during periods when cryptocurrency is not prominent in the news cycle are certainly needed.

To the extent that losses in value occur cyclically or predictably and are followed by value recovery, regardless of the long-term prospects for cryptocurrencies, these events could present opportunities for diversification into cryptocurrency, as long as purchases and sales are well-timed. Since there are multiple competing cryptocurrencies, selecting a cryptocurrency for purchase at points of value loss presents a challenge -- it is possible that not all competing cryptocurrencies will recover in a competitive market. Further financial analysis together with Twitter data tracking is needed to fully explore this question and evaluate the role of social media in predicting and driving cryptocurrency values.

REFERENCES

1. https://blockgeeks.com/guides/what-is-blockchain-technology/, accessed 09/20/2017.
2. Armknecht F. Karame GO. Ripple: overview and outlook. *International Conference on Trust and Trustworthy Computation* (2015).
3. Dwyer, GP. The economics of Bitcoin and similar private digital currencies. https://poseidon01.ssrn.com/delivery.php?ID=668089000120104018090007002080105106020020059065037078000112092006105103006096108071022118037001014005040069004078006124119078052021093009085104076094113098118004015012084070080100113105025065081016113087078093118082065024078029065085104002110003&EXT=pdf , accessed 09/18/2017
4. Russell, MA. *Mining the Social Web, 2nd Edition.* Copyright 2014 Matthew A. Russell, ISBN 978-1-449-36761-9.
5. TextBlob Sentiment: Calculating Polarity and Subjectivity (2015). http://planspace.org/20150607-textblob_sentiment/ , accessed 09/2017.
6. https://www.bitstamp.net/market/tradeview/ (Price data for BTC, ETH, ETH, XRP, LTC)
7. https://coinmarketcap.com/currencies/namecoin/ (Price data for Namecoin)
8. van Wirdum, A. “Uncertainty dominates as China continues to clamp down on cryptocurrencies” Bitcoin Magazine: http://www.nasdaq.com/article/uncertainty-dominates-as-china-continues-to-clamp-down-on-cryptocurrencies-cm848050/amp , accessed 09/18/2017.
9. Henry, D. Irrera, A. “JPMorgan’s Dimon says bitcoin ‘is a fraud’” Reuters: https://www.reuters.com/article/us-usa-banks-conference-jpmorgan/jpmorgans-dimon-says-bitcoin-is-a-fraud-idUSKCN1BN2KP, accessed 09/18/2017
10. https://www.quora.com/When-is-a-cryptocurrency-considered-a-ponzi-scheme, accessed 09/21/2017

Appendix



Appendix Figure I. Relative price change of selected cryptocurrencies over five-day data collection time period

Appendix Figure II. Trading view of Bitcoin (BTC) price over the data collection period (5 days)

Appendix Figure III. Trading view of Etherium (ETH) price over the five-day data collection period

Appendix Figure IV. Trading view of Litecoin (LTC) price over the five-day data collection period

Appendix Figure V. Trading view of XRP price over the five-day data collection period

Appendix Figure VI. Trading view of Namecoin (NMC) price over the five-day data collection period