Can Sentiment Be Contagious on Social Media? A Case Study on Reddit's *r/unpopularopinion*

Beatriz Borges*
Imane Jennane*
Yi Ren*
Franziska von Albedyll*
beatriz.borges@epfl.ch
imane.jennane@epfl.ch
yi.ren@epfl.ch
franziska.vonalbedyll@epfl.ch
EPFL
Lausanne, Vaud, Switzerland

Abstract

Social media plays a dual role in contemporary life: it fosters community, connection, and awareness, but it also exposes users to risks such as emotional distress, misinformation, and privacy erosion. One key mechanism underlying both its positive and negative impacts is emotion contagion—the process by which emotional expressions spread across users and influence affective states. While prior research has established this phenomenon on platforms like Facebook and Twitter, relatively little is known about its presence on Reddit, where content exposure is driven by user-curated feeds rather than algorithmic timelines. In this work, we investigate sentiment contagion within Reddit discussions, focusing on the `r/unpopularopinion` subreddit. We label the sentiment of posts and comments, and construct a regression model that accounts for users' baseline sentiment to examine whether the emotional tone of original posts influences the sentiment of subsequent replies. Our analysis reveals a strong and statistically significant association between post sentiment and comment sentiment, providing clear evidence of emotional contagion on Reddit — demonstrating that even in platforms with decentralized content discovery, emotional tone meaningfully shapes online discourse.

Keywords

Social Media, Emotion Contagion, Sentiment Contagion, Reddit, Sentiment Analysis

ACM Reference Format:
Beatriz Borges, Imane Jennane, Yi Ren, and Franziska von Albedyll. 2025. Can Sentiment Be Contagious on Social Media? A Case Study on Reddit's *r/unpopularopinion*. In *Proceedings of DH-500 Computational Social Media (DH-500 '25)*. ACM, New York, NY, USA, 8 pages. https://doi.org/NA

*All authors contributed equally to this research.

ACM ISBN NA
https://doi.org/NA

1 Introduction

Social media is a ubiquitous component of modern life, bringing with it several transformations on how individuals connect, share, and engage with the world. At its core, social media strengthens community building by fostering interpersonal connections, facilitating the exchange of knowledge, and offering emotional support through shared experiences and identities [7, 18]. It also serves as a powerful tool for global connectivity, enabling real-time communication and raising awareness about social, political, and environmental issues, while promoting meaningful cross-cultural dialogue [15, 18]. Moreover, from a commercial standpoint, social media is an effective business and marketing platform, offering cost-efficient tools for brand development, consumer engagement, and targeted advertising strategies [9, 13].

However, alongside these advantages, social media also presents substantial risks and harms. Its widespread use has been linked to mental health challenges, including increased levels of anxiety, depression, and low self-esteem—issues often exacerbated by social comparison, cyberbullying, and the addictive design of these digital platforms [1, 10, 12, 16]. Furthermore, the rapid spread of misinformation on these platforms can distort public understanding, erode trust in institutions, and disrupt democratic processes [19]. Privacy is another significant concern: users frequently share personal data without fully understanding the implications, leaving them vulnerable to data breaches, identity theft, and manipulation by third parties [8]. These concerns are compounded by the rise of surveillance capitalism, in which social media companies monetize user data and allow big corporations or even government institutions to subtly shape behavior through profiling and algorithmic targeting—undermining user autonomy and raising ethical questions about consent and control [20].

This dual nature of social media remains equally evident when focusing on emotion contagion in social media — the process by which users' emotions are influenced and propagated by the emotional tone of content they encounter online. On the upside, emotional contagion can uplift and inspire users, improving psychological well-being within supportive online communities [3, 11]. It also plays a key role in organization and collaboration during crises and social movements, where it can encourage collective action, solidarity, and civic participation [5].

Yet, emotional contagion can also have detrimental effects. Negative emotions such as anger, fear, and sadness can spread rapidly across platforms, heightening stress, hostility, and contributing to emotional fatigue among users [3, 11]. The viral diffusion of emotionally resonant content can create emotionally charged echo chambers, reinforcing bias, fueling polarization, and impeding constructive discourse [2]. In high-stakes contexts such as public health emergencies, emotionally laden content can escalate panic and undermine efforts to disseminate accurate information and maintain social order [4].

