TWEETING FOR BITCOIN: ANALYZING PRICE DYNAMICS THROUGH TRANSFER ENTROPY

Paulo Ribeiro, Mahmoud Dokmak & Rami Atassi
École Polytechnique Fédérale de Lausanne
Financial Big Data - Final Project
1015 Lausanne, Switzerland

March 11, 2025

ABSTRACT

The interplay between social media sentiment and cryptocurrency price dynamics has become a topic of significant interest due to the rising social influence of platforms like Twitter (today known as X). This study investigates the relationship between Bitcoin price fluctuations and Twitter sentiment using Transfer Entropy ($TE^{A \to B}$), a non-parametric measure capable of capturing nonlinear and time-delayed dependencies between time series A and B. We analyse two datasets: minute-level Bitcoin price data and a large collection of Bitcoin-related tweets. After preprocessing to ensure data quality, we aggregate both datasets into time series with hourly and daily granularity. Sentiment analysis using specific LLM is applied to classify tweets as *BULLISH*, *BEARISH*, or *NEUTRAL*, with confidence-weighted scores aggregated to represent sentiment dynamics. Transfer Entropy is then calculated between the Bitcoin and Twitter time series across varying time delays to detect directional information flow. Our results highlight the presence of information transfer from Twitter sentiment to Bitcoin price movements and vice versa, suggesting a bidirectional relationship. Statistical testing confirms the significance of these findings, and the optimal delay ranges for maximum information flow are identified. This study provides insights into the predictive potential of social media sentiment in cryptocurrency markets and contributes to the broader understanding of complex interactions in financial systems.

***K*eywords** Twitter · Bitcoin · Time Series · Symbolic Transfer Entropy

1 Introduction

The rapid evolution of digital cryptocurrencies and their associated markets has sparked considerable interest among researchers and practitioners in understanding the driving factors behind cryptocurrency price dynamics. Among the most prominent cryptocurrencies, Bitcoin stands out due to its market dominance and wide adoption. The volatile nature of Bitcoin returns can be linked to various factors, including macroeconomic indicators, technological developments, and maybe, sentiment expressed on social media platforms.

Twitter serves as a rich source of public sentiment data. Millions of users actively share opinions, emotions, and information related to Bitcoin and the cryptocurrency market in real-time. This user-generated content has become a valuable proxy for gauging market sentiment, offering a unique opportunity to explore the relation between social sentiment and Bitcoin price movements.

To quantify the directional flow of information between Twitter sentiment and Bitcoin returns, this project employs the concept of Transfer Entropy (TE). TE is a model-free and non-parametric measure designed to detect the transfer of information from one time series to another. Unlike correlation or linear regression methods, TE is capable of capturing non-linear and time-delayed dependencies, making it particularly suitable for analyzing complex financial systems.

The primary aim of this report is to assess whether Twitter sentiment at time t can influence Bitcoin returns at future times $t+$ delay. To achieve this objective, the project systematically investigates the information transfer across varying time delays.

This scientific report is structured as follows: Section 2 describes the data collection and preprocessing methodology, highlighting the challenges and solutions associated with handling social media and financial data. Section 3 introduces the theoretical foundation of the Transfer Entropy framework. Section 4 presents and interprets the results, while Section 5 concludes the report, summarizing the findings and discussing potential implications and avenues for future research.

2 Data

We collected two datasets from Kaggle for our analysis. The first dataset, Binance Full History, provides a comprehensive record of cryptocurrency prices. This dataset is sourced directly from Binance, one of the largest and most prominent cryptocurrency exchanges globally.

The second dataset, Bitcoin Tweets, was created using the Twitter API to gather a large volume of Bitcoin-related tweets. In addition to the tweet content, this dataset includes several metadata making it highly valuable for sentiment and trend analysis.

2.1 Data Analysis

The Bitcoin dataset consists of 2,753,182 records spanning 9 metrics, including open, high, low, and close prices. The data covers a period from August 17^{th} 2017, to November 11^{th} 2022, with minute-level granularity, making it highly suitable for high-frequency financial analysis.

