Emotional Valence Shifts and User Behavior on Twitter, Facebook, and YouTube

Ema Kušen[1], Mark Strembeck[1,2,3], and Mauro Conti[4]

[1] Vienna University of Economics and Business, Austria
[2] Secure Business Austria Research Center (SBA)
[3] Complexity Science Hub Vienna (CSH)
[4] Università di Padova, Italy

Abstract. In this paper, we present a study on 5.6 million messages that have been sent via Twitter, Facebook, and YouTube. The messages in our data-set are related to 24 systematically chosen real-world events. For each of the 5.6 million messages, we first extracted emotion scores based on the eight basic emotions according to Plutchik's wheel of emotions. Subsequently, we investigated the effects of shifts in the emotional valence on the messaging behavior of social media users. In particular, we found empirical evidence that prospectively *negative* real-world events exhibit a significant amount of shifted (i.e., *positive*) emotions in the corresponding messages. To explain this finding, we use the theory of *social connection* and *emotional contagion*. To the best of our knowledge, this is the first study that provides empirical evidence for the *undoing hypothesis* in online social networks (OSNs). The undoing hypothesis assumes that positive emotions serve as an antidote during negative events.

1 Introduction

In online social networks (OSNs), news travel fast and reach a large number of users within a short period of time [24,33]. Such a rapid information diffusion comes with valuable social benefits such as using OSNs to help save lives during the 2011 Tsunami disaster in Japan [35] as well as the Red River floods and Oklahoma fires in 2009 [43]. However, although having a great potential to do good for society, OSNs have also been recognized as a convenient tool to (positively or negatively) influence people. For example, a number of recent studies indicated that Twitter, Facebook, and YouTube have been used to spread terrorist propaganda [44] and negatively influence users (online radicalization) [2,7,26,36].

In this context, emotions have generally been recognized as an important factor in influencing or manipulating with people's opinions and beliefs [29]. In particular, recent studies indicated that emotions can be passed through online interactions from one user to another [11,21], resulting in a so-called *emotional contagion*. In addition, numerous studies reported on user reactions to emotionally-charged messages. For example, [10,38] found that users tend to pay more attention to the negative messages, while [4,37] presented a contradictory finding (i.e., positive messages receive more attention). The common denominator in either case is that emotions conveyed in OSN messages have the potential to trigger a strong emotional reaction in people [5,9,25,28].

This paper extends our prior analysis presented in [23]. In particular, we study the impact of emotions on the messaging behavior of OSN users on Twitter, Facebook, and YouTube. To this end, we performed an emotion analysis over 5.6 million social media messages that occurred in 24 systematically chosen real-world events. For each of these messages, we derived emotion scores concerning the eight basic emotions according to Plutchik's wheel of emotions [32]. In general, we found that people tend to conform to the base emotion of a particular event. However, we also found empirical evidence that in all three OSNs prospectively negative real-world events are accompanied by a substantial amount of shifted (i.e., positive) emotions in the corresponding messages. In order to explain this finding, we use the theory of *social connection* and *emotional contagion*. To the best of our knowledge, this is the first study that provides empirical evidence for the *undoing hypothesis* in online social networks (OSNs).

The remainder of this paper is organized as follows. In Section 2, we discuss related work. Next, Section 3 describes our data analysis procedure followed by a detailed report on our results in Section 4. Subsequently, Section 5 discusses our findings and Section 6 concludes the paper.

2 Related work

Prior studies predominantly examined the impact of sentiment polarities and emotions on information diffusion over OSNs. For example, Zhang and Zhang [46] examine the impact of *emojis* on message diffusion patterns over a dataset containing about 12 million Weibo messages. In particular, they found that positive and negative emojis result in the same effects with respect to re-tweets and replies. In fact, both groups of emojis have a positive effect on the number of replies a message receives and a negative effect on the re-tweet count. Other studies examined textual cues to identify a set of emotions or sentiment polarities. For example, Kim et al. [20] conducted a questionnaire-based study to examine the role of emotional valence on the diffusion of anti-tobacco messages. They found that positive emotions boost the transmission of messages, while negative ones had the opposite effect.

In [12], Ferrara and Yang extracted emotion polarities for about 19 million tweets by applying the SentiStrenght algorithm [40]. In particular, they studied four aspects of information diffusion over Twitter: re-tweet count, like count, the speed of diffusion, and the scope of the diffusion. The results of the study show a clear evidence of the *Pollyanna hypothesis* [8], which refers to the human preference to like positive messages more than negative and neutral ones. Moreover, in terms of the scope of the diffusion, the study showed that positive messages spread wider than negative and neutral ones. However, it also indicates that messages carrying negative and neutral sentiments spread faster than positive ones.

Another study that utilized SentiStrength to obtain emotion polarities [37] studied the effects of polarities on the re-tweet count and the speed of re-tweeting during the 2011 German state parliament elections. The findings suggest that emotionally-charged tweets tend to be re-tweeted more often than the neutral ones, which is in line with the findings presented in [30,41,42]. In particular, tweets carrying a negative sentiment are strongly associated with an increase in the re-tweet rate.