Previous work has documented the presence and impact of emotional contagion on platforms such as Facebook and Twitter, where algorithmically curated feeds expose users to emotionally resonant content that shapes their own expressions and behaviors [3, 11]. However, the manifestation of emotion contagion on Reddit is less explored. In addition, Reddit is a platform structured around user-driven communities and customizable content discovery — allowing users to navigate it through topic-specific subreddits and apply ranking criteria to directly influence the content they are exposed to, offering a distinct and arguably more intentional mode of content browsing than typical algorithmic feeds.

With this work, we sought to explore the phenomenon of emotional contagion in user-moderated, community-centric digital environments and to contribute to the broader discourse on the emotional dynamics of social media. We focused on understanding to what extent Reddit users are emotionally influenced by the sentiment of other users, and, specifically, how commenters' sentiments shift in response to the sentiment of the original Reddit post they are commenting on.

A recent study provided initial evidence for emotion contagion on Reddit, reinforcing the notion that social influence through emotional tone extends broadly to the online landscape [6]. This thesis is supported by our own findings. We summarize the main contribution of our work as the conduction of a large-scale quantitative analysis of sentiment contagion in Reddit, focusing on how the sentiment of original posts influences the sentiment of subsequent comments. We find strong and statistical evidence for the presence of sentiment contagion in Reddit, corroborating previous findings and further supporting the hypothesis of emotion contagion's existence in online social media.

2 Data

2.1 Source

We utilized the Pushshift Reddit dataset, a comprehensive archive of Reddit posts and comments. To extract the data specifically for the subreddit `r/unpopularopinion`, we followed the procedure described in [17].

2.2 Filtering

We applied three filtering steps to our data. First, we restricted our dataset to top-level (TL) comments—comments made directly in response to a post rather than as replies to other comments. This was necessary because we aimed to measure how the sentiment of the post influences the sentiment of the reply; replies to other comments might instead reflect a reaction to those comments.

Table 1: Overview of the `r/unpopularopinion` dataset extracted from Pushshift.

Type	Count	Authors	Time Range
Posts	2,394,871	703,426	02.2012 – 12.2024
TL-Comments	19,883,897	1,918,328	08.2013 – 12.2024

Second, we only included users who had made at least one top-level comment on at least 100 different posts. This ensured that each user had sufficient data to analyze their reactions across a range of post sentiments. Finally, to establish a sentiment baseline for each user, we required that they had made at least one top-level comment on at least 10 different posts classified as having neutral sentiment. This allowed us to interpret sentiment shifts relative to each user's typical expression in emotionally neutral contexts.

3 Methods

3.1 Sentiment Analysis

We used `twitter-roberta-base-sentiment-latest` by Cardiff NLP to analyze the sentiment of user comments [14]. This model is based on RoBERTa, a robustly optimized BERT architecture, and was fine-tuned on 124 million tweets to perform sentiment analysis in social media contexts. We consider tweets and Reddit comments sufficiently similar in structure and linguistic style-both being short, informal, user-generated texts-which makes the model appropriate for our Reddit-based sentiment analysis. The model outputs one of three sentiment labels: `positive`, `neutral`, or `negative`, together with its probability score.

3.2 User Sentiment Baseline

To control for user-specific tendencies in sentiment, we compute a sentiment baseline for each user. This baseline is calculated by averaging the sentiment scores of the user's top-level comments on posts labeled as `neutral`. These comments are assumed to represent the user's typical sentiment in contexts where the post itself does not carry strong emotional cues, providing a stable reference point for subsequent analysis.

3.3 Statistical Modelling

To investigate whether the sentiment of a Reddit comment is influenced by the sentiment of the corresponding post, we model the relationship using multiple linear regression. The dependent variable y is the sentiment score of the user's top-level comment, while the independent variables are:

- x_1: sentiment of the post title
- x_2: sentiment of the post body
- x_3: baseline sentiment of the user

The regression takes the form:

$$y = ax_1 + bx_2 + cx_3 + d \tag{1}$$

where a, b, and c represent the contribution of each variable to the predicted comment sentiment, and d is the intercept.

