GRAND ACADEMIC PORTAL
RESEARCH JOURNALS

GAP iNTERDISCIPLINARITIES
A Global Journal of Interdisciplinary Studies
(ISSN – 2581-5628)
Impact Factor: SJIF - 5.363, IIFS - 4.875
Globally peer-reviewed and open access journal.

DYNAMIC TOPIC MODELING OF BITCOIN DISCUSSIONS: AI-DRIVEN TOPIC EVOLUTION AND ENGAGEMENT ANALYSIS ON X

Saifullah Rahimi[1], Dr. Hiren Joshi[2], Dr. Hardik Joshi[3]

[1]PhD Scholar, Department of Computer Science, Gujarat University, India
[1]Saifullah.r@gujaratuniversity.ac.in, +919714535459
[2,3]Professor, Department of Computer Science, Gujarat University, India
[2]hdjoshi@gujaratuniversity.ac.in, [3]hardikjoshi@gujaratuniversity.ac.in

Abstract

The world is today more interconnected than ever due to the increasing influence of digital currents. The internet, particularly social media, enables the rapid spread of vast amounts of information, highlighting the importance of speed in business. This study aims to provide a new perspective on the conversation around Bitcoin on X (formerly Twitter), employing dynamic topic modeling to identify shifting Bitcoin discourse trends, sentiment, and user interaction patterns. The rise of sophisticated artificial intelligence tools, such as Grok AI, which uses real-time Twitter data to offer insights and forecasts, emphasizes even more Twitter's relevance as a data source. Although this study does not directly utilize Grok AI its mere existence demonstrates the increasing relevance of Twitter as a real-time analytics tool, presenting the opportunity to merge social media data with best-of-breed financial research.

Leveraging AI-powered analytical methods, the present work addresses the problem of scalability in traditional approaches of insights extraction from large-scale social media data and its bridging with financial research. Analyzing a dataset of 4.8 million Bitcoin tweets utilizing advanced computational techniques, such as sentiment analysis, introduces a new phase in the Bitcoin conversation by using Dynamic Latent Dirichlet Allocation (dLDA) to explore the nuances of Bitcoin discourse. The results indicate discussions primarily revolve around price speculation & regulatory developments, technological innovations, and security concerns. Key findings include the sentiment fluctuations and how user traits affect engagement with Bitcoin content. Also, a comprehensive sentiment analysis shows that the evolution of sentiment in the community is well correlated with the volatility of Bitcoin price, highlighting the possible predictive power of social media analytics in financial markets. Additionally, this research investigates the relationship of discussions via hashtag analysis to form an understanding of how the public perceives trends under rapidly evolving financial circumstances.

Keywords: *Dynamic Latent Dirichlet Allocation (dLDA), Twitter-based Bitcoin Sentiment Analysis, AI-driven financial analytics*

INTRODUCTION

The rise of social media has fundamentally changed how information, opinions and trends are distributed especially in industries as sensitive to public sentiment as cryptocurrency. X (formerly Twitter) provides a barometer for public sentiment, and tweets frequently reflect real time shifts in user perceptions and market dynamics. Bitcoin, being the first cryptocurrency, has seen more debate, speculation, and research interest on social media sites than any other asset (Julianto et al., 2022). The resulting data is vast and extremely dynamic, and while they can certainly be represented in numeric terms, sophisticated computational methods need to be used to make sense of the topics, contexts, as well as the sentiments behind the contexts, and also how they change over time. So, sentiment analysis is a computational technique for assessing the emotional tone of a text. This is important to have an understanding of the attitudes, opinions and emotions covered in the individual Bitcoin-related tweets. Sentiment analysis tools like VADER and TextBlob classify sentiments into negative, positive, or neutral with a mix of natural language processing (NLP) and machine learning (Hutto & Gilbert, 2014). Sentiment analysis has been widely used, by these tools (VADER and SentiWordNet) to understand public sentiment towards several issues from health services(Dikiyanti et al., 2021) to electric vehicles (Suresha & Kumar Tiwari, 2021).

