The Impact of News Media on Cryptocurrency Prices: Modelling Data Driven Discourses in the Crypto-Economy

Kelly Ann Coulter

Orcid ID: 0000-0003-3426-0962

DOI: 10.5526/r0a3-7n46

Centre for Computational Social Science, Department of Sociology, University of Essex, Wivenhoe Park, Essex, UK

Keywords: Cryptocurrency, Topic Modelling, Latent Dirichlet Allocation, News Media, Text Analysis, Sentiment

1. Summary

Natural Language Processing was adopted in this study to model data driven discourses in the crypto economy. Utilising topic modelling, specifically Latent Dirichlet Allocation (LDA), a text analysis of cryptocurrency articles (N=4218) published from over 60 countries in international news media, identified key topics associated with cryptocurrency in the international news media from 2018 to 2020. This study provides empirical evidence that 18 key topics were framed around the following key macro discourses: crypto related crime, financial speculation and investment, financial governance and regulation, political economy (with reference to specific geographical areas), cryptocurrency actors and communities and specific crypto projects and their respective markets. Analysis showed that the identified cryptocurrency macro discourses may have had a 'social signal' effect on movements in the crypto financial markets, including potential effects of crypto price volatility. Further in some cases, that the source of the news may have amplified the effect, particularly in terms of geographical region, relative to broader market conditions.

2. Introduction

Cryptocurrency research has primarily focused on Bitcoin, where there has previously been extensive research in analysing the link between Google search volumes and Bitcoin metrics (1–8) and further research developed on the words and sentiment that underly search terms (9). Whilst past research has focused on Google searches and specifically social media, this article shows that alternatively, media such as digital news articles from online news outlets (for example such as Bloomberg and FT among others), can identify not only popular single worded terms, but contextualised cryptocurrency market sentiment too. By linking words to other common and popular crypto-economic terms (or topics), narratives emerge that express a range of micro ideologies and speculation. The combinations of these narratives, cumulate into macro discourses within the crypto economy, providing a clearer and broader view of overall sentiment.

Bitcoin was the original blockchain. It was the first of a peer to peer digital currency that eliminated the need for a third party (such as a central bank) to validate transactions, by utilising a distributed ledger system where the decentralised consensus mechanism - Proof of Work - ensured validity and trust within the network (10). Since Bitcoin's inception, other cryptocurrencies have entered the crypto market. As of February 2021 there were 4,501 crypto coins (11). Cryptocurrencies, or 'crypto-

*Author for correspondence (kc18540@essex.ac.uk).
†Present address: Computational Sociology, University of Essex, Colchester, Essex

assets' as they are commonly referred to (12), are along with Bitcoin, part of a wider crypto market (13). It is a market that is growing in capacity of coins and in trading volume (11) and is an economy with a growing market capital (14).

2.1 Why News Media?

This paper puts forward that crypto market activity should not be restricted to studies of Bitcoin but must consider the array of other cryptocurrencies on the market with macro discourses in the news and their interrelated sentiment connections. What discourses the media present to their audiences (its construction) and how the news media presents cryptocurrency (its analysis), are as important and influential to the crypto-economy, as Google searches and social media Bitcoin metrics (15). Research has confirmed the suggestion that movements in financial markets, and movements in financial news are intrinsically interlinked (16).

The function of the media, is a source of information and sentiment in the financial market (17). Bloomberg or Reuters for example, as reputable financial media outlets can affect the markets as investor behaviour can be in response to company news and events which gain high media coverage (18,19). Further, social signal effect research of the social media platform Twitter, has revealed that increases in opinion polarization and exchange volume precede rising Bitcoin prices, and that emotional valence precedes opinion polarization and rising exchange volumes (6). Media coverage that exhibits varying optimism and pessimism may be captured through the fixed effects, as well as article length, writing style, and availability of information to different journalists (17).

Sentiment analysis in the media can be particularly useful for computational finance, where digital traces of human behaviour offer a great potential to drive trading strategies (6). This research therefore examines the news media and their reporting of cryptocurrency to identify popular discourses and underlying sentiment. By undertaking text analysis of N=4218 cryptocurrency articles published between 2018 and 2020, discourses in the international news media from over 60 countries are modelled utilising a Natural Language Processing method, specifically Latent Dirichlet Allocation (LDA) Topic Modelling (20). LDA is adopted to identify cryptocurrency discourses that may have a social signal effect on movements in the financial markets.

2.2 Contributions of this article

This study provides empirical evidence that key topics associated with cryptocurrency in the international news media from 2018-2020, are framed around the following key macro discourses: crypto related crime[1], financial speculation and investment[2], financial governance and regulation[3], political economy (with reference to specific geographical areas)[4], cryptocurrency actors and communities[5] and specific crypto projects and their respective markets[6].

This research builds upon Burnie and Yilmaz's (9) research into social media signals and Bitcoin metrics, but instead of determining which words on reddit matter as the bitcoin pricing dynamic

[1] (topics 1, 4, 5, 8, 13 and 14).
[2] (topics 5 and 11).
[3] (topics 3, 5, 8, 11).
[4] (topics 5 and 17).
[5] (topics 7 and 9).
[6] (topics 6 and 9).

changes from one phase to another, this study takes a complementary next step looking at the broader perspective of overall sentiment in terms of which words matter[7], and are involved in supporting macro discourses in international news media. Where past research has shown a clear connection between the role of the media and the market i.e. commonly used words and market behaviour (16–19), discourses where words present context may have differing effects. This research identifies those key crypto discourses in a novel, and yet unexplored way.

2.3 LDA Topic Modelling

Latent Dirichlet Allocation (LDA) topic modelling is one technique in the field of natural language processing text mining, that is a process to automatically identify topics present in a text object and to derive hidden patterns exhibited by a text corpus, based on probabilistic latent semantic analysis (21). Using Bayesian inference, a topic therefore, is a distribution over a fixed vocabulary, where unobserved (or latent) topics are assumed to be generated first before documents. Documents are generated from a mix of topics in different proportions, in this way it is understood as a generative process (22). A key assumption of LDA topic modelling is that documents can exhibit multiple topics but only the number of topics is specified in advance, it is hence a generative process.