The Twitter dataset comprises 16,889,765 tweets, providing detailed metadata such as likes, replies, retweets, and user information. These tweets were collected over a period from April 19^{th} 2007, to November 23^{th} 2019. A quick exploratory analysis revealed that the 99^{th} quantile for likes, replies, and retweets are relatively low, as shown in Table 1, indicating that the dataset predominantly contains tweets with limited social impact.

Quantile	Likes	Replies	Retweets
99^{th}	43	4	12

Table 1: Summary of social interaction metrics in the Twitter dataset.

This analysis underscores the necessity of cleaning the dataset to remove spam and low-quality content, such as promotional messages or bots advertising unrelated products or services. These steps are crucial to ensure the remaining data is meaningful and relevant to the study.

The effective dataset used for the analysis corresponds to the overlapping time range of the Bitcoin and Twitter datasets. This intersection from August 17^{th} 2017 to November 23^{th} 2019, results in a time span of 828 days, or approximately 2 years and 98 days. As illustrated in Figure 1, this period captures a phase of significant volatility in Bitcoin prices, offering an excellent opportunity to investigate the relationship between Bitcoin's market behavior and Twitter sentiment.

2.2 Preprocessing

In this section, we outline the preprocessing steps undertaken to prepare the data for analysis. First, for the Bitcoin dataset, we reduced the granularity of the price data from minute-level to hourly intervals. This change was made to align with the aggregation of relevant tweets at an hourly interval, ensuring consistency between the two datasets. After aligning the data, we computed the hourly return for Bitcoin.

The preprocessing of the Twitter dataset involved several steps to filter and clean the data. Initially, we removed all tweets deemed non-pertinent, specifically those with fewer than 100 likes, replies, or retweets. This filtering step eliminated spam or scam tweets, such as promotions for Telegram channels, which do not provide meaningful insights and are unlikely to influence Bitcoin price movements. Additionally, we found at this step the presence of foreign language tweets, that we wanted to exclud as our sentiment classifier is primarily trained on English data, and those tweets were representing an insignificant proportion of our dataset (less than 1%). To facilitate accurate language detection by the model, we cleaned the text by removing user mentions, URLs, and hashtag symbols. We then used a language detection model[1] from HuggingFace to identify and retain only English tweets.

For sentiment analysis, we employed a specified – for financial tweets – sentiment classifier model[2] to classify tweets into three categories: *BULLISH*, *NEUTRAL*, and *BEARISH*. The sentiment scores were retained and later used to construct time series data. The entire preprocessing pipeline took approximately 22 minutes to execute using a T4 GPU on Google Colab.

Table 2: Sentiment Analysis Results

Sentiment Category	**Number of Tweets**	**Average Sentiment Score**
BULLISH	26,359 (44.4%)	0.80
NEUTRAL	16,495 (27.8%)	0.77
BEARISH	16,472 (27.8%)	0.83
Total	59,326	–

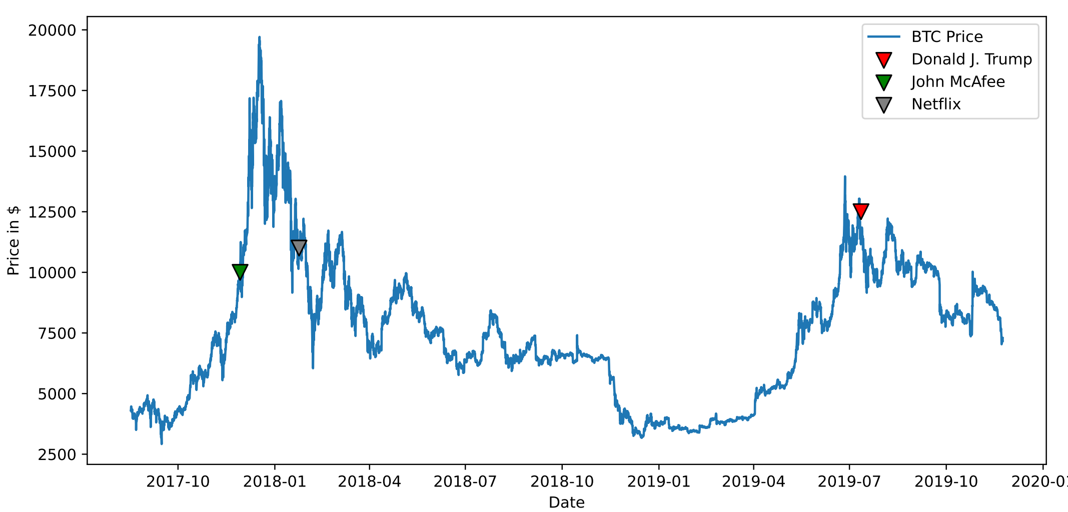
These steps produced a clean dataset that provided preliminary qualitative insights into the relationship between Twitter sentiment and Bitcoin prices. For example, three tweets from influential accounts were selected to illustrate the relationship, categorized as *Bullish*, *Bearish*, and *Neutral*, and marked with green, red, and gray respectively.