(Phillips & Gorse, 2018) suggested that increased social media activity surrounding Bitcoin correlates with sharp changes in its price, as they demonstrated that spikes in Twitter activity tend to precede large price movements. According to (Rossi & Magnani, 2021), network analysis techniques also used in studying the evolution of hashtag communities, and provided us insight into how topics are distributed across time and how topics emerge. (Kwak et al., 2010) when he studied Twitter in-depth and found that there was a small community of very connected users which acts as a key player in the propagation of information. Recent research has highlighted the need for dynamic topic modeling approaches like dLDA that allow the model to

adapt through fine-grained updates for the topics as they change over time, giving precise insight into how public discourse adapts in relation to new events (Nguyen & Shirai, 2015). Such methodologies have been used in areas ranging from pandemic relevant discourse (Jang et al., 2021), and public reactions(Habibi et al., 2021) to economic crises, revealing their versatility and quiet depth. High-volume tweets have been useful in identifying relatively common themes, with topic modeling methods such as Latent Dirichlet Allocation (LDA) (Mishra et al., 2021). The dLDA is a generalization of the LDA, a generative statistical model for finding the topics in a set of abstracts by modeling the documents from the perspective of unobserved groups which gives context. These unseen clusters explain why certain data points have so much in common. dLDA allows not only capturing the topics but also their changes over time, suggesting it is a practical approach for analyzing temporal variations in extensive textual repositories such as Twitter streams (Blei & Lafferty, 2006).
However, considering the acknowledged amount of existing works, the analysis of Twitter's Bitcoin-related conversations is still a relatively under explored territory where dynamic topic modeling can offer more unprecedented results. In an attempt to do so, we use dynamic topic modeling in conjunction with sentiment analysis in order to reveal the complex interplay of topics and sentiments that form Bitcoin trends.

RESEARCH OBJECTIVES AND QUESTIONS

The ultimate purpose of this research is to provide an overview of the trends, issues, and sentiments that fuel Bitcoin-oriented conversations, thus serving a rather invaluable insight for investors, analysts, and devotees. Utilizing the latest computational linguistics and machine learning methods, this study aims to:

1. Systematically investigate temporal topics evolution in the Bitcoin dialogue on Twitter to help better understanding of changing nature of the Bitcoin dialogue associated with the leading cryptocurrency.
2. To Find out how sentiment changes over time & explore Correlations Between Bitcoin Sentiment Trends and Real-World events.
3. Analyze how user engagement facilitates the spread and visibility of these topics according to the followers by the users on topic Engagement that serves as an implicit measure of the credibility and popularity of the information.
4. Explore the interconnectedness of Bitcoin-related subjects via network analysis of hashtags, revealing the elaborate mesh of conversation and the centrality of each subject to the ongoing discourse.

This research advances current research in the area of topic modeling and sentiment analysis (Dikiyanti et al., 2021; Julianto et al., 2022) and sets a foundation for its application in the cryptocurrency space (Mishra et al., 2021; Nguyen & Shirai, 2015). In addition, this study contributes to financial discourse investigation through machine learning, an area where quick interpretation of sentiment data can result in notable financial implications (Jang et al., 2021). This research proves important in computational finance, social media analysis, and linguistic studies. This study aims to assist stakeholders (financial analysts, investors, policymakers, etc.) with actionable insights and a better comprehension of bitcoin market sentiments by utilizing public discourse and sentiment on an economically and socially important topic, which would be helpful due to its high textual relevance and potential hindsight before the real market moves.

METHODOLOGY

In this study computational analysis performed on the tweets related to Bitcoin. This methodology implement steps for the dynamic topic modeling method, from data collection to implementing the dynamic topic modeling method, which are described in detail.
Data collection: Kaggle, a well-known platform for data science and machine learning, is used extensively in the study's first data collection phase. The dataset consisted of about 4.8 million tweets (Bitcoin Tweets, n.d.) containing tweet text, metadata (e.g., date, user verification status, follower count), and engagement metrics, referenced to Bitcoin-related topics. Specifically, the data focuses on tweets with relevant keywords surrounding Bitcoin, ensuring data pertinence and emphasis.
Data Preprocessing: The dataset preprocessed at a large scale to enhance and prepare it for risk prediction and analysis. This includes multiple steps:
Data Cleaning: The text data is cleaned up at first by removing stop words, URLs, user handles, digits, and special characters that usually add more noise than the value in text analysis.
Lemmatization: Natural language processing techniques used for this study to reduce words to their base/root. This allows the model to treat multiple variants of the same word as a single entity, which makes analysis easier.
Tokenization: Next, the processed text data broken down into its respective tokens. This is a critical step when performing text analysis since it converts the unstructured input into a format that can be used by machine learning models. Figure 1 illustrates before and after data preprocessing of Bitcoin-related tweet texts.

