JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 06/06 pp168−180
Volume 4, Number 2, June 2023
DOI: 10.23919/JSC.2023.0014

Negative Sentiment Shift on a Chinese Movie-Rating Website

Hongkai Mao*

Abstract: Shifting to negativity is more and more prevalent in online communities and may play a key role in group polarization. While current research indicates a close relationship between group polarization and negative sentiment, they often link negative sentiment shifts with echo chambers and misinformation within echo chambers. In this work, we explore the sentiment drift using over 4 million comments from a Chinese online movie-rating community that is less affected by misinformation than other mainstream online communities and has no echo chamber structures. We measure the sentiment shift of the community and users of different engagement levels. Our analysis reveals that while the community does not show a tendency toward negativity, users of higher engagement levels are generally more negative, considering factors like the different movies they consume. The results indicate a fitting-in process, suggesting the possible mechanism of group identity on sentiment shift on social media platforms. These findings also provide guidance on web design to tackle the negativity issue and expand sentiment shift analysis to non-English contexts.

Key words: computational social science; online community; group polarization; sentiment analysis; user engagement; negative sentiment

1 Introduction

People may show negative behaviors on social media platforms for reasons from the user level and the platform level. At the user level, people may post negative content for personal considerations. As negative messages could spread widely and rapidly compared to positive ones[1], political figures may post emotionally negative comments and fuel the negative trend[2]. Although not intended for attention, ordinary users may also show negative behaviors due to factors like high information load and social load, leading to quitting that platform[3, 4]. At the platform level, on the other hand, people may post negative content because they are influenced by platform designs. Some platforms' feed or recommendation systems may lead users toward extreme content, which in turn causes users to express more outrage and negative expressions[5]. What is more, some platforms' features like anonymity may induce users to show higher negativity in their posts[6]. These findings suggest that the "bad is stronger than good" psychological phenomenon is prevalent in the digital age[7].

In addition to the above factors, community users' negative behaviors can also be interpreted from the perspective of group polarization. According to Sunstein[8], group polarization could form when people interact with like-minded people in a group. Group members could abandon mild opinions and express more extreme opinions. Even though they may be exposed to broader viewpoints in the digital age, group polarization still exists. People may opt to discuss with like-minded people or expose to like-minded information sources, and such behaviors strengthen their group identity[9, 10]. On many occasions, polarized sentiment is a part of group polarization in the form of being overly positive or negative, of which being

• Hongkai Mao is with the Social Sciences Division, University of Chicago, Chicago, IL 60637, USA. E-mail: hongkai@uchicago.edu.
* To whom correspondence should be addressed.
Manuscript received: 2023-02-21; revised: 2023-06-06; accepted: 2023-07-22

negative is more common[11, 12]. Sometimes, polarized sentiment, in turn, even helps create polarized opinions, leading to group polarization[13, 14]. Therefore, these findings suggest that negative sentiment is closely related to group polarization, further implying that being negative may relate to users' group identity.

Nevertheless, previous works focusing on negative sentiment dynamics and group polarization usually emphasize its intertwined relationship with misinformation in echo chambers. This focus makes it unable to clearly identify the possible role group identity plays in users' negative sentiments. According to Cacciatore et al.[15], echo chambers are places in online communities where people segregate and consume similar information. However, users with more extended involvement in a homogeneous community could have more negative sentiments, leading to a possible negative emotional contagion[11, 12]. Nevertheless, such behaviors are often mixed with misinformation, as people in echo chambers may preferably "like" intentionally false information, making other group members consume misinformation[16]. As a result, it remains unclear whether negativity is related to group identity because previous related work concluded from the empirical analysis of homogeneous communities that misinformation is prevalent and does not control for related variables that may influence users' sentiment change.

To study the sentiment shift in an online environment less affected by misinformation and echo chambers, we analyzed over 4 million comments from Douban Movie, a Chinese movie-rating website with a strong community culture[♯]. We use this dataset for several reasons. First, when people comment on a movie, the content is usually constrained by the movie and is based on the facts of the movie. So, less misinformation is included in movie comments. Second, although movie comments are usually personal opinions toward movies, we can disentangle and control movie-related variables, meaning our results could be robust. Finally, millions of movie comments are usually publicly viewable on different movie-rating websites. So, we could scale up for comparative analysis across different websites and even control the same set of movies.

Following group polarization and previous research, we propose the following two hypotheses. First, since negative comments are more prevalent, we propose the following hypothesis for the overall sentiment trend:

H1: Douban Movie tends to be negative in comments over time.

Then, since users' group identity may grow as they engage in the community deeper, we propose the following hypothesis for the users' sentimental behaviors:

H2: Active users tend to be more negative than less active users.

We measure sentiment shift at the community and user levels using multiple approaches. First, we analyze the sentiment in the community over time using trend visualizations and formal regressions with different specifications. Since the sentiment encoded in comments is greatly influenced by movies other than users' subjective perceptions, we control different movie-related variables in regression models. Then, using users' comments count to approximate their engagement level in the community, we measure how users' sentiment shift in comments relates to their community engagement. We use formal regressions and supportive visualizations jointly to identify users' sentiment shift. In regressions, we first define a metric called sentiment polarity to identify users' preference of using extremely sentimental comments. Then, we regress sentiment polarity scores against users' community engagement while controlling for factors such as the time users entered the community and movie quality, in order to observe the differences among users with varying levels of engagement. In visualizations, we also check how users of different engagement levels differ by year and how they update the sentiment they expressed in comments in their user-lifecycles.

This work contributes to online community sentiment analysis in the following ways. First, it unveils sentiment dynamics in communities when misinformation is not a major concern. Hence, the results are meaningful for further exploring the general mechanisms that drive such dynamics. Second, it provides guidance on web design to tackle the negativity issue, especially with regard to strategies to motivate different users. Thus, it can help build healthy and sustainable online communities. Third, it expands sentiment shift analysis to non-English contexts. The work used data from a Chinese movie-rating website. Currently, there is less research on such topics using

[♯]Website: https://movie.douban.com

data from Chinese social media platforms due to reasons like the language barrier and data accessibility. So, this work also helps the study of large-scale sentiment dynamics on Chinese social media platforms.

