

Emoji Reactions on Telegram: Unreliable Indicators of Emotional Resonance

Serena Tardelli¹, Lorenzo Alvisi^{1,2}, Lorenzo Cima^{1,3}, Stefano Cresci¹, Maurizio Tesconi¹

¹Institute of Informatics and Telematics, National Research Council, Italy

²IMT School for Advanced Studies Lucca, Italy

³Dept. of Information Engineering, University of Pisa, Italy

{firstname.lastname}@iit.cnr.it

Abstract

Emoji reactions are a frequently used feature of messaging platforms, yet their communicative role remains understudied. Prior work on emojis has focused predominantly on in-text usage, showing that emojis embedded in messages tend to amplify and mirror the author’s affective tone. This evidence has often been extended to emoji reactions, treating them as indicators of emotional resonance or user sentiment. However, they may reflect broader social dynamics. Here, we investigate the communicative function of emoji reactions on Telegram. We analyze over 650k crypto-related messages that received at least one reaction, annotating each with sentiment, emotion, persuasion strategy, and speech act labels, and inferring the sentiment and emotion of emoji reactions using both lexicons and LLMs. We uncover a systematic mismatch between message and reaction sentiment, with positive reactions dominating even for neutral or negative content. This pattern persists across rhetorical strategies and emotional tones, indicating that emojis used as reactions do not reliably function as indicators of emotional mirroring or resonance of the content, in contrast to findings reported for in-text emojis. Finally, we identify the features that most predict emoji engagement. Overall, our findings caution against treating emoji reactions as sentiment labels, highlighting the need for more nuanced approaches in sentiment and engagement analysis.

1 Introduction

Emoji reactions are a central feature of digital platforms, enabling users to respond quickly and visually to content. While much research has focused on in-text emojis (i.e., linguistic elements embedded within authored messages), emoji reactions (i.e., reactions appended by readers as lightweight feedback mechanisms), remain comparatively underexplored. Yet they are equally revealing, as they

are instantaneous, costless, and a quick way to interact with content without composing a reply. Moreover, while prior work has examined in-text emojis, showing they amplify sentiment (Shiha and Ayvaz, 2017), improve sentiment classification (Liu et al., 2021), and provide emotional nuance (Lou et al., 2024; Khan et al., 2025), this understanding has also been implicitly extended to emoji reactions, which have been used as proxies for emotional labels in distant supervision settings (Pool and Nissim, 2016; Plaza Del Arco et al., 2018; Raad et al., 2018). However, recent work suggests that emoji reactions may instead express pragmatic or social meanings, such as to approve or normalize toxic discourse (Morales et al., 2025).

Building on this idea, we provide the first large-scale study examining whether emoji reactions actually reflect emotional resonance, an assumption often implicitly adopted when reactions are used as proxy labels in sentiment analysis. To do so, we investigate the communicative function of emoji reactions on Telegram, a messaging platform widely used for news, politics, and activism. On Telegram, emoji reactions are publicly visible and frequently used, making them a rich signal of collective response. In detail, we analyze over 650k crypto-related Telegram messages and their associated emoji reactions. We combine large-scale emotion classification with rhetorical analysis to examine the alignment between message tone and reaction type. In particular, we address three research questions:

- **RQ1:** Do emoji reactions align with the sentiment expressed in the original message?
- **RQ2:** Do emoji reactions reflect the emotional tone of the original message, creating emotional resonance?
- **RQ3:** Which message features predict emoji reactions?

We adopt a mixed-methods approach to examine messages and their reactions across multiple dimensions, including sentiment, emotional tone, persuasion strategy, and speech act. We find that positive emoji reactions dominate, even in response to messages that are emotionally neutral or negative. This consistent mismatch holds across rhetorical strategies and emotional categories, suggesting a pragmatic function for some emoji reactions, more closely tied to social approval or support than to shared emotional resonance. Finally, we show that certain rhetorical strategies are strong predictors of emoji engagement, and that emotionally neutral or negative messages tend to attract more emoji reactions overall. Our contributions are manifold:

- We provide the first large-scale analysis of emoji reactions on Telegram to determine whether they distort or amplify emotional alignment.
- We conduct our analysis on a large dataset of 650k crypto-related Telegram channels, a domain characterized by strong community identity and high engagement, where in-text emojis are known to play a meaningful role in shaping financial sentiment (Mahrous et al., 2023; Zuo et al., 2024), while emoji reactions remain largely unexplored.
- We show a consistent misalignment between message sentiment and emoji reactions, indicating that, unlike in-text emojis, reactions do not reliably act as emotional mirroring.
- We identify specific message features, such as emotional framing or persuasive tactics, that predict user engagement in terms of emoji reactions.

Our findings suggest that emoji reactions are not reliable proxies for sentiment content, as they weakly mirror or resonate with the emotion evoked by the message. Together, these results challenge the assumption that emoji reactions reliably reflect emotional response, as commonly observed for in-text emojis. These patterns should also be interpreted in light of Telegram’s specific communicative context, where community norms may shape reaction behavior. More broadly, these findings have important implications for research on misinformation, particularly in political and financial contexts, where reaction-based signals can be in-

terpreted as indicators of emotional resonance or public sentiment.