text	hashtags	is_retweet	processed_text
Blue Ridge Bank shares halted by NYSE after #b...	['bitcoin']	False	[blue, ridge, bank, shares, halted, by, nyse, ...
😎 Today, that's this #Thursday, we will do a "...	['Thursday', 'Btc', 'wallet', 'security']	False	[today, that, s, this, thursday, we, will, do,...
Guys evening, I have read this article about B...	NaN	False	[guys, evening, i, have, read, this, article, ...
$BTC A big chance in a billion! Price: \487264...	['Bitcoin', 'FX', 'BTC', 'crypto']	False	[btc, a, big, chance, in, a, billion, price, b...
This network is secured by 9 508 nodes as of t...	['BTC']	False	[this, network, is, secured, by, nodes, as, of...

Fig. 1. Before and After data preprocessing of tweet texts comparison

dLDA Implementation: The foundation of our analysis relies on the implementation of the Dynamic Latent Dirichlet Allocation (dLDA) model, utilizing the LdaSeqModel class from the gensim library. This dynamic aspect is crucial to the modelling of the time variation of public views and opinions regarding Bitcoin.

Model Training: Training the dLDA model required a lot of time and thought. The model is built with specific hyper-parameters as per the research requirements and dataset characteristics. These included the number of subjects, the temporal spread of those subjects, and the temporal evolution of the word distributions associated with the topics. The generated corpus utilized for training the model, thus permitted us to track and trace the evolution of different topics over time. In this study we used Gensim'sLDASeqModel, which is an implementation of Dynamic LDA, to discover topics while keeping track of time-based variability. The model trained on monthly segmented corpus, to adjust topic distributions of the different time period. A robust methodology underlies this work, allowing for deep, exploratory analysis of Bitcoin tweets.

RESULT AND ANALYSIS

This study examined the collaborative landscape of the online world through an analysis of the dynamic and constant Twitter conversations surrounding Bitcoin through the lens of social media analytics. The advanced computational tools, the experimental and analytical segments systematically uncovered the dynamic interplay of Bitcoin-related topics, trends, sentiment analysis, and user engagement. The study began with collection of a large dataset of tweets, and then applied Dynamic Latent Dirichlet Allocation (dLDA) to track how topics evolved over time. Our study measured the occurrence of terms in the social media discourse, along with describing the dynamics and evolution of the social media discourse in the fast-growing Bitcoin ecosystem.

Trends:

A. The Evolution of Key Topics in Bitcoin Discussions

In this section, we explore how topics evolved over varying time frames. A line graph displays the output of this model, which shows how often words like 'price', 'crypto', 'bitcoin', and 'btc' are prevalent. In Figure 2, the peaks on the graph indicate the sensitivity of the issues both with respect to outside influences, which correlate with peaks of market activity as well as peaks of events.