We also explore how the strength of this relationship varies across different temporal and user-specific subsets:

- **Temporal variation:** To assess whether users have become more or less responsive to post sentiment over time, we divide the dataset into five intervals, with each interval representing 27 months, and compare regression coefficients across periods. In particular, for each time interval, we only include users with at least 10 top-level comments on neutral posts, in order to have reliable estimations on the users' base emotion.
- **User group comparison:** To examine whether certain user groups are more susceptible to post sentiment, we segment users based on their average sentiment (predominantly negative vs. predominantly positive) and compare the resulting regression models.

This analysis aims to quantify the extent to which post content—both title and body—correlates with the emotional tone of user responses, while controlling for individual baseline tendencies.

4 Manual validation

To evaluate the reliability of our automated sentiment classification pipeline and to ensure that the model predictions are consistent with human interpretation in the context of Reddit data, we conducted a structured manual validation. This process aimed to assess the accuracy of the model's predictions on posts and comments and to measure the agreement between human annotators as a proxy for the task's inherent subjectivity.

Sampling Strategy and Annotation Design.
We selected a total of 200 data points from the processed `r/unpopularopinion` dataset. This sample was stratified across different types of content to ensure broad coverage. Annotator 1 reviewed 100 data points from Reddit posts, comprising 50 post titles and 50 post bodies. Annotator 2 labeled 100 top-level comments. The distinction between post content and reactive commentary allowed us to assess whether the model performs consistently across both original and responsive discourse, which differ in tone and structure.

The primary objective of this phase was to assess whether the sentiment model, originally trained on Twitter data, can be reliably applied to Reddit, whose posts are structurally longer, topic-specific, and often more sarcastic or ironic in tone. Given the informal and opinionated nature of both platforms, we hypothesized that a transfer of model accuracy would be viable, but required empirical confirmation.

Annotation Protocol.
Each annotator independently labeled their assigned 100 data points without access to model outputs or the other annotator's labels. Sentiment categories were defined as:

- `Positive` — content expressing approval, optimism, appreciation, or joy.
- `Neutral` — emotionally flat, factual, or ambiguous content.
- `Negative` — content indicating disapproval, frustration, anger, or pessimism.

A brief calibration session was conducted beforehand to align understanding of ambiguous cases and reduce systematic labeling bias.

Figure 1 presents the agreement and disagreement rates between the model and each annotator over their respective 100 data points.

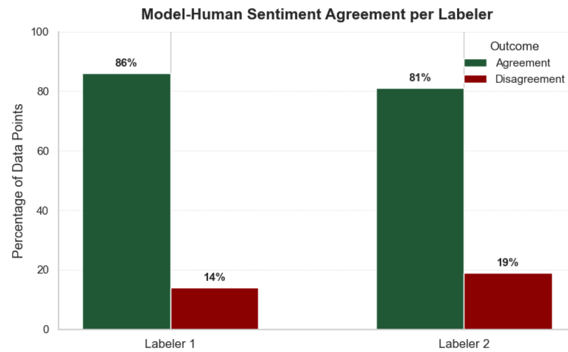


Figure 1: Model-human agreement for each labeler across 100 annotated data points. Green bars represent agreement, red bars represent disagreement.

Cross-Validation Between Annotators.
To evaluate internal consistency, a second round of labeling was conducted. Annotator 1 reviewed 100 comments originally labeled by Annotator 2, while Annotator 2 cross-validated 100 post-related samples from Annotator 1. This cross-review allowed us to evaluate the degree of intersubjective agreement on sentiment interpretation, particularly across content types that differ in length, tone, and intention.

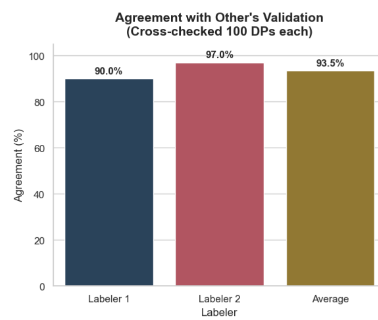


Figure 2: Agreement levels when each annotator validated the other's original labels, showing inter-annotator consistency.