2 Related work

Emojis have been extensively studied as markers of affective communication in digital discourse. Early work recognized their role as nonverbal cues that supplement text with emotional tone (Miller et al., 2016; Kralj Novak et al., 2015), and many researchers treated them as reliable proxies for emotion. Subsequent studies showed that emojis can amplify sentiment in social media messages (Shiha and Ayvaz, 2017). Other research found that incorporating emojis as features can significantly improve sentiment classification accuracy (Liu et al., 2021). More recent work highlighted how emojis provide emotional context in posts highlighting the role of emojis as emotional indicators (Lou et al., 2024; Zhou et al., 2025). Similar work showed that emojis play a vital role in sentiment expression, often conveying more explicit and nuanced information than the text alone (Khan et al., 2025). As a result, datasets such as the one introduced by Kralj Novak et al. (2015), which assign sentiment scores to individual emojis based on crowd-sourced annotations, have been widely used for downstream sentiment classification. Similarly, the DeepMoji model (Felbo et al., 2017) was trained on large-scale Twitter data labeled with emojis in order to learn emotional representations, relying on the assumption that emoji usage reflects the author’s emotional intent.

This assumption has been extended to emoji reactions, i.e., predefined emoji responses users can select to express a reaction to content on platforms such as Facebook. These reactions have often been interpreted as ground truth labels for emotion classification tasks (Pool and Nissim, 2016; Plaza Del Arco et al., 2018; Raad et al., 2018), enabling models to learn emotion categories from reaction distributions. However, more recent studies caution against this approach. Graziani et al. (2019) observed that reactions can be semantically ambiguous or perform multiple pragmatic functions. Paolillo (2023) noted that reactions can operate not only as affective signals but as interactional cues, marking approval, affiliation, or irony depending on context.

Our work extends these critiques by examining emoji reactions on Telegram, a platform where reactions are public, optional, and highly community-

driven. Unlike prior research that treats emoji reactions as emotional response, we find systematic mismatch between the sentiment expressed in a message and the sentiment inferred from emoji reactions. Messages with negative or neutral tone frequently receive positive emoji reactions, such as the thumb up or love emoji. This suggests that reactions function more as social endorsement than as expressions of shared affect. The result challenges the methodological practice of using reaction emojis as emotional supervision signals, at least in environments where group identity and approval play a stronger role than emotion mirroring.

On a similar line, recent work (Morales et al., 2025) explored how emoji reactions are used in toxic versus non-toxic contexts on Telegram, showing that even toxic comments often receive positive emoji reactions, signaling approval or normalization of harmful content. Our findings complement and generalize this pattern: while they focused specifically on toxicity, we observe that positive reaction bias holds across a broader range of communicative intentions, including persuasive tactics, emotional tone, and speech acts. Moreover, while our results do not exclude the possibility of emotional mirroring, they suggest that social signaling frequently overrides it in practice.

Further, prior work on emotional dynamics in engagement showed that emotionally charged content tends to propagate faster (Vosoughi et al., 2018), and that moral-emotional language increases virality within ideologically homogeneous groups (Brady et al., 2017). Our findings complement this literature by showing that the type of emotional response conveyed through reactions is not always aligned with message tone, and instead may reinforce in-group signaling mechanisms. This distinction between emotion expression and interactional feedback is especially relevant for studying engagement in polarized or activist spaces, where affective reactions might serve to signal allegiance more than to reflect emotional resonance.

By focusing on Telegram, which remains comparatively understudied despite its growing role in political and financial discourse and in information diffusion, we contribute new evidence on communication dynamics within this platform. In fact, while recent studies have examined toxicity (Alvisi et al., 2025b,a) and misinformation (Urman and Katz, 2022; Alvisi et al., 2024) on Telegram, as well as their association with emoji usage (Morales et al., 2025), the communicative function of emoji

reactions remains largely unexplored, a gap that we address in the present study.

3 Dataset

We construct our dataset by collecting Telegram channels listed on TGStat¹, a third-party analytics platform that indexes and categorizes public channels by topic, language, and popularity. From this source, we select channels related to the cryptocurrency topic, as it is an area where emojis are known to play a meaningful role in shaping financial sentiment (Mahrous et al., 2023; Zuo et al., 2024). We collect those English-language channels with at least 5,000 subscribers, resulting in an initial sample of 20,695 channels. For each channel, we retrieve all publicly available messages posted between January and December 2024. The distribution of messages follows a long-tail pattern consistent with a power-law distribution, where a small number of channels account for the majority of messages.

Since our analysis focuses on messages that received at least one emoji reaction, we further filter the dataset accordingly. In particular, we retain only messages that include at least one standard (default) emoji reaction, excluding custom emojis, which may be enabled depending on channel settings. After this filtering, the final dataset consists of 647,879 messages from 993 channels.

4 Methods

We analyze Telegram messages using a multi-stage pipeline that combines NLP-based linguistic annotation with behavioral cues derived from user reaction data (i.e., emoji-based reactions). First, we automatically annotate each message with linguistic features, including sentiment, persuasion strategies, emotion, and speech act categories. Then, we automatically label the emoji reactions to infer both their sentiment polarity and emotional content. This allows capturing both linguistic content and social responses.

Message Annotation We enrich each Telegram message with linguistic and psychological features through automatic labeling using multiple classification layers. For each annotation task, although multiple labels may co-occur within a single message, we adopt a single-label (multiclass) formu-

¹<https://tgstat.com/>

lation and retain only the highest-scoring label returned by each classifier.

