Congruent or Polarized? Mining Opinion Dynamics towards Cryptocurrency on Twitter

Mintao Wei

May 31, 2022

1 Introduction

This study aims to complement the existing literature by examining the sentiment dynamics and opinion polarization towards cryptocurrency using 4.75 million online social media data from Twitter. Inspired by the statement that information cascade and opinion dissemination can be momentary while very fragile, meaning that the individual decisions such as whether to adopt new technology are contagious and tend to affect the network groups quickly, but it could potentially collapse rapidly as well (Easley and Kleinberg, 2010), I hypothesize that the opinions toward cryptocurrency inherit similar patterns. Therefore, I propose three research questions as below:

1. How polarized are people's opinions towards cryptocurrency?

2. How does the polarity of cryptocurrency opinions on Twitter evolve over time?

3. Can we effectively predict future cryptocurrency opinions on Twitter?

Probing into sentiments, texts, as well as the network structure, I provided empirical evidence regarding how polarized the public opinions are with regard to adopting cryptocurrency as well as how the polarity evolves over time. Additionally, this paper evaluated 5 time series based models to predict the future polarity landscape, consisting of Facebook Prophet, a modern industrial application, and traditional methods including Exponential Smoothing, Double Exponential Smoothing, Triple Exponential Smoothing, and Autoregressive–moving-average (ARMA). The empirical results imply that it could be difficult to accurately predict future opinion dynamics merely based on its historical values.

The remainder of the paper is organized as follows. Section 2 describes the data samples, measurement methodologies, and method identifications. Section 3 explains the empirical findings related to the dynamics of cryptocurrency sentiments. Section 4 discusses the current limitations. Section 5 concludes the study.

2 Data and Methods

2.1 Data

2.1.1 Cryptocurrency Relevant Tweets

Guided by the interpretation from Hasan et al. (2022), I collected 4.75 million crypto-related tweets from August 21, 2019 to November 30, 2019 using tweepy 4.9.0. Specifically, I used the *search_all_tweets* method from Twitter developer API to access the entire historical Twitter archive and retrieved the following information: user identification, number of likes received, number of retweets, text/content of the tweet, the associated hashtag (if any), timestamp upon creation. These 4.75 million tweets contain at least one of the following keywords (case insensitive): "bitcoin", "BTC", "crypto", "ETH", "memecoin". The highlighted keywords are determined based on the current market capitalization shares and topic trending. Bitcoin and Ethereum are the first and second-largest digital currencies in the world by market value as of 2022/05, which is more than three times the third Tether (USDT) and together take up approximately 64% of the market value worth, according to Slickcharts. I also included "memecoin" as one of the keywords to capture those 'ridicules' that receives growing attention in recent years. Additionally, I filtered out tweets with non-English content and empty texts, as suggested by the Twitter documentation.

2.1.2 Bitcoin Market Price

To understand the potential impact of market fluctuation on the crypto-relevant sentiments on social media platforms, I collected daily Bitcoin USD pricing index from Blockchain.com from August 21, 2019 to November 30, 2019 and mapped to each relevant tweet by its date of posting.

It is widely believed that the price volatility and drifting are to some extent synchronized and Bitcoin price has become the primary indicator due to its long-standing market position. Therefore, we expect the price index trend for Bitcoin to reflect the general market movement for cryptocurrencies.

2.2 Methods

2.2.1 Sentiment Classification

Mooijman et al., 2018 uses VADER (Valence Aware Dictionary and sEntiment Reasoner) toolkit to extract the sentiments, which is a lexicon and rule-based sentiment analysis tool that is specifically attuned to the sentiments expressed in social media. The distinctive advantage of VADER over other dictionary packages is that it performs very well even with emojis, slang, and acronyms in sentences, which are common in the informal and unstructured language of Twitter. However, I found that this 'word-counting and score-aggregating' approach sometimes performs below expectation by mistaking the overall sentiments for the opinions toward a specific entity of interest. For instance, VADER will classify a greeting followed by a pure neutral statement such as 'Good morning! Bitcoin's price today is $40,000' as positive. To account for this potential flaw, I first generated a sentiment classification using **VADER**, and then fine-tuned a pre-trained **BERT**(Bidirectional Encoder Representations from Transformers) model to utilize the contextual embedding features from BERT. Including a deep-learning based transformer model can effectively capture the semantics to remedy the drawbacks of the lexical outputs. In the end, the predictions from VADER and BERT, which are in the probability form, are averaged to produce the final sentiment classification.

As for the textual cleaning process, I only performed light processing to retain richer information such as punctuation and merely remove the entity mentions (i.e. @username), hyperlinks, and trailing white spaces before applying VADER and BERT. Using the integrated sentiment classifications from the two, this paper constructs the **percentage of tweets with positive/negative sentiments over all cryptocurrency-related tweets** as the proxy for opinions, where the cryptocurrency-related tweets are defined as any tweets that contain at least one relevant keywords (case insensitive): "bitcoin", "BTC", "crypto", "ETH", "memecoin". Furthermore, I obtained the sentiment intensity scores from VADER to measure the strength of the emotions.

$$Positive\ Opinion = \frac{Number\ of\ Positive\ Tweets(Including\ Retweets)}{All\ Cryptocurrency\ Relevant\ Tweets(Excluding\ Neutral\ Tweets)}$$

$$Negative\ Opinion = \frac{Number\ of\ Negative\ Tweets(Including\ Retweets)}{All\ Cryptocurrency\ Relevant\ Tweets(Excluding\ Neutral\ Tweets)}$$

$$Opinion\ Polarity = Positive\ Opinion - Negative\ Opinion$$

It is worth noting that the opinion measures include retweets because retweet behaviors indicate the person agrees with the opinion of the original tweet. I also exclude neutral tweets because neutral tweets usually consist of factual statements rather than subjective opinions. This indicator tends to increase when the public has divergent opinions toward cryptocurrency and declines as the opinions converge to consensus regardless of standpoints. Using these proxies, this study investigates the answers to the above-mentioned two research questions, what the current opinion landscape looks like, and how it has changed over time.

2.2.2 Representing Tweets

To provide a more straightforward and vivid representation of the opinion landscape, I use t-SNE(t-distributed Stochastic Neighbor Embedding) and PCA to project every tweet to a 3-dimensional space and disentangled by its sentiment labels (i.e. positive or negative). Specifically, I first conducted TF-TDF (Term Frequency-Inverse Document Frequency) to obtain a 1000 dimensional word embedding for the text of each tweet, and reduced it to a 3d world using the Tensorflow Projector. All the tweet representations are identified as either positive or negative per previous classifications and the spatial distribution can assist in answering how polarized people's beliefs are regarding the crypto topic.