The design of the rest of this paper is as follows. First, we introduce the data and methods. Second, we present the results. Finally, we discuss and conclude with findings, limitations, and possible future work.

2 Material and Method

2.1 Data

This study uses over four million comments from Douban Movie, a large Chinese movie-rating website. Douban Movie is part of the Douban website, an online community for people who love reading, listening to music, watching movies, etc. Hence, the Douban Movie shares the Douban community culture. This study uses comments from Douban Movie as an entry point for studying the Douban community.

When the Douban Movie has both short and long comments, this study uses short comments for analysis as they express sentiment straightforwardly and get more users involved. Short and long comments are placed in two different sections on the webpage. Long comments are like reviews given by critics with a length of several paragraphs, typically offering an in-depth analysis. Therefore, the sentiment in long comments could be complicated. On the other hand, in the short movie comments section, users can simply leave one sentence telling others their opinions about the movie. So, it is more suitable to extract sentiments for aggregated user sentiment analysis. Besides, the number of short comments is much higher than that of long comments. Typically, the number of short comments surpasses that of long comments by one or even two orders of magnitude. This discrepancy suggests that only a minor fraction of users participate in generating long comments in comparison to short ones. Therefore, using short comments for analysis also guarantees wider user coverage.

We used a dataset of short movie comments on Douban Movie collected in August 2019※. The dataset contains over 4.4 million comments from more than 68 thousand movies given by more than 638 thousand users, and a detailed distribution is shown in Fig. 1. More than 94% of the total comments come with rating information. The comments start from June 2005, when the Douban movie was established. Since Douban Movie set a max viewing limit of 220 comments for unlogged users, for movies with less than 220 comments, all comments were scraped, and for movies with thousands of comments, the first 220 comments were scraped¤.

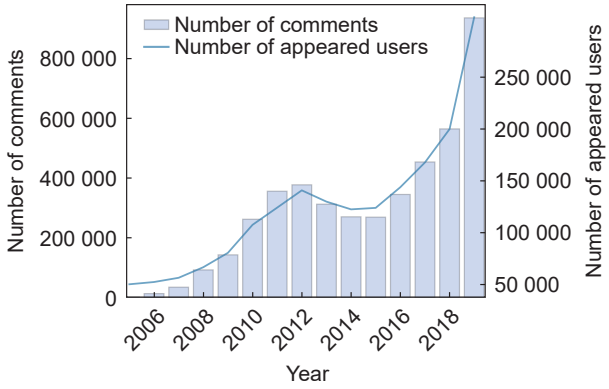


Fig. 1 Number of comments and appeared users in the dataset by year. Both the number of comments and users in our dataset kept increasing from 2005 to 2012, then suddenly started decreasing until 2015, and finally began to increase faster than ever before. This indicates that many new users have swarmed into the community in recent year.

2.2 Sentiment labeling

We used BERT to label movie comments. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a deep language model that can process the text both from left to right and right to left to generate high-quality embeddings[17]. It is a sizeable pre-trained language model released by Google that can perform various tasks, including sentiment classification. As BERT is pre-trained with commonly crawled datasets over different tasks, it performs well in various scenarios. Once being fine-tuned with labeled data, it can reach even better performance on the targeted dataset. Therefore, given its superiority in sentiment classification, we fine-tuned a pre-trained BERT for this task.

To make use of all comments and since comments do not have ground-truth sentiment labels, we use human-labeled comments for fine-tuning a BERT model that can classify a comment as positive, neutral, or negative.

※ Data are accessible at: https://github.com/csuldw/AntSpider. You may contact the repository's owner or follow the guidance to get the dataset.

¤ The dataset was collected by a group of developers. When they scraped the comments, they scraped several times in August and September of 2019, and they merged the dataset. So, some movies have more than 220 comments. This does not influence our further analysis.

Previous research involving movie comments typically used ratings directly as sentiment labels to train binary classifiers[18]. The method proved successful on long comments, as movie ratings are usually highly correlated with the sentiment contained in the text of the comments. However, the rating and sentiment of a short comment could diverge. Thus, ratings cannot be considered as ground-truth labels for sentiment in comments. While users usually use ratings to assess the overall quality of the movies, they may mention one character or a specific plot in the movie that impresses them in the body of the short comments. Moreover, a binary classifier means that only comments with high and low ratings are kept and labeled as positive and negative, respectively. Comments with middle rating ranges are discarded for the equivocal sentiment they entail.

In this study, two human experts labeled random sampled 2000 comments for fine-tuning. Human experts did not know the rating information of the comments. They are only accessible to the body of comments for the labeling processing, making sure the rating information does not influence the labels they give. They used 1, 0, and −1 to denote negative, neutral, and positive sentiments. We also hired two additional human raters to validate the labeling results. We got 0.716 and 0.702 Krippendorffs' alpha scores separately for the two samples, confirming good annotator agreement. After finishing labeling, we split the 2000 labeled comments into training, validation, and test set, which contains 1200, 400, and 400 comments, separately. Then, using human-labeled sentiments as ground-truth labels, a pre-trained BERT model is fine-tuned and validated with the training and validation sets. Finally, we test the performance of the fine-tuned BERT model on our test set. The Matthews correlation score on the test set is 0.571, confirming a good model performance*. See Appendix A for other details about fine-tuning BERT.

2.3 Measurement

Sentiment polarization. We used the metric sentiment polarization from Vicario et al.[12] to describe how users shift toward positivity or negativity‡. The metric is as follows:

$$\rho_{\sigma}(j) = \frac{\left(N_j - 2k_j - h_j\right)\left(N_j - h_j\right)}{N_j^2},$$

where N_j, k_j, and h_j are the numbers of all comments, negative comments, and neutral comments of a user j, respectively. The metric ranges from −1 to 1, where −1 means extremely negative and 1 means pure positive. The metric puts less weight on small changes and emphasizes extreme shifts more.

Mean sentiment. Since the sentiment polarization metric is specific to users, we also used mean sentiment when sentiment polarization is unavailable, like the sentiment trend in the community or the sentiment trend of different groups of users.

User engagement. We used users' number of comments to approximate their engagement level in the community. For ordinal comparison, we define the top 5% of users as active users. For regression, we use the log of the number of comments to represent engagement. We use the week users first commented as the time they entered the community.