Blei (22) argued for the use of topics modelling, claiming that the "utility of topic models stems from the property that the inferred hidden structure resembles the thematic structure of the collection". Adopting probabilistic topic models for text analysis as an algorithmic solution, can be therefore very useful for the purpose for document clustering, organizing large blocks of textual data, information retrieval from unstructured text and feature selection (23). This study adopted computational text analysis and human qualitative interpretation for the discourse analysis. In the first instance, the quantitative work was performed using the Python programming language utilising Natural Language Processing (NLP) to undertake the unsupervised learning technique of Latent Dirichlet Allocation (LDA) (24) topic modelling. No topics were given or 'fixed' to the model, as would have been the case in an alternative supervised approach.

The unsupervised approach was used in this crypto study for finding and observing co-occurring groups of words, understood as thematically coherent "topics" in large clusters of texts. Topics can be defined as a repeating pattern of co-occurring terms in a corpus; however, the name can be misleading. As Jacobs & Tschötschel (2019, p. 471) explained: "topics are clusters of words that reappear across texts, but the interpretation of these clusters as themes, frames, issues, or other latent concepts (such as discourses) depends on the methodological and theoretical choices made by the analyst". Iterative qualitative interpretation was then required from the researcher to undertake a discourse analysis, which presented as the most frequently co-occurring across the corpus; understood as the key topics. Both computationally and qualitatively, the data was pre-processed before analysis, which broadly involved cleaning and text processing the data.

[7] This paper does not intend to undertake a strict sentiment analysis in terms of methodology but aims to provide a conceptual public understanding of cryptocurrency sentiment in the loosest way; or to put it another way, in a more general 'every-day language' sense.

3. Data Preparation

3.1 Data Collection and preparation

To construct a suitable corpus of documents for analysis, the researcher manually collected and downloaded media articles in the form of text files from traditional media outlets. The articles were retrieved from across 60 countries globally, covering the broad theme of 'cryptocurrency'. 4218 news articles written in the English language were drawn from the Nexis news database and 'News API' (26), using the query 'cryptocurrency'.

3.2 Natural Language Pre-Processing Stage

After the text had been collected and collated, the text was pre-processed in Python using the SpaCy (27), Gensim (28) and Pandas (29) python packages. Pre-processing was a prerequisite prior to conducting the Natural Language Processing (NLP) on the text. The NLP stage essentially consisted of four broad steps; 1. to load the input data (crypto text articles), 2. to pre-process the data, 3. to transform documents into bag-of-words vectors and finally 4. to train the LDA model.

SpaCy is a free, open-source library for advanced Natural Language Processing (NLP) in Python. SpaCy explains that it is "...designed specifically for production use and helps you build applications that process and "understand" large volumes of text. It can be used to build information extraction or natural language understanding systems, or to pre-process text for deep learning" (27). SpaCy was used to 'parse and tag' a given document. This was where the trained pipeline and its statistical models were applied, enabling SpaCy to make predictions of which tag or label was most likely to apply in the context. One of SpaCy trained components included binary data that was produced by showing the corpus enough examples for it to make predictions that generalized across the language – for example, a word following "the" in English was most likely to be a noun.

Part of the pre-processing stage was to train the phraser which automatically detected common phrases (multiword expressions) from a stream of sentences. This process included lemmatizing the text articles (assigning the base forms of words (30)) using SpaCy, tokenizing the text articles (segmenting the text into words and punctuation marks etc (31)) and to compute bigrams (multi word expressions or common phrases) using Gensim (28).

4 Methodology

4.1 NLP Stage

Gensim was used to vectorise the sets of tokens into a doc-term matrix which were then used to determine the LDA model. Therefore, after the text was converted from text to tokenised documents, the 'tokens' were stored in a dictionary format to create a map between words and their integer id's, in an ID-to-word fashion. This then allowed for the tokenised documents to be converted to vectors. An algorithm (doc2bow) was then applied to count the number of occurrences of each distinct word, converting the word to its integer and returning the result as a sparse vector. In the first trial, the LDA model was then set up and run over an arbitrary 20 topics in the first instance, (Source code available at github.com/kellyann88/Crypto_NLP) (32).

4.2 Determining the Number of Topics in the LDA Model

An assumption about Latent Dirichlet Allocation is that the number of topics is assumed known and fixed (22). The Bayesian non-parametric topic model (33) provides a solution where the number of

topics is determined by the collection during posterior inference, and, new documents can exhibit previously unseen topics (22). Therefore, the Gensim implementation of LDA was adopted. Topic coherence was used as an intrinsic evaluation metric in this study. This metric was used to quantitatively justify the model selection. Topic coherence measures score a single topic by measuring the degree of semantic similarity between high scoring words in the topic These measurements help distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference (34).

Human inference is superior for topic interpretability, as "human topic rankings serve as the gold standard for coherence evaluation" (35). However human evaluations can be a lengthy and time-consuming process. Therefore, quantitative topic coherence measures aided in the process for this big data study. As there is no "correct" number of topics per se in topic modelling, some results may identify better topics than others. In this study, it was then a reiterative process to conduct various trials of number of topics to identify the optimal number. To find the optimal number of topics, LDA trials based on different values of number of topics were undertaken, in order to select the one that produced the highest coherence score alongside initiative assessment.

4.3 Coherence Scores

A graph of coherence scores is presented in figure 1. It was qualitatively apparent from topic model results that the optimal number of topics in the LDA model trials was 18 topics; this was the 3rd highest coherence score (see figure 1). The top three coherence scores were extremely close in value[8] and a decision was taken by the researcher to commence with the 18-topic model. The LDA model was then run, passing in the default alpha and beta values to calculate coherence.