2.2.3 Topic Modeling

I further model the topic distribution and excavate the most representative words for the theme using Latent Dirichlet Allocation (LDA) for all the crypto-relevant tweets. LDA is a widely adopted generative probabilistic model that assumes each topic is a mixture over an underlying set of words, and each document is a mixture of over a set of topic probabilities. I then visualize the LDA output using the *pyLDAvis version 3.3.1* in Python to demonstrate how topics are distributed in an Intertopic Distance Map (via multidimensional scaling) and what are the top salient words. Next, I performed word cloud analysis on the positive and negative corpora, respective, to explore the underlying topic distinction. The findings from topic modeling provide granular insights and complement the answer to the first research question regarding the polarization landscape.

2.2.4 Community Detection on Twitter Hashtag Co-occurrence Network

To address the first research questions more sufficiently and **uncover potential subtopics** under the general cryptocurrency, guided by Bovet et al., 2018 and Martinez-Romo et al., 2011, I construct a co-occurrence network graph with a each node and edge representing a hashtag and whether two hashtags appear on the same tweet, respectively. The size of the node reflects the number of tweets associated with the hashtag, which describes the magnitude of influence. The distribution of hashtags clusters provides important insights from the perspective of social network structure.

I then partitioned the network graph using Newman Modularity into two sub-communities, extracted and visualized the subgraphs centering the "#crypto" hashtag, which retains the highest degree in the network, following the procedure suggested in Bovet et al., 2018. The separation of the two communities reflects how polarized the opinions toward digital currencies are from the perspective of the hashtag co-occurrence network structure. The partition is evaluated using the Newman Modularity metric.

To capture the essence of the co-occurrence network relationship in a computationally-efficient manner, the network graph needs to be **pruned**. I manage to process the co-occurrence network and reduce the structural complexity through the following two measures:

First, I employ a statistical technique from Martinez-Romo et al., 2011 to assign a significance of the co-occurrence and eliminate those edges that mean the co-occurring of the two hashtags is purely by chance. Generally, Matinez-romo et al. propose that the co-occurrence of two elements (i.e. node) in the entire collection (i.e. network) is akin to statistical hypothesis testing, with the hypothesis being that the two elements co-occur because of semantic relatedness. The null hypothesis outlines that elements are randomly and independently distributed among the sets of the collection. Co-occurrence will be considered statistically significant if it is unlikely that it arises by pure chance. Mathematically, let N denote the number of sets in the collection, which is the total number of tweets in the study context, n_1 and n_2 represent the number of each node, or hashtag, occurs, mutually or independently, given the number of co-occurrence r, we could calculate the p-value p_{val} for the above null hypothesis using the formula below, as suggested by Martinez-Romo et al., 2011:

$$p(k) = \prod_{j=0}^{n_2-k-1} (1 - \frac{n_1}{N-j}) * \prod_{j=0}^{k-1} \frac{(n_1 - j)(n_2 - j)}{(N - n_2 + k - j)(k - j)}$$

$$p_{val} = \sum_{k>=r} p(k)$$

Please refer to the original paper for a more detailed derivation.

Second, I use the *SequenceMatcher* algorithm from the *difflib* package in Python to combine hashtags that looks very similar and thus very highly originate from typos. The *SequenceMatcher* method identifies two strings as similar if one can be altered into the other through very few simple manipulations such as adding blank lines, symbols, or white spaces. This method could help emphasize meaningful hashtags, or nodes, and effectively make the network graph condensed.

2.2.5 Time Series Analysis and Prediction

I mainly use the time series data analysis to answer the second research question, which centers on the historical dynamics of opinion polarity. As for the third research question about predicting future opinions, I built 5 time series models: Simple Exponential Smoothing (SES), Double Exponential Smoothing (or Holt's Method), Triple Exponential Smoothing (or Holt-Winters Method), ARMA (AutoRegressive Moving Average), Facebook Prophet. The above five, traditional or newly-emerged, are all time series methods rather than machine learning-based models. The reason that I use them is that I would like to test the predictability using the historical time serial sequence itself, that is, whether the characteristics of the past opinion dynamics can effectively predict future trends. Particularly, ARMA is used rather than ARIMA because the sentiment proxy is validated as stationary according to the Augmented Dickey-Fuller test result. All five methods are evaluated using Root-mean-square-error (RMSE).

3 Results

3.1 Opinion Polarization

3.1.1 Sentiment Analysis

Overview I first aggregated all the relevant tweets posted between Aug 21, 2019 and Nov 30, 2019 to explore the sentiment polarity from a holistic perspective. As demonstrated in Table 1, during this time period, there are around 4.75 million tweets (including retweet count) under the cryptocurrency topic that contain subjective emotions, among which the positive tweets constitute around 73% and the negative tweets make up the rest 27%, leading to a relatively large polarity of around 46%. In addition, I observe that tweets that express positiveness toward digital currencies are approximately 1.5x (0.56 over 0.38) stronger than the negative tweets in their emotional intensity. Furthermore, the positive tweets receive more recognition, agreement, and user engagement as the average number of likes and retweets per crypto-favoring post are 3.73 and 0.93, both of which are higher than those of the negative tweets.