- *Persuasion strategies*: We detect rhetorical appeals of each message (e.g., emotion, authority, scarcity, social proof, fear, popularity, logic, reciprocity, fear of missing out) plus a neutral and a no persuasion categories using the *valhalla/distilbart-mnli-12-1* zero-shot model (Valhalla, 2020) and the taxonomy by Qachfar and Verma (2023).
- *Emotion analysis*: We assign each message an emotion from (Plutchik, 1994)’s taxonomy using the *bhadresh-savani/bert-base-uncased-emotion* model (Savani, 2021).
- *Sentiment analysis*: We label messages as positive, neutral, or negative using *cardiffnlp/twitter-roberta-base-sentiment* (CardiffNLP, 2021), widely adopted in computational social science (Jahin et al., 2024).
- *Speech acts*: We classify messages as assertives, directives, commissives, expressives, and declarations based on Searle (1969) taxonomy (Jegade, 2024; Saha et al., 2020) using the *valhalla/distilbart-mnli-12-1* zero-shot model (Valhalla, 2020).

Emoji Reaction Interpretation We infer the emotional and affective content of reactions using two complementary methods:

- *Emoji sentiment score*: We use the emoji sentiment lexicon (Kralj Novak et al., 2015)², which assigns each emoji a polarity (−1, 0, +1). For each message, we compute a weighted average score (−1 to +1) based on reaction frequencies, then map it to positive (≥ 0.2), negative (≤ -0.2), or neutral, as done in prior work (Chifu et al., 2015; Chaithra et al., 2019) showing that defining a neutral range around zero reduces label noise and improves the robustness of downstream sentiment analyses. This mapping allows us to assign a discrete sentiment label to the emoji reaction set for each message.
- *LLM-based emoji emotion*: We prompt OpenAI GPT-4o model with the list of reaction

²<https://www.clarin.si/repository/xmlui/handle/11356/1048>

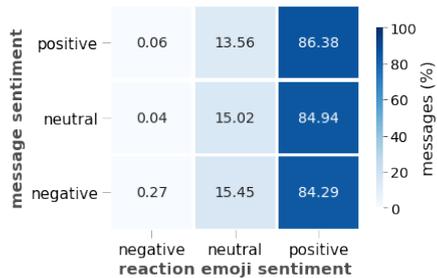


Figure 1: Message sentiment vs emoji reactions sentiment. Positive emoji reactions dominate independently of message sentiment.

emojis to assign an emotion from Plutchik (1994)’s taxonomy (coarse analysis) or a fine-grained emotional label, allowing for more nuanced and free interpretation beyond the core categories. In this way, we can also assign each message the emotion of its most frequent reaction emoji.

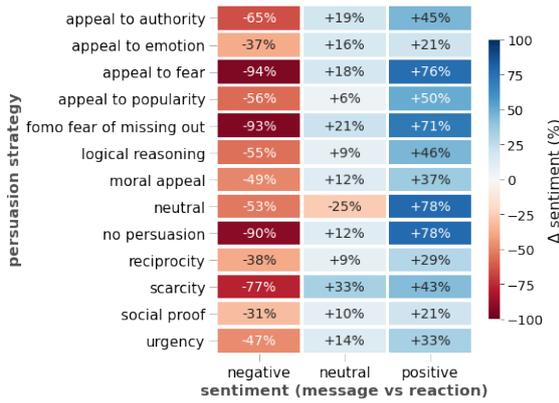
Predictive Modeling and Feature Interpretation

We investigate which aspects of a message predict its likelihood of receiving a large number of emoji reactions. We frame the task as a binary classification problem. We train a binary Random Forest classifier (Breiman, 2001) with 28 features to distinguish between “high engagement” posts and “low engagement” ones, using the median number of emoji reactions as threshold. The 28 features correspond to the automatically inferred linguistic attributes described above (i.e., sentiment, emotion, persuasion strategy, and speech act), encoded as categorical features. While the classification task is not the core goal, this predictive setup allows us to quantify the impact of individual message-level features. To understand the contribution of each feature to engagement, we leverage SHAP, a game-theoretic method that has become standard in explainable AI (Lundberg and Lee, 2017). SHAP assigns each feature a local importance score for every prediction, allowing us to rank linguistic and emotional cues according to their predictive power, as done in multiple related social media analysis tasks (Gambini et al., 2024; Tessa et al., 2025).

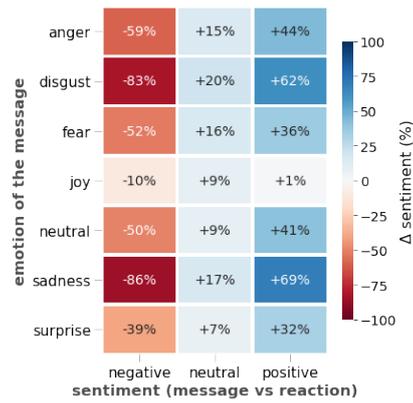
5 Results

5.1 RQ1 - Alignment of emoji reactions with sentiment

In Figure 1 we compare the sentiment conveyed in the message with the sentiment of the emoji reactions it receives. The majority of messages, regard-



(a) Message sentiment vs emoji reactions sentiment, for different persuasive strategies. Most strategies show a consistent shift toward more positive sentiment in reactions. Sentiment distributions in messages and reactions differ significantly across all persuasion strategies (Chi-squared test, $p < 0.001$), revealing a systematic misalignment.



(b) Message sentiment vs emoji reactions sentiment, for different emotions. Negative emotions (e.g., sadness, disgust) receive positive reactions. Sentiment distributions in messages and reactions differ significantly across all emotions (Chi-squared test, $p < 0.001$).