> I am not a fan of **Bitcoin** and other Cryptocurrencies, which are not money, and whose value is highly volatile and based on thin air. Unregulated Crypto Assets can facilitate unlawful behavior, including drug trade and other illegal activity.
>
> (*Donald J. Trump, 2019-07-12, 70,298 likes*)

> FRIENDS 2018 PLOTLINES - Joey eats Tide pods, goes to hospital (guest star Ellen Pompeo) - Chandler won't shut up about how funny his Vines were - Ed Sheeran cameo interrupts Phoebe's Central Perk set - Ross gets everyone into **Bitcoin** - The One Where The Gang Realizes Their Privilege.
>
> (*Netflix US, 2018-01-24, 282,959 likes*) –

> When I predicted **Bitcoin** at $500,000 by the end of 2020, it used a model that predicted $5,000 at the end of 2017. BTC has accelerated much faster than my model assumptions. I now predict Bitcoin at $1 million by the end of 2020. I will still eat my d*ck if wrong.
>
> (*John McAfee, 2017-11-29, 12,519 likes*)

Figure 1: BTC closing price over time



[1] papluca/xlm-roberta-base-language-detection
[2] StephanAkkerman/FinTwitBERT-sentiment

Figure 1 shows the selected tweets corresponding to significant moments, illustrating the potential influence of public sentiment on price changes. For example, the *Bearish* sentiment expressed by Donald J. Trump coincides with a period of declining prices. The *Neutral* sentiment shared by Netflix US highlights the limitations of our approach, where marketing tweets, lacking meaningful information, can bring unwanted influence to our analysis. Conversely, the *Bullish* prediction by John McAfee aligns with a sharp increase in Bitcoin's value, emphasizing the impact of optimistic forecasts.

2.3 Time Series Creation

After completing the preprocessing, the next step was to construct time series data for both Bitcoin and Twitter. For Bitcoin, we used the return as the primary metric to capture price fluctuations and classified it as UP, NEUTRAL or DOWN. We then define a return threshold that determines the appropriate class.

$$\begin{aligned} BTC_{return} &> threshold_{BTC} \rightarrow UP(2) \\ |BTC_{return}| &\leq threshold_{BTC} \rightarrow NEUTRAL(1) \\ BTC_{return} &< threshhold_{BTC} \rightarrow DOWN(0) \end{aligned}$$

For Twitter, we aggregated tweets within each hour to form a single sentiment class representing that hour. To achieve this, we performed a weighted average of the sentiment scores for each hour. Specifically, we mapped *BULLISH* to 1, *BEARISH* to -1, and *NEUTRAL* to 0, multiplying the sentiment label and the confidence of our model. The final label was determined based on how close the resulting score was to -1, 0, or 1, corresponding to *BEARISH*, *NEUTRAL*, and *BULLISH*, respectively.

$$Tweets_{score} = \frac{\sum_{tweet \in hour} \text{Sentiment score}_{tweet} \times \text{Sentiment label}_{tweet}}{\text{Number of tweet} \in hour}$$

$$\begin{aligned} Tweets_{score} &> threshold_{Twitter} \rightarrow BULLISH(2) \\ |Tweets_{score}| &\leq threshold_{Twitter} \rightarrow NEUTRAL(1) \\ Tweets_{score} &< threshhold_{Twitter} \rightarrow BEARISH(0) \end{aligned}$$

This method was chosen to ensure that tweets with opposing sentiments (e.g., a highly confident *BULLISH* tweet and an equally confident *BEARISH* tweet) could effectively balance each other, resulting in a *NEUTRAL* sentiment representation. This balance allows for a more accurate reflection of the overall sentiment within each hour.