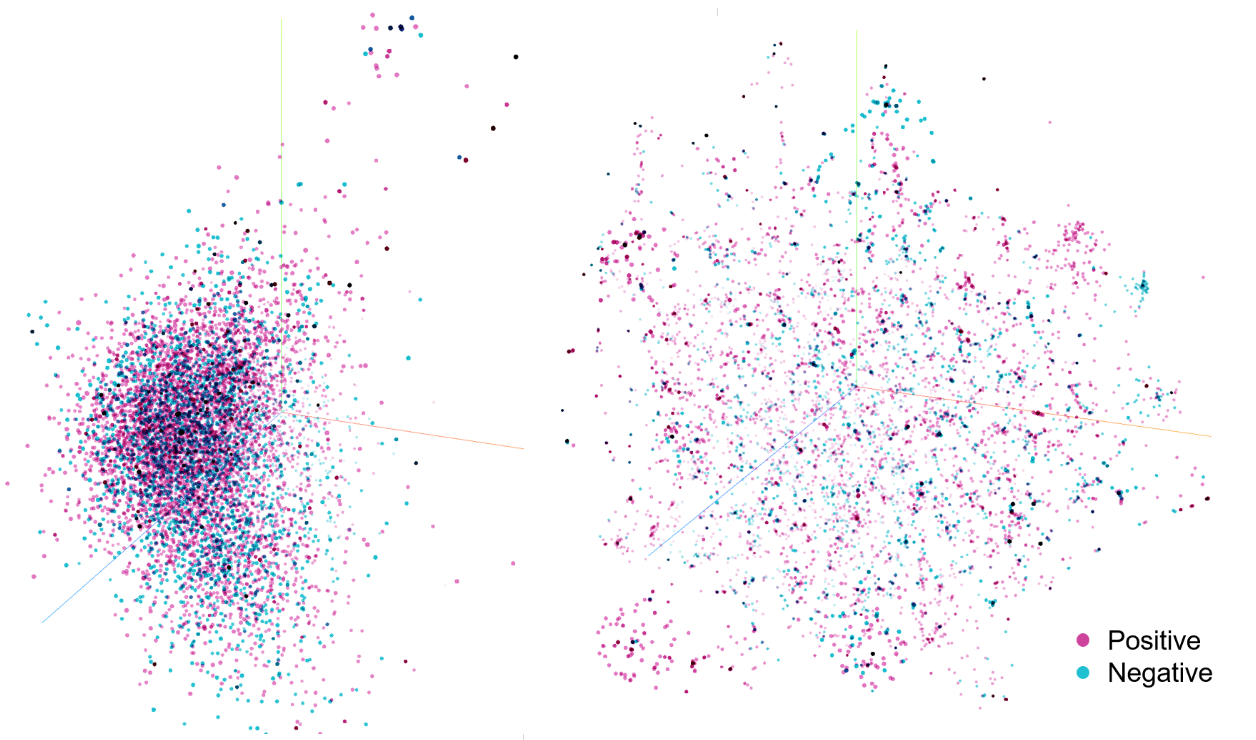
Table 1: Overview of Cryptocurrency Sentiments and Opinions

Sentiment Class	# Tweets(RT)	% Tweets(RT)	Sentiment Intensity	Avg. Likes per Tweet	Avg. Retweet per Tweet
Negative	1,273,736	27%	0.38	3.44	0.70
Positive	3,476,861	73%	0.56	3.85	1.03
Total	4,750,597	100%	0.47	3.73	0.93
Polarity	2,203,125	46%	0.18	0.41	0.33

Note: Time Window: Aug 21, 2019 Ñov 30, 2019. The total number of tweets metrics exclude the original neutral tweets and include the retweet count. The sentiment intensity score produced by VADER ranges between 0 and 1 and has a larger value when the sentiment is strong.

Representing Tweet Delving deeper, I visualized the vectorized representation of each tweets to understand the sentiment distribution of crypto-related tweets in Figure 1. In particular, I randomly sampled 10,000 tweets, represented each of them into TF-IDF embeddings, and project them into a 3-dimensional space using PCA and t-SNE, as depicted in Figure 3. The three principal components collectively explain 6.9% of the total variance, which is considered a fair amount in textual and image representation tasks. The t-SNE is created with a perplexity of 25, learning rate of 10 after 500 iterations.

Among these 10,000 subsets, there are 6,939 positive tweets and 3,061 negative tweets, and we also visually observed a more dominant scattering of positive tweets. However, while TF-IDF is usually effective in vectorizing tweets/documents as a collection of tokens that captures the main subject, we fail to observe distinct clusters by sentiment classifications in both the linear PCA and non-linear t-SNE projections. The 6,939 positive tweets and the 3,061 negative tweets tend to spread in a disordered manner and the two groups interweave with each other, which suggests that (1) the positive tweets have a variety ways of expressing supportive opinions and so do the negative tweets; (2) the underlying semantics of the crypto-favoring and opposing tweets are not significantly different, making it hardly to infer the sentiment polarity merely from the word-frequency-based features that describe the importance of each token to its corpus.



(a) PCA (b) t-SNE

Figure 1: Total Number of Cryptocurrency-related Tweets, Tweets with Positive Sentiments and Negative Sentiments

3.1.2 Content Analysis

Topic Modeling This study aims to provide finer insights at the topic and word level to demonstrate the opinion dynamics in complementary to the sentiment analysis. I employed LDA to disentangle all cryptocurrency related tweets into 3 topics with corresponding representing subject word components. The result is visualized in Figure 1. The number of topics is determined by grid-searching the optimal log-likelihood score and manually evaluating the distinction among clusters. Topic 1 comprises of over 68% tokens and it is mainly related to daily usage when discussing cryptocurrencies, such as “free”, “news”, “money”, “buy” and “wallet”. Topic 2 primarily stands for the names of a variety of digital currencies, namely, “btc”, “ltc”, “eth” and “bch”. Comparably, topic 3 incorporates a more general and wider domains. One interesting discovery is that topic 3 highlights many terms relevant to technology, firms, and millennium. For instance, “millennium firm”, “topmost technology”, “human evolution”, “technology company”. For more details, please refer to the Appendix.