Figure 2 illustrates the agreement rates when each labeler independently re-evaluated the other's annotations.

Measuring Model-Human Agreement: Cohen's Kappa Coefficient.
To evaluate the extent to which the model's sentiment predictions align with human interpretation, we employed *Cohen's Kappa coefficient* (κ). Although traditionally used to assess agreement between two human annotators, Cohen's Kappa can also be used to evaluate how closely a model replicates human-like categorization in tasks involving discrete class labels.

Unlike simple accuracy, which measures the raw percentage of matching labels, Cohen's Kappa corrects for chance agreement by considering the expected agreement that would occur if both annotator and model were assigning labels randomly based on their own distribution. This is particularly important in multi-class classification settings with class imbalance, which is typical in real-world sentiment data.

Formally, Cohen's Kappa is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (2)$$

where:

- p_o is the observed agreement proportion between the human annotator and the model,
- p_e is the expected agreement by chance, based on the marginal distributions of the labels.

A value of $\kappa = 1.0$ indicates perfect agreement beyond chance, $\kappa = 0.0$ corresponds to random labeling, and negative values indicate less agreement than would be expected by chance.

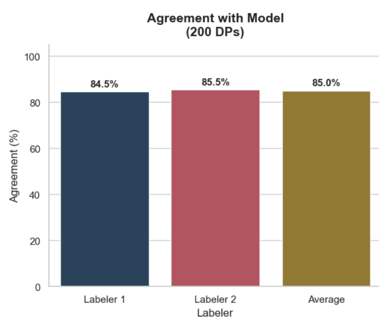


Figure 3: Agreement between the annotators and the model, along with their average agreement. Based on the 200 samples per anotator.

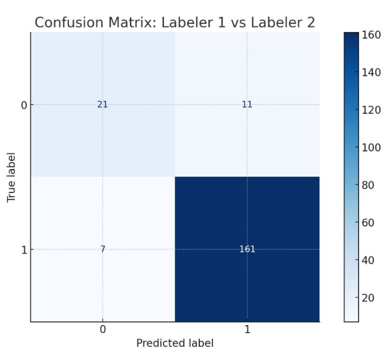


Figure 4: The confusion matrix of the two annotators' agreement (1) or disagreement (0) with the sentiment classification model outputs.

Figure 4 summarizes how well each annotator individually agreed with the model.

Interpretation of Results.

Across the 200 annotated samples, the computed Cohen's Kappa between the model predictions and the human annotators was $\kappa = 0.65$, which corresponds to "substantial agreement" on the commonly accepted interpretation scale from Landis and Koch (1977):

- $\kappa < 0.00$: Poor agreement
- 0.00–0.20: Slight agreement
- 0.21–0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–1.00: Almost perfect agreement

This result strongly suggests that the model is capable of approximating human sentiment classification with high fidelity. This is further visualized in Figure 3, which shows the individual and average agreement levels between the annotators and the model. Moreover, a qualitative review of mismatches revealed that the majority occurred on borderline cases, most often between `neutral` and either `positive` or `negative`. Importantly, no instances of full polarity inversion (e.g., model: `positive`, human: `negative`) were observed. This confirms that the model not only captures directionality of sentiment but also avoids catastrophic misclassification.

Validation Summary.

Metric	Description	Value
Model Evaluated	Sentiment classifier	`twitter-roberta-base-sentiment-latest`
Agreement Metric	Comparison metric	Cohen's Kappa (κ)
Validated Samples	Total manually annotated	200
Annotator 1	Post titles and bodies	50 + 50
Annotator 2	Top-level comments	100
Sentiment Classes	Label options used	Positive, Neutral, Negative
Kappa Score	Model vs. human agreement	0.65 (Substantial)
Polarity Inversions	Strong misclassifications	Very rare
Typical Disagreement	Borderline misclassifications	Neutral vs. Pos/Neg

Table 2: Overview of model-human validation with class definitions, annotator roles, and agreement results.