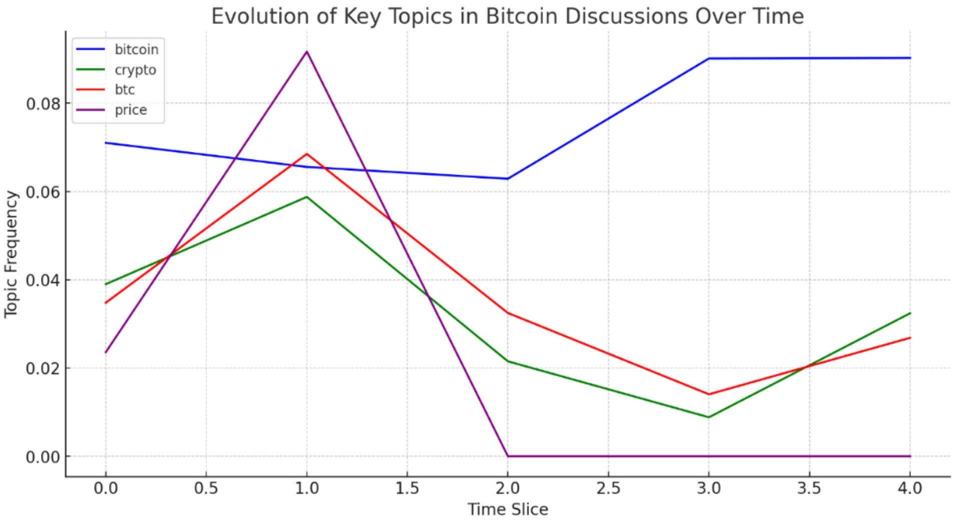


Fig. 2. Key Topics in Bitcoin Discussion Over Time

The illustrated above figure is directly tied to recognizing trends in discussions about Bitcoin. Understanding the shifting importance of these themes is made easier by the representation of each line, which shows the frequency of a particular topic over several time slices.

Sentiment Analysis:

GRAND ACADEMIC PORTAL
RESEARCH JOURNALS
GAP iNTERDISCIPLINARITIES
A Global Journal of Interdisciplinary Studies
(ISSN – 2581-5628)
Impact Factor: SJIF - 5.363, IIFS - 4.875
Globally peer-reviewed and open access journal.

In this study we initialize VADER Sentiment Analyzer to assign sentiment scores to tweets and aggregate them by topic.

B. Sentiment Score per Bitcoin Topic

Sentiment associated with Bitcoin topics is analysed in this study, to determine whether people discussing Bitcoin-related topics are positive, negative, or neutral. Figure 3 demonstrates the mean sentiment score on multiple Bitcoin-related subjects, in which Altcoins & Trends have the highest sentiment score, indicating that people talk positively about Altcoins & Trends. Likewise, Social Media& Promotions and NFTs & Sales have high sentiment scores respectively, suggesting continued interest both for social engagement and digital asset transactions. Meanwhile, Bitcoin Price & Trading and the Other category scored lower in terms of sentiment, which may be attributed to some volatility in the market as well as more general apprehensiveness in conversations. In Figure 3, we see more details on the state of public understanding and interest across aspects of the Bitcoin ecosystem. These results are important, especially for investors, researchers, and analysts who want to gain insights into the market psychology and trends.