5 Results

5.1 18-Topic LDA Model

In a two-step process, this topic modelling algorithm estimated the distribution of topics over a set of documents, and a probability distribution of words for each of the 18 topics shown in figure 2. Therefore, the number next to each topic represents the topic-word probability distribution across the corpus. For every word then there is a proportion expressed as a score aligned to each topic, this is a distinguishing characteristic of Latent Dirichlet Allocation, the documents in the selection share the same set of topics, but each document exhibits those topics in different proportion (22). The 18-topic model was run using the default alpha and beta scores. The model produced top topics, as presented in figure 2.

6 Discussion

How can Discourse Sentiment Affect Price?

Google trend popularity along with social media (reddit) analysis has shown the link between words and Bitcoin metrics (1–8). Google search term frequency has been used as a proxy for attractiveness (or popularity) of crypto to discover potential price drivers. Sovbetov's research (36) in particular, observed the attractiveness proxied by Google search term frequency finding significant coefficients for Bitcoin, Ethereum, Litecoin and Monero at 10% significant level. It indicated that 1 unit increase in Google trend popularity of Bitcoin, Ethereum, Litecoin, and Monero leads 1.27, 0.24, 0.07, and 0.05 units increases in their prices in long-run respectively (36). If google search terms alone can prove to have this affect, it is likely that use of terms together can provide a narrative effect when words are linked in a contextualised manner. In the following discussion, the NLP text analysis results are explored thematically in terms of their discourse and their respective market reactions

6.1 Cryptocurrency and Crime Discourse

Topics relating to crypto-crime were identified in topics 1, 5, 8, 13 and 14[9].

6.1.1 The Quadriga Scandal (Topic 1)

The LDA output shown below identified terms which relate to crypto crime within topic 1.

TOPIC 1 | 0.002*"quadrigacx" + 0.001*"court_appoint" + 0.001*"payment_processor" + 0.001*"bank_draft" + 0.001*"vancouver_base" + 0.001*"scotia_supreme" + 0.001*"owe" + 0.001*"ceo_and_sole" + 0.001*"pass_code" + 0.001*"quadrigacx_user"

The first word; 'quadrigacx' refers to media reports on Canada's largest cryptocurrency exchange. In 2019 the exchange ceased operations and the company was declared bankrupt with C$215.7 million in liabilities and about C$28 million in assets, with the FBI and Royal Canadian Mounted Police investigating due to the mysterious death of Quadriga's CEO-Gerald Cotton (37).

This particular crime-related cryptocurrency case associated with Quadriga, was heavily reported on by the media some 70 times by Canadian Press outlets including 'The Globe and Mail' between the dates of November 2018 and May 2020. As an example of potential market effect of this negative sentiment in terms of crypto-crime and the crypto-markets, 3 of these 70 articles were published on 27/08/2019, where Bitcoin price appears to fall -1.80% on publication date and -4.47% the following day after publication of the articles.

Date	Price	Open	High	Low	Vol.	Change %
Aug 28, 2019	9,729.4	10,184.7	10,271.3	9,629.6	580.29K	-4.47%
Aug 27, 2019	10,184.8	10,372.2	10,387.6	10,060.2	419.81K	-1.80%
Aug 26, 2019	10,371.8	10,136.0	10,568.2	10,136.0	568.77K	2.32%

Highest: 10,568.2 Lowest: 9,629.6 Difference: 938.7 Average: 10,095.3 Change %: -4.0 (38).

Similarly, 3 articles published in Canadian press outlets earlier during the year of 2019, on 23/02/2019 was preceded by a -8.86% drop in Bitcoin price on 24/02/2019.

Date	Price	Open	High	Low	Vol.	Change %
Feb 24, 2019	3,755.2	4,120.5	4,194.2	3,738.7	977.78K	-8.86%
Feb 23, 2019	4,120.4	3,965.2	4,152.6	3,939.4	727.85K	3.91%
Feb 22, 2019	3,965.2	3,937.4	3,983.1	3,931.7	649.55K	0.73%

Highest: 4,194.2 Lowest: 3,738.7 Difference: 455.5 Average: 3,946.9 Change %: -4.6 (38).

3 Canadian articles pertaining to the Quadriga case were published earlier in the same month, on the 05/02/2019, where the following day the price of Bitcoin dropped -1.85%.

Date	Price	Open	High	Low	Vol.	Change %
Feb 06, 2019	3,404.3	3,468.5	3,478.0	3,383.9	514.21K	-1.85%
Feb 05, 2019	3,468.4	3,463.0	3,485.9	3,450.3	460.95K	0.16%
Feb 04, 2019	3,462.8	3,459.0	3,479.7	3,437.1	503.92K	0.11%

(38).

These examples serve to illustrate that whilst a statistical correlation should not be drawn with such limited data[10], there is some evidence that cases such as the Quadriga crypto scandal can play into narratives of crypto-crime which may cause negative market confidence. After each group of articles were published in these specific instances in the Canadian press, the price of Bitcoin dropped each time in varying percentages. While there are many factors that can cause price fluctuations in financial markets, the function of the media, is a consistent source of information and sentiment (17). Sentiment therefore remains a candidate for causes of crypto volatility, due to investor behaviour reacting to negative news gaining high media coverage (18,19). The weight of the negative sentiment in the Canadian press may reflect the fact that North America is the third-most active region by cryptocurrency volume moved on-chain, just behind Northern & Western Europe (NWE) (39). Therefore, news concentrated in this area may target large audiences in the crypto sector, which in turn affect a larger amount of investor behaviour in these regions.

6.1.2 Crypto Ransom (Topic 8)

The LDA output shown below identified terms which relate to crypto crime within topic 8.