In [16], Gruzd et al. analyzed sentiment polarities from tweets related to the 2010 Winter Olympics. They found that a user's position in the social network can be regarded as an indicator of the user's tendency to post positive or negative messages. In specific, Gruzd et al. showed that users who tweeted predominantly positive messages generally have more followers on Twitter, while users who tweeted more negative messages exhibited a higher tweet-per-user rate.

In [41], Trung et al. assigned sentiment polarities to a data-set of about 11.000 tweets by using a Bayesian classifier trained on the annotated tweets from three domains: news, industry, and entertainment. In particular, they studied three aspects of information diffusion: the number of re-tweets, speed of diffusion, and the scope of diffusion. In contrast to the findings from [12], Trung et al. found that all emotionally charged messages (i.e., positive as well as negative ones) spread wider (i.e., to more users) than the neutral ones. However, in terms of the speed of diffusion, they found no significant difference among the neutral, positive, and negative messages.

Such a dissonance in the findings might result from the fact that existing papers predominantly study the diffusion patterns only with respect to one particular domain of interest (such as health-care, politics, popular culture, or sports; see, e.g., [20,37]) which makes it difficult to generalize the respective findings across domain borders. While some papers report on diffusion patterns of messages belonging to different domains, the corresponding papers do not follow a systematic approach for studying the differences between domains with respect to information diffusion patterns (see, e.g., [12,41]).

Even though a number of effects relating to sentiment polarities have been studied in the related work, aspects beyond the effects of sentiment polarities on the information diffusion in OSNs have rarely been investigated. For example, Berger [6] discusses the effects of emotional arousal on information sharing. In particular, the study distinguishes between dimensions of emotions other than emotion polarities only. Berger found that arousal increases the likelihood for sharing an information, regardless of whether the respective information conveys a positive or a negative sentiment. Two other studies [39,18] consider anger, anxiety, awe, and sadness, as annotated by human encoders. The results of both studies indicate that anger and awe increase the content sharing behavior, while sadness and anxiety were negatively associated with content diffusion.

3 Data analysis procedure

In this section, we outline the four main phases of our study (see Figure 1). In Section 3.1 we describe our data extraction procedure. Section 3.2 provides more details on cleaning the raw data-set followed by Section 3.3 in which we outline the heuristics used to identify emotions and their corresponding emotions scores. Finally, Section 3.4 provides details on our data analysis procedure and the scope of this paper.

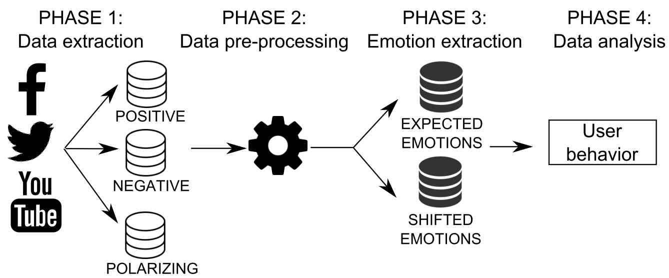

Fig. 1. Research phases.

3.1 Data extraction

In order to study the impact of emotional valence shifts on information diffusion, we systematically identified 24 real-world events that belong to 5 different domains (sports, politics, popular culture, war and terrorism, and other) and collected more than 5.6 million corresponding messages published on Twitter, Facebook, and YouTube. The 24 events have been selected such that they fall in one of the following categories:

1. events that potentially trigger positive emotions (e.g., birthday celebrations, festivities)
2. events that potentially trigger negative emotions (e.g., war, terror, death)
3. emotionally polarizing events (e.g., presidential elections, controversial topics)

We summarize the events extracted for each category in Table 1, where N refers to the number of messages in each category, followed by the relative size of each category in our data-set (in percent).

To extract the public messages, we used the corresponding API's offered by Twitter, Facebook, and YouTube. In particular, we used Twitter's Search API[5] to extract publicly available tweets. For this extraction, we used a pre-defined list

[5] https://dev.twitter.com/overview/api

Event	T messages	F messages	Y messages
Negative	(N=1,490,495; 34%)	(N=161,898; 14%)	(N=33,563; 22%)
1) Erdogan's threats to EU	804	36	1,160
2) Anti-Trump protests	381,982	218	5,270
3) Death of Leonard Cohen	89,619	43,808	11,820
4) Death of Colonel Abrams	1,253	6	18
5) Aleppo bombings	995,561	94,116	8,129
6) Seattle shooting	73	2,085	660
7) Lufthansa strikes	3,387	156	26
8) Ransomware incidents	2,564	1,012	3,724
9) Yellowstone hotpot	15	9,980	2,156
10) Earthquake central Italy	15,237	10,481	600
Positive	(N=1,115,587; 25%)	(N=328,792; 29%)	(N=83,394; 55%)
11) Rosberg wins F1	215,703	586	804
12) Murray wins ATP	62,184	169	495
13) Rosberg retires	34,201	10,817	353
14) "Beauty and the Beast" trailer	138,979	73,399	42,180
15) Fantastic beasts trailer	64,264	51,468	17,829
16) Vienna ComiCon	704	4,693	57
17) Miley Cyrus birthday	76,270	3,014	197
18) Pentatonix album release	9,341	70	15,159
19) Ellen Degeneres medal of freedom	73,854	184,519	4,450
20) Thanksgiving	440,087	57	1,870
Polarizing	(N=1,812,573; 41%)	(N=637,945; 57%)	(N=35,657; 23%)
21) Death of Fidel Castro	720,548	21,938	2,068
22) Austrian elections 2016	2,558	3,351	1,096
23) The Walking Dead S07 premiere	198,042	34,486	5,136
24) US elections 2016	891,425	578,170	27,357