Figure 2: Sentiment shift from messages to reactions by persuasion strategy (left) and emotion category (right).

less of whether they express positive, neutral, or negative sentiment, receive predominantly positive reactions. For instance, over 84% of negative and neutral messages are met with positive emoji sentiment, with less than 0.3% of negative messages receiving negative emoji feedback. This suggests a systematic mismatch between the emotional tone of the message and the sentiment expressed in the reactions, an inconsistency in line with previous work on other platforms (Wang et al., 2023).

To further explore this phenomenon, in Figure 2 we break down the analysis by persuasion strategy and emotion detected in the message. In both cases, we plot the delta in sentiment, that is, the sentiment shift computed as the difference between the sentiment score of the original message and the weighted average sentiment score of the emoji reactions it received. These differential heatmaps reinforce the pattern we previously observed. While the original sentiment expressed in messages spans a wide range, the received sentiment, as inferred from emojis, is skewed toward positive or neutral categories. This suggests a general positivity bias in emoji usage, potentially reflecting users’ preference for agreement, support, or acknowledgment over dissent.

In Figure 2a, this effect holds across most persuasive strategies. Techniques like appeal to fear, fear of missing out, and scarcity, which typically rely on negatively framed content, still elicit predominantly positive emoji reactions, with net sentiment shifts exceeding +70%. However, the magnitude

of this shift varies across strategies. For instance, techniques such as reciprocity, fear of missing out, or appeal to emotion show more balanced distributions, with a smaller difference between message tone and reaction sentiment. This suggests that not all forms of persuasion equally trigger endorsement, and that emotional proximity between message and reaction may play a role in moderating the shift. A similar pattern emerges in Figure 2b, where messages expressing sadness, disgust, or fear are still met with disproportionately positive reactions. However, the shift is not uniform, as messages labeled with joy or surprise show only modest changes, indicating that positive messages are not necessarily amplified through emoji reactions. Interestingly, this tendency is especially evident in response to emotionally neutral content, where reactions may reflect a desire to affirm or engage with others, convey presence, support, or as a form of interpersonal acknowledgment.

5.2 RQ2 - Emoji reactions and emotional resonance

While RQ1 revealed a tendency for messages to receive positive emoji reactions regardless of their sentiment, the analysis was based on a coarse-grained polarity framework (positive, neutral, negative). To better understand which emotions drive this misalignment, RQ2 shifts focus from sentiment to discrete emotion categories. We ask whether certain emoji reactions systematically distort alignment with the emotional tone of the message, and

whether removing them improves interpretability. Specifically, we hypothesize that emoji reactions are not always used to mirror the emotional tone of a message, but instead often serve as signals of social approval, engagement, or banter. To evaluate this, we measure how often the emotion of the message appears among the top-ranked emoji reactions. In other words, we compute the percentage of messages where the top-ranked emoji reaction shares the same emotion category as the message itself. To assess the impact of individual emoji reactions, we perform an ablation analysis. We remove one emotion at a time from the reaction distribution and we re-calculate the top-ranked emoji reaction. If removing a specific emotion significantly increases the number of matches (i.e., if it makes the dominant emoji reaction more likely to align with the emotion of the message), we interpret that category as disruptive to emotional alignment (i.e., the removed reaction was often selected even when it did not match the emotional tone of the message). Figure 3 presents the relative change in alignment match after removing each emoji reaction category, grouped by the emotion label of the message. Each cell shows the percentage change in the proportion of messages where the emotion of the top-ranked emoji matches the emotion of the message. Statistically significant changes, defined as those with a standardized effect size (Cohen’s h) ≥ 0.2 , are marked with an asterisk. Cohen’s h quantifies the magnitude of difference between proportions before and after ablation, providing an interpretable measure of effect strength beyond percentages.

As expected, the lowest scores appear along the diagonal, where the removed emoji category matches the labeled emotion of the message. In fact, removing the correct category naturally lowers alignment, simply because it eliminates true matches. In these cases, the drop does not indicate interference from other emotions, but simply reflects the expected loss in alignment from removing the correct category. The main takeaway from our results is that removing joy consistently leads to large and statistically significant improvements in alignment, especially for messages labeled as sadness, neutral, surprise, and fear. In contrast, removing other reaction emotions has minimal impact. For instance, while removing surprise reactions results in a large relative increase in matches for disgust messages (i.e., +125%), the number of messages labeled with disgust is very small, making this effect statistically negligible.

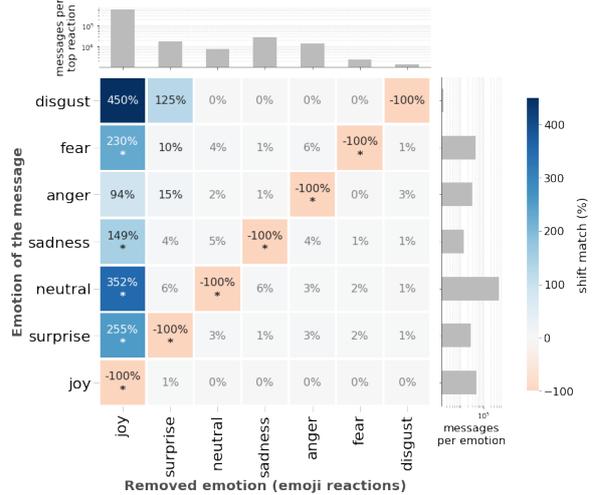


Figure 3: Percentage change in alignment between message emotions and top emoji reactions after removing each emotion. Removing joy leads to large and statistically significant improvements.