Movie quality. We used the average movie rating to approximate the movie quality. This approach or variations of this approach are adopted by many movie-rating websites and are consistent with our intuitions. We used the average rating as a control variable when comparing sentiment across comments since the ratings are untouched when fine-tuning BERT.

2.4 Model

Ordinary Least Squares (OLS) estimation. Other than checking the sentiment trend at the community level and user level with visualizations, we also formally evaluate the trend by examining how the sentiment in comments can be predicted by the comment time and user engagement with OLS regressions. For predicting the sentiment in comments at the community level, the model is formulated as

$$\text{comment sentiment} = \alpha + \beta \cdot \text{commented week} + \gamma X_{\text{mov}} + \varepsilon,$$

where comment sentiment is obtained using our fine-tuned BERT model and encoded as −1, 0, or 1 to represent negative, neutral, or positive sentiment; commented week is the week the comment was posted to the short comments section, and the week Douban

* Matthews correlation coefficient could be used as a measure of the quality of classification even if the dataset is unbalanced. It ranges from −1 to 1, where 1 means perfect prediction. See https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html for details.

‡ This is transcribed from their paper. However, the paper contained an error and wrongly described the polarity. Here we corrected this error.

Movie was established is encoded as 1 to represent the first week; X_{mov} is a vector of control variables that relate to movies, including the movie genre, movie rating, movie region, and movie release year; α, β, and γ are parameters to be estimated; and ε is the error term. To support our first hypothesis, the coefficient for the commented week is expected to be significantly negative.

For predicting users' sentiment polarity score as a function of their community engagement, the model is formulated as

$$\text{sentiment polarization} = \alpha + \beta \cdot \log(\text{number of comments}) + \gamma X_{\text{usr,mov}} + \varepsilon,$$

where sentiment polarization is calculated using the pre-defined metric; log (number of comments) is employed given considerations of the great variations of users and users' insensibilty to comments as the number of comments increases; and $X_{\text{usr,mov}}$ is a vector of control variables that relate to users and movies, including the enter week, which represents the first time a user gave a comment, and average movie rating, which represents the quality of the movies a user consume. ε is the error term. To support our second hypothesis, the coefficient for the log (number of comments) is expected to be significantly negative.

Apart from formal regression models, we also have supplementary visual analysis to support our conclusions. For brevity, we describe how we implement supportive analysis in the results section along with the findings[♯].

3 Result

Overall trend. The average sentiment consistently went down from 2005 to 2015, then went up, similar to the mean rating change (see Fig. 2). This indicates that movie quality may be a significant contributor to sentiment at the community level. We also visualized the detailed distribution of the trend for different sentiments and the popular and less popular movies. See Appendix D for details.

Further controlling movie-related variables, including movie rating, for regression, we found the community sentiment did not go negative over time. The coefficient for the commented week is slightly over 0 when controlling additional variables. Instead, movie rating, which could be considered the quality of movies, dominates the sentiment change in comments, and a higher movie rating corresponds to a higher mean sentiment score. See Table A1 in Appendix B for details of the regression.

Seeing the increase in sentiment echoes the increase in comments and appeared users, we also tested how the increase in sentiment related to new users. We use the proportion of comments from new users in a week to measure the influx of new users. As Fig. 3 shows, mean sentiment by week is highly correlated with the proportion of comments from new users by week. Furthermore, starting from 2015, although there are fluctuations, there is a trend of having a higher proportion of comments from new users in a week. See Table A2 in Appendix B for details of the regression.

We also checked the ratings of movies released after 2005 to see if the rating trend corresponds to the sentiment trend, as newer movies, like box movies, tend to be popular. For example, suppose a movie was released in 2018. In that case, we use all comments under this movie to calculate the average rating of this movie as an approximation of the quality of this movie. This movie is considered a movie from 2018. As Fig. 4 shows, the movie rating kept decreasing even after 2014. Hence the increasing trend after 2014 was less likely to be caused by the increase in movie quality.

Difference between active and non-active users. Comments of active users tend to have lower sentiment scores than other users (see Fig. 5). Here, active users are the top 5% of users. In our dataset, the threshold is 22 comments. According to this criterion, there are 32 464 active users, and they contributed over 2.7 million comments. As our dataset is incomplete, users who have less than 22 records in the dataset may have more than 22 comments on Douban Movie, so we set users who have 2−21 comments as less active users and users who only have 1 comment as the least active users to make an ordinal comparison to compensate for the deficiency of dataset. Using this approach, we compared the sentiment disparity between active, less active, and the least active users. We find that although, in some years, comments from active users were more positive than those of less active users, they were always lower than comments from the least active users.

♯ The analysis scripts and models are provided at: https://github.com/Hongkai040/Negative_Sentiment_Shift_on_a_Chinese_Movie-Rating_Website.

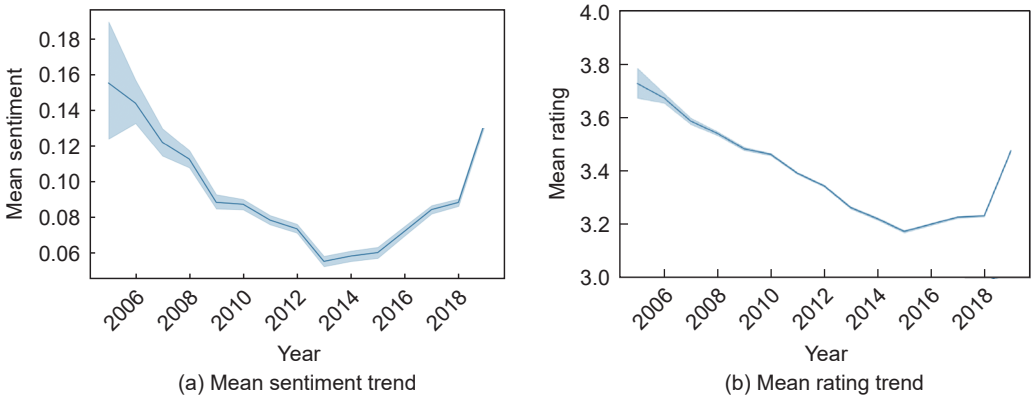


Fig. 2 Mean sentiment trend and mean rating trend. The mean sentiment trend reached its valley in 2013 and then went up; the mean rating reached its valley in 2015 and then bounced up. Otherwise specified, all error bars in this figure and the following figures represent a 95% Confidence Interval (CI).