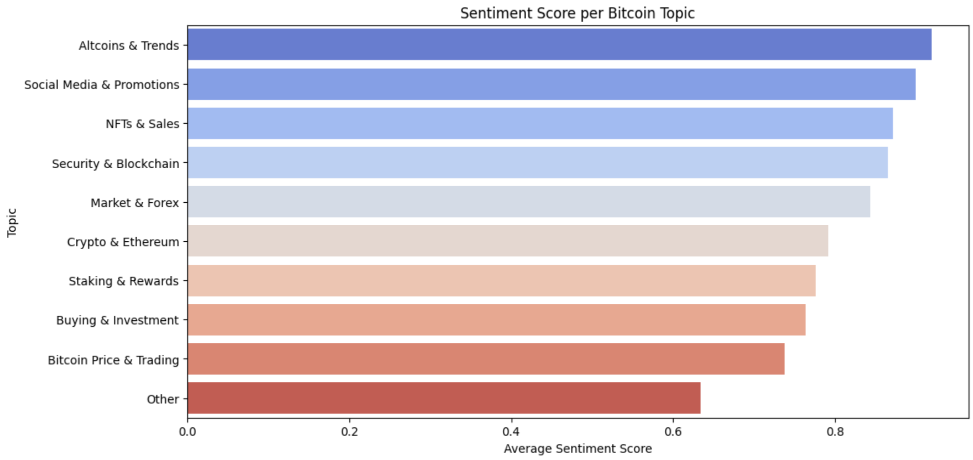
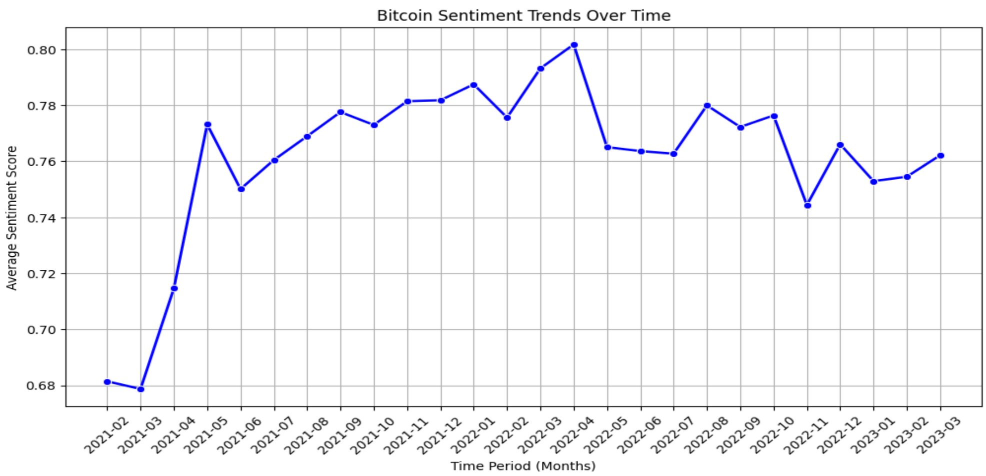


Fig. 3. Sentiment Score per Bitcoin Topic

C. Sentiments of Bitcoin Trends Over Time

This study investigates sentiment (attitude or perception) of Bitcoin-related topics over time. The blue line with markers in Figure 4, shows the fluctuations in sentiment from February 2021 to March 2023. The initial score was relatively low with 0.68 but by April 2021, it witnessed a massive increase to about 0.77. Although it saw a small decline, sentiment rigorously climbed at the beginning of 2021 and 2022, culminating at 0.80 in March 2022. This was however followed by a drop in mid-2022 and the sentiment has remained around the 0.76-0.78 mark toward the end of it. This indicates that the sentiment has largely been positive for Bitcoin, though peaks and troughs largely depend on external factors.

Now that we have Bitcoin sentiment trends over time, this study cross-examines major Bitcoin events to see how sentiment changed in response to those events. To help us assess the trends and gauge how key developments are received by the public. We took historical events between February 2021 and March 2023 that might have an impact on Bitcoin sentiment including Market crashes & rallies, Regulatory decisions, Institutional adoption, Security breaches & hacks, New technology (e.g., Taproot upgrade, ETF approvals).



https://www.gapinterdisciplinarities.org/

GRAND ACADEMIC PORTAL
RESEARCH JOURNALS
GAP iNTERDISCIPLINARITIES
A Global Journal of Interdisciplinary Studies
(ISSN – 2581-5628)
Impact Factor: SJIF - 5.363, IIFS - 4.875
Globally peer-reviewed and open access journal.