TOPIC 8 | 0.004*"hagen" + 0.001*"anne_elisabeth" + 0.001*"tom_hagen" + 0.001*"ransom_note" + 0.001*"disappearance" + 0.000*"falkevik_hagen" + 0.000*"mr_hagen" + 0.000*"oslo" + 0.000*"hagen_lawyer" + 0.000*"char"

This topic solely relates to the media coverage of a criminal case involving ‘Mr Tom Hagen’, a wealthy businessman from Oslo, Norway whose wife ‘Anne Elisabeth’, who reportedly disappeared under mysterious circumstances during 2018. A ransom note was allegedly left at the scene demanding 9 million euros in the cryptocurrency Monero (40). Investigators believed Hagen invented a cryptocurrency ransom to cover up the murder of his wife and he was subsequently arrested (41).

[10] Research would benefit from further statistical analysis on the relationship between themed discourse, sentiment, and crypto price.

News outlets, Huffpost.com and the United Kingdom ‘the times’ news outlets covered the Hagen story during April and May 2020 with 3 articles. Unlike the Bitcoin negative sentiment market impact from the Quadriga scandal, the Hagen story appeared to have little impact on the Monero market, with in fact a slight increase in Monero’s value the day after publication on the 29/04/2020. Indicating that the story gained public attention and visibility but didn’t infer a devaluation of cryptocurrency.

Date	Open	High	Low	Close*	Adj Close**	Volume
Apr 29, 2020	62.53	66.98	62.08	66.55	66.55	146,559,090
Apr 28, 2020	62.39	63.79	61.61	62.50	62.50	110,007,751
Apr 27, 2020	61.22	62.51	61.22	62.39	62.39	97,849,93

(42).

The cryptocurrency Monero, was covered 13 times by journalists, in a range of countries including Mexico, South Africa, Kenya, UK, Thailand, Bahrain and Nigeria. Renowned for its privacy functions, criminals often use Monero in different kinds of malware and DDOS extortion attacks to launder money (40). Of the crime news articles published, Monero was implicated for its use on the platform ‘Wall Street Market”, allegedly the secondly largest illegal sales market on the dark web where drugs such as cocaine and heroin, cannabis and amphetamines were traded, as well as forged documents and malicious software (43). The German police arrested three men in April 2019, suspected of managing the platform, where agents seized bitcoin and Monero, as well as seizing 550,000 euros in cash (44). After release of this article on 03/05/2019, the following day the price of Monero rose from $64.37 to $67.02.

Date	Open	High	Low	Close*	Adj Close**	Volume
May 04, 2019	$67.02	$68.35	$65.23	$67.79	$37,095,216	$1,149,800,512
May 03, 2019	$64.37	$68.40	$64.17	$67.02	$42,073,269	$1,136,650,882
May 02, 2019	$64.99	$65.82	$63.34	$64.41	$40,766,958	$1,092,119,266

(42)

6.1.3 Bitcoin Extortion (Topic 14)

The LDA output shown below identified terms which relate to crypto crime within topic 14.

TOPIC 14 | 0.003*"accuse" + 0.001*"crore" + 0.001*"arrest" + 0.001*"r_crore" + 0.001*"kotadiya" + 0.001*"bhatt" + 0.001*"surat" + 0.001*"patel" + 0.001*"police" + 0.001*"bhardwaj"

This topic relates to a criminal case covered by the ‘Indian Express’ news outlet in India. A Surat-based businessman Shailesh Bhatt was kidnapped by suspect Inspector Anant Patel and who extorted Bhatt for Bitcoins worth over Rs 9.45 crore (45). Kotadiva who was also declared a proclaimed offender in the Rs 9 crore bitcoin extortion case of Surat, was subsequently arrested by the Ahmedabad Crime Branch (46). This case was published in an article on the 08/07/2018. Again, as in the previous crime related Monero example, the day after publication of this negative sentiment, the price of Monero increased from $135.28 to $137.93.

Date	Open	High	Low	Close*	Adj Close**	Volume
Jul 09, 2018	$137.93	$141.16	$135.48	$135.73	$33,160,900	$2,200,247,171
Jul 08, 2018	$135.28	$140.01	$134.49	$138.12	$25,550,900	$2,238,474,381
Jul 07, 2018	$133.83	$135.63	$130.73	$135.61	$27,346,900	$2,197,415,127

6.2 Cryptocurrency Financial Speculation and Investment

6.2.1 Crypto Valley of Asia – CEZA (Topic 5)

The LDA output shown below identified terms which relate to financial speculation and investment within topic 5.

TOPIC 5 | 0.001*"ceza" + 0.001*"iceland" + 0.001*"lambino" + 0.000*"salerno" + 0.000*"char" + 0.000*"economic_zone" + 0.000*"cagayan_economic" + 0.000*"zone_authority" + 0.000*"raul_lambino" + 0.000*"ihe"

Cagayan Economic Zone Authority (*CEZA*) is a freeport area that offers a wide spectrum of business undertakings from beach resorts, world-class golf courses and modern township projects to manufacturing, online gaming and financial technology services for cryptocurrency and bitcoin companies (47). Raul Lambino, administrator, and CEO of CEZA, pledged transparency, no corruption and smooth operations in a bid to attract investment and companies to the area. Asian media outlets such as the Manila Bulletin, Businessworld and the Philippine Star reported on the development of the CEZA zone during 2018 until 20th September 2019, where the various media outlets reported the rise in tax payments to the government owned corporation, due to development and plans for Chinese businesses to invest $3.9 billion into the CEZA economy. However, in September 2019 the Philippine Star reported that Mike Gerald David, spokesperson and chief fintech and cryptocurrency business officer at CEZA, had stated in a press conference that the agency was suspending the operation of all crypto licensees in Manila by freezing them (48). Nine articles were published by Asian press covering the CEZA development between the dates of 27/06/18 and 11/01/2020.