Conclusion

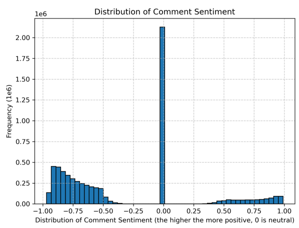
Overall, the manual validation confirms that the sentiment classification model generalizes well to Reddit data despite being trained on Twitter content. The classifier achieved a Cohen's Kappa of 0.65 and an average agreement rate of 85% with human annotators, indicating substantial reliability. Importantly, no instances of complete polarity inversion were observed (e.g., cases where the model labeled a clearly negative post as positive, or vice versa) except for rare cases of irony. Most disagreements occurred on borderline cases, particularly between `neutral` and `positive`/`negative` classes.

Takeaway: The model demonstrates high accuracy, with most mismatches attributable to subjective interpretation rather than major classification errors, validating its suitability for large-scale sentiment analysis on Reddit.

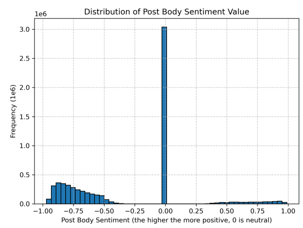
5 Results

5.1 Emotion Distribution

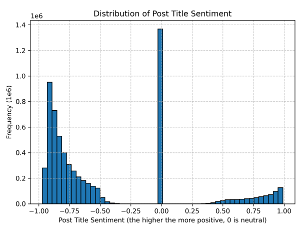
We did some illustration on the emotion distribution of post title, post body and user baseline emotion.



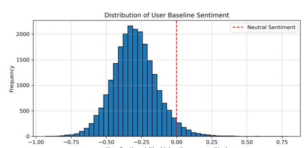
(a) Comment Sentiment



(b) Post Body Sentiment



(c) Post Title Sentiment



(d) User Baseline Sentiment

Figure 5: Sentiment distributions across comments, post bodies, post titles and user baseline

As shown in Figure 5, the sentiment distributions of comments, post bodies, and post titles are quite similar, with a noticeable skew toward the negative side. In both the positive and negative ranges, sentiment becomes increasingly concentrated as emotional intensity rises. A prominent bar at 0.00 reflects the presence of neutral sentiment in the data.

We also have the distribution of user baseline sentiment in Figure 5(d). It is close to a normal distribution, with mean value between -0.50 and -0.25. Only 3.5% users have baseline sentiment larger than 0. The interesting similarity between the user baseline sentiment distribution and normal distribution could be attributed to the central limit theorem, given the fact that each user has enough comments in our dataset.

5.2 Regression on the whole processed dataset

We conducted the linear regression to examine the relationship between comment sentiment, post sentiment and user baseline sentiment, as is explained in 3.3. The regression statistics is shown in Table3, and the regression result is Equation 3.

Table 3: Statistics for regression on the whole processed dataset

Variable	Coef	t-value	P>|t|
user_base_emotion	0.5643	368.087	0.000
post_title	0.1047	263.864	0.000
post_body	0.0381	79.993	0.000
const	-0.0888	-165.875	0.000

No. Observations: 6,447,646 **R-squared**: 0.037

$$\begin{aligned}\text{comment\_emotion} = &+ 0.5643 \cdot \text{user\_base\_emotion} \\ &+ 0.0381 \cdot \text{post\_body} \\ &+ 0.1047 \cdot \text{post\_title} \\ &- 0.0888\end{aligned} \tag{3}$$

From Table 3, we observe that the R-squared value is quite low, indicating that the model does not provide highly accurate predictions. However, all coefficients are statistically significant (p-value $\approx$ 0.000), which demonstrates the validity of our model.

Regarding the coefficient values, we observe that `user_base_emotion` has a much larger magnitude than `post_title` and `post_body`, indicating that a user's baseline emotion is the most influential factor in determining comment sentiment. Furthermore, the positive sign of the `user_base_emotion` coefficient aligns with intuition: users with generally positive dispositions tend to leave more positive comments, and vice versa. Additionally, the coefficient for `post_title` is larger than that for `post_body`, which suggests that the title of a post generally has a greater impact on comment sentiment than its content.