It is worth noting that score of tweets labeled as *NEUTRAL* do not directly contribute to the weighted average, as their scores are multiplied by 0. This design choice is intentional, as low confidence in being *NEUTRAL* does not provide meaningful information regarding whether the sentiment leans toward *BULLISH* or *BEARISH*. Thus, their exclusion is both logical and appropriate. One potential concern is that weighting sentiment scores by confidence levels might introduce bias toward the sentiment category with higher average confidence. However, as shown in Table 2, the average confidence scores across the three sentiment classes (*BULLISH*, *BEARISH*, and *NEUTRAL*) are nearly identical. Therefore, this approach does not introduce any significant bias.

To ensure accurate classification, we optimized the thresholds using a grid search, selecting the pair of thresholds that maximized the mean transfer entropy between the Bitcoin and Twitter time series. This process resulted in the following thresholds: $threshold_{BTC} = 1.1[\%]$ and $threshold_{Twitter} = 0.1[-]$. These thresholds yield the time series class distribution shown in Table 3.

Labeling	Bullish/Up	Neutral	Bearish/Down
Twitter	5,327 (26.8%)	12,290 (61.8%)	2,265 (11.4%)
Bitcoin	1,436 (7.2%)	17,021 (85.6%)	1,436 (7.2%)

Table 3: Distribution of labels in the constructed time series.

Interestingly, Twitter appears to provide the most significant transfer of information for relatively large Bitcoin price movements, as its threshold only considers Bitcoin's upward or downward phases for 14% of the movements (7% for upward and 7% for downward). This result seems intuitive, as it is unlikely that Twitter would have a substantial impact on very small Bitcoin price fluctuations.

3 Transfer Entropy

3.1 Definition

Transfer Entropy ($TE^{B\to A}$) quantifies the information gain from the past of B in predicting the future of A, while accounting for the information already contained in the past of A. It is computed as the difference between the conditional entropy of A's future, conditioned on its past, and the conditional entropy of A's future, conditioned on both its past and B:

$$TE^{B\to A} = H(A|A_{\text{past}}) - H(A|A_{\text{past}}, B).$$

With the two time series prepared, we begin by calculating the transfer entropy for the entire time series with a history length of 1. Next, we systematically shift the time series and analyze the evolution of TE, allowing us to identify the delay in information transfer and its duration.

3.2 Significance

To test for the statistical significance of a computed transfer entropy, we verify that the score given by two times the length of the time series multiplied by the transfer entropy ($2T \cdot TE_{B\to A}$) diverges from a chi-square distribution, as demonstrated in paper [1]. It states that the transfer entropy between random time series has this score following a chi-square distribution of degree $\nu = n^A \cdot (n^{A^+} - 1) \cdot (n^B - 1)$ with :

- $\boldsymbol{n^A}$ representing the number of possible states that encode the past of time series A (here Bitcoin time series) at any time t.
- $\boldsymbol{n^{A^+}}$ denoting the number of possible states that describe the future of time series A.
- $\boldsymbol{n^B}$ representing the number of possible states that describe time series B (Tweets time series) at time t.