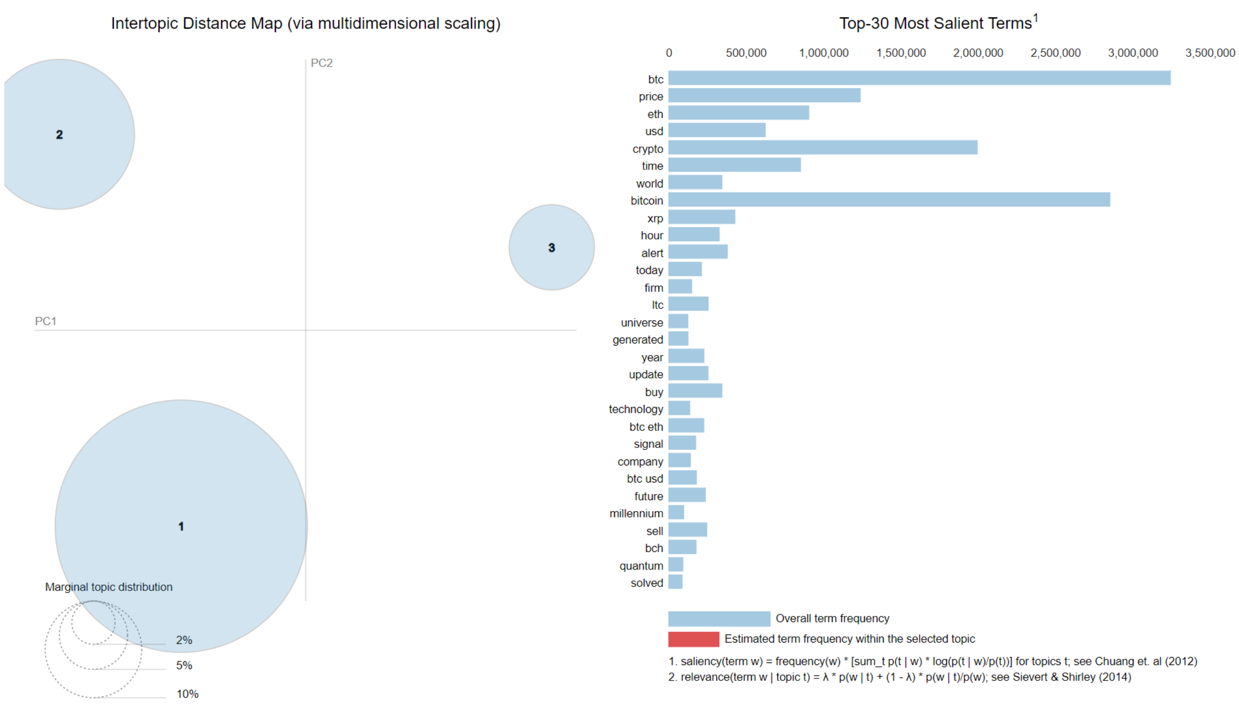


Figure 2: LDA Topic Modeling: Intertopic Distance Map and Top-30 Most Salient Words

Word Frequency In addition, to investigate the difference in the most frequently-used words between positive tweets and negative tweets, I created two word clouds for the two sentiment classes after aggregating all relevant tweets, as demonstrated in Figure 2. To emphasize subjective expressions, I removed non-interesting objective nouns that are commonly-used in cryptocurrency context such as names of the currencies and titles of the exchange markets. The results suggest that notable dissimilarity exists in the regularly-used words between the two contrasting opinions. Tweets that contain positive emotions tend to include “free”,“thanks”,“play”, “easy”, “mobile”, “app”, “daily”,“faucet”, from which we could picture Twitter users propagating the advantages of adopting digital currencies as well as referrals and simple tasks to earn some coins (e.g. Bitcoin faucet). We also expect to capture an increasingly-popular integration of video games and crypto components, which has recently earned its name as Crypto Game. In contrast, though the negative corpus also highlights “earn”, “usd”, we observe there are many pessimistic utterance underneath, including “stoploss”, “bubble”, “stop”, “hack”, “low”, “loss”. It is also interesting to point out the significant presence of big tech when people state their opinion toward cryptocurrencies, among which “Google” has been detected in the positive community while “Facebook” is frequently mentioned in negative tweets.

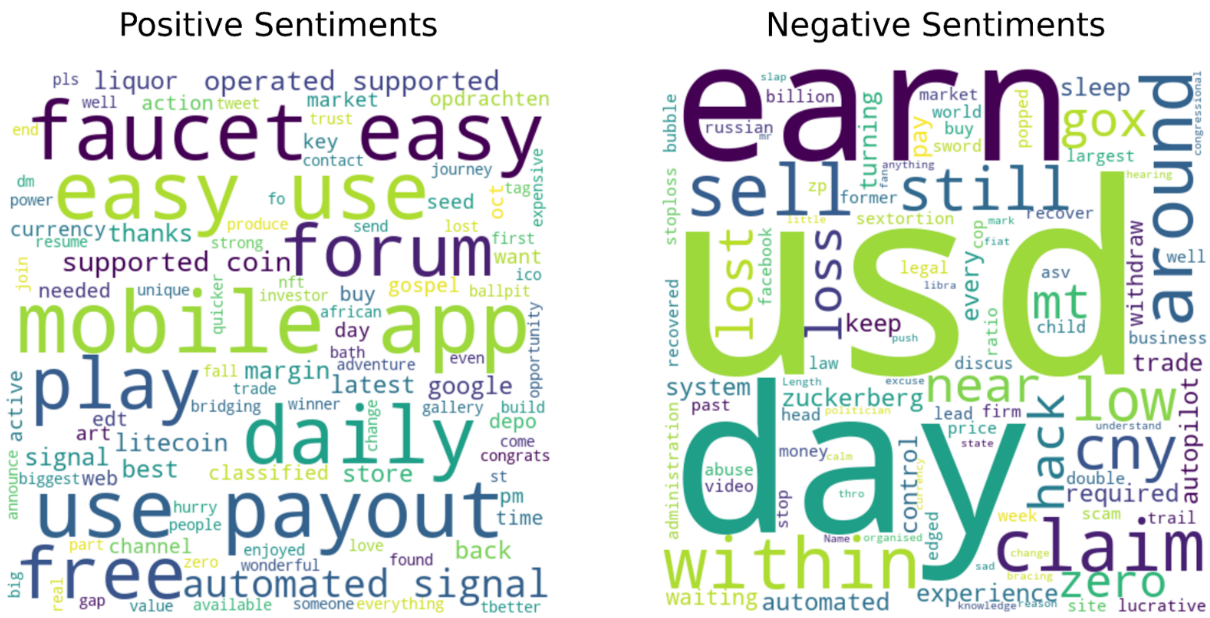


Figure 3: Word Clouds for Positive Tweets and Negative Tweets

3.1.3 Hashtag Co-occurrence Network Analysis

Community Detection on Twitter Hashtag Co-occurrence Network Figure 4 depicts the partitions of hashtag co-occurrence network communities as well as sentiment classifications at the hashtag level. Interesting, the co-occurrence network is highly modularized and contains a few apparently closely-connected subsets, to name a few, the "financial-news" centered cluster comprising of "#bloomberg", "#nytimes" and so forth, that lies on the upper-right of the graph and the lower-right cluster that highlights hashtags such as "#401K", "#savings", "#retirement", and "#neverworkagain". This unique network structure suggests that digital currency is a complex topic that incorporates various sub-topics.

Meanwhile, we notice that almost every hashtag has been labeled as positive. This is within our expectation because each hashtag is essentially a collection of tweets and the approximate percentage of positive tweets in most large samples is around 73%, according to our exploratory analysis above. The comparative landscapes of sentiment distribution and co-occurrence clustering imply that it is difficult to infer opinion dynamics from network-based information regarding whether the two hashtags occur together.

