Dec 12th, 12:00 AM

Social Media Sentiment Contagion

Hyungsoo Lim
Hong Kong University of Science and Technology, hyungsoolim@ust.hk

Jing Wang
Hong Kong University of Science and Technology, jwang@ust.hk

Allen Huang
HKUST, allen.huang@ust.hk

Social Media Sentiment Contagion

Completed Research Paper

Hyungsoo Lim
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
hyungsoolim@ust.hk

Jing Wang
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
jwang@ust.hk

Allen Huang
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
allen.huang@ust.hk

Abstract

We propose an empirical setting to discover sentiment contagion in social media. We find that, after controlling for concurrent events, sentiment contagion exists in social media. We conduct additional analyses to explore how the source and valence of exposure contents and individual heterogeneity affect the degree of sentiment contagion. We find robust evidence of sentiment contagion not only in contents under the same thread but also under different threads of the same forum. Additional analysis provides evidence of negativity bias. In terms of individual heterogeneity, we find that more experienced social media users are less sensitive to sentiments in social media. Last, we find that social media users are more likely to become inactive in the long run after being exposed to more negative contents. Managerial and practical implications are discussed.

Keywords: Sentiment contagion, Individual Heterogeneity, Negativity Bias, Social Media

Introduction

The number of active social media users is rapidly growing to 4.2 billion users around the world. This number has grown by 490 million over the past year and is now equivalent to more than 53 percent of the world's total population (Statista 2021). According to data published by Kepios Analysis in January 2022, the amount of time an average person spends on social media is close to two and a half hours per day. Along with the enormous usage of social media, social media has also become a ubiquitous platform for people to interact with strangers and acquaintances, which is very different from traditional social networks where people are likely influenced by their friends, families, or co-workers (Kane et al. 2014, Kaplan and Haenlein 2010). Social media users are interact with people outside of their networks in a variety of ways, for instance, exchanging and spreading news (Vosoughi et al. 2018), collective sense making (Oh et al. 2015), and giving and receiving support (Yan et al. 2015). Anecdotally, Frances Haugen, a former employee of Facebook, recently revealed that Facebook's algorithms aiming to improve user engagement leads to the amplification of hateful, divisive, and polarizing contents (Mac and Kang 2021). According to survey results, 58% of Americans say that social media negatively affects their mental health such as distress, anxiety, and Fear of Missing Out (FOMO) according to Onlinetherapy. [1] Recent academic studies based on randomized experiments also show that deactivating social media account can improve subjective well-being (Allcott et al. 2020).

[1] https://www.onlinetherapy.com/6-in-10-americans-say-social-media-negatively-affects-their-mental-health/

Sentiment contagion in social media has gained increasing attention in the literature. Sentiment contagion refers to a process in which the sentiments of a perceiver become more like the sentiments that he/she is exposed to (Hatfield et al. 1994; the authors used the term "emotion contagion"). The sentiment contagion plays a major role in spreading sentiments (Goldenberg et al. 2020), which could lead to several online social movements such as moral outrage (Crockett 2017) and protests (Van Zomeren et al. 2012).

However, there is scant evidence on the immediate sentiment contagion effect of social media. That is, does exposure to sentiments on social media affects users' sentiments? This question has public policy implications. If social media messages' sentiment can affect users' sentiment, then the mitigation strategy should be access time restriction (e.g., conducted by the Government of the Republic of Indonesia)[2] instead of content censorship—the approach adopted by most social media so far.

Despite the importance of the sentiment contagion, there are several challenges present in the empirical investigation of sentiment contagion due to the nature of data that researchers are able to access.[3] To properly assess the effect, researchers need to 1) measure the emotion of a focal user (perceiver), 2) define the treatment contents likely to be exposed to the focal user, and 3) control for the concurrent events that may affect the focal user's emotion. Furthermore, complex dynamics of online communication makes it much more challenging to explore the sentiment contagion, which could involve the consideration of the history of the focal user (e.g., latest emotion of the focal user, social media experience).

Existing research on sentiment contagion is subject to several limitations. In terms of coverage, the literature focuses on limited types of social media activities. For instance, empirical studies using Facebook only cover the posts (also refer to "status change") without considering comments under the posts (Coviello et al. 2014, Kramer et al. 2014). Moreover, prior literature mostly provides except Ferrara and Yang (2015) correlational evidence of sentiment contagion instead of a causal link. It is challenging to tease out potential confounding factors and establish causality in empirical settings. Lastly, prior literature cannot incorporate individual heterogeneity as the unit of analysis is not at the individual level.

The goal of this study is to propose a novel identification strategy to examine the sentiment contagion and provide empirical evidence of sentiment contagion in social media. By controlling for concurrent events and interdependency of user's past behavior (e.g., the sentiment expressed in the latest content, the location where the focal user generated the latest content), we establish stronger causal evidence of the sentiment contagion. Furthermore, we scrutinize whether and to what degree the features of social media users (e.g., individual heterogeneity) can strengthen or weaken the sentiment contagion since little literature has studied its impact. We also investigate the negativity bias on the sentiment contagion address the question whether the sentiment affects the user's long-term activity.

We use English Premier League (EPL) soccer community data from Reddit, which include over 3.6 million pieces of contents from 20 EPL subreddits in the 2018/19 season. Using a deep learning algorithm to measure the sentiment of the contents, we find robust evidence of sentiment contagion.

Next, we explore whether the source of the content and users' individual heterogeneity can influence the magnitude of sentiment contagion. We find that social media users are likely to be affected by the contents in a thread where they are currently involved rather than the ones in a thread where they were previously involved. We also find evidence of sentiment contagion both in the same thread of the users' previous content, in different threads, and even in different subreddits. In terms of individual heterogeneity, we find that more experienced social media users, i.e., those who have generated more contents in the past, are less sensitive to the treatment sentiment, likely because they have had more exposure to social media contents. Furthermore, we discover that a negativity bias in the sentiment contagion, i.e., negative contents have stronger effects on users' sentiments. Our results also show that users are less likely to be active in the long run during the next two years after exposure to negative sentiment.

Our paper makes several important contributions. First, we contribute to the literature on the sentiment contagion by proposing a novel identification strategy to better examine the causal effect of the sentiment contagion. Second, this study enriches the vast literature on sentiment contagion by shedding light on its

[2] The Government of the Republic of Indonesia restricted the public from accessing social media due to rapid increase of false news before 2019 election. This action could be applied by social media platform due to emotion and sentiment of social media users. Sometimes, social media users tried to restrict accessing social media.
[3] See Goldenberg and Gross (2020) for more details.

heterogeneous impact with various treatment contents and individual heterogeneity. Third, this paper extends our knowledge of negativity bias by providing empirical evidence in the context of sentiment contagion.

Literature Review

This study is related to two streams of research. The first related stream focuses on the emotion contagion proposed by Hatfield et al. (1994). Emotion contagion is defined as "the process by which the emotions of a perceiver become more similar to those of others as a result of exposure to these emotions" (Goldenberg and Gross 2020, p. 317; which is known under various names, such as emotion transfer, emotion contagion, and emotional contagion; in this study, we refer to it as 'sentiment contagion').

