

Cryptocurrency Sentiment Waves: Analyzing Dogecoin and Reddit Interactions

Aruem Cho ¹ Carleigh Marie Chiavello ² Merve Yilmaz ³ Tristan Gourmelen Raimbault ⁴

¹Aruem.Cho@etu.univ-paris1.fr , ²Carleigh-Marie.Carleigh@etu.univ-paris1.fr , ³Merve.Yilmaz@etu.univ-paris1.fr , ⁴Tristan.Gourmelen-Raimbault@etu.univ-paris1.fr



Research Question

The purpose of this analysis is to examine how the sentiment of dogecoin-related posts on the cryptocur-rency and dogecoin subreddit pages, r/Cryptocurrency and r/Dogecoin, correlate with daily dogecoin returns using Natural Language Processing (“NLP”) techniques. We employed FinBERT, a pre-trained NLP model designed to analyze the sentiment of financial text, along with a social media-specific lexicon developed by Thomas Renault. Given the large and active dogecoin community on social media, particularly Reddit, we hypothesize that the sentiment of Reddit posts influences its price. Additionally, we compare dogecoin returns with those of bitcoin and the SP to determine whether dogecoin returns correlate with broader cryp-tocurrency and stock market trends.

Why does this question matter?

Unlike established cryptocurrencies such as bitcoin or ethereum, dogecoin has no intrinsic value; its price is determined purely by popularity and community interest. Launched in 2013 as the first “meme coin,” the volatility of dogecoin volatility makes it an intriguing case for analyzing how sentiment on platforms like Reddit drives prices.

Related Work

The rise of cryptocurrencies that are highly influenced by social media hype and online discussions like dogecoin, highlights the need to study the relationship between public sentiment and price movements. Traditional financial models often fail to capture the sentiment-driven nature of cryptocurrency markets, a gap addressed by FinBERT (Araci, 2019).

- **Social Media and Sentiment Analysis in Cryptocurrency Markets** Several studies explore sentiment’s impact on cryptocurrency prices. Kristoufek (2013) links Google and Wikipedia search trends to bitcoin returns. Anamika et al. (2021) focus on sentiment indices like the Happiness Sentiment index derived from Twitter. Wooley et al. (2019) demonstrate that Reddit sentiment predicts bitcoin and ethereum prices with 74.2% and 73.1% accuracy, respectively.

- **FinBERT Applications for Sentiment Analysis** Huang et al. (2022) show that FinBERT, fine-tuned on financial data, outperforms traditional models in sentiment tasks. Gu et al. (2024) combine FinBERT with Long Short Term Memory (LSTM) to analyze financial news, achieving superior performance. Similarly, Girsang et al. (2023) predict ethereum and solana prices by integrating FinBERT-driven social media sentiment and historical price data.

Data and Descriptive Statistics

The analysis focuses on data spanning 1 January 2021 to 31 December 2021, a period characterized by the dogecoin’s rapid rise in popularity. Notably, on 8 May 2021, dogecoin achieved its highest market capitalization of \$88.8 billion, reaching a price of \$0.6818 per coin [1]. By concentrating on this timeframe, during which dogecoin garnered significant online attention, we aim to better understand the correlation between social media activity and fluctuations in its price.

We found Reddit posts archived and packaged using Reddit Pushlift on an academic torrent web. From this, we selected relevant Reddit submissions for our study, focusing on the “dogecoin” and “cryptocurrency” subreddits, which had significant mentions of dogecoin during the study period. The number of observa-tions for the subreddits are 982,365 and 26,015, respectively. To track only dogecoin-related discussions, r/Cryptocurrency was filtered for posts containing “doge.”

Additionally, the price data for all assets was sourced from Investing.com. Instead of using the raw prices of the coins and index, we calculated the return values. These were computed by subtracting the price of the asset at time (t-1) from the price at time t, and then dividing that by the price at time (t-1). This return value was used in our analysis.