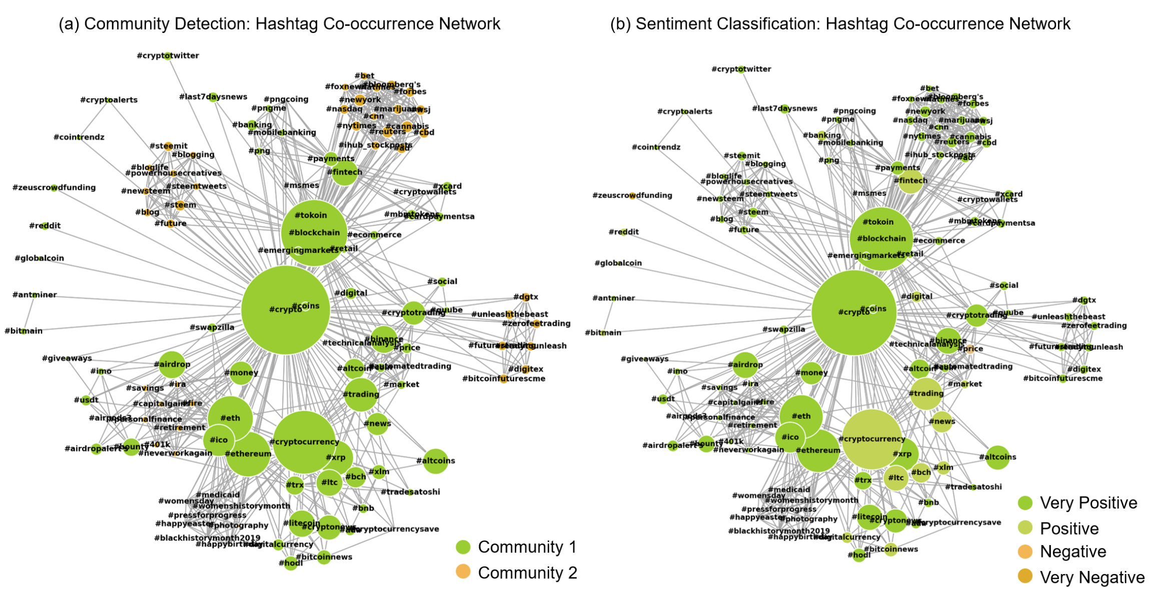


Figure 4: Hashtag Co-occurrence Network Graph and Sentiment Classification

3.2 Historical Dynamics

I depicted the historical polarity dynamics in two aspects: Sentiment Proportion and Sentiment Intensity. Figure 5 and Figure 6(a) demonstrates that how the positive and negative sentiment classes are segmentized. Specifically, figure (5) provides a bird-eye overview with regard to the absolute number of tweets per day. At first glance, we discover that the number of tweets exhibit a certain degree of weekly seasonality, and the landscape of polarization is relatively stable all the time. This preliminary finding is further evaluated in Figure 6(a), where the percentage trend is delineated. The weekly pattern is no longer noticeable in the percentage form and we identify drastic change to the polarization in certain time period, for instance, Sep 29, 2019, when the price excessively plunges.

For the sentiment intensity fluctuations in the past periods, we find that the average intensity for positive tweets is almost higher than that of the negative tweet along with the time frame, which reiterates the initial exploratory finding from Table 1 that positive tweets retain a stronger emotion. Additionally, it is interesting that I identify a negative correlation in the average intensity between the positive tweets and negative tweets. The Pearson coefficient is -0.59 and the corresponding p-value is smaller than 0.001.

This test result, together with the time series trend delineated in Figure 6, reveal that the relationship between the two sentiment classes in terms of the emotion intensities can be non-trivial. This emphasizes that the negativeness will be alleviated when the social media platform is populated with strong positive tweets, and vice versa.

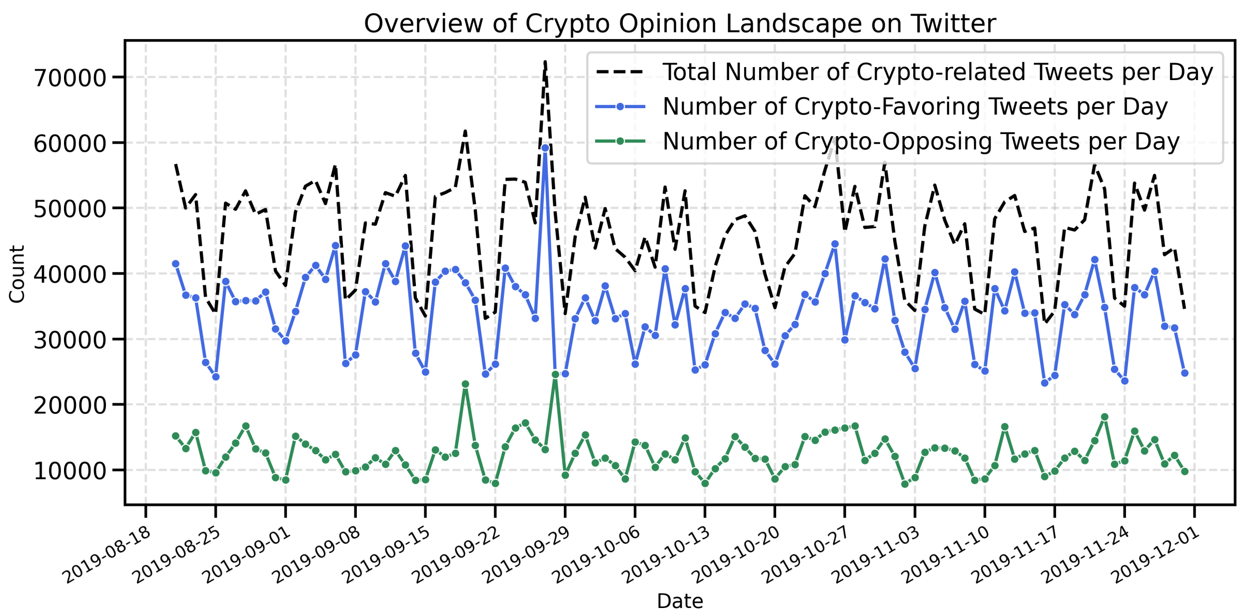


Figure 5: Total Number of Cryptocurrency-related Tweets, Tweets with Positive Sentiments and Negative Sentiments