Several underlying mechanisms may explain sentiment contagion: 1) mimicry, 2) category activation, and 3) social appraisal (see Goldenberg and Gross (2020) for details). The first, mimicry, refers to the fact that emotional expression activates synchronous behaviors which could be either nonverbal (e.g., facial expression, body postures; Hatfield et al. 1994, Prochazkova and Kret 2017, Hess 2021) or verbal (e.g., writing in the context of text-based community; Chartrand and Van Baaren 2009). Recent studies show that offensive speech online is contagious through mimicry (Kwon and Gruzd 2017, Song et al. 2022). The second, category activation, is that exposure to people primes specific emotional categories (e.g., Peters and Kashima 2015, Niedenthal et al. 2009), which leads to the activation of a specific emotion. Category activation is different from mimicry in that it can occur merely with exposed texts in the absence of imitating a behavior. The third, social appraisal, means that people tend to use others' emotions as a guide for their own emotion appraisals (Manstead and Fischer 2001, Parkinson and Simons 2009, Parkinson 2020, Bruder et al. 2014), thus generating similar responses to a specific event. These three mechanisms can simultaneously contribute to sentiment contagion (e.g., Wróbel and Imbir 2019).

Prior literature investigates sentiment contagion in social networks with various treatments. Some studies show the existence of sentiment contagion induced by other people's posts on the focal user's post (Coviello et al. 2014, Ferrara and Yang 2015, Kramer et al. 2014). Online social media users are easily emotionally synchronized by contents related to an external event (e.g., rainfall in other cities) even if they are not experiencing it (Coviello et al. 2014). Ferrara and Yang (2015) trace the posts from authors' friend list before one hour of new content and find sentiment contagion at large group level (e.g., null, positive, neutral, and negative group). Prior research in this stream has also found evidence of sentiment contagion of a post to its corresponding comments (Stieglitz and Dang-Xuan 2012, Kwon and Gruzd 2017, Dang-Xuan and Stieglitz 2012). In addition, Kwon and Gruzd (2017) show that offensive commenting and swearing are contagious to corresponding comments.

The second stream of research investigates the factors that can strengthen or weaken the sentiment contagion (see Goldenberg and Gross 2020 for more details). The first factor is the contents that users see. Prior studies suggest that negative sentiments lead to stronger sentiment contagion (e.g., Kramer et al. 2014, Cacioppo et al. 2014, Soroka et al. 2019). The second factor is a relation between the expressers and the focal user. It is reported that strong ties lead to stronger sentiment contagion (Lin and Utz 2015, Zeng and Zhu 2019). Individual characteristic of the perceiver is also an important factor but received little attention from researchers. Recently, a study shows that personality traits of users (e.g., Conscientiousness, Agreeableness) can affect the degree of the sentiment contagion (Gruda et al. 2022). Prior literature reveals that frequent exposure to emotions can lead to habituation or fatigue, making each exposure to emotional expression online less impactful (Wilson and Gilbert 2008, Crockett 2017). However, it is not clear whether and to what degree frequent exposure can weaken the sentiment contagion.

The Setting

In this section, we describe our empirical data and explain our empirical approach to establish the existence of sentiment contagion on social media.

Empirical Data

Reddit is an American social news aggregation, web content rating, and discussion website. Registered users can submit content such as links and text posts to the site, which are then voted up or down by other users. Posts are organized by subject into user-created boards called "subreddits", which cover a variety of topics such as sports, news, politics, and music. Among various subreddits, we collected contents on subreddits representing 20 teams in the EPL during the 2018/19 season (from August 1, 2018, to May 31, 2019) since EPL is undoubtedly the most popular football league in the world and their fans are notorious for their enthusiasm especially during the league seasons. The dataset consists of 63,988 users who generated at least one content and 3,608,654 contents, which shows that these subreddits are very active. The huge volume of this dataset allows us to gain useful insights into the dynamics of online communications.

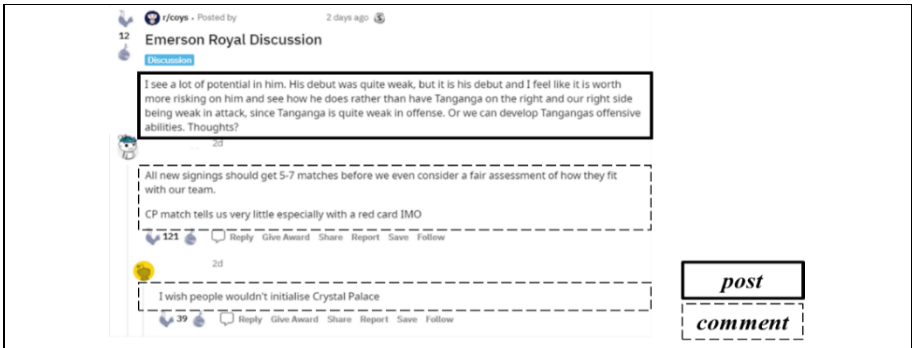

Figure 1. Example of Contents on Subreddit *coys*

Our dataset includes detailed information of given contents (e.g., user id, type, posted time, and subreddit name). Each user is identified by a unique ID so we can trace every content created by the users. Reddit has two types of content: posts or comments. A post refers to a main post and comments are replies to posts or comments created by other users. Figure 1 shows examples on subreddit "coys" representing Tottenham Hotspur. Table 1 summarizes the statistics of the number of posts and comments on 20 teams' subreddits. There are 93,395 posts and 3,515,259 comments in our dataset.

Team	Subreddit	Posts	Comments	Total	(%)
Arsenal	Gunners	17,041	664,397	681,438	18.88%
Bournemouth	AFCBournemouth	331	469	800	0.02%
Brighton & Hove Albion	BrightonHoveAlbion	116	870	986	0.03%
Burnley	Burnley	114	375	489	0.01%
Cardiff City	bluebirds	279	3,443	3,722	0.10%
Chelsea	chelseafc	10,390	390,369	400,759	11.11%
Crystal Palace	crystalpalace	696	7,120	7,816	0.22%
Everton	Everton	2,408	58,490	60,898	1.69%
Fulham	fulhamfc	469	2,882	3,351	0.09%
Huddersfield Town	HuddersfieldTownFC	165	423	588	0.02%
Leicester City	lcfc	1,184	7,742	8,926	0.25%
Liverpool	LiverpoolFC	21,699	889,592	911,291	25.25%
Manchester City	MCFC	4,704	112,519	117,223	3.25%
Manchester United	reddevils	20,496	994,103	1,014,599	28.12%
Newcastle United	NUFC	1,484	45,598	47,082	1.30%
Southampton	SaintsFC	726	13,897	14,623	0.41%
Tottenham Hotspur	coys	8,043	277,906	285,949	7.92%
Watford	Watford_FC	1,081	4,555	5,636	0.16%
West Ham United	Hammers	1,520	38,988	40,508	1.12%
Wolverhampton Wanderers	WWFC	449	1,521	1,970	0.05%
Total		93,395	3,515,259	3,608,654	100%

Table 1. Number of Posts and Comments

Sentiment Analysis

To capture the sentiment, we apply a model specialized in Twitter-specific classification tasks proposed by Barbieri et al. (2020).[4] We perform sentiment analysis on all the contents in our dataset. The analysis outputs the likelihood of each piece of content being positive, neutral, or negative. We assign each content to the sentiment category with the highest likelihood. The overall sentiments are summarized in Table 2. Nearly half of the contents are neutral (44.26%), followed by negative ones which rank the second (35.15%) and positive ones which rank the lowest (20.59%). It is not surprising that football fans are more likely to release their negative emotions compared to celebrating and cheering up each other. However, the general emotion is not as extreme as we expected.