Fig. 4. Bitcoin sentiment trends over time

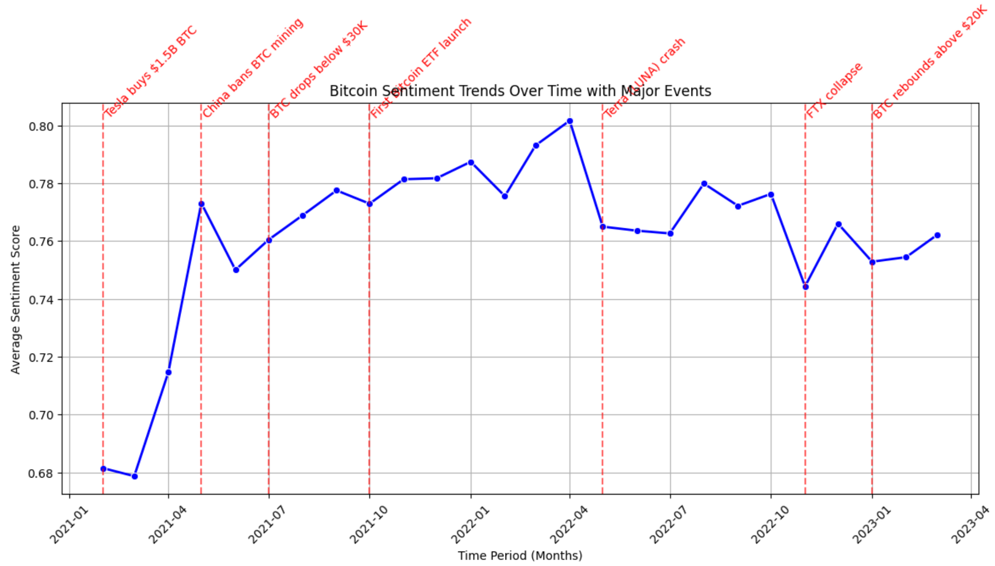


Fig. 5. Bitcoin sentiment trends over time with Major Events

Figure 5, shows the historical data along with important market events providing a better view of their effect on the sentiment. Bitcoin sentiment fluctuations are indicated by the blue line with markers, while important market events are flagged with red dashed lines and annotations. Major events comprise of Tesla's acquisition of $1.5 billion worth of Bitcoin in the first half of 2021, which matches with an extensive growth of sentiment, and China's prohibition on Bitcoin mining follows a decrease in sentiment. Likewise, the launch of the first Bitcoin ETF coincides with a peak, while the Terra-LUNA crash and 2022 FTX collapse correspond with steep sentiment declines. As shown in above graph, external factors have a prominent effect on Bitcoin sentiments. Positive news, such as institutional adoption and regulatory milestones, correlate with higher sentiment scores, while negative news like market crashes and regulatory crackdowns, leads to a steep drop. That also helps understand investor psychology and market dynamics and how sentiment changes in response to positive and negative news in the cryptocurrency sector.

User Engagement Patterns Analysis:

D. The Impact of User Followers on Topic Engagement

This study implemented on small chunk of the dataset to find the relationship between follower counts, user influence and topic engagement.The most striking observation, other than the fact that bitcoin promotion is present in the holdings of all follower categories, is that the intensity is particularly strong among users. Figure 6, shows the distribution of topics between different follower segments, from 0-100 followers to 1M+. The heatmap implies that users within the medium range followers are the most active Bitcoin users, likely a combination of retail investors, traders, or influencers in smaller crypto communities. The lack of notable engagement on terms such as 'Crypto' and 'BTC' suggests that it is general cryptocurrency discussions, which is less popular than discussions surrounding the Bitcoin itself.

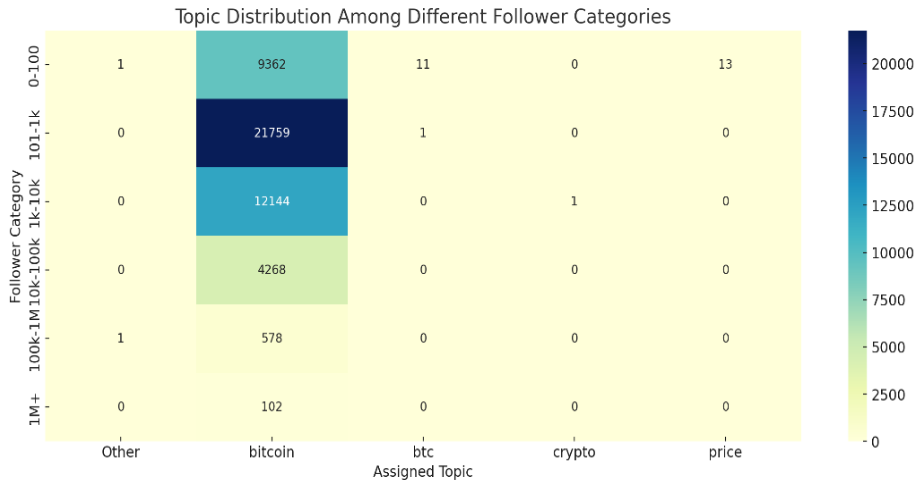


Fig. 6. The Impact of User Followers on Topic Engagement

Discussions Related to Bitcoin:

E. Network Analysis of Hashtag Co-Occurrence

The analysis of the hashtag network used in the tweets reveals insightful information on how the various Bitcoin related topics are connected to each other. It enables the discovery of clusters or groups of hashtags that co-occur together and can be indicative of themes or topics that are relevant in the context of Bitcoin. Hashtag network analysis, Helps identify relationships between topics in the Bitcoin space and how they evolve over time.

So, because full hashtag network is too large and complex, visualizing the whole network would not bring effective conclusions due to the number of nodes and elements that connect them. By working with a subset of the data we were able to develop a more tidy and consumable graph, which explores words that are more central or belonging to a particular community identified previously and limiting to a manageable number of nodes.

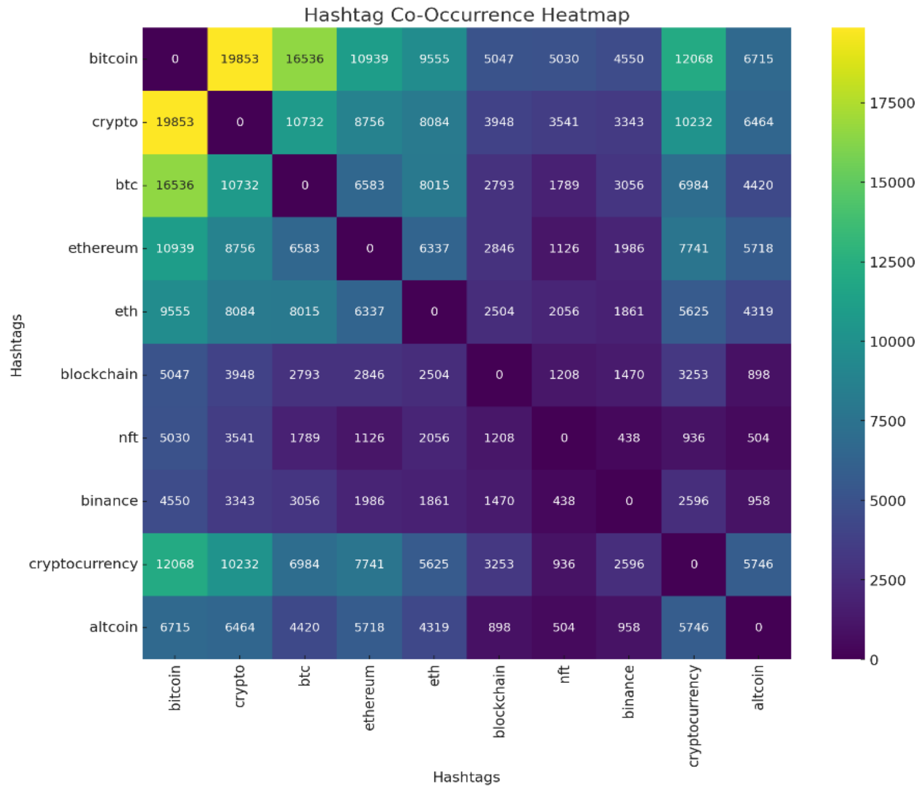


Fig. 7. Hashtags Co-Occurrence

As shown in Figure 7, each cell indicates how often a given pair of hashtags (one on the x-axis and one on the y-axis) co-occurred in the tweets. Darker colours indicate higher frequencies, the diagonal elements in the figure are set to zero as they only indicates the individual occurrence of each hashtag, which is uninteresting in the context of co-occurrence. A high co-occurrence rate between two hashtags, suggests that the themes of those hashtags are more likely to be addressed in conjunction, and therefore that they are closely connected. The above figure illustrates that #bitcoin, #crypto, and #btc are common hashtag mentions in the Bitcoin conversation arena, they are also strongly associated with a range of themes. Indicating that the debates over Bitcoin are a critical part of the cryptocurrency conversation on Twitter, and often bleed into more general crypto topics. The study emphasises the most widely-discussed Bitcoin-related topics and demonstrates the interactions between different topics within the Bitcoin Twitter discourse.