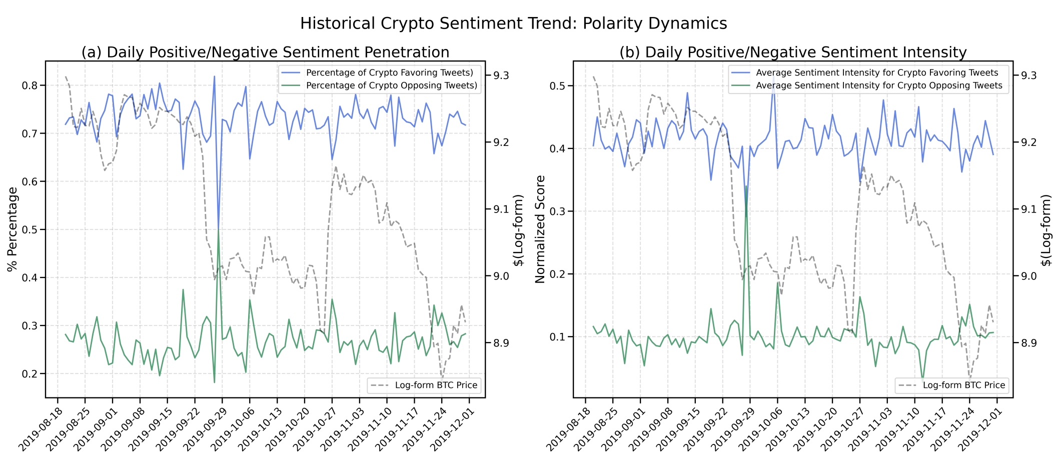


Figure 6: Opinion and Sentiment Historical Dynamics with Market Price Fluctuations

3.3 Prediction

As described in section 2.2.5, I aim to predict future crypto opinions using historical sentiment values through 5 time series models: Simple Exponential Smoothing (SES), Double Exponential Smoothing (or Holt's Method), Triple Exponential Smoothing (or Holt-Winters Method), ARMA (AutoRegressive Moving Average), Facebook Prophet. The hyper-parameters for the exponential smoothing family are automatically tuned by specifying the *initialization_method* as *estimated*. The lag order and moving average term are determined as 4 and 1 by exploring the Partial Autocorrelation and the Autocorrelation graphs. Figure 7 displays the predicted values in the next 14 days using the above 5 methods in dotted lines as well as the actual values in scattering forms. It is evident to notice that all the five time-series predictions are not flexible enough and thus fail to capture the underlying volatility in the opinion dynamics. This implies that the opinion dynamics can be too fickle to difficult to predict from the historical values.

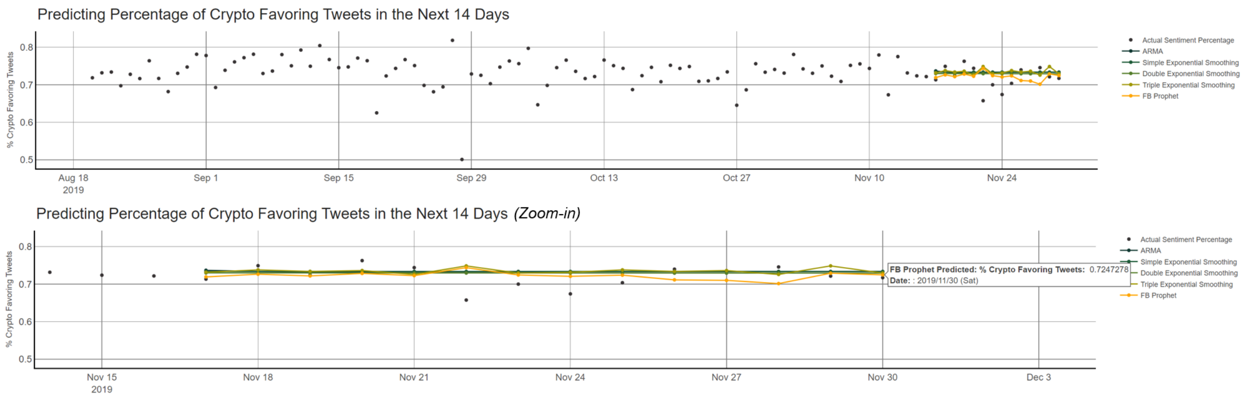


Figure 7: Predicting Sentiments and Opinions toward Cryptocurrency: (1) Simple ES; (2) Double ES; (3) Triple ES; (4) ARMA; (5) FB Prophet

The Root-mean-square-error (RMSE) criteria in Table 2 reveal that the Double Exponential Smoothing, with only the level and trend components in the time series model, generates the most accurate predictions. Interestingly, the Simple Exponential Smoothing has the second-lowest RMSE, which is possibly due to the underperformance of all the five models. The weak predictability of Triple Exponential Smoothing is mainly attributed to the trivial presence of seasonality pattern, which also explains partially the incompetency of the famous Facebook Prophet. Another explanation for this is that the Facebook Prophet is particularly capable of long-range forecasts based on a large collection of time series data, which is not the case in the scenario of this study.

Based on the above empirical results, I would conclude that it is hard to predict the cryptocurrency opinion dynamics in the future 14 days using merely the historical values in the past 89 days due to the lack of seasonality and sample constraint.

Table 2: Evaluations of the 5 Time Series Methods using RMSE

Method	RMSE
Double ES	0.029762
Simple ES (Baseline)	0.031296
ARMA(4,1)	0.031342
FB Prophet	0.034071
Triple ES	0.0344

4 Discussions

The study provides empirical evidence to complement the behavioral research on the cryptocurrency topic, which has been narrowly focused on financial applications such as analyzing volatility and predicting price. Additionally, this Crypto-concentrating research enriches the variety of the topics in the computational social media analysis, where the majority of the research covers political affiliation and voting behaviors.

Nevertheless, there do exist limitations in this study that could be improved. Firstly, due to the data constraint, I don't yet have ground truth labels of whether a Bitcoin-related tweet contains positive sentiments. Consequently, the process of fine-tuning BERT is essentially a compromised weak-form supervised method that couldn't eliminate the bias introduced by VADER. Meanwhile, lacking these authentic labels prevents us from evaluating any model performance during the natural language processing procedure.

Furthermore, sentiments do not sufficiently reflect one's true opinion affiliation. While emotional polarity is more fickle and context-dependent, opinions tend to be long-lasting and consistent across different tweets. In this case, correctly capturing the context of the events is crucial to measuring people's opinions. Therefore, more in-depth textual mining analysis, contextual word embedding, NLP models are needed to excavate the actual opinions. Meanwhile, I also observe there are many anomalies that receive extremely high retweets and likes. A closer look at the specific tweet text revealed that these viral spread tweets are mostly business campaigns introducing their new Crypto Games or propagating their registration referrals for cryptocurrency exchange apps. The attached positive sentiment labels no double fail to reflect the genuine opinions toward cryptocurrency, and thus call for a more thorough and meticulous textual cleaning. However, the empirical result regarding sentiment dynamics provided in this research could function as a starting point for more detailed studies in the future.