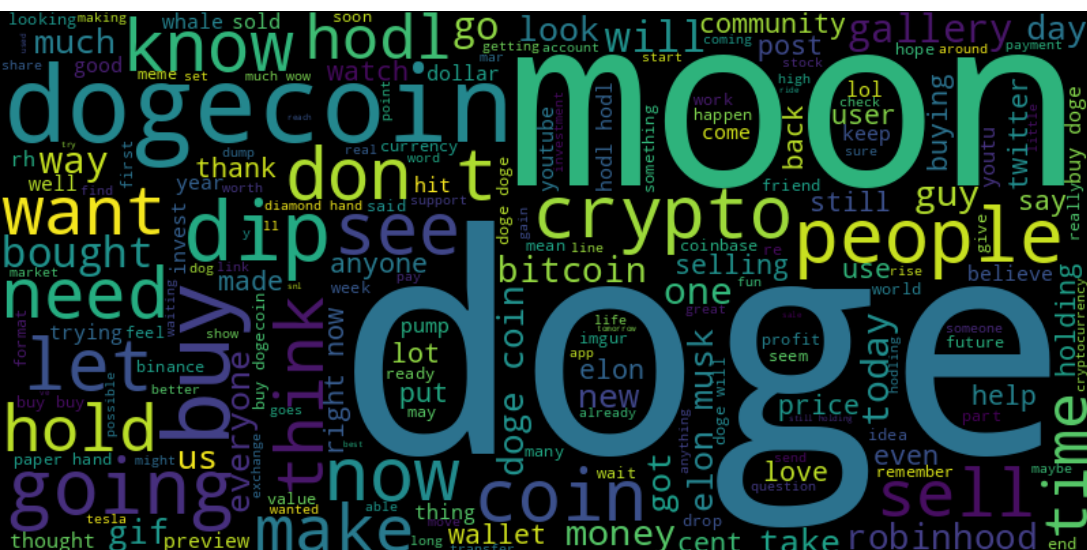


Figure 1 reveals a Wordcloud map visualizing the most frequent words from r/Cryptocurrency and r/Dogecoin. It clearly demonstrates that some of the most striking words are “moon” and “want”, which all imply positive sentiment. Furthermore, action words, such as “hold” and “bought” show direct relevance to market activity.

Figure 1. Wordcloud Map of the Two Subreddits.

Model and apporaches

To investigate the factors driving dogecoin performance, we conducted three distinct analyses. First, we assessed the sentiment scores of relevant posts. Second, we analyzed trends in other financial market instruments, such as bitcoin and the SP 500, which we hypothesized might influence dogecoin behavior. Lastly, we employed regression analysis to model the combined impact of these factors on dogecoin returns.

1.Sentiment Analysis: To analyze sentiment, we used FinBERT, a pre-trained NLP model specifically de-signed for financial texts, to classify Reddit posts into positive, neutral, or negative sentiments. To enhance the accuracy of the analysis, we incorporated a social media-specific lexicon developed by T. Renault, which provided a more nuanced sentiment evaluation. First, sentiment scores for each subreddit post were calcu-lated using FinBERT. Next, the social media-focused lexicon was applied to refine the sentiment estimation. Finally, average daily sentiment scores were calculated for each subreddit and compared with dogecoin daily returns.

2. Analyzing Returns to doge, BTC, S&P: We analyzed the returns of dogecoin, bitcoin, and the SP 500. Bit-coin and the SP 500 were selected as benchmarks because they serve as major market indicators: bitcoin, as the dominant cryptocurrency, and the SP 500, as a key stock market index. Both are likely to influence dogecoin price movements due to their market significance and the correlations observed in investor be-havior.

3. Regressions: To enhance the analysis, we employed two OLS regression models. The first model simplifies the approach by using bitcoin and SP 500 returns as independent variables to analyze dogecoin returns. This model aims to capture the influence of these major market assets on dogecoin price movements. In the second model, a more comprehensive regression was conducted, with dogecoin returns as the de-pendent variable. The independent variables, represented by the X matrix, included lagged dogecoin returns (one- and two-period lags), bitcoin returns, and SP 500 returns. Additionally, daily sentiment scores from the dogecoin and cryptocurrency subreddits were incorporated as explanatory variables to assess the impact of social media sentiment on dogecoin returns. Following this methodology, estimated regressions are as follows:

$$\text{Doge Return}_t = \beta_0 + \beta_1\text{BTC}_t + \beta_2\text{SPY}_t + \varepsilon_t \quad (1)$$

$$\text{Doge Return}_t = \lambda_0 + \lambda_1\text{Sentiment Score}_{it} + \mathbf{X}_t\boldsymbol{\lambda} + v_t \quad (2)$$

Results

1.Sentiment Analysis:

We analyzed the correlation between daily sentiment scores of the two subreddits and dogecoin returns using the Pearson correlation coefficient, which measures the strength and direction of a linear relationship between two variables. The results showed weak correlations: for r/Dogecoin the correlation was 0.0960 (p-value = 0.0673), and for r/Cryptocurrency it was 0.0579 (p-value = 0.2704). Although neither result was statistically significant, the analysis offers insights into how public sentiment might influence market returns, laying the groundwork for further research into additional contributing factors.