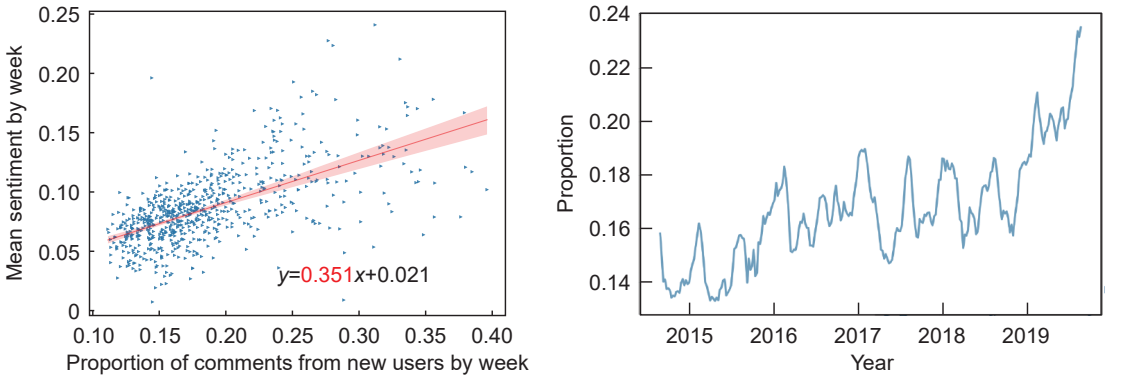


Fig. 3 Regression of mean sentiment and proportion of comments from new users. A higher proportion of new users' comments correlates with a higher mean sentiment in the community, and we also see an increased influx of new users into the community overtime staring from August 2014.

Sentiment polarization of users. Users of higher involvement are more likely to shift to the negative side according to our regression result (see Fig. 6). Users with more comments are likely to have a lower sentiment polarization score. We also regressed the average movie rating as a function of watching sequences for users. The regression result shows that movies watched later tend to have higher ratings, indicating that sentiment polarization is more likely out of users' inner motivation rather than a decrease in movie quality. The other two OLS models also showed that the average movie rating and the time at which a user enters the community do not influence the trend. See Table A3 in Appendix B for details.

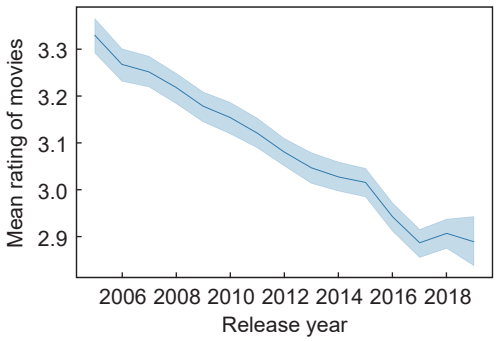


Fig. 4 Mean rating of movies by year. The mean rating of movies keeps decreasing, even after 2015. Although there was a slight increase between 2017 and 2018, the rating later decreased again.

[b] Note that for some weeks in 2005 there is a very high proportion of new users because that is the establishment year of Douban Movie. Most comments left at that time were usually from new users by definition, which does not align with our purpose of regression. Hence, the regression only considers weeks that have less than 40% comments from new users to filter out the weeks greatly influenced by the establishment of Douban Movie. The figure omitted several outlier data points to improve the readability, but those data points are considered by the regression model.

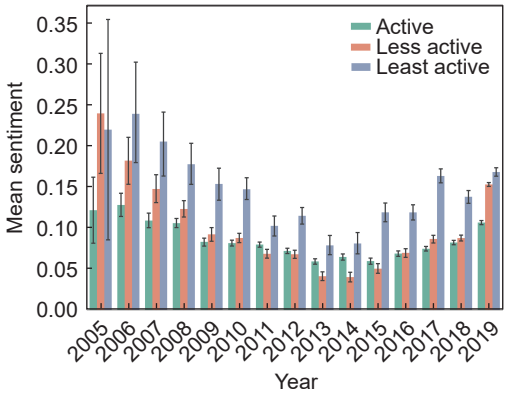


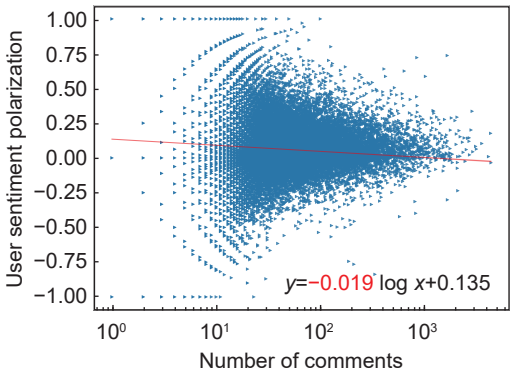
Fig. 5 Mean sentiment of comments by users with varying levels of engagement. Before 2008, the number of comments each year was relatively small. So, we can see a wide confidence interval. The interval shrank as the number of comments increased.

4 Discussion

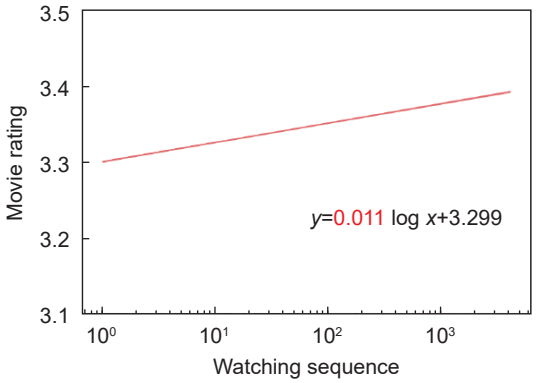
In this section, first, we report the sentiment trend on Douban Movie and analyze how the reversed positive trend after 2015 may relate to the release of Douban app, and the user-level finding is that users with higher engagement tend to be more negative; second, we compare this work to similar studies, elucidating its contribution; third, we discuss how group identity might have functioned behind it on a movie-rating website; fourth, we conclude how these findings can help us focus on and motivate those negative users; finally, we discuss in what ways the study is limited and can be improved in future work.