The consistent improvement in alignment when joy reactions are removed suggests that these emojis are often used even when they do not match the emotional tone of the message. Unlike other categories, joy appears to systematically distort emotional alignment, pointing to a distinct functional role. Rather than reflecting shared affect, joy reactions may serve pragmatic or social purposes, such as signaling encouragement, approval, banter, or support, regardless of the emotional content of the message. The disruptive effect of joy, revealed through targeted ablation, highlights its unique communicative function in digital contexts, one that is orthogonal to emotional resonance.

Our previous analysis relied a binary metric: whether the top-ranked emoji reaction matched the emotion expressed in the message. While this provides an intuitive signal of alignment, it reduces the full distribution of emoji reactions to a binary outcome. To capture a more nuanced and robust measure of alignment, we compute the cosine similarity between the distribution of emoji reaction emotions and the message emotion represented as a one-hot vector. This allows us to detect subtle changes in alignment, not just categorical shifts. Unlike the previous binary approach, cosine similarity reflects the degree to which the overall emotional tone of emoji reactions matches the intended emotion of the message, making it more sensitive and interpretable. This is particularly important when testing whether the presence of specific emoji emotions obscures rather than reflects affective res-

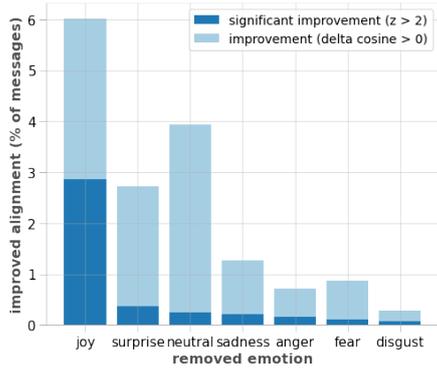


Figure 4: Percentage of messages with improved alignment after removing each emoji reaction emotion. Joy most frequently obscures alignment between message and reaction tone.

onance. To identify which emotions most affect this alignment, we systematically removed each emotion category from the emoji distribution and measured the resulting change in cosine similarity (Δ). A negative effect indicates that removing the emotion improves alignment, suggesting that its presence may distort the emotional signal conveyed by emoji reactions.

Among all categories, joy produced the most pronounced negative effect ($\Delta = -0.107$), followed by surprise, with a much smaller change ($\Delta = -0.006$). Other emotions, such as disgust ($\Delta = -0.0007$), had negligible impact. This suggests that joy-related emojis, more than any other category, systematically misalign with the emotional tone of the message. We then examined, for each emotion removed from the emoji reaction distribution, how often this ablation led to improved alignment with the emotion of the message. Figure 4 shows the results. While all emotions showed low rates of improvement in cosine similarity (i.e., $< 4\%$, with most $< 1\%$), removing joy led to a substantial increase in alignment in over 6% of the messages. Moreover, joy accounted for over 17,000 statistically significant improvements (z-score > 2), surpassing any other emotion. This pattern is not explained simply by frequency. While joy appears in over 95% of emoji reactions, its removal leads to disproportionately many and significantly large alignment improvements, suggesting that joy plays a qualitatively different role. In fact, while sadness appears in over 26% of emoji reactions, removing it leads to improved alignment in just 1.27% of messages, compared to 6.02% for joy. No other emotion, regardless of frequency,

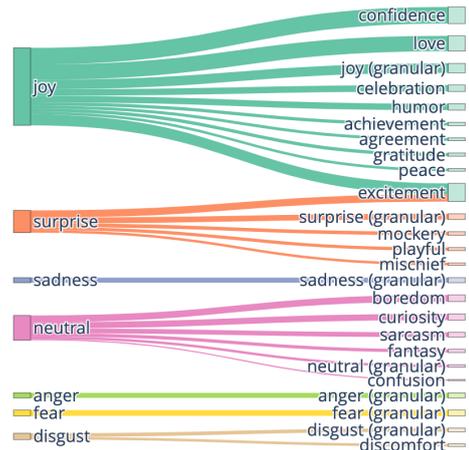


Figure 5: Mapping of coarse-grained emotions (left) to their finer-grained subcategories (right). Even when emojis reflect joy or neutrality, they frequently encode more specific sentiments of approval like agreement, gratitude, and excitement.

produces such a consistently large and disproportionate improvement in alignment when removed. These findings reinforce the idea that joy reactions are not merely emotional mirroring, but often mask or override the underlying sentiment of a message.

To shed light on why joy behaves differently from other emotions, we examine the fine-grained emotional subtypes associated with each coarse emoji emotion category. Figure 5 shows a Sankey diagram mapping each coarse emoji emotion to its fine-grained variants. Notably, while categories like anger or disgust map to narrowly defined emotional expressions, joy includes a wide range of sub-emotions, such as confidence, love, celebration, agreement, and gratitude. These variants often signal social approval, support, or encouragement, rather than shared affect. This broader functional range helps explain why joy reactions are frequently misaligned with the emotional content of messages, suggesting that they do not function as an emotional response but rather as a vehicle for social feedback. The mismatch becomes evident in real examples from the dataset, shown in Figure 6, where joy reactions are frequently used in response to emotionally negative messages such as those expressing sadness, fear, or anger. For example, reacting with a heart emoji to a sad post might mean “I’m here for you” more than “I feel love,” reflecting group dynamics and social norms.