Sentiment	N	%
Negative	1,268,576	35.15%
Neutral	1,597,019	44.26%
Positive	743,059	20.59%
Total	3,608,654	100.00%

Table 2. Summary of Sentiment

Empirical Approach

We first provide details of our empirical setting. To examine the sentiment contagion, it is necessary to specify the treatment contents that the focal user[5] would read or care about before he/she creates the next content. These treatment contents eventually affect the sentiment of the focal user's next content. As we cannot observe whether the focal user has read a specific piece of content, we need to make several assumptions. First, we assume that the treatment contents that affect the focal user's next content are generated in-between the focal user's current and next content. Second, we assume that the treatment contents exist only in threads that the focal users are involved, that is, source and target threads. Source threads refer to threads where the focal users currently appear. A focal user can get involved in the source thread by either initiating the source thread or leaving a comment in the source thread already initiated by others. Once the focal user is involved in the source thread, he/she will likely pay attention to the comments in the source thread until he/she creates the next content. Target threads refer to threads where the focal users appear next. Like the source thread, the focal user can get involved in the target thread by either initiating the target thread or leaving a comment in the target thread already initiated by others. The focal user may pay attention to the target thread as well because some contents in the target thread may have captured his/her attention before he/she generates the next content. In sum, we regard the contents in these two threads as the treatment contents which might influence the sentiment of the focal user's next content.

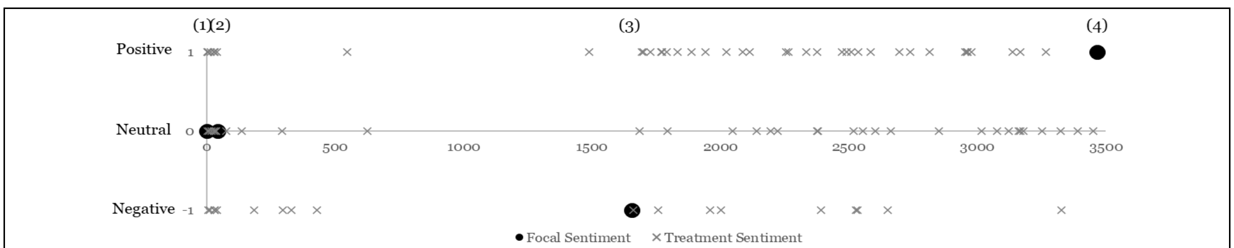

Figure 2. Example of Focal Sentiments with the Sentiment of Treatments over Time

[4] The authors used the Semeval2017 data set for Subtask A Rosenthal et al. 2019, which includes data from previous runs (2013, 2014, 2015, and 2016) of the same SemEval task. The task is that annotators, given a tweet, decide whether it expresses positive, negative, or neutral sentiment. More than 50,000 tweets were annotated by them through Mechanical Turk and ClowdFlower. Among several models they applied (e.g., SVM, LSTM, three variants of RoBERTa), we use one of the variants of RoBERTa model (RoBERTa-base) since RoBERTa showed better performance than SVM and LSTM and similar performance to other variants of RoBERTa show similar performance in sentiment analysis. Accuracy of the sentiment analysis is 71.4±1.9.

[5] The term "focal user" denotes a user of interest in a network (e.g., the subject of this study) and is used to explain the relationship between users in social network studies.

To show a brief example of the empirical approach, we randomly sample a focal user and depict his/her four consecutive contents on football subreddits in Figure. The X-axis represents time[6] and the Y-axis represents the sentiment of a piece of content. The four contents generated by the user are marked as dots and the sentiments are neutral, neutral, negative, and positive for 1st, 2nd, 3rd, and 4th contents, respectively. Treatment contents are marked as stars. As shown in Figure, there are differences in the number and sentiment of treatments across the four focal contents. For example, before the focal user generated 3rd content, there were 10 treatment contents while there were 63 treatment contents for the 4th content. In terms of the sentiment, 40% of the treatment contents are negative for 3rd content whereas 49% of the treatment contents are positive for 4th content.

If the focal user keeps generating contents in the same thread, then the contents in-between that thread are the treatment contents. When the focal user starts a new post (i.e., initiating the target thread) as the next content, the treatment contents are only in the source thread as any content in the target thread will be created after the focal user's next content. If the focal user responds to a target thread with a comment, then the contents in-between in both the source thread and the target thread are the treatment contents.

We define a sentiment index to properly measure the treatment sentiments using contents in both the source and the target thread as "sentiment index of treatment" (*SIT*, hereafter), adapted from Antweiler and Frank (2004), Cookson and Niessner (2020), and Shanthikumar et al. (2020). *SIT* is a key measure of this study as it captures the sentiment of the treatment contents. *SIT* is calculated as follows

$$SIT = \frac{N_{positive} - N_{negative}}{N_{total}}, \quad (1)$$

where N_{total} is the total number of treatment contents, $N_{positive}$ is the number of positive treatment contents, and $N_{negative}$ is the number of negative treatment contents. Note that we assign each content to its most likely sentiment category among positive, neutral, or negative. *SIT* ranges from -1 to 1 and its absolute value indicates sentiment extremity. We also consider other factors such as the number of treatment contents, time difference between the current and next content generated by each user.

Model-Free Evidence

We trace the contents created by each user and provide a model-free evidence of how their future sentiments will change based on the contents they see. We use the transition matrix to illustrate how the sentiment of each user's next content changes from the current one by the contents created by others in between. Note that the transition matrix, an estimate of first-order Markov model, is a simple and powerful metric which summarizes the carryover and spillover effects between groups (Montgomery et al. 2004, Cadez et al. 2000). We provide a transition matrix to demonstrate model-free evidence of the sentiment contagion.

To check the sentiment contagion, we calculate the sentiment index (i.e., *SIT*) of the treatment contents—all the contents in the source thread and the target thread generated in between the current and the next content following the discussions in Subsection—Empirical Approach. Then, we categorize the whole dataset into three treatment groups based on the *SIT* to clearly show how *SIT* affects the sentiment of each user's next content across the different treatment groups. If the *SIT* is below the 20th percentile, they are assigned to the *negative* treatment group. The upper 20th percentile is assigned to *positive* treatment group. All others are in the *baseline* treatment group.

Table 3 summarizes the transition probability across the three treatment groups. If there are no impacts of treatments, then the transition probability should be similar across the groups. However, Table 3 shows that the treatments exert an influence on the sentiment of the next content. The sentiment of next content in the *negative* group is more likely to be negative (47.61%; row (8) and column (1)) compared to the *baseline* group (34.48%; row (4) and column (1)). Similarly, the sentiment of next content in the *positive* group is more likely to be positive (29.57%; row (12) and column (3)) compared to the *baseline* group (20.08%; row (4) and column (3)).

Interestingly, when we compare the differences in probabilities between the *positive* and *negative* groups, the impact of negative treatment is larger (47.61% - 34.48% = 13.13% vs 29.57% - 20.08% = 9.50%),

[6] The time when the focal user generated the first content is set to zero.

meaning that negative contents have a much stronger effect on users' emotion compared to positive contents. Such an observation could be explained by the negativity bias, which refers to the tendency for negative information to have a greater impact on the brain than equally extreme positive information (Cacioppo et al. 2014). This may be the reason why online disinhibition is often regarded as negative or toxic, which results in a deterioration of our well-being, such as lower levels of social connectedness and flourishing (Stuart and Scott 2021).