At the community level, while we can see a shift toward negativity on Douban Movie before 2015, the trend was negligible after controlling movie-related variable and even reversed after 2015. We believe that it might relate to Douban Movie's transformation from a web-based community to a mobile app, which brought new users into the community. In late 2014, Douban, the company that owns Douban Movie, released its mobile app so that users of Douban Movie can easily give comments on their mobile devices, share their thoughts about movies, books, and songs under corresponding sections, or join groups and talk to people with the same interests or from the same places. At the same time, our analysis shows that regardless of the decrease in movie quality, the shift toward positivity highly correlates with the increase of new users after 2015. Hence, a possible and reasonable explanation is that the release of the Douban app successfully brought many new users and positivity into the community.

At the user level, considering some temporal factors like the sequence of movies they watched, our analysis shows that active or highly engaged users tend to be more negative, suggesting a fitting-in process. Some may argue that users tend to watch good movies or movies they prefer more at first, then watch less good or less preferred movies later, and that is why users become negative over time. When this may be true for some users, our result that users tended to watch better



(a) Sentiment polarization as a function of the number of comments for all users



(b) Movie rating as a function of the user's watching sequence. Specifically, the regression considers the rating of the first, second, ..., and last movies a userwatched to see whether they tend to watch better or worse movies over time.

Fig. 6 Sentiment polarization score and the regression of movie rating against movie watching sequence. Users who have made a larger number of comments correlate with a lower sentiment polarization score, and we can also see that the watching sequence positively correlates with the movie ratings.

quality movies later suggests that this is not a general pattern. What is more, since our definition and operationalization of user engagement involve a temporal process, it is for sure that a proportion of the non-active users are future active users who just joined the community. So, it is possible for us to identify that users in their later community life stage are, in general, more negative than users who just entered the community. In fact, we did find such a pattern. Defining users in the first stage when they are giving their first 50% of comments and the second stage when giving the last 50% of comments, we found that comments from users in the second stage are more negative than those from users in the first stage. See Appendix C for visualizations of how comments change during users' community life.

This study aligns with previous similar studies and offers new insights into the relationship between user engagement and negativity in the digital age. Previous studies such as the users' emotional behaviors on Facebook or BBC forums reveal that active users tend to be more negative and predominantly contribute to the negativity of the community[11, 12, 19]. However, such previous studies emphasized the role of interactions among users in controversial topic discussions where misinformation or echo chambers may play a role. Nevertheless, our results show that, in an online environment that lacks such interactions, active users may still show a tendency toward negativity. What is more, this study also expands such negativity analysis to non-English contexts, where how users emotionally behave remains less developed and sometimes may be controversial[20, 21]. Hence, this study also helps provide a generally accountable user engagement and negativity pattern that transcend language differences.

According to the results, the driving force of the users' sentiment shift may be group identity mixed with a desire of being perceived as intelligent. As Wallace[22] claimed in her book, online settings might help form group polarization as group members can easily get a sense of being surrounded by like-minded people. Such a feeling of belongingness can be captured by user participation. Therefore, the more active a user behaves in a community, the more likely the user becomes negative if the community does not have a favor of positivity. What is more, a previous study showed that, while negative evaluators are less likable as positive evaluators, they are perceived as more intelligent[23]. Hence, in the context of an online movie-rating website, user engagement in the community may shape users' behaviors in a negative direction because users want to be perceived as intelligent, and negative evaluators are not necessarily less likable for the sake of group identity. Thus, being a negative evaluator may prevail in the community. In other words, the reason why active users of Douban Movie were more negative compared to less active users might be that they have a stronger feeling about group identity through frequent participation, during which their negative comments displaying their intelligence are not necessarily disliked.

This study provides implications on designs for online communities to tackle the negativity issue. When many online communities have similar mechanisms, like spam detectors, auto-moderators, etc., to deal with toxic and malicious content in groups, less straightforward efforts are made with regard to this aspect. However, our findings, along with previous studies, suggest that users can have a tendency toward negativity. Hence, if communities prioritize content produced by old active users for their high credibility compared to new less active users, they may encourage tacit agreement among users that negative evaluators could be likable and thereby exacerbate the negativity trend among active users. Hopefully, this research can bring attention to the negativity back to the table in the digital age and inspire methods like motivating users showing a tendency toward negative to become more positive.

This study has several limitations. First, sentiment labeling simplifies the information conveyed by comments. Users may express various emotions in comments. This study only makes use of the polarity of sentiments. A finer-granularity analysis could be done by analyzing various emotions and incorporating sentiment intensity. Second, the results could be influenced by fake comments. It is not unusual to see that users deliberately give abnormally positive or negative comments. Hence those spoilers could influence our results. However, since commercial benefits mostly drive such behaviors, comments are

more likely to be highly positive, and we would expect that the actual negativity may be underestimated. Finally, the current findings could not entirely exclude algorithm confounding issues. Our dataset is a sample of all comments on the Douban Movie, and the ranking algorithm makes it a non-random one. However, the algorithm is less likely to favor selecting more negative comments over more positive comments for displaying due to reasons like maintaining good community culture. In fact, we test the trend of sentiment with decreasing rank of the comments and do not find evidence to support a claim that the ranking algorithm specifically favors positive or negative comments[∷]. Hence, our conclusion is unlikely to be distorted by the ranking algorithm. What is more, Douban Movie claimed that the algorithm would fold comments containing personal attacks. Hence, the negativity trend on Douban Movie may be underestimated. Nevertheless, using a more representative dataset for analysis would be appreciated.

Some possible improvements and future work can be made. First, more factors should be taken into consideration to better understand sentiment dynamics in online communities. In particular, it would be helpful to consider the influence of upvotes on comments, the impact of network topology on users' sentiment, the influence of moderators, etc. Second, future work can focus on the diffusion process of negativity among users by identifying strategies users employ to show negativity and how they acquire those strategies. A process of gaining group identity might also be a process of learning. Users might learn how to critically comment on a movie by imitating others. It might be out of a strategy of gaining attention from others, and active users might be better at winning attention by giving critically negative comments. Third, future work can scale up to larger datasets or multiple data sources. If we access users' full comments history, we can better analyze the interplay between users and the community. Moreover, as Douban Movie is only a part of the Douban community, the sentiment dynamics of the Douban community could be better understood using multiple datasets from Douban.