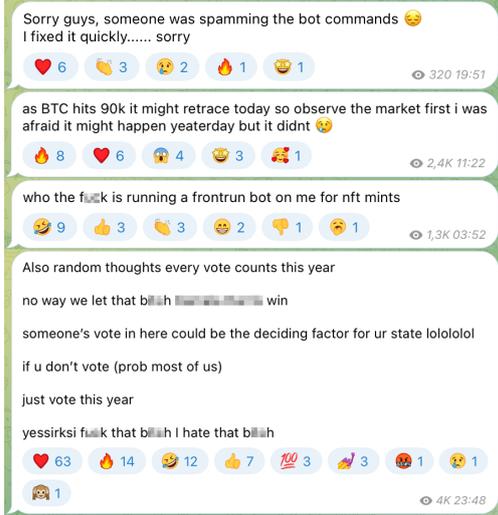


Figure 6: Examples from our Telegram dataset showing that messages expressing negative or neutral emotions receive predominantly positive reactions.

5.3 RQ3 - Prediction of emoji reactions

This research question investigates which message-level features predict the presence of emoji reactions, with the goal of understanding how rhetorical and linguistic strategies influence the engagement of the users through emoji reactions.

To this end, we formulate the task as a binary classification problem, where the model predicts whether a message elicits a high level of emoji reactions based on its textual, emotional, and rhetorical features. We employ a Random Forest classifier (Breiman, 2001) with class balancing, achieving a moderate yet consistent performance, with accuracy = 0.71 and F1 = 0.71. This provides a solid base for further interpreting results through SHAP, which estimates the marginal impact of each input feature on the predictions of the model.

Figure 7 shows the top 15 contributing features (red-colored points represent high, while blue low) and their influence on the prediction toward high engagement. The top predictors of emoji reactions include Emotion Neutral, Sentiment Negative, and (to a lesser extent) Sentiment Positive, suggesting that the emotional valence of a message, whether positive or negative, plays a key role in shaping reaction dynamics. Interestingly, Emotion Neutral emerges as the strongest individual predictor, indicating that even emotionally flat messages can elicit strong social responses. Notably, stronger negative sentiment correlates with higher reaction rates, while less intensely positive sentiment is also linked to higher reactions. This pattern suggests

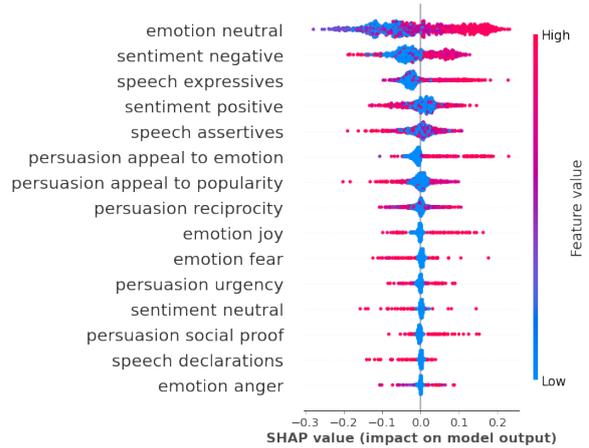


Figure 7: Impact of linguistic features on predicting high emoji reaction engagement.

that both negative emotional content and muted positivity may drive greater engagement, possibly because they elicit stronger social responses or invite interpretation. These findings align with prior research on negativity bias in online platforms, which shows that negatively valenced messages tend to attract more interaction on social platforms (Schöne et al., 2023; Watson et al., 2024). Our novel results indicate that this dynamic extends to emoji reactions as well.

Linguistic categories also contribute. For example, Speech Expressives and Speech Assertives are among the most impactful predictors, suggesting that stylistic elements of the message may influence emotional engagement beyond what is said. Several persuasion strategies such as Appeal to Emotion, Appeal to Popularity, and Reciprocity also emerge as relevant predictors. In contrast, fine-grained emotion labels such as joy, anger, or fear showed comparatively modest influence.

6 Discussion and Conclusions

Our analysis reveals that emoji reactions on Telegram do not reliably act as emotional mirroring and may instead signal social support. While previous studies have largely examined emojis as elements embedded within text, demonstrating that they can intensify expressed sentiment (Shiha and Ayvaz, 2017), improve the performance of sentiment analysis models (Liu et al., 2021), and convey additional affective information (Lou et al., 2024; Khan et al., 2025), this perspective has subsequently been extended, often implicitly, to emojis used as reactions as well, treating them as proxies for sentiment (Pool and Nissim, 2016; Plaza Del Arco et al.,

2018; Raad et al., 2018).

Our findings challenge this assumption. In particular, through ablation analysis, we show that joy is the most disruptive emotion in terms of alignment. Unlike other categories, joy reactions frequently appear in response to emotionally negative or neutral messages. This suggests that joy serves a pragmatic role, signaling approval, support, or social bonding. These patterns indicate that emoji reactions on Telegram may operate more as social cues following human biases, than as emotional reflections. This urges caution against using emoji reactions as ground truth for sentiment analysis and raises concerns about AI models that rely on such signals to estimate public opinion. Our results also help explain why negative or controversial content online often receives higher engagement, as reactions may reflect participation or support for the person posting, complementing work by Morales et al. (2025), where authors highlighted positive emoji reactions in the context of toxic discourse as approval signal. Moreover, we show that this bias toward positive reactions holds across a wider range of communicative intents and emotional tones. Finally, we show that stronger negative sentiment and muted positivity are linked to higher emoji reaction rates, extending the negativity bias observed in online platforms (Schöne et al., 2023; Watson et al., 2024) to emoji-based interactions. Overall, since our analysis focuses on crypto-related messages, a domain where emoji reactions are highly visible and central to community interaction and where in-text emojis are known to shape financial sentiment (Mahrous et al., 2023; Zuo et al., 2024), our findings shed light on a largely unexplored aspect of communication, namely the use of emoji reactions, while also highlighting the need to assess whether these patterns generalize beyond this context.