Date	Open	High	Low	Close*	Adj Close**	Volume
Sep 21, 2019	$10,183.65	$10,188.10	$10,000.71	$10,019.72	$13,425,266,806	$179,853,287,294
Sep 20, 2019	$10,266.32	$10,285.87	$10,132.19	$10,181.64	$14,734,189,639	$182,738,947,696
Sep 19, 2019	$10,200.50	$10,295.67	$9,851.69	$10,266.41	$19,937,691,247	$184,240,949,577

The day after publication of the CEZA suspension of crypto licences, the price of Bitcoin fell from $10,266.32 to $10,183.65. Since the freezing of licences, the CEZA website claims that CSEZFP will be the first economic zone in Asia to regulate, license and propagate offshore financial technology solutions enterprises and offshore virtual currency exchanges (49). The 2019 CEZA coverage of the development of the economic zone donning headlines such as “Crypto Valley of Asia is a haven for foreign investors” (47) would have engendered trust and provided positive sentiment to potential foreign crypto based investors investing in the area with the promise of state governance, transparency and protection (48). However, undermining this speculative narrative with one of fear of devaluation and suspicion with the state’s freezing of crypto licenses some months later, appeared to correspond with a market downturn in Bitcoin price.

The growth of all crypto activity however in Central and Southern Asia[11] from December 2019 to June 2020 rose by an increase in crypto transactions. The value received by Central and Southern

[11] Including Oceania.

Asia[12] rose from 2 billion in December 2019 to over 4 billion in June 2020 (39). This two-fold increase in crypto trading value occurred whilst investor behaviour responded to a regulatory narrative of crypto licence suspension which had an influential effect on crypto markets, in this case specifically Bitcoin's price. A regional increase in trading at this level could suggest a disproportionate level of media influence on investor behaviour compared with other regions. If for example, there is an increase in crypto trading in a particular region coinciding with a regional media narrative acting as an amplified social signal affecting regional investor sentiment, this could result in a disproportionate market effect, for example on price formation which appears to be evidenced by Bitcoin's price volatility in the case of Asia's reporting on the CEZA crypto friendly investment zone.

The market effect may have been regional but in fact could go beyond that region affecting the wider crypto market globally. Just as in the Canadian case with the Quadriga Scandal, the weight of the negative sentiment correlated with a downturn in Bitcoin price which may have been driven by the fact that North America is the third-most active region by cryptocurrency volume moved on-chain (39). Thus, this may have also been true in the Asian case where negative sentiment again emanated and amplified from regional press coverage, in a geographical area where crypto activity is rising in volume, contributed to falling Bitcoin prices post publication of each article on the CEZA suspension of crypto licences. Under reaction of stock prices to news such as earnings announcements, and overreaction of stock prices to a series of good or bad news is well documented by Barberis et al (50) as regularities among investor behaviour in how beliefs are formed. As with equities, in a similar vein this could also hold for the crypto markets, where an over-reaction to media reporting as in the case studies above led to a volatile price drop in the Bitcoin market.

5 Conclusion

To conclude, this study provided empirical evidence that key topics associated with cryptocurrency in the international news media from 2018-2020, were framed around the following key macro discourses: crypto related crime[13], financial speculation and investment[14], financial governance and regulation[15], political economy (with reference to specific geographical areas)[16], cryptocurrency actors and communities[17] and specific crypto projects and their respective markets[18]. LDA topic modelling was used as a computational methodology to identify and model data driven discourses of cryptocurrency in the news media and the potential 'social signal' effects this had on the cryptocurrency markets during the given period. This contributes to the current research and understandings around Cryptocurrency (9,36), where past research has revealed a clear connection between the role of the media and the market i.e. commonly used words and market behaviour (16–19).

Developing upon previous crypto studies, this article provided some specific examples from a limited number of discourses, showing how they might have differing effects on the crypto markets, with

[12] Including Oceania.
[13] (topics 1, 4, 5, 8, 13 and 14).
[14] (topics 5 and 11).
[15] (topics 3, 5, 8, 11).
[16] (topics 5 and 17).
[17] (topics 7 and 9).
[18] (topics 6 and 9).

possible correlations between discourse sentiment and crypto price volatility, particularly in the case of Bitcoin over Monero. Whilst discourse may have played a role in crypto markets, this study discovered that a potential important factor in the effect of the media on crypto markets may be driven and amplified dependent upon the geographical source of the news. Use of existing data from the 'Geography of Cryptocurrency' report and historical price data complemented the analysis to contextualise discourses and consider the potential weight of sentiment depending on crypto activity in their respective regions. Further research could give more time and consideration to modelling theses crypto discourses to generate statistical trends among the specific discourses identified, to uncover relationships between discourse, source of news and price volatility based upon sentiment. This would also positively add to the budding literature on the role of the media and crypto markets.