Treatment group	**Sentiment of current content**		**Sentiment of next content**					
			Transition probability			Difference in probabilities (Treatment group – *Baseline*)		
			(1)	(2)	(3)	(4)	(5)	(6)
			Negative	Neutral	Positive	Negative	Neutral	Positive
Baseline (N: 2,165,197)	(1)	Negative	40.48%	41.98%	17.53%	0.00%	0.00%	0.00%
	(2)	Neutral	31.80%	48.77%	19.43%	0.00%	0.00%	0.00%
	(3)	Positive	30.27%	43.82%	25.91%	0.00%	0.00%	0.00%
	(4)	Total	34.48%	45.44%	20.08%	0.00%	0.00%	0.00%
Negative (N: 722,050)	(5)	Negative	53.02%	35.66%	11.32%	12.53%	-6.32%	-6.21%
	(6)	Neutral	43.10%	43.25%	13.65%	11.31%	-5.52%	-5.78%
	(7)	Positive	41.13%	40.30%	18.57%	10.86%	-3.52%	-7.34%
	(8)	Total	47.61%	39.22%	13.17%	13.13%	-6.22%	-6.91%
Positive (N: 721,407)	(9)	Negative	29.80%	43.80%	26.39%	-10.68%	1.82%	8.86%
	(10)	Neutral	23.59%	48.62%	27.79%	-8.21%	-0.15%	8.36%
	(11)	Positive	22.06%	42.88%	35.06%	-8.21%	-0.95%	9.15%
	(12)	Total	24.69%	45.74%	29.57%	-9.79%	0.30%	9.50%

Table 3. Model-Free Evidence of Sentiment Contagion

Furthermore, Table 3 shows how the next sentiment differs depending on the current sentiment regardless of the treatment groups. For instance, in the *negative* treatment group, the sentiment of next content is likely negative when the current content is negative (53.02%; row (5) and column (1)). However, the transition probability drops by 11.89% when the sentiment of current content is positive (41.13%; row (7) and column (1)). We find similar patterns for the other groups (40.48% - 30.27% = 10.21% for the *baseline* group in rows (1) and (3) and column (1); 29.80% - 22.06% = 7.74% for the *positive* group in rows (9) and (11) and column (1)). Based on this finding, we include the sentiment of the current content as one of the controls in our model.

Identification

A potential concern in our setting is that the treatment sentiment (i.e., *SIT*) could be endogenous. For example, there could be high association with concurrent events—EPL game results (e.g., a team won the game against a rival team) or news about major players' injuries. If a team won the game, especially against rivals, then the overall sentiments may be positive. On the other hand, when a team lost the game, the contents may likely be negative.

To address those issues, we use two identification strategies. First, we include control variables representing concurrent events. Given a period between the current content and the next content of the focal user, we construct two control variables. The first, *Control PC*, is measured by all posts and comments within a subreddit where the target thread exists, *excluding* the contents in both the source and the target thread. The second, *Control NP*, is similar to *Control PC*, but it only contains the "new" posts. Specifically, *Control NP* only considers the posts that any user who is not the focal user generated their first contents as the post during the empirical period. We use *Control NP* as the main control variable as it is highly likely to contain pure sentiments induced by the concurrent events without any interference from other contents.

The other approach to control for concurrent events is to use residuals that regress *SIT* on all control variables (i.e., *Control NP* and *Control PC*) as the main treatment variable. Hereafter, we define the residuals that regress *SIT* on *Control NP* and *Control PC* as *RSIT_NP*, *RSIT_PC*, respectively. This specification allows us to capture the impact of "abnormal sentiments". For instance, if there is sudden news about the injury of an important player, then most of the contents in a subreddit supporting the team would be mostly negative. It is then difficult to argue whether the contents are affected by negative sentiments or concurrent events. By controlling for the sentiments of the overall contents on the subreddit, we can clearly examine the impact of the treatment contents.

Empirical Model and Estimation Results

We specify our empirical model as follows:

$$y_{il} = \beta_1 RSIT\_NP_{il} + \beta_2 \ln(N_{il}) + \beta_3 \ln(TD_{il}) + \alpha X_{i,l-1} + \gamma_i + \tau_l + \varepsilon_{il}, \quad (2)$$

where y_{il} denotes the dependent variable with i indicating individual user and l indicating user i's l-th content; $RSIT\_NP_{il}$ denotes the main independent variable discussed in the previous Subsection of user i's l-th content; N_{il} denotes the number of treatment contents for user i's l-th content; TD_{il} denotes the time difference between user i's l-th content and (l-1)th content ($TD_{il} = Time_{il} - Time_{i,l-1}$); $X_{i,l-1}$ denotes user i's sentiment of (l-1)th content; γ_i captures user fixed effects; τ_l captures the joint day-team fixed effects.

The dependent variable is the sentiment of a piece of content, defined as the difference between its likelihood of being positive or negative. If the value is negative, then the content is likely negative as well. This measure allows us to capture both the direction and magnitude/extremity of the sentiment. We specify all non-index continuous variables (e.g., N_{il} and TD_{il}) in logarithms since they have a skewed distribution. As appropriate, we add one to the variables to avoid logarithms of zero. Table 4 and Table 5 report the summary statistics and correlations of the variables.

Variables	Unit	1st Q.	Median	Mean	3rd Q.	Min.	Max.
Dependent variable		-0.594	-0.135	-0.117	0.251	-0.982	0.993
SIT		-0.333	-0.104	-0.138	0.000	-1.000	1.000
Control NP		0.000	0.000	0.032	0.000	-1.000	1.000
Control PC		-0.237	-0.070	-0.106	0.000	-1.000	1.000
Time difference (*TD*)	Second	204	1299	27,366	23,613	0	259,199
Number of treatments (*N*)		3	23	165	101	0	11,222
Negative lag (dummy)		0	0	0.353	1	0	1
Positive lag (dummy)		0	0	0.205	0	0	1

Table 4. Summary Statistics

Note that we discard the contents 1) without their own precedent contents (i.e., discard the first contents of users);[7] 2) where the time difference (TD_{il}) between his/her two consecutive contents exceeds three days ($TD_{il} = Time_{il} - Time_{i,l-1} \geq 72\ hours$) to exclude less active users. To control the impact of lagged sentiment, we define negative and positive lags. *Negative lag* indicates whether the lagged sentiment was negative (1) or not (0). *Positive lag* indicates whether the lagged sentiment was positive (1) or not (0).

Correlation	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Dependent variable	1.000	0.184	0.046	0.087	0.045	-0.057	-0.115	0.097
2. *SIT*	0.184	1.000	0.094	0.185	0.057	-0.162	-0.149	0.122
3. *Control NP*	0.046	0.094	1.000	0.130	0.144	0.002	-0.041	0.031
4. *Control PC*	0.087	0.185	0.130	1.000	0.019	-0.083	-0.077	0.063
5. Log (TD + 1)	0.045	0.057	0.144	0.019	1.000	0.425	-0.055	0.034
6. Log (N + 1)	-0.057	-0.162	0.002	-0.083	0.425	1.000	0.050	-0.002
7. Negative lag (dummy)	-0.115	-0.149	-0.041	-0.077	-0.055	0.050	1.000	-0.375
8. Positive lag (dummy)	0.097	0.122	0.031	0.063	0.034	-0.002	-0.375	1.000

Table 5. Correlation[8]

Main Results

We report our main results in Table 6. We include user and joint day-team fixed effects in all regressions. Recall the key independent variable, *RSIT_NP*, as specified in the previous Subsection controls for the effect

[7] The reason is that we use sentiment of precedent content as a control variable.
[8] Pearson's correlation test show that all variables are significantly correlated at p=0.05.

of concurrent events.[9] We fit Equation (1) to all observations. As reported in the first row of Table 6 (*RSIT_NP*), the estimates are significantly positive across all models, which means that the sentiment of the focal user's next content becomes more positive or negative when the treatment content sentiments are more positive or negative. This finding is consistent with prior study studies that show textual content is a sufficient channel for contagion without additional behavior (Kramer et al. 2014). In the sixth row of Table 6, *TD* has a positive impact on the dependent variable. It means that users are negatively biased (the average of the dependent variable is -0.117) but it will oscillate to the neutral side as time goes by. The coefficient of *N* is negative ($p < 0.01$), meaning that the more treatment contents a focal user sees, the more negative his/her sentiment becomes. We also include indicators to control the impact of lagged sentiment.