Lastly, the size of the data set is constrained due to the scale limit of Twitter Developer API. This leads to the weak predicting power of the time series models presented in the end. Meanwhile, one may also argue about the generalized ability to infer opinion dynamics based on Twitter data only. Since most of the people who are interested in or have invested in decentralized financial assets are the younger generation who are relatively more willing to adopt new technology, despite the popularity and inclusiveness of Twitter, discussions about cryptocurrencies tend to be more active on other more youth-oriented and specialized communities.

Another potential direction to enrich this research is to break down the opinion polarity dynamics by key dimensions such as account affiliation to depict a more unique picture of the landscape on social media platforms. Authors of different identities are not likely to speak the same language. Organization accounts are believed to be more neutral. Similarly, influencers and other individual authors have their own unique patterns, which are worth detailed investigation.

5 Conclusion

The study began by demonstrating the polarization landscape by projecting the tweet representation embedding into the 3-dimensional space, modeling dominant topic words, and detecting hashtag co-occurrence network communities. I found that the opinion polarity toward cryptocurrency can be non-trivial - the positive tweets constitute around 73% and the negative tweets make up the rest 27%, leading to a relatively large polarity of around 46%. Additionally, the intensity of the positiveness expressed in crypto-favoring tweets is approximately 1.5x (0.56 over 0.38) stronger than the negative tweets. Meanwhile, the most frequently-used topic words are also different between the two tweet corpora with contrast sentiment classifications. The positive tweets tend to include “free”,“play”, “easy”, “mobile”, and “app”, from which we could picture Twitter users actively propagating the advantages of adopting digital currencies and integrating digital currencies as part of their living. While the most representative words in the negative tweets are remarkably characterized by pessimistic adjectives and verbs including “stoploss”, “bubble”, “stop”, “hack”, “low”, “loss”.

The time-series data analysis on historical trends echoes the previous finding regarding substantial opinion polarization. The percentage of positive tweets remains higher than the percentage of negative tweets and centers around a proportion of 73%. Though vastly volatile, generally, the sentiment intensity is also higher for tweets that are favorable to digital currencies.

Nevertheless, the study revealed that it can be difficult to effectively predict the future cryptocurrency opinion dynamics using merely its historical values of itself. However, the modeling analysis is potentially limited due to the sample constraint imposed by Twitter Developer API.

6 Data and Code Availability Statement

The code used in the current study is available through the Google Colab Notebook. To abide by Twitter Developer Policy, the raw tweet archive will not be shared publicly. More information is available upon reasonable request from the author.

7 Appendix

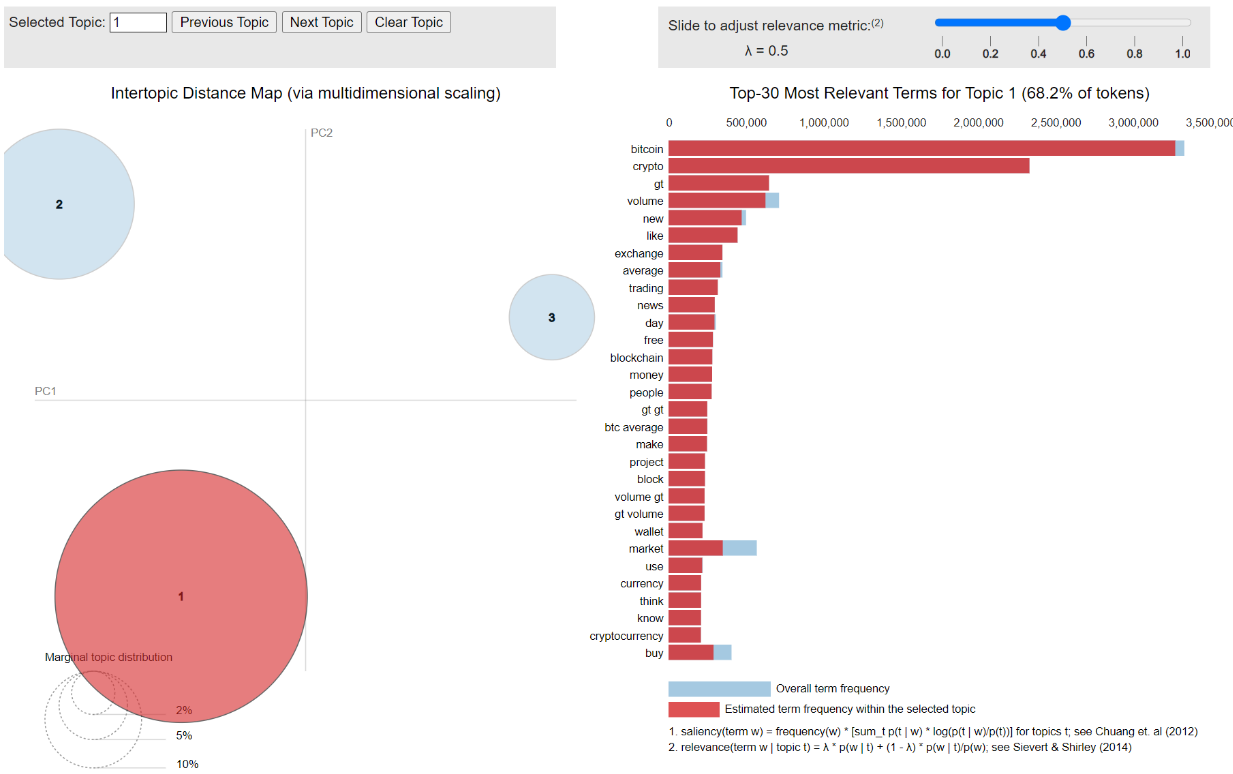


Figure 8: LDA Topic Modeling - Topic 1: Intertopic Distance Map and Top-30 Most Salient Words