6 References

1. Kristoufek L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Sci Rep. 2013;3(1):1–7.
2. Garcia D, Tessone CJ, Mavrodiev P, Perony N. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. J R Soc Interface. 2014;11(99):20140623.
3. Kristoufek L. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PLoS One. 2015;10(4):e0123923.
4. Matta M, Lunesu I, Marchesi M. Bitcoin Spread Prediction Using Social and Web Search Media. In: UMAP workshops. 2015. p. 1–10.
5. Georgoula I, Pournarakis D, Bilanakos C, Sotiropoulos D, Giaglis GM. Using time-series and sentiment analysis to detect the determinants of bitcoin prices. Available SSRN 2607167. 2015;
6. Garcia D, Schweitzer F. Social signals and algorithmic trading of Bitcoin. R Soc open Sci. 2015;2(9):150288.
7. Bouoiyour J, Selmi R, Tiwari AK. Is Bitcoin business income or speculative foolery? New ideas through an improved frequency domain analysis. Ann Financ Econ. 2015;10(01):1550002.
8. Ciaian P, Rajcaniova M, Kancs d'Artis. The economics of BitCoin price formation. Appl Econ. 2016;
9. Burnie A, Yilmaz E. Social media and bitcoin metrics: which words matter. R Soc open Sci. 2019;6(10):191068.
10. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008;
11. Best R. Number of crypto coins 2013-2021 [Internet]. Statista. 2021 [cited 2021 May 18]. Available from: https://www.statista.com/statistics/863917/number-crypto-coins-tokens/
12. Treasury Committee H of C. Crypto-assets Twenty-Second Report of Session 2017-19 Report [Internet]. London; 2018. Available from: www.parliament.uk/treascom
13. Martin J, Cunliffe J, Munksgaard R. Cryptomarkets: A Research Companion [Internet]. Emerald Group Publishing; 2019. Available from: http://sdc-evs.ebscohost.com/EbscoViewerService/print?an=2255567&db=nlebk&format=EB&lang=eng&ppIds=[%22pp_22%22,%22pp_23%22,%22pp_24%22,%22pp_25%22,%22pp_26%22,%22pp_27%22,%22pp_28%22,%22pp_29%22,%22pp_30%22,%22pp_31%22,%22pp_32%22,%22pp_33%22,%22pp_34%22
14. Kharif O. Bitcoin (BTC USD) Cryptocurrency Price Rise Leads $2 Trillion Crypto Market Cap. Bloomberg [Internet]. 2021 Apr [cited 2021 May 13];1. Available from: https://www.bloomberg.com/news/articles/2021-04-05/crypto-market-cap-doubles-past-2-trillion-after-two-month-surge
15. Burnie A, Yilmaz E. Social media and bitcoin metrics: which words matter. 2019 [cited 2021 May 13]; Available from: http://dx.doi.org/10.1098/rsos.191068
16. Alanyali M, Moat HS, Preis T. Quantifying the relationship between financial news and the stock market. Sci Rep [Internet]. 2013 Dec 20 [cited 2021 May 21];3(1):1–6. Available from: www.nature.com/scientificreports
17. Walker CB. The direction of media influence: Real-estate news and the stock market. J Behav Exp Financ. 2016 Jun 1;10:20–31.
18. Baker M, Wurgler J. Investor sentiment in the stock market. In: Journal of Economic Perspectives. 2007. p. 129–51.
19. Tetlock PC. Giving content to investor sentiment: The role of media in the stock market. J

Finance [Internet]. 2007 Jun 1 [cited 2021 May 18];62(3):1139–68. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.2007.01232.x
20. Blei DM, McAuliffe JD. Supervised topic models. In: Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference [Internet]. 2009 [cited 2021 Apr 22]. Available from: https://arxiv.org/abs/1003.0783v1
21. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. J Inf Sci. 2017;43(1):88–102.
22. Blei DM. Probabilistic topic models. Commun ACM. 2012;55(4):77–84.
23. Ding C, Li T, Peng W. Nonnegative Matrix Factorization and Probabilistic Latent Semantic Indexing: Equivalence, Chi-square Statistic, and a Hybrid Method [Internet]. 2005 [cited 2021 Apr 20]. Available from: www.aaai.org
24. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003;3:993–1022.
25. Jacobs T, Tschötschel R. Topic models meet discourse analysis: a quantitative tool for a qualitative approach. Int J Soc Res Methodol. 2019;22(5):469–85.
26. NewsAPI. News API – Search News and Blog Articles on the Web [Internet]. [cited 2021 Jul 12]. Available from: https://newsapi.org/
27. Honnibal M MI. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing [Internet]. 2017 [cited 2021 Apr 9]. Available from: https://spacy.io/usage/spacy-101
28. Rehurek, R. & Sojka P. Gensim–python framework for vector space modelling [Internet]. Brno: NLP Centre, Faculty of Informatics, Masaryk University,; 2011 [cited 2021 Apr 22]. Available from: https://pypi.org/project/gensim/
29. Mckinney W. Data Structures for Statistical Computing in Python. 2010.
30. spaCy. Lemmatizer · spaCy API Documentation [Internet]. [cited 2021 Apr 9]. Available from: https://spacy.io/api/lemmatizer
31. spaCy. Tokenizer · spaCy API Documentation [Internet]. [cited 2021 Apr 9]. Available from: https://spacy.io/api/tokenizer
32. Coulter K. Crypto_NLP_Model [Internet]. GitHub Repository; 2021. Available from: https://github.com/kellyann88/Crypto_NLP
33. Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical dirichlet processes. J Am Stat Assoc. 2006;101(476):1566–81.
34. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring Topic Coherence over many models and many topics [Internet]. Association for Computational Linguistics; 2012 [cited 2021 Apr 28]. Available from: http://mallet.cs.umass.edu/
35. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery, Inc; 2015 [cited 2021 Apr 26]. p. 399–408. Available from: https://dl.acm.org/doi/10.1145/2684822.2685324
36. Sovbetov Y, Sovbetov Y. Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero Journal of Economics and Financial Analysis. J Econ Financ Anal. 2018;2(2):1–27.
37. Roberts J. FBI and RCMP Probing Quadriga Exchange Over Missing Funds, Source Alleges | Fortune [Internet]. Fortune. 2019 [cited 2021 May 24]. p. 1. Available from: https://fortune.com/2019/03/04/quadriga-fbi-bitcoin/
38. Investing.com. Bitcoin Historical Data - Investing.com UK [Internet]. [cited 2021 May 24]. Available from: https://uk.investing.com/crypto/bitcoin/historical-data
39. Chainalysis. The 2020 Geography of Cryptocurrency Report Analysis of Geographic Trends in Cryptocurrency Adoption, Usage, and Regulation [Internet]. 2020 [cited 2021 May 28]. Available from: https://go.chainalysis.com/2021-Crypto-Crime-Report.html
40. Koerhuis W, Kechadi T, Le-Khac N-A. Forensic analysis of privacy-oriented cryptocurrencies. Forensic Sci Int Digit Investig. 2020 Jun 1;33:200891.