CONCLUSION

This study investigated AI driven analysis of Bitcoin on X, powered by dynamic topic modeling, sentiment analysis, and metrics to track user engagement around this contemporary trend. The findings indicate a high degree of correlation between changes in sentiment and price fluctuations in Bitcoin, showcasing the possible predictive utility of social media analytics in regards to financial markets. In order to gain insight into the changing nature of public conversation surrounding Bitcoin, Dynamic Latent Dirichlet Allocation was used to determine how topics linked to the cryptocurrency evolve over time. Additionally, by showing how particular topics arise and proliferate within the cryptocurrency landscape, hashtag network analysis illuminates the interconnectedness of discussions pertaining to Bitcoin. Additionally, an analysis of user involvement shows that consumers significantly influence the conversation around Bitcoin.

Such integration of computational intelligence with the field of finance not only opens up the scope of research at the intersection of AI, social media analytics, finance, but also extends by addressing in greater detail the implications of such analytics in financial markets. The methodologies presented in this work can be further expanded in future research by the integration of multimodal data sources, such as news articles and financial reports, for improved predictive power and decision-making strategies. This research expands on the understanding of Bitcoin-related social media conversations and provides a basis for using AI to a broader analysis of financial discourse. As digital conversations have a greater impact on the financial markets, our research highlights the growing demand for scalable and reliable analytical frameworks.

REFERENCES

[1] Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. Proceedings of the 23rd International Conference on Machine Learning - ICML '06, 113–120. https://doi.org/10.1145/1143844.1143859

[2] Dikiyanti, T. D., Rukmi, A. M., & Irawan, M. I. (2021). Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm. Journal of Physics: Conference Series, 1821(1), 012054. https://doi.org/10.1088/1742-6596/1821/1/012054

[3] Habibi, M., Priadana, A., & Rifqi Ma'arif, M. (2021). Sentiment Analysis and Topic Modeling of Indonesian Public Conversation about COVID-19 Epidemics on Twitter. IJID (International Journal on Informatics for Development), 10(1), 23–30. https://doi.org/10.14421/ijid.2021.2400

[4] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

[5] Bitcoin Tweets. (n.d.). [Dataset]. https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets/data

[6] Jang, H., Rempel, E., Roth, D., Carenini, G., & Janjua, N. Z. (2021). Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. Journal of Medical Internet Research, 23(2), e25431. https://doi.org/10.2196/25431

[7] Julianto, I. T., Kurniadi, D., Nashrulloh, M. R., & Mulyani, A. (2022). TWITTER SOCIAL MEDIA SENTIMENT ANALYSIS AGAINST BITCOIN CRYPTOCURRENCY TRENDS USING RAPIDMINER. Jurnal Teknik Informatika (Jutif), 3(5), 1183–1187. https://doi.org/10.20884/1.jutif.2022.3.5.289

[8] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, 591–600. https://doi.org/10.1145/1772690.1772751

[9] Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., & Nawaz, N. (2021). Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic. Frontiers in Computer Science, 3, 775368. https://doi.org/10.3389/fcomp.2021.775368

[10] Nguyen, T. H., & Shirai, K. (2015). Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 1354–1364. https://doi.org/10.3115/v1/P15-1131

[11] Phillips, R. C., & Gorse, D. (2018). Mutual-Excitation of Cryptocurrency Market Returns and Social Media Topics. Proceedings of the 4th International Conference on Frontiers of Educational Technologies, 80–86. https://doi.org/10.1145/3233347.3233370

[12] Rossi, L., & Magnani, M. (2021). Conversation Practices and Network Structure in Twitter. Proceedings of the International AAAI Conference on Web and Social Media, 6(1), 563–566. https://doi.org/10.1609/icwsm.v6i1.14313

[13] Suresha, H. P., & Kumar Tiwari, K. (2021). Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data. Asian Journal of Research in Computer Science, 13–29. https://doi.org/10.9734/ajrcos/2021/v12i230278