5.3 Regression with temporal variation

In order to analyze whether the comment sentiment of users is becoming more susceptible to post sentiment through time, we divided out dataset into 5 bins with same time interval. The bins range and the regression coefficients are illustrated in Table 4. To better illustrate the regression coefficients of `post_title` and `post_body`, we plot Figure 6.

As is shown in Figure 6, there is no obvious or consistent trend in the coefficient of `post_title` and `post_body` across time intervals. However, it's noticeable that in the 2018–2020 and 2020–2022 intervals, the coefficients of `post_title` are particularly high. In the same time intervals, the count of comments, posts and unique commentators is also much higher than before and after. One possible explanation is that during the COVID-19, people have to turn to online platforms due to social distance restrictions. During this period, people send more comments and are easier to be swayed by the `post_title`. However, this explanation is not by the coefficients of `post_body`, and we have not verified if the content of the posts and comments is related to COVID-19.

In Figure 6, except for the 2013–2015 interval, the other four intervals show that `post_title` has a larger impact on the comment sentiment than `post_body`. In Table 4, the coefficients for `user_base_emotion` are positive and larger than the coefficients

Table 4: Regression Coefficients by Time Interval

Variable	2013–2015	2015–2018	2018–2020	2020–2022	2022–2024
Time Range	2013-08-27 2015-12-04	2015-12-04 2018-03-12	2018-03-12 2020-06-18	2020-06-18 2022-09-25	2022-09-25 2024-12-31
user_base_emotion	0.1535 (0.086)	0.5628 (0.014)	0.5852 (0.003)	0.6255 (0.003)	0.5522 (0.004)
post_title	0.0503 (0.024)	0.0953 (0.004)	0.1103 (0.001)	0.1054 (0.001)	0.0688 (0.001)
post_body	0.0650 (0.025)	0.0677 (0.004)	0.0410 (0.001)	0.0207 (0.001)	0.0592 (0.001)
const	-0.3181 (0.036)	-0.1078 (0.006)	-0.0949 (0.001)	-0.0656 (0.001)	-0.0874 (0.001)
R-squared	0.011	0.040	0.038	0.036	0.031
Posts	1,099	26,229	520,035	393,879	116,225
Unique Commentators	8	204	6,742	9,628	3,472
Comments (Sample Size in Regression)	1,575	69,943	2,096,821	2,322,633	895,643

***Note:* All regression coefficients have p-value ≈ 0.000, except in the 2013–2015 interval, where only `post_title` (p = 0.039) and `post_body` (p = 0.009) are significant; `user_base_emotion` is not significant (p = 0.074). Standard errors are in parentheses.**

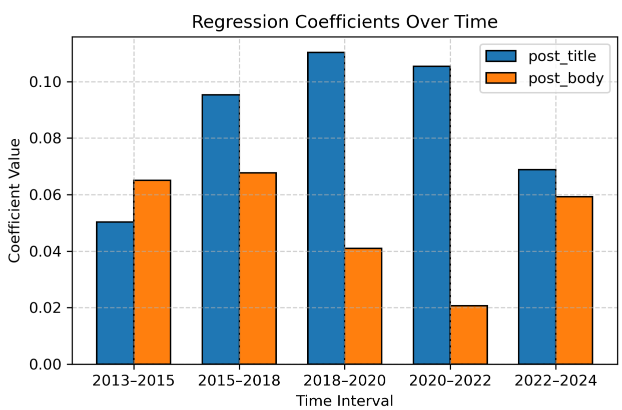


Figure 6: Regression coefficients of `post_title` and `post_body` across time intervals.

of `post_title` and `post_body`. These observations are in accordance with the previous results.

In 2013-2015 interval, `post_title` (p = 0.039) and `post_body` (p = 0.009) are significant ($p<0.05$), the `user_base_emotion` is not significant (p = 0.074). This could be explained that there are only 8 unique commentators in during this period, contributing to less valid `user_base_emotion` data. The post and comment count is much smaller than other time periods, making the p-value of these two coefficients larger.