41. Adler D. Did This Norwegian Multimillionaire Invent a Cryptocurrency Ransom to Cover Up the Murder of His Wife? | Vanity Fair [Internet]. Vanity Fair. 2020 [cited 2021 May 24]. p. 1. Available from: https://www.vanityfair.com/style/2020/04/tom-hagen-norway-murder-arrest
42. CoinMarketCap. Monero price today, XMR live marketcap, chart, and info [Internet]. [cited 2021 May 27]. Available from: https://coinmarketcap.com/currencies/monero/historical-data/
43. Coldewey D. How German and US authorities took down the owners of darknet drug emporium Wall Street Market | TechCrunch. TechCrunch [Internet]. 2019 May 3 [cited 2021 May 28];1. Available from: https://techcrunch.com/2019/05/03/how-german-and-us-authorities-took-down-the-owners-of-darknet-drug-emporium-wall-street-market/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAM49GKFIhidoNzOrpt15tc0D1rA-BsUNCCZFjw2Lvmlh
44. DW. Germany: 3 arrests in darknet 'Wall Street Market' probe | News | DW | 03.05.2019 [Internet]. DW. 2019 [cited 2021 May 27]. Available from: https://www.dw.com/en/germany-3-arrests-in-darknet-wall-street-market-probe/a-48583560
45. Service EN. Gujarat: Suspended SP gets bail in Bitcoin extortion case [Internet]. The Indian Express. 2019 [cited 2021 May 27]. Available from: https://indianexpress.com/article/cities/ahmedabad/gujarat-suspended-sp-gets-bail-in-bitcoin-extortion-case-5549266/
46. Express TI. Bitcoin extortion case: Ex-BJP MLA Nalin Kotadiya declared proclaimed offender [Internet]. [cited 2021 May 28]. Available from: https://indianexpress.com/article/cities/ahmedabad/bitcoin-extortion-case-ex-bjp-mla-nalin-kotadiya-declared-proclaimed-offender-5222429/
47. South China Morning Post. CEZA's Crypto Valley of Asia, is a haven for foreign investors | South China Morning Post [Internet]. 2019 [cited 2021 Jun 1]. p. 1. Available from: https://www.scmp.com/country-reports/country-reports/topics/philippines-business-report-2019/article/3013776/cezas
48. South China Morning Post. CEZA's Crypto Valley of Asia, is a haven for foreign investors | South China Morning Post. 2019. p. 1.
49. Cagayan Economic Zone Authority. Fintech Solutions and OVCE [Internet]. [cited 2021 Jun 1]. Available from: https://ceza.gov.ph/fintech-solutions-and-ovce
50. Barberis N, Shleifer A, Vishny R. A model of investor sentiment. J financ econ. 1998 Sep 1;49(3):307–43.

7 Figures

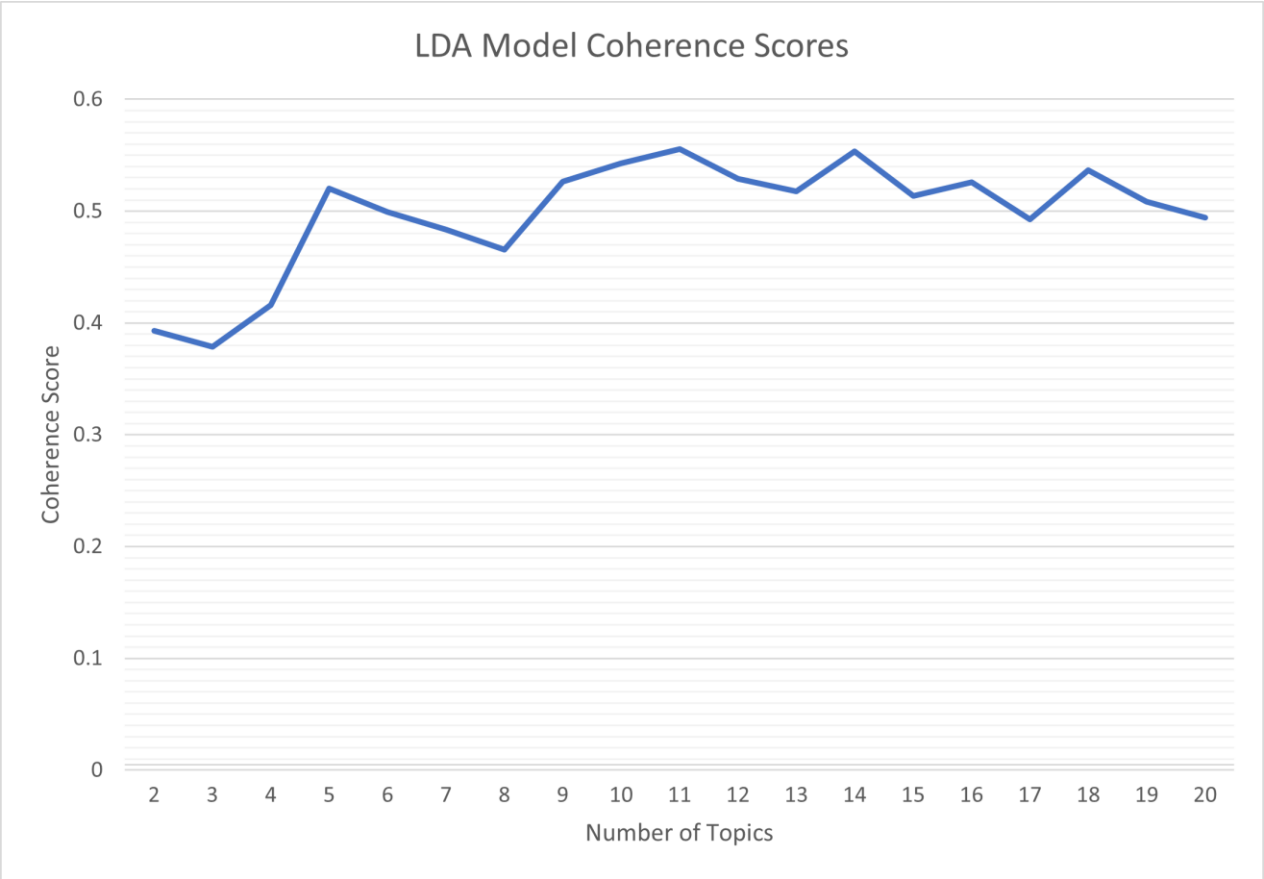

Figure 1. LDA Model Coherence Scores.