	(1)	(2)	(3)	(4)	(5)
	RSIT_NP	Interactions	*Control NP*	*Control PC*	*RSIT_PC*
RSIT_NP	0.2002*** (0.0018)	0.0481*** (0.0036)			
SIT			0.0501*** (0.0036)	0.0481*** (0.0036)	
RSIT_PC					0.0258*** (0.0035)
Time difference (*TD*)		0.0056*** (0.0002)	0.0045*** (0.0002)	0.0047*** (0.0002)	0.0051*** (0.0002)
Number of treatments (*N*)		-0.0095*** (0.0003)	0.0030*** (0.0003)	0.0029*** (0.0003)	-0.0097*** (0.0003)
RSIT_NP x *TD*		-0.0022*** (0.0005)			
SIT_NP x *TD*			-0.0021*** (0.0005)	-0.0016*** (0.0005)	
RSIT_PC x *TD*					-0.0002 (0.0005)
RSIT_NP x *N*		0.0865*** (0.0009)			
SIT_NP x *N*			0.0860*** (0.0009)	0.0848*** (0.0009)	
RSIT_PC x *N*					0.0900*** (0.0009)
Control NP			0.0012 (0.0011)		
Control PC				0.0161*** (0.0011)	
Negative lag	-0.0365*** (0.0008)	-0.0307*** (0.0007)	-0.0304*** (0.0007)	-0.0303*** (0.0007)	-0.0314*** (0.0007)
Positive lag	0.0240*** (0.0009)	0.0198*** (0.0009)	0.0195*** (0.0009)	0.0194*** (0.0009)	0.0204*** (0.0009)
User FE	Yes	Yes	Yes	Yes	Yes
Team x day FE	Yes	Yes	Yes	Yes	Yes
Observations	3,608,654	3,608,654	3,608,654	3,608,654	3,608,654
R-squared	0.111	0.118	0.118	0.118	0.117
Adjusted R-Squared	0.094	0.101	0.101	0.101	0.100

Table 6. Main Table[10]

As previously discussed, treatment sentiments could differ depending on the number of treatment contents and time difference. We incorporate interaction effects (*RSIT_NP* x *N* and *RSIT_NP* x *TD*) and report the results in columns (2) of Table 6. The estimates are consistent with our expectations. We find strong

[9] We report the results with other control variables in columns (3)-(5) of Table 6.
[10] Robust standard errors clustered by user are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01. Hereafter, all the tables are summarized as under these setting. Variables are defined in Equation (2) and Subsection—Main Results.

evidence that treatment sentiments affect the sentiment of the focal user's next content, and the impact is intensified when there are more treatment contents (i.e., *N* is larger). If the focal user generates the next content slowly (i.e., *TD* is larger), then the impact is weakened.

Robustness Checks and Falsification Tests

We conduct several robustness checks. First, we apply several specifications to control for the treatment sentiments of omitted confounders. *Control NP*, *Control PC*, and *RSIT_PC* defined in the previous Subsection are used as control variables in columns (3), (4), and (5) of Table 6, respectively. The results are consistent with the main findings in column (2) of Table 6 regardless of specifications. Second, we use the subsample excluding the contents on major subreddits representing Arsenal, Chelsea, Liverpool, and Manchester United. As shown in column (1) of Table 7, the result with subsample (27.7% of all observations) is once again consistent with the main results in column (2) of Table 6. Third, we use another subsample excluding the contents generated by heavy users who frequently posted contents during the empirical period.[11] The result with this subsample (76.5% of all observations) is reported in column (2) of Table 7, which remains similar to the main results. Fourth, we use two alternative sentiment indexes—*SIT*2 and *SIT*3—following Shanthikumar et al. (2020). They are measured as

$$SIT2 = \frac{N_{positive} - N_{negative}}{N_{positive} + N_{negative}}, \tag{3}$$

$$SIT3 = \frac{\sum net\ score}{\sum abs(net\ score)}, \tag{4}$$

where *net score* is the difference between the likelihood of a piece of content being positive and negative; *abs*(*net score*) is the absolute value of the *net score*. We regress *SIT2*, *SIT3* on the *Control NP* and use the residuals as the independent variables: *RSIT2_NP*, *RSIT3_NP*, respectively. The results based on *RSIT2_NP* and *RSIT3_NP* are reported in column (3) and (4) of Table 7, respectively. The estimates are consistent with our main results.

Next, we conduct falsification tests to show that our main findings do not hold when we use wrong treatments in terms of timing and contents. The first falsification is using the treatments in the "incorrect" period. We set an incorrect period to one month after the true period. For instance, if the focal user generated a content on coys at 9 PM, 24th, September 2021 and his/her prior content had been generated one hour ago (8 PM, 24th, September 2021), then the contents generated on coys from 8 PM, 24th, October 2021 to 9 PM, 24th, October 2021 are used as treatment contents. The correlation between the true and the false sentiment of treatments is 0.023 and the correlation between the true and false number of treatments is 0.428. As shown in column (5) of Table 7, the coefficient of *RSIT_NP* and the interaction between *RSIT_NP* and *TD* are no longer significant. We exclude *N* and the interactions with *N* as the correlation between *TD* and *N* is too high (0.909). [12] The second falsification is using "irrelevant" contents. We choose "BabyBumps" as the "irrelevant" subreddit considering following conditions: 1) this subreddit is active enough,[13] and 2) they are not related to sports, games, and entertainment which may be correlated with football. We apply the same sentiment analysis used for football subreddits to contents on BabyBumps. The total number of contents on BaByBumps are 313,099 and nearly 40% of them are positive (118,676; 37.90%). Negative ones rank the second (99,552; 31.80%) and neutral ones rank the lowest (30.30%).

We calculate placebo treatment sentiment using the whole contents on the subreddit given the same period as the actual period and apply it to the model. The correlation between the actual and the irrelevant treatment sentiment is -0.002 and the correlation between the actual and irrelevant number of treatments is 0.388. As shown in column (6) of Table 6, the coefficient of *RSIT_NP* (0.0007) and the interaction between *RSIT_NP* and *TD* (-0.0004) are no longer significant. Again, we exclude the variables as *N* and the interactions with *N*, due to the high correlation between *TD* and *N* (0.907).

---

[11] We define 473 users who post more than 1,000 contents as heavy users and exclude 848,370 contents generated by them.

[12] Correlation between *TD* and *N* in the main specification is 0.43 as shown in Table 5. For this falsification test, we consider all the contents on a subreddit to which the target thread belongs during the "incorrect" period. In other words, false treatments are not limited to contents in the source and target thread. That is why *N* would be higher when *TD* is larger.

[13] Total number of contents on BabyBumps is 313,099 during the empirical period.