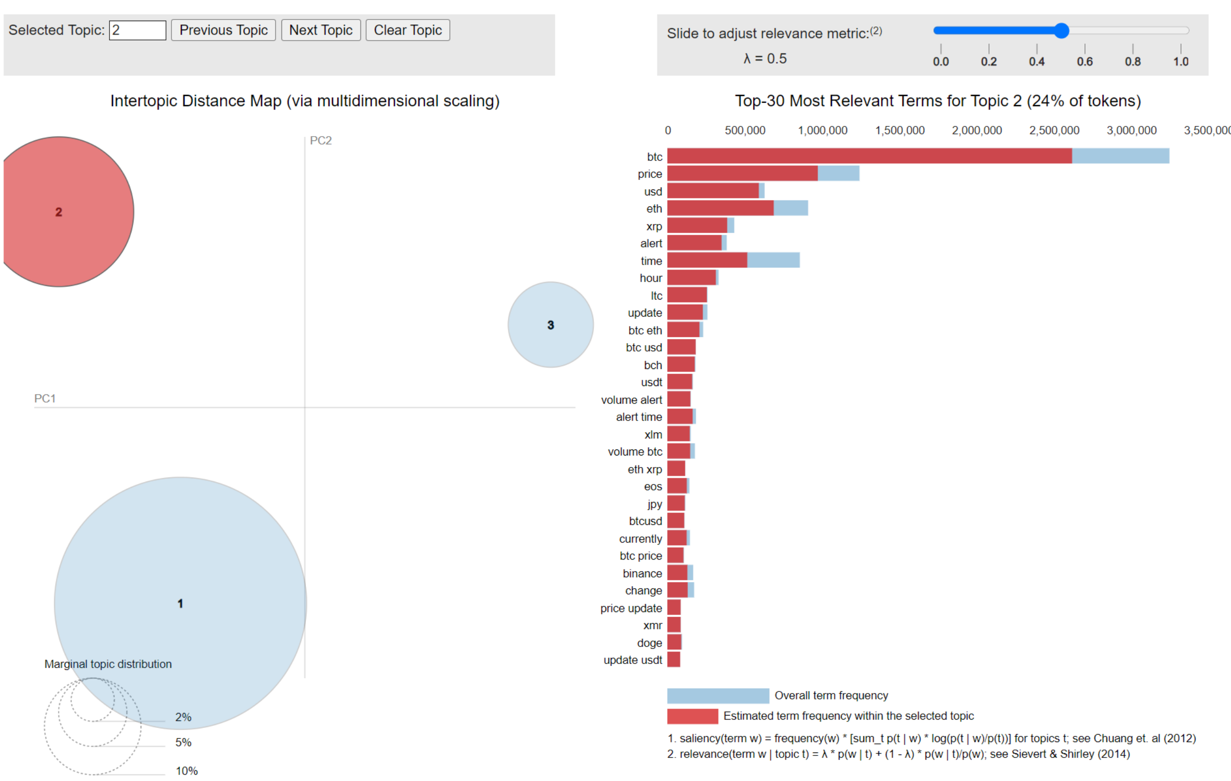


Figure 9: LDA Topic Modeling - Topic 2: Intertopic Distance Map and Top-30 Most Salient Words

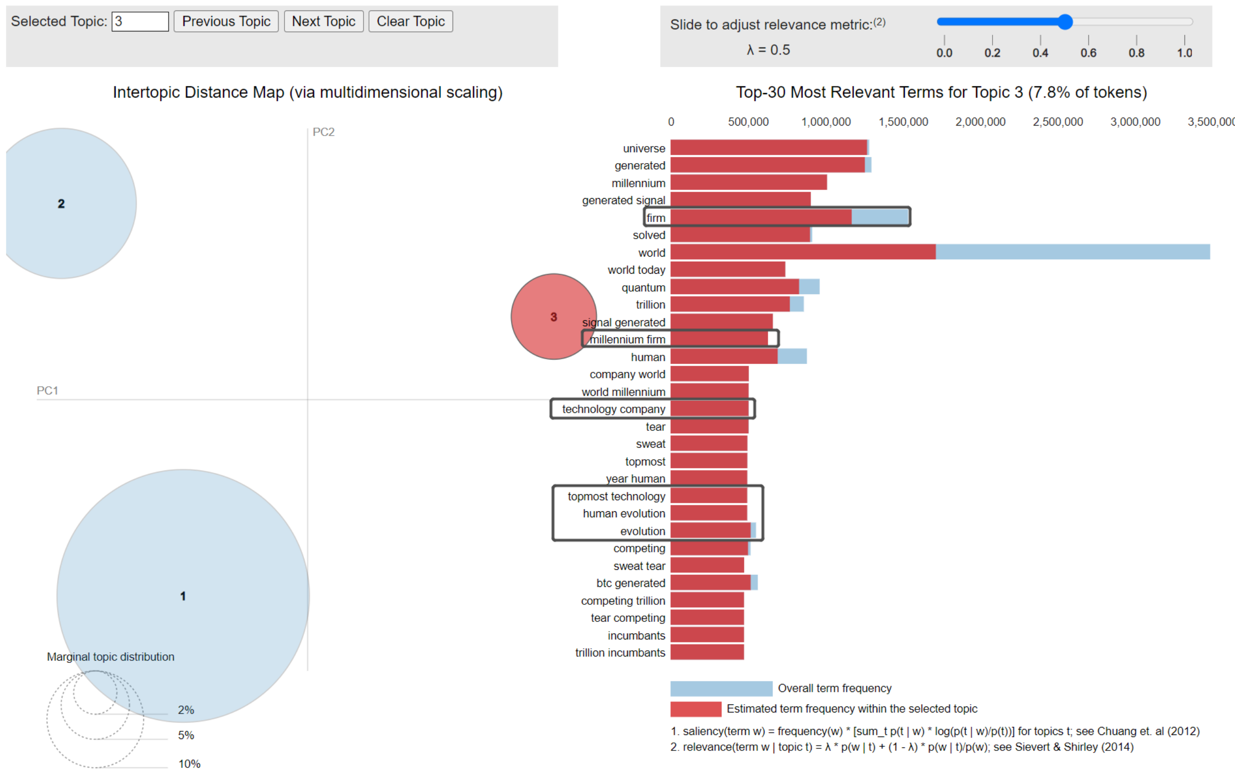


Figure 10: LDA Topic Modeling - Topic 3: Intertopic Distance Map and Top-30 Most Salient Words

References

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge university press.

Martinez-Romo, J., Araujo, L., Borge-Holthoefer, J., Arenas, A., Capitán, J. A., & Cuesta, J. A. (2011). Disentangling categorical relationships through a graph of co-occurrences [Publisher: American Physical Society]. *Physical Review E*, *84*(4), 046108. https://doi.org/10.1103/PhysRevE.84.046108

Bovet, A., Morone, F., & Makse, H. A. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *8*(1), 8673. https://doi.org/10.1038/s41598-018-26951-y

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396. https://doi.org/10.1038/s41562-018-0353-0