TOPIC 0 | 0.027*"ph" + 0.003*"vh" + 0.002*"group" + 0.002*"victor_harbor" + 0.002*"club" + 0.002*"centre" + 0.002*"carrickalinga_house" + 0.002*"goolwa" + 0.002*"market" + 0.001*"pt_elliot"

TOPIC 1 | 0.002*"quadrigacx" + 0.001*"court_appoint" + 0.001*"payment_processor" + 0.001*"bank_draft" + 0.001*"vancouver_base" + 0.001*"scotia_supreme" + 0.001*"owe" + 0.001*"ceo_and_sole" + 0.001*"pass_code" + 0.001*"quadrigacx_user"

TOPIC 2 | 0.006*"year" + 0.005*"people" + 0.005*"company" + 0.005*"use" + 0.004*"work" + 0.004*"time" + 0.004*"money" + 0.003*"know" + 0.003*"day" + 0.003*"like"

TOPIC 3 | 0.000*"patent" + 0.000*"commend" + 0.000*"philips" + 0.000*"hku_space" + 0.000*"asia_leadership" + 0.000*"legal_team" + 0.000*"wuxi" + 0.000*"chong_sing" + 0.000*"jeepney" + 0.000*"kokila_alagh"

TOPIC 4 | 0.001*"maren" + 0.000*"axe" + 0.000*"ueland" + 0.000*"bobby" + 0.000*"khayali" + 0.000*"esalen" + 0.000*"rodney" + 0.000*"ejjoud" + 0.000*"jespersen" + 0.000*"ueland_and_jespersen"

TOPIC 5 | 0.001*"ceza" + 0.001*"iceland" + 0.001*"lambino" + 0.000*"salerno" + 0.000*"char" + 0.000*"economic_zone" + 0.000*"cagayan_economic" + 0.000*"zone_authority" + 0.000*"raul_lambino" + 0.000*"ihe"

TOPIC 6 | 0.016*"bitcoin" + 0.008*"cryptocurrency" + 0.008*"blockchain" + 0.007*"use" + 0.006*"cryptocurrencie" + 0.006*"company" + 0.006*"year" + 0.006*"market" + 0.006*"new" + 0.005*"technology"

TOPIC 7 | 0.016*"wright" + 0.005*"nakamoto" + 0.004*"north_korea" + 0.002*"north_korean" + 0.002*"craig_wright" + 0.001*"satoshi" + 0.001*"mr_freeman" + 0.001*"andresen" + 0.001*"wright_claim" + 0.001*"pty_ltd"

TOPIC 8 | 0.004*"hagen" + 0.001*"anne_elisabeth" + 0.001*"tom_hagen" + 0.001*"ransom_note" + 0.001*"disappearance" + 0.000*"falkevik_hagen" + 0.000*"mr_hagen" + 0.000*"oslo" + 0.000*"hagen_lawyer" + 0.000*"char"

TOPIC 9 | 0.000*"char" + 0.000*"bitcoin_btc" + 0.000*"let_have_a_baby" + 0.000*"million_yuan" + 0.000*"global_stablecoin" + 0.000*"eur_million" + 0.000*"facebooks" + 0.000*"week_edition" + 0.000*"today_where_satoshi" + 0.000*"satoshi_nakaboto"

TOPIC 10 | 0.014*"good_morning" + 0.010*"property" + 0.009*"euro" + 0.008*"income" + 0.008*"thank" + 0.008*"box" + 0.007*"rent" + 0.006*"greeting_and_a_lot" + 0.006*"declare" + 0.005*"return"

TOPIC 11 | 0.001*"oil_and_gas" + 0.001*"intercontinental_exchange" + 0.000*"loeffler" + 0.000*"char" + 0.000*"energy_sector" + 0.000*"sugarbud" + 0.000*"security_filing" + 0.000*"kolochuk" + 0.000*"kelly_loeffler" + 0.000*"corporate_governance"

TOPIC 12 | 0.000*"char" + 0.000*"shop_locally" + 0.000*"teach_young" + 0.000*"stitcher" + 0.000*"second_be_decentralisation" + 0.000*"retailer_which_own_no_inventory" + 0.000*"undermine_by_communication" + 0.000*"chairman_of_wandisco" + 0.000*"unassailable_have_be_superseded" + 0.000*"bygone_age"

TOPIC 13 | 0.000*"â" + 0.000*"manifesto" + 0.000*"tarrant" + 0.000*"char" + 0.000*"mass_murderer" + 0.000*"australian_academic" + 0.000*"pseudocommando" + 0.000*"ammunition_belt" + 0.000*"tarrant_life" + 0.000*"regular_guy"

TOPIC 14 | 0.003*"accuse" + 0.001*"crore" + 0.001*"arrest" + 0.001*"r_crore" + 0.001*"kotadiya" + 0.001*"bhatt" + 0.001*"surat" + 0.001*"patel" + 0.001*"police" + 0.001*"bhardwaj"

TOPIC 15 | 0.000*"char" + 0.000*"gramatik" + 0.000*"epigram" + 0.000*"data_source" + 0.000*"muvhango" + 0.000*"napier" + 0.000*"bitcoin_btc" + 0.000*"ada" + 0.000*"skeem_saam" + 0.000*"isidingo"

TOPIC 16 | 0.000*"gv" + 0.000*"sept" + 0.000*"oct" + 0.000*"char" + 0.000*"nando" + 0.000*"camping" + 0.000*"macaron" + 0.000*"hulme" + 0.000*"info" + 0.000*"admission"

TOPIC 17 | 0.009*"venezuela" + 0.007*"petro" + 0.004*"maduro" + 0.003*"bolivar" + 0.003*"oil" + 0.003*"venezuelan" + 0.002*"venezuelans" + 0.002*"sovereign_bolivar" + 0.002*"economic" + 0.002*"hyperinflation"

Figure 2. LDA Topics.