5.4 Regression with User Group Comparison

To investigate whether users with different baseline sentiments are differently influenced by post sentiment, we divided users into three groups—positive, neutral, and negative—based on their baseline sentiment scores. Using a threshold of 0.005, users with a baseline sentiment greater than this value were assigned to the positive group, those with sentiment below −0.005 to the negative group, and the remaining users to the neutral group. This threshold-based classification ensures a broader and more balanced neutral group, which is both analytically advantageous and intuitively reasonable. This grouping is derived from the distribution of user baseline sentiment, as illustrated in Figure 5(d). The statistics of the user groups are illustrated in Table 5. However, the negative group still holds the most post, comments and commentators. The unbalance in the sentiment group is a direct result of unbalanced user baseline sentiment Figure5(d).

Table 5: Statistics of Sentiment Group

Metric	Positive	Neutral	Negative
#Post	88,902	19,283	1,073,600
#Comment	121,056	20,643	6,305,947
#Commentator	768	112	21,891

After dividing the user groups, we conducted separate linear regressions for the positive, neutral, and negative groups, as shown in Table 6. The variable `user_base_emotion` was excluded from the regression for the neutral group because users in this group exhibit baseline emotion values very close to zero (absolute value < 0.005). Including this variable would offer little explanatory power, potentially reduce the numerical stability of the estimation, and yield an inflated or meaningless coefficient due to the lack of variance.

As is shown in Table 6, in positive and negative group, the coefficient for `user_base_emotion` is positive and remains the largest term compared to the coefficients of `post_title` and `post_body`. Across all groups, `post_title` has larger impact than `post_body`. These results align well with previous regressions.

Particularly, it could be inferred from the coefficients of `post_title` and `post_body` that the more positive a user is, the more susceptible she/he is to the sentiment of post. And the coefficient of `user_base_emotion` in negative group is larger than that in the

Table 6: OLS Regression Results by User Group

Variable	Positive Group	Neutral Group	Negative Group
user_base_emotion	0.3647 (0.019)	–	0.5894 (0.002)
post_title	0.1604 (0.003)	0.1263 (0.007)	0.1033 (0.000)
post_body	0.0806 (0.004)	0.0620 (0.008)	0.0368 (0.000)
const	-0.0744 (0.003)	-0.0934 (0.005)	-0.0812 (0.001)
R-squared	0.042	0.029	0.034
No. Observations	121,056	20,643	6,305,947

Note: **All coefficients are statistically significant with $p < 0.001$. Standard errors are in parentheses. The variable `user_base_emotion` is not included in the regression for the neutral group.**

positive group, indicating that negative users may tend to keep their original comment sentiment than the positive users.

Previous work has found that users were significantly more likely to adopt positive sentiments than negative ones [3], which seems to contradict our findings. However, this contrast makes sense in the broader context of the `r/unpopularopinion` subreddit, where there exists a strong imbalance between posts with positive and negative sentiments, the latter being much more common. In [3], researchers found "that on average a negative post follows an over-exposure to 4.34% more negative content than baseline." Given that in our data, 90.85% of posts have a negative sentiment, this overwhelming exposure to negative valence content provides a likely explanation for the larger susceptibility of users with positive sentiment baselines compared to the users with negative sentiment baselines, which maintain their negative valence more often.

6 Discussion

Our analysis offers robust evidence for the presence of sentiment contagion on Reddit. This effect remained consistent across the full dataset and within various data segments, spefically across different time periods and different user baseline sentiment categories. These findings suggest that users are emotionally influenced by the sentiment of original posts, even in a platform structure that emphasizes user agency in content exposure.