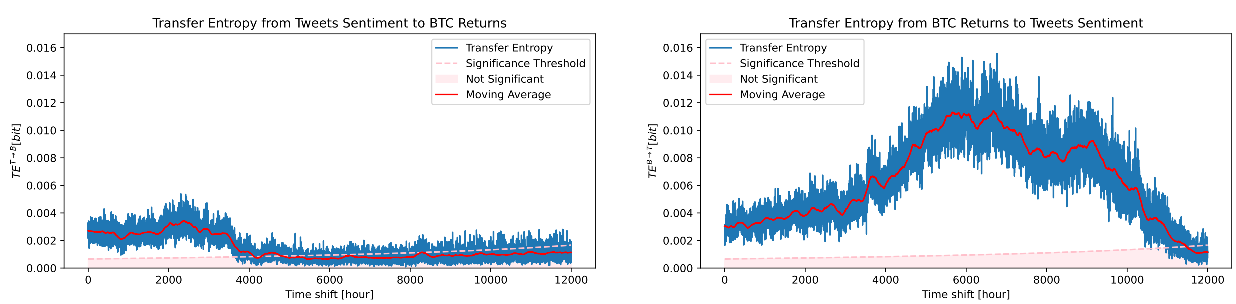
In our case, all n^A, n^{A^+}, and n^B equal 3 because we are considering three distinct states: *BULLISH*, *BEARISH* and *NEUTRAL*. The total degree of freedom for the system is calculated as:

$$\nu = 3 \cdot (3 - 1) \cdot (3 - 1) = 3 \cdot 2 \cdot 2 = 12.$$

Thus, we can perform a chi-square (χ^2) test with p-value = 0.01 to establish the significance of the TE for each time shift. Note that with fixed starting and end date of our two time series and with the time shift progressively increasing during the analysis, the significance threshold will also be increasing since the time series length used for TE is actually decreasing.

4 Results

Figure 2a illustrates the transfer entropy from Tweets to Bitcoin returns, while Figure 2b shows the transfer entropy from Bitcoin prices to Tweet sentiment, calculated with hourly granularity over a delay range of 0 to 12,000 hours (500 days).



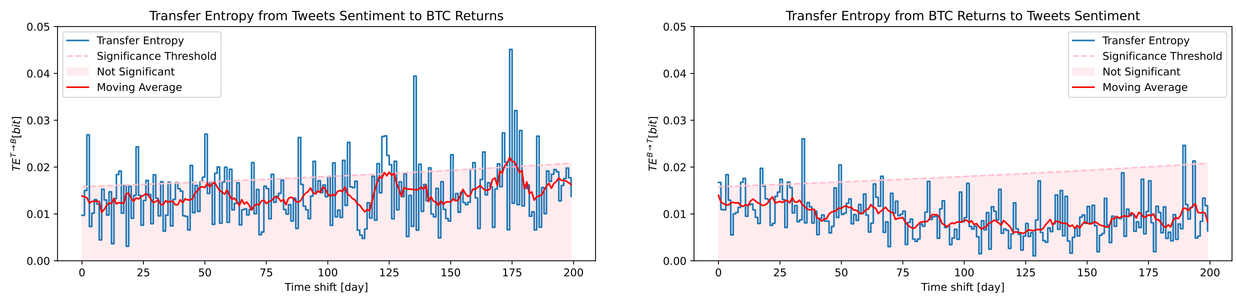
(a) $TE^{T\to B}_{hour}$: from Tweets to Bitcoin

(b) $TE^{B\to T}_{hour}$: from Bitcoin to Tweets

Figure 2: Comparison of Transfer Entropy on an hourly granularity

Figure 2 provides insights into the bidirectional relationship between Twitter sentiment and Bitcoin returns. On the right-hand side, the transfer entropy from Bitcoin returns to Twitter sentiment highlights the substantial impact of price changes on public sentiment. Interestingly, the peak transfer entropy, approximately 1.1%, occurs around a time shift of 6,000 hours (250 days). This result is unexpected, as we initially hypothesized that Bitcoin price changes would influence Twitter sentiment most strongly within a few hours after a major price fluctuation. Instead, the transfer entropy gradually increases over time, reflecting the growing trend surrounding Bitcoin price changes until it reaches its peak, after which it follows a diminishing trend.