We then plotted the sentiment scores of each subreddit alongside dogecoin returns. Figure 2 illustrates that the sentiment score of r/Cryptocurrency tends to be larger than that of r/Dogecoin throughout the period. However, this is likely due to the significantly lower number of posts on r/Cryptocurrency. Nevertheless, despite not explaining all price fluctuations, it can be observed that the sentiment scores and returns fluc-tuate together overall.

2.Analyzing Returns to Doge, BTC, S&P:

The returns of dogecoin, bitcoin, and the SP 500 were analyzed, revealing similar trends through graphical comparisons. As shown below, dogecoin exhibits the highest volatility among the three, followed by bitcoin, with the SP 500 displaying the lowest level of volatility. Interestingly, although the amplitude differs, the overall trend remains strikingly similar.

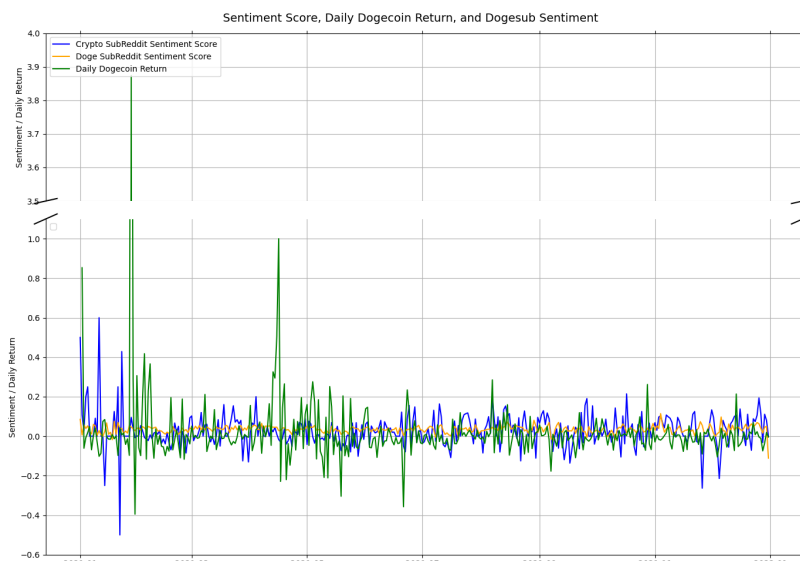


Figure 2. Daily return of dogecoin and sentiment scores of subreddits

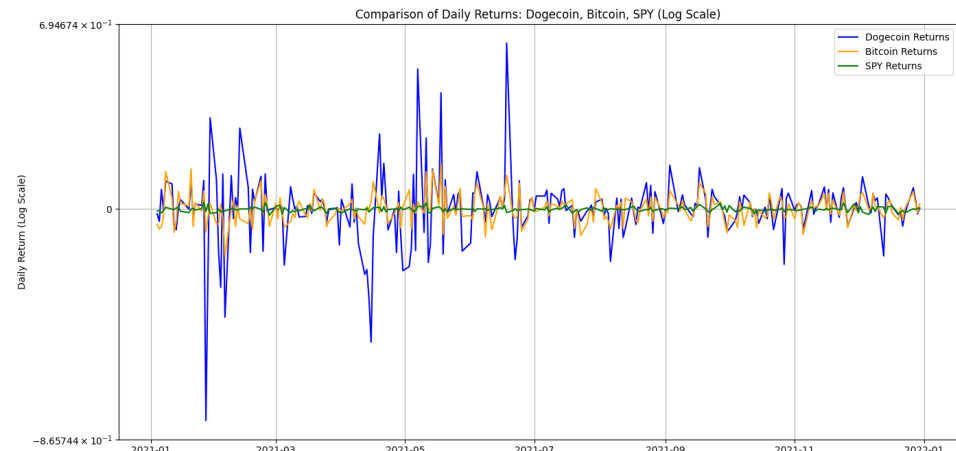


Figure 3. Daily returns with dogecoin on a secondary axis

3. Regressions:

The two tables below present the results of OLS regressions with dogecoin returns as the dependent variable. Table 1 indicates that bitcoin returns are significant at the 1% level when regressed alone and with the SP, while SP returns are only significant when regressed alone. Table 2, column 1 shows that the dogecoin sentiment score is significant at the 5% level, meaning a positive sentiment score increases dogecoin returns by 1.56 units on average. Additionally, SP returns are significant at the 1% level in columns (2) and (3), and at the 5% level in column (5).

	BTC Only	S&P Only	BTC+S&P
Intercept	-0.0012	0.0008	-0.0009
BTC return	1.2591***		1.2481***
S&P return		2.2915***	0.2412
R-squared	0.244	0.022	0.244
Adjusted R-squared	0.241	0.018	0.238

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1. OLS Regression using bitcoin and S&P returns to predict dogecoin returns

	(1)	(2)	(3)	(4)	(5)
Intercept	-0.0518	-0.0669	-0.0602	0.0209	0.0258
r/Dogecoin sentiment score	1.5638**	1.7736	1.7354		
r/Cryptocurrency sentiment score			0.0879	0.1647	0.1276
Lagged sentiment	0.5247	1.0808	0.8393		
Lagged return 1	0.0215	0.0742	0.0705	0.0617	0.0855
Lagged return 2	-0.0687	-0.0067	0.0795	0.1280	0.1275
BTC return	-0.1451	-0.2788	-0.3374	-0.1244	-0.3529
S&P return		5.8100***	5.7952***		5.3040**
R-squared	0.023	0.049	0.048	0.008	0.032
Adjusted R-squared	0.009	0.025	0.021	-0.008	0.012