However, we acknowledge several limitations. First, our study is confined to a single platform — Reddit — and within it, a single subreddit: `r/unpopularopinion`. While this subreddit is conversationally rich, it may not reflect broader Reddit usage patterns or general online discourse. Second, Reddit's user base is not demographically representative of the global population. It is predominantly composed of individuals who are Western, Educated, Industrialized, Rich, and Democratic (WEIRD), and English-speaking, which limits the generalizability of our findings across cultural and linguistic contexts. Additionally, our methodology for establishing user baseline sentiment required filtering for users with high activity levels. As a result, the analysis may disproportionately reflect the behavior of particularly engaged or prolific users, rather than casual participants. Finally, our dataset exhibits imbalances in sentiment distribution and posting frequency across different time intervals, which may influence model sensitivity and limit temporal generalization.

Despite these constraints, our results offer meaningful insight into sentiment contagion dynamics on Reddit and contribute towards understanding emotional contagion on community-driven platforms — corroborating the existence of emotional contagion in social media, and reflecting the need for designing emotionally-aware online communities.

7 Conclusion

In this work, we examined the degree to which the sentiment of Reddit posts influences the sentiment of subsequent user comments. Focusing on the `r/unpopularopinion` subreddit, we labeled post and comment sentiments and constructed a regression model that accounted for each user's baseline sentiment tendencies.

Our analysis revealed a strong and statistically significant association between the sentiment of original posts and the sentiments expressed in replies. These findings provide clear evidence of sentiment contagion on Reddit, underscoring that even in user-curated environments, emotional tone can meaningfully shape online discourse. This work contributes to a growing body of research supporting the existence of emotional contagion in the online landscape in general and in online social media in particular.

References

[1] Cecilie Andreassen, Torbjørn Torsheim, Geir Brunborg, and Ståle Pallesen. Development of a facebook addiction scale. *Psychological reports*, 110:501–17, 04 2012.
[2] William J. Brady, Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318, 2017.
[3] Emilio Ferrara and Zeyao Yang. Measuring emotional contagion in social media. *PLOS ONE*, 10(11):1–14, 11 2015.
[4] Dana Garfin, Roxane Silver, and Alison Holman. The novel coronavirus (covid-2019) outbreak: Amplification of public health consequences by media exposure. *Health Psychology*, 39:355–357, 03 2020.
[5] Paolo Gerbaudo. Constructing public space| rousing the facebook crowd: Digital enthusiasm and emotional contagion in the 2011 protests in egypt and spain. *International Journal of Communication*, 10(0), 2016.
[6] Ram D. Gopal, Afrouz Hojati, and Raymond A. Patterson and. A little bit goes a long way: Indirect effects of content moderation on online social media. *International Journal of Electronic Commerce*, 29(1):39–64, 2025.
[7] Anatoliy Gruzd, Barry Wellman, and Yuri Takhteyev. Imagining twitter as an imagined community. *American Behavioral Scientist*, 55(10):1294–1318, 2011.
[8] Ankit Kumar Jain, Somya Ranjan Sahoo, and Jyoti Kaubiyal. Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems*, 7(5):2157–2177, Oct 2021.
[9] Andreas M. Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53(1):59–68, 2010.

[10] Robin Kowalski, Gary Giumetti, Amber Schroeder, and Micah Lattanner. Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological bulletin*, 140, 02 2014.

[11] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[12] Ethan Kross, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. Facebook use predicts declines in subjective well-being in young adults. *PLOS ONE*, 8(8):1–6, 08 2013.

[13] W. Glynn Mangold and David J. Faulds. Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4):357–365, 2009.

[14] Cardiff NLP. twitter-roberta-base-sentiment-latest. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest, 2024. Accessed: 2025-05-29.

[15] Luke Sloan and Anabel Quan-Haase. *The SAGE Handbook of Social Media Research Methods*. SAGE Publications Ltd, 55 City Road, London, Jun 2016.

[16] Jean M. Twenge and W. Keith Campbell. Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports*, 12:271–283, 2018.

[17] u/Watchful1. Separate dump files for the top 40k subreddits. https://www.reddit.com/r/pushshift/comments/1itme1k/separate_dump_files_for_the_top_40k_subreddits/, 2024. Accessed: 2025-05-29.

[18] Jose van Dijck. *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press, 01 2013.

[19] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[20] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st edition, 2018.

Received 10 June 2025