	(1)	(2)	(3)	(4)	(5)	(6)
	Subsample at team level	Subsample at user level	*RSIT2_NP*	*RSIT3_NP*	Falsification : Incorrect time	Falsification : Irrelevant subreddit
RSIT_NP	0.0586*** (0.0060)	0.0622*** (0.0037)			0.0029 (0.0038)	0.0007 (0.0030)
RSIT2_NP			0.0609*** (0.0025)			
RSIT3_NP				0.0460*** (0.0024)		
Time difference (*TD*)	0.0049*** (0.0004)	0.0056*** (0.0002)	0.0068*** (0.0002)	0.0062*** (0.0002)	0.0025*** (0.0002)	0.0023*** (0.0001)
Number of treatments (*N*)	-0.0094*** (0.0006)	-0.0098*** (0.0003)	-0.0099*** (0.0003)	-0.0103*** (0.0003)		
RSIT_NP x *TD*	-0.0040*** (0.0008)	-0.0027*** (0.0005)			-0.0002 (0.0007)	-0.0004 (0.0005)
RSIT2_NP x *TD*			-0.0059*** (0.0003)			
RSIT3_NP x *TD*				-0.0037*** (0.0004)		
RSIT_NP x *N*	0.0887*** (0.0016)	0.0835*** (0.0009)				
RSIT2_NP x *N*			0.0523*** (0.0006)			
RSIT3_NP x *N*				0.0565*** (0.0006)		
Negative lag	-0.0275*** (0.0014)	-0.0294*** (0.0008)	-0.0323*** (0.0007)	-0.0315*** (0.0007)	-0.0458*** (0.0008)	-0.0461*** (0.0008)
Positive lag	0.0169*** (0.0016)	0.0176*** (0.0009)	0.0201*** (0.0009)	0.0190*** (0.0009)	0.0317*** (0.0009)	0.0323*** (0.0009)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Team x day FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,001,326	2,760,284	3,608,654	3,608,654	3,142,942	3,608,654
R-squared	0.125	0.128	0.116	0.118	0.102	0.100
Adjusted R-Squared	0.101	0.106	0.100	0.101	0.084	0.083

Table 7. Robustness Checks and Falsification Tests[14]

Individual Dynamics[15]

We next conduct additional analyses to explore the impact of individual dynamics on sentiment contagion. First, we check whether our main findings would differ depending on where the next content is generated. Specifically, users can generate content in the same thread or in different threads. Column (1) of Table 8 reports the results with subsamples where users generate contents in the same thread.[16] The results are consistent with results in column (2) of Table 6 except that the main effect of *RSIT_NP* becomes insignificant. Columns (2) and (3) of Table 8 report the results with observations where the users generate their subsequent contents in different threads. The coefficients of both treatment sentiments in the source thread and the target thread are positive and significant in column (2) of Table 8. The interaction effects

[14] Variables are defined in Equation (2) and Subsection—Robustness Checks and Falsification Tests.

[15] Individual dynamics refers to dynamic behavior of users, which is whether they generate the consecutive content in the same thread or in different threads.

[16] We report the estimate of *RSIT_NP* as *RSIT_ST_NP* in column (1) of Table 8. Note that both variables are identical when the focal user consecutively generates contents in the same thread.

between treatment sentiments and number of treatment contents, as summarized in column (3) of Table 8, are positive and significant.

Note that the coefficient of *RSIT_TT_NP* (*RSIT_NP* in the target thread) is 0.2606 but the coefficient of *RSIT_ST_NP* (*RSIT_NP* in the source thread) is only 0.0096 in column (2). It is challenging to understand why users generate their next content in the target thread different from the source thread since people can be affected by either relatively positive or negative sentiments than their prior sentiment depending on their situation-specific motivation (Goldenberg et al. 2020) and researchers can hardly observe the motivation. However, we can investigate how the contents in the target thread are different from contents in other possible threads where they may generate their next content (e.g., source thread, other threads on a subreddit where the focal user pays attention to). If there is a significant difference, then we can argue that social media users are more likely to pay attention to the contents in the target thread rather than content in other threads including the source thread. To explore the difference, we focus on *SIT* and compare *SIT*s in the target thread, source thread, and the other threads (i.e., *Control_PC*). With Welch's two sample *t*-test, we find significant differences not only between the target thread and the source thread ($t= 79.186$; $p < 0.01$) but also between the target thread and the other threads ($t = 94.479$; $p < 0.01$) while the difference between the source thread and other threads is insignificant ($t=-0.3547$; $p = 0.7228$).

	(1)	(2)	(3)
	Same thread	Different threads (1)	Different threads (2)
RSIT_ST_NP (in the source thread)	0.0075 (0.0050)	0.0096*** (0.0012)	0.0748*** (0.0043)
RSIT_TT_NP (in the target thread)		0.2606*** (0.0022)	-0.0135*** (0.0048)
Time difference (*TD*)	0.0015*** (0.0004)	0.0055*** (0.0002)	0.0029*** (0.0002)
Number of treatments in the source thread (*N_ST*)	-0.0112*** (0.0005)	0.0004* (0.0003)	0.0006** (0.0003)
Number of treatments in the target thread (*N_TT*)		-0.0156*** (0.0003)	-0.0080*** (0.0003)
RSIT_ST_NP x *TD*	0.0050*** (0.0008)		-0.0095*** (0.0005)
RSIT_TT_NP x *TD*			0.0116*** (0.0005)
RSIT_ST_NP x *N_ST*	0.0879*** (0.0012)		0.0096*** (0.0009)
RSIT_TT_NP x *N_TT*			0.0999*** (0.0011)
Negative lag	-0.0567*** (0.0011)	-0.0117*** (0.0009)	-0.0114*** (0.0009)
Positive lag	0.0425*** (0.0014)	0.0073*** (0.0011)	0.0072*** (0.0011)
User FE	Yes	Yes	Yes
Team x day FE	Yes	Yes	Yes
Observations	1,645,648	1,963,006	1,963,006
R-squared	0.146	0.130	0.138
Adjusted R-Squared	0.113	0.104	0.112

Table 8. Individual Dynamics[17]

Individual Heterogeneity

We next explore the impact of individual heterogeneity on the sentiment contagion. Prior literature reveals the habituation—frequent exposure to emotions can lead to habituation or fatigue, making each exposure

[17] Variables are defined in Equation (2) and Subsection—Individual Dynamics.

to emotional expression online less impactful (Wilson and Gilbert 2008). To test the possibility, we set an individual sequence variable (*SN*) as a proxy of the individual heterogeneity. *SN* indicates the sequence number of generated contents by the focal user considering whole contents on 20 subreddits since August 2013. For instance, if a user generated 50 contents before the empirical period, 51 is assigned to *SN* when a user generates the first content during the empirical period.

	(1)	(2)	(3)
	SN	Same thread with *SN*	Different threads with *SN*
RSIT_NP	0.0799*** (0.0054)		
RSIT_ST_NP (in the source thread)		0.0378*** (0.0074)	0.0845*** (0.0065)
RSIT_TT_NP (in the target thread)			0.0148* (0.0079)
Time difference (*TD*)	0.0056*** (0.0002)	0.0015*** (0.0004)	0.0028*** (0.0002)
Number of treatments (*N*)	-0.0095*** (0.0003)		
Number of treatments in the source thread (*N_ST*)		-0.0112*** (0.0005)	0.0006** (0.0003)
Number of treatments in the target thread (*N_TT*)			-0.0080*** (0.0003)
Sequence number (*SN*)	-0.0090*** (0.0010)	-0.0068*** (0.0013)	-0.0110*** (0.0012)
RSIT_NP x *TD*	-0.0027*** (0.0005)		
RSIT_ST_NP x *TD*		0.0043*** (0.0008)	-0.0097*** (0.0005)
RSIT_TT_NP x *TD*			0.0111*** (0.0005)
RSIT_NP x *N*	0.0866*** (0.0009)		
RSIT_ST_NP x *N_ST*		0.0880*** (0.0012)	0.0095*** (0.0009)
RSIT_TT_NP x *N_TT*			0.0996*** (0.0011)
RSIT_NP x *SN*	-0.0046*** (0.0008)		
RSIT_ST_NP x *SN*		-0.0043*** (0.0009)	-0.0013** (0.0007)
RSIT_TT_NP x *SN*			-0.0038*** (0.0010)
Negative lag	-0.0307*** (0.0007)	-0.0567*** (0.0011)	-0.0114*** (0.0009)
Positive lag	0.0197*** (0.0009)	0.0424*** (0.0014)	0.0071*** (0.0011)
User FE	Yes	Yes	Yes
Team x day FE	Yes	Yes	Yes
Observations	3,608,654	1,645,648	1,963,006
R-squared	0.118	0.146	0.138
Adjusted R-Squared	0.101	0.113	0.112

Table 9. Individual Heterogeneity[18]

[18] Variables are defined in Equation (2) and Subsection—Individual Heterogeneity.