5 Conclusion

This study expands the boundary of large-scale sentiment analysis in online communities. This study finds active users' tendency to post negative comments in a Chinese online community, echoing research conducted in English-based communities. Unlike research focused on emotion contagion related to misinformation and echo chambers, this study showed that a similar process of getting sentimentally polarized could also form in a community less affected by fake news and information isolation. These findings suggest that a possible mechanism of group identity drives these behaviors.

It is our hope that although this study starts with focusing on negativity, it, along with future work, can help explain the mechanisms behind it and provide guidance for promoting positivity and healthy communities in the digital age.

Appendix

A Model Fine-Tuning

The BERT model is fine-tuned using 4 epochs with trainer Application Programming Interface (API) provided by Huggingface[☼]. Since most comments are short, we use a max length of 256 to fine-tune BERT. The training set contains 1200 comments, and both the validation and test set have 400 comments. Validating the fine-tuned BERT on the test set yields the confusion matrix in Fig. A1, confirming good alignment between the model and human labelers.

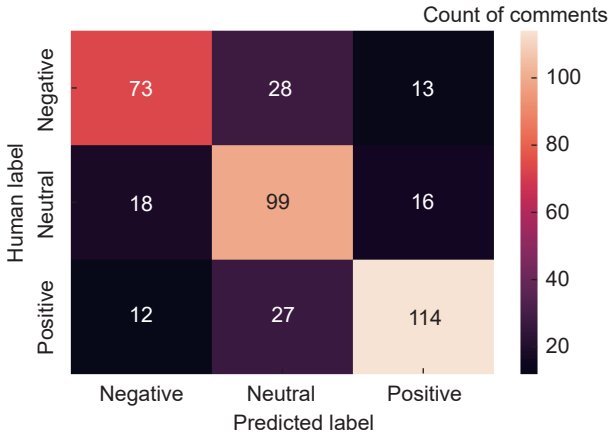


Fig. A1 Fine-tuned BERT confusion matrix on the test set. In each cell, the number represents the count of comments that was classified by the human checkers and the model. Larger numbers/lighter colors on the diagonal and smaller numbers/darker colors off the diagonal mean good alignment between the model and human labelers.

[∷] See Table A4 in Appendix B for the regression details.

[☼] https://huggingface.co

B Tables of Regression Results

Table A1 reports the results of multiple OLS regressions for the variable comment sentiment on the independent variable, comment week, and various control variables, including movie release year, movie rating, movie genres, and movie regions. In Table A1, observations are comments. Note that in the third model, 10% data points randomly were sampled for memory efficiency and categories in movie genre and movie region are not listed for brevity. Otherwise specified, coefficients in this table and the following tables are rounded to three significant digits.

Table A2 reports the results of OLS regression for the variable mean sentiment on the independent variable, comments from new users proportion. In Table A2, observations are weeks.

Table A3 reports the results of multiple OLS regressions for the variable sentiment polarization on the independent variable, log (number of comments), and various control variables, including average movie rating and enter time. In Table A3, observations are users.

Table A4 reports the results of OLS regression for the variable comment sentiment on the independent

Table A1 OLS for comment sentiment.

Variable	Dependent variable: comment sentiment		
	Model 1	Model 2	Model 3
Comment week	−0.007*** (−0.008, −0.007)	0.005*** (0.004, 0.004)	0.006*** (0.005, 0.008)
Movie release year	–	0.000 (0.000, 0.000)	0.000*** (0.000, 0.000)
Movie rating	–	0.288*** (0.286, 0.289)	0.301*** (0.295, 0.307)
Movie genre	–	–	(omitted)
Movie region	–	–	(omitted)
Constant	14.472*** (13.443, 15.500)	−9.859*** (−10.867, −8.850)	−13.186*** (−16.300, −10.073)
Observation	2 064 925	2 064 925	206 205
R^2	0.000	0.057	0.061
Adjusted R^2	0.000	0.057	0.060
Residual standard error	0.787	0.764	0.763
F statistic	752.993***	41 434.738***	66.942***

Note: ***$p<0.01$.

Table A2 OLS for mean sentiment as a function of comments from new users proportion.

Variable	Dependent variable: mean sentiment
Comments from new users proportion	0.351*** (0.316, 0.386)
Constant	0.021*** (0.014, 0.027)
Observation	699
R^2	0.354
Adjusted R^2	0.353
Residual standard error	0.026
F statistic	382.098***

Note: ***$p<0.01$.

Table A3 OLS for sentiment polarization.

Variable	Dependent variable: sentiment polarization		
	Model 1	Model 2	Model 3
log (number of comments)	−0.019*** (−0.021, −0.018)	−0.029*** (−0.030, −0.028)	−0.026*** (−0.027, −0.024)
Average movie rating	–	0.267 (0.265, 0.269)	0.266*** (0.264, 0.268)
Enter time	–	–	0.000*** (0.000, 0.000)
Constant	0.135*** (0.133, 0.137)	−0.729*** (−0.736, −0.721)	−0.766*** (−0.775, −0.758)
Observation	628 832	628 832	628 832
R^2	0.001	0.085	0.085
Adjusted R^2	0.001	0.085	0.085
Residual standard error	0.630	0.603	0.603
F statistic	704.390***	29 095.183***	19 488.684***

Note: ***$p<0.01$.

Table A4 OLS for comment sentiment as a function of the decreasing rank of the comments.

Variable	Dependent variable: comment sentiment
Rank	0.000*** (0.000, 0.000)
Constant	0.082*** (0.080, 0.083)
Observation	4 428 395
R^2	0.000
Adjusted R^2	0.000
Residual standard error	0.788
F statistic	305.680***

Note: ***$p<0.01$.

variable, rank, which refers to the decreasing rank of the comments under movies' short comments entry. In Table A4, observations are comments.

C Change Within Users' Life Cycle

We explored the time dependency of the mean sentiment of users. We defined that users are in the first stage when they are giving their first 50% of comments, and the second stage when they are giving the last 50% of comments. Then we calculate the average sentiment of comments from users that are in their first stages and the average sentiment of comments from users that are in their second stages by year. We find the average sentiment of the second stage tends to be lower than that of the first stage, considering a 95% CI (Fig. A2), and users tend to be less positive in the second stage.