In conclusion, we provide insights into the social meaning of emoji reactions on Telegram, suggesting that they serve a distinct communicative purpose compared to in-text emojis. Recognizing this distinction is especially important for research on online engagement and misinformation, not only in financial discourse but also in political (Tretakov et al., 2025) and infodemic (Calamusa et al., 2020; Ng and Loke, 2020) contexts, as well as in moderation settings (Zhou et al., 2025; Cima et al., 2025; Tessa et al., 2025), where emoji reactions may be used as signals or indicators of emotional resonance or public sentiment.

7 Limitations

First, our findings rely on automated classifiers, including transformer-based models (e.g., BERT, RoBERTa, DistilBART) and a large language model (GPT-4o), which were not specifically fine-tuned or validated on our dataset. As a result, their predictions may introduce a degree of uncertainty and noise typical of automated annotation pipelines, particularly for nuanced phenomena such as emotion, persuasion, and speech acts. Second, as our dataset consists mostly of public cryptocurrency-related Telegram channels, the communicative dynamics we observe may reflect domain-specific norms, which limits the generalizability of our findings and highlights the need to examine other domains in future work. Doing so would also clarify our interpretation of the positivity bias: although we view it as consistent with social-approval, engagement, or banter signaling, other explanations remain possible, including community norms specific to this domain that may encourage more positive reactions. Nonetheless, evidence that emoji reactions do not reliably reflect emotional resonance, at least in this setting, remains informative. At the same time, our analytical approach is general and can be applied to other domains and platforms. Beyond these interpretative factors, the Telegram reaction system itself introduces structural constraints. The set of standard emoji reactions is finite, and our study does not consider custom emoji sets. In addition, not all channels enable reactions, and when they do, administrators may restrict the available options to a small and often predominantly positive set. These features can introduce selection bias and promote approval-oriented responses by design. Future work could extend these insights by examining how reaction emojis operate within a broader range of digital communication practices.

Acknowledgements

This work was partly supported by SoBigData.it which receives funding from European Union – NextGenerationEU – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) – Project: “SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics” – Prot. IR0000013 – Avviso n. 3264 del 28/12/2021.; and by project SERICS (PE00000014) under the NRRP MUR program funded by the EU – NGEU.

References

- Lorenzo Alvisi, Victoria Popa, Guglielmo Cola, Serena Tardelli, and Maurizio Tesconi. 2025a. From Toxicity to Conformity: Adaptive user behavior to social norms in Telegram communities. *arXiv preprint arXiv:2511.17333*.
- Lorenzo Alvisi, Serena Tardelli, and Maurizio Tesconi. 2024. Unraveling the Italian and English Telegram Conspiracy Spheres through Message Forwarding. In *International Conference on Advances in Social Networks Analysis and Mining*, pages 204–213. Springer.
- Lorenzo Alvisi, Serena Tardelli, and Maurizio Tesconi. 2025b. Mapping the Italian Telegram Ecosystem: Communities, Toxicity, and Hate Speech. *arXiv preprint arXiv:2504.19594*.
- William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318.
- Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.
- Andrea Calamusa, Serena Tardelli, Marco Avvenuti, Stefano Cresci, Iliaria Federigi, Maurizio Tesconi, Marco Verani, and Annalaura Carducci. 2020. Twitter monitoring evidence of Covid-19 infodemic in Italy. *European Journal of Public Health*, 30(Supplement_5):ckaa165–066.
- CardiffNLP. 2021. twitter-roberta-base-sentiment. <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment>. Accessed: July 22, 2025.
- V. D. Chaithra and 1 others. 2019. Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. *International Journal of Electrical and Computer Engineering*, 9(5):4034–4041.
- Emil Chifu, Roxana Hatos, Monica Borda, and Alexandra Terec. 2015. Web harvesting and sentiment analysis of consumer feedback. *Acta Technica Napocensis: Electronics and Telecommunications*, 56(3):27–32.
- Lorenzo Cima, Benedetta Tessa, Amaury Trujillo, Stefano Cresci, and Marco Avvenuti. 2025. Investigating the heterogeneous effects of a massive content moderation intervention via Difference-in-Differences. *Online Social Networks and Media*, 48:100320.
- Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625.
- Margherita Gambini, Serena Tardelli, and Maurizio Tesconi. 2024. The anatomy of conspiracy theorists: Unveiling traits using a comprehensive Twitter dataset. *Computer Communications*, 217:25–40.
- Lisa Graziani, Stefano Melacci, and Marco Gori. 2019. Jointly Learning to Detect Emotions and Predict Facebook Reactions. In *International Conference on Artificial Neural Networks*, pages 185–197. Springer.
- Md Abrar Jahin, Md Sakib Hossain Shovon, MF Mridha, Md Rashedul Islam, and Yutaka Watanobe. 2024. A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *Scientific Reports*, 14(1):24882.
- Olusegun Oladele Jegede. 2024. Speech Acts in the Digital Sphere: A Corpus-Based Exploration of Interactional Dynamics on Social Media Platforms. *Corpus-based Studies across Humanities*, 2(2):359–383.
- Amit Khan, Dipankar Majumdar, and Bikromaditya Mondal. 2025. Sentiment analysis of emoji fused reviews using machine learning and Bert. *Scientific Reports*, 15(1):7538.
- Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PLoS one*, 10(12):e0144296.
- Chuchu Liu, Fan Fang, Xu Lin, Tie Cai, Xu Tan, Jianguo Liu, and Xin Lu. 2021. Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, 2(4):246–252.
- Yinxia Lou, Junxiang Zhou, Jun Zhou, Donghong Ji, and Qing Zhang. 2024. Emoji multimodal microblog sentiment analysis based on mutual attention mechanism. *Scientific Reports*, 14(1):29314.
- Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, 30.
- Ahmed Mahrous, Jens Schneider, and Roberto Di Pietro. 2023. The Role of Emojis in Sentiment Analysis of Financial Microblogs. In *2023 Fourth International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 76–84. IEEE.
- Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. “Blissfully Happy” or “Ready to Fight”: Varying Interpretations of Emoji. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1):259–268.
- Esteban Morales, Jaigris Hodson, Anatoliy Gruzd, and Philip Mai. 2025. I [thumbs-up] your Hate: Emojis as Infrastructural Platform Violence on Telegram. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 2316–2325. University of Hawai’i Press.