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2. OLS Regression where the dependent variable is dogecoin returns

Conclusion

This study examines whether social media and other major cryptocurrencies influence dogecoin prices from various perspectives. By utilizing advanced sentiment analysis tools like FinBERT and social media lexicons to analyze Reddit data, we demonstrated the potential of sentiment-driven models to capture trends and correlations in dogecoin returns. The findings suggest that sentiment scores can serve as useful predictors of price dynamics for dogecoin.

However, the analysis also revealed the limitations of relying solely on sentiment data, as dogecoin’s price movements were found to align closely with broader market trends, such as those of bitcoin and the SP 500. This suggests that while sentiment plays a significant role, other market-wide factors and economic conditions also influence cryptocurrency returns.

In conclusion, while sentiment analysis offers valuable insights into cryptocurrency behavior, future research should focus on incorporating additional market and economic variables to better understand and predict price movements in this rapidly evolving space.

References

- [1] Anamika, Anamika and Sowmya Subramaniam. Do news headlines matter in the cryptocurrency market? *Applied Economics*, 54(54):6322–6338, 2022.
- [2] Araci, D. FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.
- [3] Girsang, Abba Suganda, et al. Hybrid LSTM and GRU for cryptocurrency price forecasting based on social network sentiment analysis using FinBERT. *IEEE Access*, 11:120530–120540, 2023.
- [4] GlobalData. Dogecoin’s market capitalization history, 2021. Accessed: 2024-12-20.
- [5] Huang, Allen H., Wang, Hui, Yang, Yi. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2):806–841, 2023.
- [6] Gu, Wen jun, Zhong, Yi hao, Li, Shi zun, Wei, Chang song, Dong, Li ting, Wang, Zhao yue, Yan, Chao. Predicting stock prices with finbert-1stm: Integrating news sentiment analysis. *In Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing*, pages 67–72, 2024.
- [7] Kristoufek, Ladislav. Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3(1):3415, 2013.
- [8] Wooley, Stephen, Edmonds, Andrew, Bagavathi, Arunkumar, Krishnan, Siddharth. Extracting cryptocurrency price movements from the Reddit network sentiment. *In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 500–505, IEEE, 2019.
- [9] Renault, Thomas. Extracting StockTwits Lexicon for Sentiment Analysis. <http://thomas-renault.com/data.php>