The effect of being negative may even be underestimated by the mean method. For example, according to the definition, the last 20 comments for user A having 40 comments in our dataset are comments from the second stage, and the last 200 comments for user B having 400 comments in our dataset are also from the second stage. However, if users do become negative over time, we would expect that user B may already be very negative when giving the 41st comment. So, the meaning of checking sentiment change is not telling the magnitude of sentiment change but telling the time dependency of the sentiment of comments, and our results are robust and consistent.

D Detailed Distribution

As shown in Fig. A3, in the sentiment distribution, we can see an increase of negative comments, but a decrease of positive and neutral comments from 2005 to 2015. The trend reversed after 2015. The rating distribution is consistent with the sentiment distribution, where a reversal in 2015 can also be observed. Additionally, the rating distribution indicates a trend of reducing missing ratings. Finally, the mean sentiment trend between popular and less popular movies is slightly different yet consistent as well.

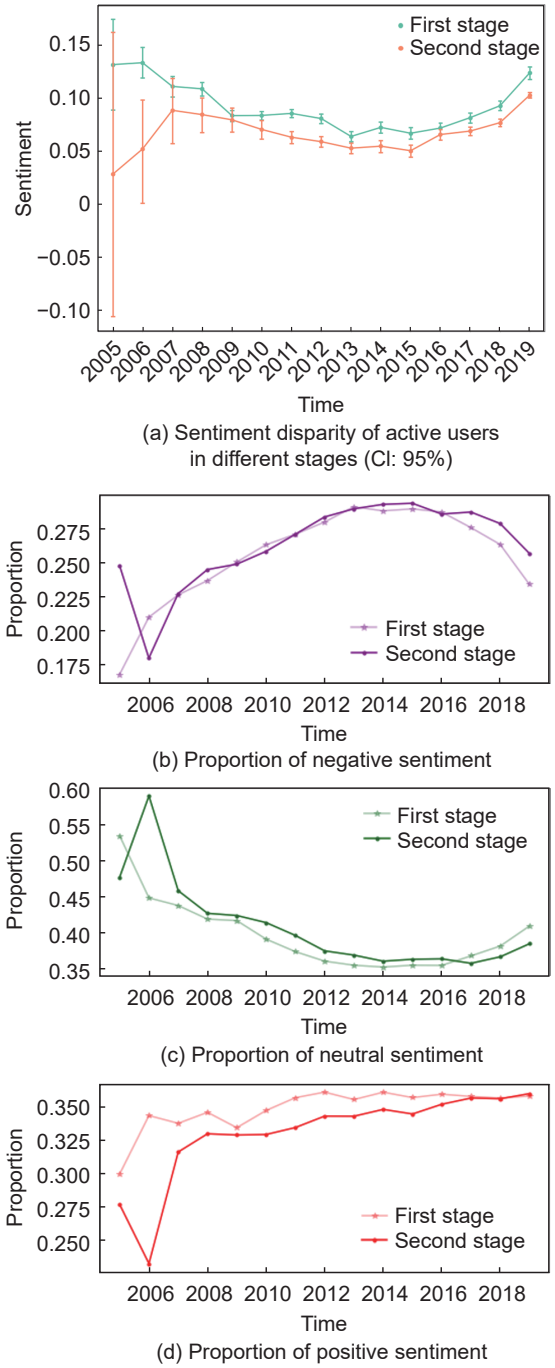
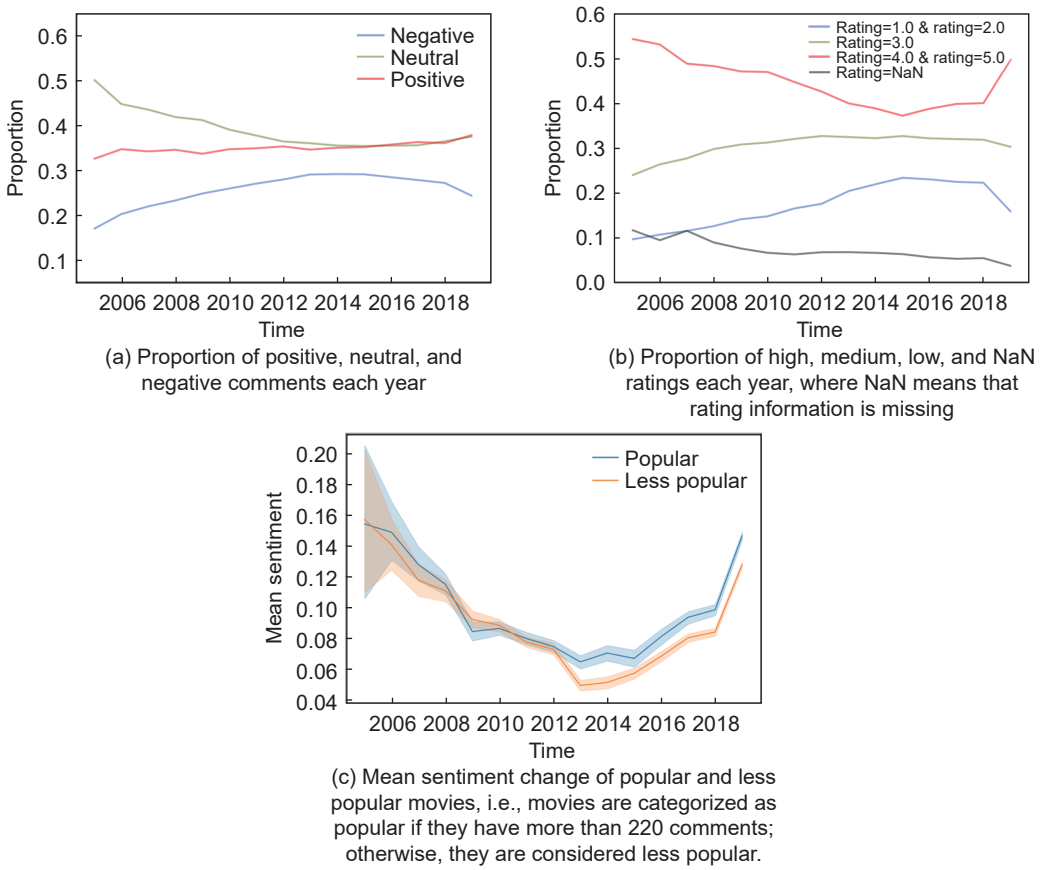


Fig. A2 Sentiment disparity of active users in different stages. Active users' second stage comments always have a lower mean sentiment score.