On the left-hand side, the transfer entropy from Twitter sentiment to Bitcoin returns demonstrates that the information available on Twitter is significant in determining future Bitcoin price movements. However, this influence is less pronounced than the reverse relationship, with a peak transfer entropy of approximately 0.4% with a delay of approximately 2,500 hours. Moreover, this influence is limited to a maximum time shift of 4,000 hours (approximately 6 months), beyond which the predictability diminishes down to insignificancy. This result is also surprising, as we initially expected the predictability gain to vanish after a few days or weeks rather than persisting for months. However, in another sense, it aligns with Bitcoin's well-known volatility, making it reasonable to observe some tweets influencing its price dynamics over an extended period. Overall, these findings reveal a logical asymmetry in the strength and temporal dynamics of information transfer, with Bitcoin returns having a larger and more prolonged influence on Twitter sentiment than vice versa.



(a) $TE_{day}^{T\to B}$: from Tweets to Bitcoin

(b) $TE_{day}^{B\to T}$: from Bitcoin to Tweets

Figure 3: Comparison of Transfer Entropy on a daily granularity

Figure 3a illustrates the transfer entropy from Tweets to Bitcoin returns, while Figure 3b depicts the transfer entropy from Bitcoin prices to Tweet sentiment. Both are calculated with daily granularity over a delay range of 0 to 200 days.

When analysing transfer entropy with a daily granularity, the results differ significantly from those observed at an hourly granularity. In both directions, Tweets sentiment to Bitcoin returns ($TE_{day}^{T\to B}$) and Bitcoin prices to Tweets sentiment ($TE_{day}^{B\to T}$), the transfer entropy values are notably higher but it fails to rise meaningfully above the significance threshold in both direction. This lack of statistical significance could be due to the reduced amount of data at the daily granularity; while we had approximately 20,000 hours in the hourly analysis, this translates to only 828 data points when aggregated daily, not allowing the transfer entropy to capture any significant information. Another potential explanation is that daily aggregation smooths out finer-grained nuances, eliminating the transfer entropy signals that are evident at hourly intervals.

These observations emphasize the importance of granularity but also of the size of the dataset in transfer entropy analysis, with hourly granularity capturing significant bidirectional information flows that it cannot when the data is aggregated and diminished in daily information.

5 Conclusions

By maintaining an hourly granularity, the results show that Twitter sentiment has a statistically significant impact, at the 1% level, in predicting future Bitcoin prices. Conversely, Bitcoin price changes also significantly influence Twitter sentiment, with a stronger magnitude.

However, several disclaimers should be considered when interpreting these findings. First, only tweets with interactions were analyzed, focusing on the sentiment of influential individuals rather than the overall Twitter sentiment. Second, the analysis covers the period between 2017 and 2019, during Bitcoin's early years. This may explain the observed long delays and durations of influence, as Bitcoin was less well-known and trusted at the time. Analyzing more recent data could lead to different insights.

This study contributes to understanding the interaction between public sentiment and financial markets, offering a foundation for future research and applications in algorithmic trading and market analysis. Future work could expand on this analysis and validate the findings with machine learning models capable of predicting Bitcoin returns.

References

[1] Christian Bongiorno and Damien Challet. Statistical inference of lead-lag at various timescales between asynchronous time series from p-values of transfer entropy, 2022.

[2] J.V. Critien, A. Gatt, and J. Ellul. Bitcoin price change and trend prediction through twitter sentiment and data volume. *Financial Innovation*, 8:45, 2022.

[3] Kelechi M. Ikegwu, Jacob Trauger, Jeff McMullin, and Robert J. Brunner. Pyif: A fast and light weight implementation to estimate bivariate transfer entropy for big data. In *2020 SoutheastCon*, pages 1–6, 2020.

[4] Luca Papariello. xlm-roberta-base-language-detection (revision 9865598), 2024.

[5] David Neto. Examining interconnectedness between media attention and cryptocurrency markets: A transfer entropy story. *Economics Letters*, 214:110460, 2022.

[6] Yanzhao Zou and Dorien Herremans. Prebit — a multimodal model with twitter finbert embeddings for extreme price movement prediction of bitcoin. *Expert Systems with Applications*, 233:120838, 2023.