We specify *SN* in logarithms and add one to the variables to avoid a logarithm of zero as appropriate.[19] We add *SN* and interactions with *SN* to the model specified in column (2) of Table 6 and present the results in column (1) of Table 9. Under the same setting in columns (1) and (3) of Table 8, we investigate the impact of *SN* and summarize results in columns (2) and (3) of Table 9, respectively. Regardless of variants of cases presented in Table 9, we find robust evidence of both sentiment contagion and habituation. The coefficient of *SN* is significant and negative, meaning that experienced users are more likely to generate negative content than new users. The coefficient of the interaction between *RSIT_NP* and *SN* is also significant and negative, which suggests that users become less affected by the exposed emotions as they are more experienced. Using the estimates in column (1) of Table 9, the magnitude of being less impacted by the treatment sentiment is plausible because the impact is reduced by 57.7% ($\frac{0.0799-0.0046\times\ln(1545+1)}{0.0799}\times 100$) after users generate 1,545 contents, which is the mean of *SN*.

Other Specifications

We now conduct other analyses to enrich our main findings.

	(1) Without similarity	(2) Long-term effect: Active	(3) Negativity bias
RSIT_NP	0.0314 (0.0540)	0.3568*** (0.0895)	0.0365*** (0.0055)
Time difference (*TD*)	-0.0047 (0.0034)	0.1640*** (0.0091)	0.0049*** (0.0003)
Number of treatments (*N*)	0.0012 (0.0044)	0.1720*** (0.0040)	-0.0091*** (0.0004)
RSIT_NP x *TD*	-0.0035 (0.0083)	-0.1928*** (0.0425)	0.0003 (0.0008)
RSIT_NP x *N*	0.0611*** (0.0149)	0.1038*** (0.0192)	0.0851*** (0.0015)
RSIT_NP x I (Neg)			0.0258*** (0.0085)
RSIT_NP x I (Neg) x *TD*			-0.0057*** (0.0013)
RSIT_NP x I (Neg) x *N*			0.0032 (0.0022)
Constant		0.2116*** (0.0193)	
Lagged negative	-0.0458*** (0.0008)	-0.0000 (0.0000)	-0.0308*** (0.0007)
Lagged positive	0.0317*** (0.0009)	0.0009*** (0.0001)	0.0198*** (0.0009)
User fixed effects	Yes	No	Yes
Team x day fixed effects	Yes	No	Yes
Observations	12,116	63,988	3,608,654
R-squared	0.470	0.051	0.118
Adjusted R-Squared	0.168	0.051	0.101

Table 10. Other Specifications[20]

One potential concern with our finding is that it is hard to differentiate between the sentiment contagion and homophily driven behavior (Aral et al. 2009). In our context, homophily driven behavior indicates that users generate contents with similar level of sentiment in a specific thread since they are just similar to each other instead of being influenced by each other. To address this concern, we make subsamples by using each user's loyalty. Premier league supporters commonly have lifelong loyalty to their team (Newson et al. 2016). Therefore, we can assume that users who generate their contents mostly in a specific subreddit are

[19] 1,545 is the mean of *SN* and 504 is the median of *SN*.
[20] Variables are defined in Equation (2) and Subsection—Other Specifications.

likely to support a team representing that subreddit.[21] Also, users who generate contents in the same subreddit are likely to be similar as they support the same team by discussing same players, managers, and tactics. Thus, if they generate any content in other subreddits, it alleviates the concern in the homophily driven behaviors to a large extent. We define each user's supporting team based on their past records from August 2013 to July 2018 as the team representing a subreddit where he/she generated most of the contents. Then, we sample users who generated at least 10 contents, and more than 90% of the contents should have been generated in their supporting subreddit to make sure that selected users are enough to be loyal supporters. Then, we make a subsample where the selected users generate their content in a subreddit different from their loyal subreddit. The result with this subsample is reported in column (1) of Table 10. Although main coefficients are insignificant (i.e., *RSIT_NP*, *TD*, and *N*), the coefficient of interaction between *RSIT_NP* and *N* remains significant and positive (0.0611; $p < 0.01$), which shows the existence of the sentiment contagion in the absence of the homophily driven behavior.

In column (2) of Table 10, we check whether the treatment sentiment has a long-term effect on user behavior. We collected additional dataset containing the 2-year contents on 20 subreddits after the empirical dataset (i.e., from June 1, 2019 to May 31, 2021) to check the long-term effect. We define a variable at user level that indicates whether a user is still active during the additional period (1), meaning he/she generates at least one content during the additional period on any 20 subreddits, or not (0). Also, each independent variable for a user is measured as the mean of his/her corresponding variable. We use 63,988 users who generate at least one content during the empirical period and run the same analysis in column (2) of Table 6. As shown in column (2) of Table 10, we find strong evidence of long-term effect of treatment sentiments on users. This means that the more positive contents users read, the more likely they are to be active on social media. This result is consistent with the empirical evidence that treatment sentiment has both short-term and long-term effects on social media users.

In column (3) of Table 10, we check whether the negativity bias (Cacioppo et al. 2014) exists in our setting. If there is negative bias, we should expect that the impact of *RSIT_NP* is enhanced when *RSIT_NP* is negative. Otherwise, the impact would be the same regardless of the sign of *RSIT_NP*. To check, we define a new variable, I (Neg), indicating whether *RSIT_NP* is negative (1) or not (0) and add an interaction variable between I (Neg) and RSIT_NP. As shown in column (3) of Table 10, estimate of the interaction variable is positive and significant (0.0258; $p < 0.01$), which suggests the existence of negativity bias.

Conclusion

In this paper, we find robust evidence of the sentiment contagion by applying a novel identification strategy with massive social media contents across many users. Social media users become more positive or negative when they are exposed to more positive or negative contents. Furthermore, the more contents a focal user sees, the more negative his/her sentiment becomes. A series of analyses provide robust evidence that sentiment contagion occurs no matter when a focal user generates the subsequent content in the same thread, in different threads, even in different subreddits. In terms of heterogeneity, a set of analyses suggest that experienced users, who generate more contents and are likely to be exposed to many treatment contents, become less sensitive to what they see, which is a new empirical evidence in line with Wilson and Gilbert 2008. Furthermore, we find that the negativity bias exists in social media (Cacioppo et al. 2014). We also find a long-term effect on user activity, which shows that users are more likely to be inactive in the long run when they are exposed to more negative treatment sentiment.

Our findings clearly convey a clear message that social media users can be affected by just sentiment embedded in social media contents and provide important implications for social media platforms. For example, content censorship, which is the most of social media platforms currently use, might not be the best option for preventing problems related to users' sentiment and mental health. Access time restriction or promoting the positive and healthy environment of social media platforms would be potential solutions for the social media platforms.