Acknowledgment

The author would like to sincerely thank Professor Sanja Miklin for providing many constructive comments on an earlier draft, Professor Zhao Wang and James Evans for their valuable suggestions on methodological improvements, Stephen Parkin for refining the language, and anonymous reviewers for their constructive feedback. The author would also like to thank Hongding Zhu, Henry Lin, Rui Pan, and Peihan Gao for their passionate discussions and encouragement and the developers who generously shared the dataset.



(a) Proportion of positive, neutral, and negative comments each year

(b) Proportion of high, medium, low, and NaN ratings each year, where NaN means that rating information is missing

(c) Mean sentiment change of popular and less popular movies, i.e., movies are categorized as popular if they have more than 220 comments; otherwise, they are considered less popular.

Fig. A3 Comments' sentiment and rating distribution. Despite minor differences, the distribution of comment sentiments is similar to comment ratings. Furthermore, while popular movies tend to have higher ratings, there is no distinguishable rating difference between popular and less popular movies.

References

[1] S. Tsugawa and H. Ohsaki, Negative messages spread rapidly and widely on social media, in *Proc. 2015 ACM Conf. Online Social Networks*, Palo Alto, CA, USA, 2015, pp. 151–160.

[2] J. A. Fine and M. F. Hunt, Negativity and elite message diffusion on social media, *Political Behav.*, doi: https://doi.org/10.1007/s11109-021-09740-8.

[3] L. Teng, D. Liu, and J. Luo, Explicating user negative behavior toward social media: An exploratory examination based on stressor–strain–outcome model, *Cogn. Technol. Work.*, vol. 24, no. 1, pp. 183–194, 2022.

[4] X. Zhang, X. Ding, and L. Ma, The influences of information overload and social overload on intention to switch in social media, *Behav. Inf. Technol.*, vol. 41, no. 2, pp. 228–241, 2022.

[5] L. Munn, Angry by design: Toxic communication and technical architectures, *Humanit. Soc. Sci. Commun.*, vol. 7, no. 1, pp. 1–11, 2020.

[6] E. Omernick and S. O. Sood, The impact of anonymity in online communities, in *Proc. 2013 Int. Conf. Social Computing*, Alexandria, VA, USA, 2014, pp. 526–535.

[7] R. F. Baumeister, E. Bratslavsky, C. Finkenauer, and K. D. Vohs, Bad is stronger than good, *Rev. Gen. Psychol.*, vol. 5, no. 4, pp. 323–370, 2001.

[8] C. R. Sunstein, The law of group polarization, https://chicagounbound.uchicago.edu/law_and_economics/542/, 1999.

[9] J. K. Lee, J. Choi, C. Kim, and Y. Kim, Social media, network heterogeneity, and opinion polarization, *J. Commun.*, vol. 64, no. 4, pp. 702–722, 2014.

[10] S. Yardi and D. Boyd, Dynamic debates: An analysis of group polarization over time on twitter, *Bull. Sci. Technol. Soc.*, vol. 30, no. 5, pp. 316–327, 2010.

[11] F. Zollo, P. K. Novak, M. D. Vicario, A. Bessi, I. Mozetič, A. Scala, G. Caldarelli, and W. Quattrociocchi, Emotional dynamics in the age of misinformation, *PloS One*, vol. 10, no. 9, p. e0138740, 2015.

[12] M. D. Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on facebook, *Sci. Rep.*, vol. 6, no. 1, p. 37825, 2016.

[13] A. Abisheva, D. Garcia, and F. Schweitzer, When the filter bubble bursts: Collective evaluation dynamics in online communities, in *Proc. 8th ACM Conf. Web Science*, Hannover, Germany, 2016, pp. 307–308.

[14] J. Buder, L. Rabl, M. Feiks, M. Badermann, and G. Zurstiege, Does negatively toned language use on social media lead to attitude polarization? *Comput. Hum. Behav.*, vol. 116, p. 106663, 2021.

[15] M. A. Cacciatore, D. A. Scheufele, and S. Iyengar, The

end of framing as we know it … and the future of media effects, *Mass Commun. Soc.*, vol. 19, no. 1, pp. 7–23, 2016.

[16] A. Bessi, F. Petroni, M. D. Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, and W. Quattrociocchi, Viral misinformation: The role of homophily and polarization, in *Proc. 24th Int. Conf. World Wide Web*, Florence, Italy, 2015, pp. 355–356.

[17] J. Devlin, M. -W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv: 1810.04805, 2018.

[18] A. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, Learning word vectors for sentiment analysis, in *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, USA, 2011, pp. 142–150.

Hongkai Mao received the BS degree from Zhejiang University, China in 2021. He is currently pursuing the master degree in the Computational Social Science Program at the Social Sciences Division, University of Chicago. His research interests include network analysis, computational social science, and online toxicity and group polarization.

[19] A. Chmiel, P. Sobkowicz, J. Sienkiewicz, G. Paltoglou, K. Buckley, M. Thelwall, and J. A. Hołyst, Negative emotions boost user activity at BBC forum, *Phys. A*, vol. 390, no. 16, pp. 2936–2944, 2011.

[20] Z. Yang and W. Xu, Who post more negatively on social media? A large-scale sentiment analysis of Weibo users, *Current Psychology*, doi: 10.1007/s12144-022-03616-8.

[21] Q. Gao, F. Abel, G. -J. Houben, and Y. Yu, A comparative study of users' microblogging behavior on Sina Weibo and Twitter, in *Proc. 20th Int. Conf. User Modeling, Adaptation, and Personalization*, Montreal, Canada, 2012, pp. 88–101.

[22] P. Wallace, *The Psychology of the Internet*. Cambridge, UK: Cambridge University Press, 2015.

[23] T. M. Amabile, Brilliant but cruel: Perceptions of negative evaluators, *J. Exp. Soc. Psychol.*, vol. 19, no. 2, pp. 146–156, 1983.