- Lynnette Hui Xian Ng and Jia Yuan Loke. 2020. Analyzing Public Opinion and Misinformation in a COVID-19 Telegram Group Chat. *IEEE Internet Computing*, 25(2):84–91.
- John C Paolillo. 2023. The awkward semantics of Facebook reactions. *First Monday*.
- F Plaza Del Arco, S Jiménez-Zafra, María-Teresa Martín-Valdivia, and L Alfonso Urena-López. 2018. Using Facebook Reactions to Recognize Emotion in Political Domain. In *Proceedings of CAEPIA*.
- Robert Plutchik. 1994. *The psychology and biology of emotion*. HarperCollins College Publishers.
- Chris Pool and Malvina Nissim. 2016. Distant supervision for emotion detection using Facebook reactions. In *Proceedings of the Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 30–39.
- Fatima Zahra Qachfar and Rakesh Verma. 2023. ReDASPersuasion at SemEval-2023 Task 3: Persuasion Detection using Multilingual Transformers and Language Agnostic Features. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2124–2132.
- Bin Tareaf Raad, Berger Philipp, Hennig Patrick, and Meinel Christoph. 2018. ASEDs: Towards Automatic Social Emotion Detection System Using Facebook Reactions. In *2018 IEEE 20th international conference on high performance computing and communications; IEEE 16th international conference on Smart City; IEEE 4th international conference on data science and systems (HPCC/SmartCity/DSS)*, pages 860–866. IEEE.
- Tulika Saha, Aditya Prakash Patra, Sriparna Saha, and Pushpak Bhattacharyya. 2020. A Transformer based Approach for Identification of Tweet Acts. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Bhadresh Savani. 2021. bert-base-uncased-emotion. <https://huggingface.co/bhadresh-savani/bert-base-uncased-emotion>. Accessed: July 22, 2025.
- Jonas P Schöne, David Garcia, Brian Parkinson, and Amit Goldenberg. 2023. Negative expressions are shared more on Twitter for public figures than for ordinary users. *PNAS nexus*, 2(7):pgad219.
- John R Searle. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge university press.
- Mohammed Shiha and Serkan Ayvaz. 2017. The Effects of Emoji in Sentiment Analysis. *Int. J. Comput. Electr. Eng.(IJCEE.)*, 9(1):360–369.
- Benedetta Tessa, Lorenzo Cima, Amaury Trujillo, Marco Avvenuti, and Stefano Cresci. 2025. Beyond Trial-and-Error: Predicting User Abandonment After a Moderation Intervention. *Engineering Applications of Artificial Intelligence*, 162:112375.
- Arsenii Tretiakov, Sergio D’Antonio-Maceiras, Áurea Anguera de Sojo Hernández, and Alejandro Martín. 2025. Toxic Discourse in the Digital Battlefield: Analysing Telegram Channels During the Russia-Ukraine ‘Conflict’. *Expert Systems*, 42(7):e70081.
- Aleksandra Urman and Stefan Katz. 2022. What they do in the shadows: Examining the far-right networks on Telegram. *Information, communication & society*, 25(7):904–923.
- Valhalla. 2020. distilbart-mnli-12-1. <https://huggingface.co/valhalla/distilbart-mnli-12-1>. Accessed: July 22, 2025.
- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science*, 359(6380):1146–1151.
- Dong Wang, Tao Xiao, Teyon Son, Raula Gaikovina Kula, Takashi Ishio, Yasutaka Kamei, and Kenichi Matsumoto. 2023. More Than React: Investigating The Role of Emoji Reaction in GitHub Pull Requests. *Empirical Software Engineering*, 28(5):123.
- Joe Watson, Sander van der Linden, Michael Watson, and David Stillwell. 2024. Negative online news articles are shared more to social media. *Scientific Reports*, 14(1):21592.
- Yuhang Zhou, Yimin Xiao, Wei Ai, and Ge Gao. 2025. The Hidden Language of Harm: Examining the Role of Emojis in Harmful Online Communication and Content Moderation. *arXiv preprint arXiv:2506.00583*.
- Xiaorui Zuo, Yao-Tsung Chen, and Wolfgang Karl Härdle. 2024. Emoji driven crypto assets market reactions. *Management & Marketing*, 19(2).