There are several limitations in this research. First, this paper proposes a novel identification strategy to investigate the casual link of the sentiment contagion in social media. Although we find robust evidence of

[21] Our data shows that almost 90% (89.3%) of users generated their contents in only one subreddit during the 2018/19 season.

the sentiment contagion with a series of variants, we need to interpret this result with caution as there might be general social influences (e.g., herding) among users. Second, we assume social media users are likely to see whole contents in the source and target threads before they generate their next content. It is possible that they only read a subset of contents. In addition, we only trace the contents of social media users in 20 football subreddits. Users could visit other subreddits where they communicate about the football (e.g., r/PremierLeague) or other topics. In both cases, our estimates might be biased. However, we believe that they are unlikely to threaten the existence of sentiment contagion given the robust evidence presented.

We conclude this paper by suggesting several future research directions. First, we assume that every piece of treatment content has identical impact on the focal user's next content sentiment. However, in reality, the impacts of content generated by a friend of the focal user and stranger could be very different. Future research can explore the impact of the relation between the focal user and the author of treatment content (e.g., follower, friend, and influencer) on the sentiment contagion. Second, we study the heterogeneity of sentiment contagion between experienced and inexperienced users. Future scholars explore other individual features that could strengthen or weaken the sentiment contagion. Examples of such features may include the total number of upvotes/downvotes the focal user has obtained, the daily amount of time a person spends on social media, and loyalty to a specific platform. Finally, it would be interesting to extend our study to other social media platforms. The sentiment contagion may exist across various social media platforms. Utilizing whole social media activity by the same user would help paint a clearer picture of the sentiment contagion.

References

Allcott H, Braghieri L, Eichmeyer S, Gentzkow M (2020) The welfare effects of social media. *American Economic Review* 110(3):629-76.

Antweiler W, Frank MZ (2004) Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance* 59(3):1259-1294.

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* 106(51):21544-21549.

Barbieri F, Camacho-Collados J, Neves L, Espinosa-Anke L (2020) Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*.

Bruder M, Fischer A, Manstead AS (2014) Social appraisal as a cause of collective emotions. *Collective emotions*:141-155.

Cacioppo JT, Cacioppo S, Gollan JK (2014) The negativity bias: Conceptualization, quantification, and individual differences. *Behavioral and Brain Sciences* 37(3):309.

Cadez I, Heckerman D, Meek C, Smyth P, White S (2000) Visualization of navigation patterns on a web site using model-based clustering. *Proceedings sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 280-284.

Chartrand TL, Van Baaren R (2009) Human mimicry. *Advances in experimental social psychology* 41:219-274.

Cookson JA, Niessner M (2020) Why don't we agree? Evidence from a social network of investors. *The Journal of Finance* 75(1):173-228.

Coviello L, Sohn Y, Kramer AD, Marlow C, Franceschetti M, Christakis NA, Fowler JH (2014) Detecting emotional contagion in massive social networks. *PloS one* 9(3):e90315.

Crockett MJ (2017) Moral outrage in the digital age. *Nature human behaviour* 1(11):769-771.

Dang-Xuan L, Stieglitz S (2012) Impact and diffusion of sentiment in political communication–an empirical analysis of political weblogs. *Sixth International AAAI Conference on Weblogs and Social Media*.

Ferrara E, Yang Z (2015) Measuring emotional contagion in social media. *PloS one* 10(11):e0142390.

Goldenberg A, Garcia D, Halperin E, Zaki J, Kong D, Golarai G, Gross JJ (2020) Beyond emotional similarity: The role of situation-specific motives. *Journal of Experimental Psychology: General* 149(1):138.

Goldenberg A, Gross JJ (2020) Digital emotion contagion. *Trends in Cognitive Sciences* 24(4):316-328.

Gruda D, Ojo A, Psychogios A (2022) Don't you tweet me badly: Anxiety contagion between leaders and followers in computer-mediated communication during COVID-19. *Plos one* 17(3):e0264444.

Hatfield E, Cacioppo JT, Rapson RL (1994) Emotional contagion. *Cambridge University Press*.

Hess U (2021) Who to whom and why: The social nature of emotional mimicry. *Psychophysiology* 58(1):e13675.
Kane GC, Alavi M, Labianca G, Borgatti SP (2014) What's different about social media networks? A framework and research agenda. *MIS quarterly* 38(1):275-304.
Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53(1):59-68.
Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24):8788-8790.
Kwon KH, Gruzd A (2017) Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump's YouTube campaign videos. *Internet Research*.
Lin R, Utz S (2015) The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. *Computers in human behavior* 52:29-38.
Mac R, Kang C (2021) Whistle-Blower Says Facebook 'Chooses Profits Over Safety'. The New York Times.
Manstead A, Fischer AH (2001) Social appraisal. *Appraisal processes in emotion: Theory, methods, research*:221-232.
Montgomery AL, Li S, Srinivasan K, Liechty JC (2004) Modeling online browsing and path analysis using clickstream data. *Marketing science* 23(4):579-595.
Newson M, Buhrmester M, Whitehouse H (2016) Explaining lifelong loyalty: The role of identity fusion and self-shaping group events. *PloS one* 11(8):e0160427.
Niedenthal PM, Winkielman P, Mondillon L, Vermeulen N (2009) Embodiment of emotion concepts. *Journal of personality and social psychology* 96(6):1120.
Oh O, Eom C, Rao HR (2015) Research note—Role of social media in social change: An analysis of collective sense making during the 2011 Egypt revolution. *Information Systems Research* 26(1):210-223.
Parkinson B (2020) Intragroup emotion convergence: Beyond contagion and social appraisal. *Personality and Social Psychology Review* 24(2):121-140.
Parkinson B, Simons G (2009) Affecting others: Social appraisal and emotion contagion in everyday decision making. *Personality and social psychology bulletin* 35(8):1071-1084.
Peters K, Kashima Y (2015) A multimodal theory of affect diffusion. *Psychological Bulletin* 141(5):966.
Prochazkova E, Kret ME (2017) Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews* 80:99-114.
Rosenthal S, Farra N, Nakov P (2019) SemEval-2017 task 4: Sentiment analysis in Twitter. *arXiv preprint arXiv:1912.00741*.
Shanthikumar DM, Wang QA, Wu S (2020) Does Interaction on Social Media Increase or Moderate Extremeness? *Available at SSRN 3879175*.
Song Y, Lin Q, Kwon KH, Choy CH, Xu R (2022) Contagion of offensive speech online: An interactional analysis of political swearing. *Computers in Human Behavior* 127:107046.
Soroka S, Fournier P, Nir L (2019) Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences* 116(38):18888-18892.
Statista (2021).
Stieglitz S, Dang-Xuan L (2012) Impact and diffusion of sentiment in public communication on Facebook.
Stuart J, Scott R (2021) The Measure of Online Disinhibition (MOD): Assessing perceptions of reductions in restraint in the online environment. *Computers in Human Behavior* 114:106534.
Van Zomeren M, Leach CW, Spears R (2012) Protesters as "passionate economists" a dynamic dual pathway model of approach coping with collective disadvantage. *Personality and Social Psychology Review* 16(2):180-199.
Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Science* 359(6380):1146-1151.
Wilson TD, Gilbert DT (2008) Explaining away: A model of affective adaptation. *Perspectives on Psychological Science* 3(5):370-386.
Wróbel M, Imbir KK (2019) Broadening the perspective on emotional contagion and emotional mimicry: The correction hypothesis. *Perspectives on Psychological Science* 14(3):437-451.
Yan L, Peng J, Tan Y (2015) Network dynamics: how can we find patients like us? *Information Systems Research* 26(3):496-512.
Zeng R, Zhu D (2019) A model and simulation of the emotional contagion of netizens in the process of rumor refutation. *Scientific reports* 9(1):1-15.