Table 1. List of events extracted from Twitter (T), Facebook (F), and YouTube (Y).

of hashtags and restricted the search to English language tweets only. Moreover, we collected the tweets related to the 24 events by starting with the date of an event announcement and stopped seven days after. In total, the extraction resulted in 4,418,655 tweets. Furthermore, we used YouTube Comment API[6] to extract 152,614 publicly available YouTube comments on a set of 98 manually selected YouTube videos related to the 24 events in our study. In addition, we used Facebook's Graph API[7] to extract 1,128,635 publicly available Facebook comments on 69 manually selected Facebook posts related to the 24 events.

3.2 Data pre-processing

After the data extraction, we cleaned the raw data-set by removing entries that contained uninformative content with respect to emotion extraction (e.g., entries that consisted of URLs only). Thus, after pre-processing our data-set included messages that contained either text, emoticons, or a combination of both. Moreover, since our Facebook and YouTube data-sets included messages that were written in languages other than English, we used Python's *langdetect*[8] language detection library to identify the language of a particular message and removed messages that were not written in English.

[6] https://github.com/philbot9/youtube-comment-api
[7] https://developers.facebook.com/docs/graph-api
[8] https://pypi.python.org/pypi/langdetect

The final number of messages for each OSN after pre-processing is shown in Table 1.

3.3 Emotion extraction

After pre-processing the overall data-sets, we further processed each message by lemmatizing it and tagging words to their corresponding part-of-speech category. We then identified the presence of specific emotions conveyed in the messages and computed an emotion intensity score for each message by applying a customized emotion-extraction script (see [22] for further details on the heuristics used and an evaluation). In general, the procedure encoded in the script:

1. identifies the presence of Plutchik's eight basic emotions (anger, fear, disgust, sadness, joy, trust, anticipation, surprise) [32] by relying on the NRC word-emotion lexicon [27],
2. assigns an intensity score for each emotion in every tweet by counting the number of words in the NRC lexicon that are associated with an emotion and multiplies them with a score provided in the AFINN lexicon[9] [31],
3. deals with negation (e.g., "I am *not* happy.") by shifting the valence of a word (e.g., the term *not happy* results in a negative emotion score for joy: *joy = -1*),
4. deals with intensifiers (e.g., *very* happy), downtoners (e.g., *hardly* happy), maximizers (e.g., *absolutely* happy),
5. identifies misspellings and repeated letters to find "hidden" boosters (e.g., "I am *sooooo* happy" is regarded as "I am *so* happy."),
6. as noted in [19], emoticons are often used instead of words to express emotions. Thus, our script also identifies emoticons and categorizes them as positive (e.g., happy face :) and laughing face :D), negative (e.g., sad face :(and crying face :'(or broken heart </3), or conditional (heart <3) (see also [1]). Note that we regard a heart (<3) neither as a positive or a negative emotion-carrier because its meaning depends on the context of its use. For example, in a sentence "You will be missed <3", it is used in a negative context (sadness), while in the sentence "I love him so much <3", the emoticon is used in a positive context (joy). Thus, to correctly interpret the emoticon <3 we first identify the dominant emotion in the tweet and then assign the corresponding emotion score.

Moreover, for our analysis we also extended the NRC dictionary with a list of common acronyms used in social media (such as LOL, WTF, and YOLO).

3.4 Data analysis and research questions

First, we analyzed how social network users express specific emotions during positive, negative, and polarizing events. Next, we separated each data-set into

[9] The AFINN lexicon [31] contains scores corresponding to the emotional valence intensity of a given word. For example, words such as *sad* and *depressed* are classified as negative words, but the latter has a weaker intensity compared to the former word.

a subset that conveys *expected emotions* and a subset that conveys *shifted emotions* in terms of their valence. In particular, we treat emotions of a shifted valence as *unexpected* emotions (for example, a positive event receives messages that predominantly convey negative emotions). We then analyzed how the user behavior in each subset is influenced by expected and shifted emotions.

Our analysis was guided by the following research questions.

RQ1: Which emotions are expressed during positive, negative, and polarizing events?

For RQ1, we searched for emotions communicated during positive, negative, and polarizing events. For this research question, it was of particular interest whether emotions belonging to a specific emotional valence (i.e., positive or negative) dominate in an event category.

To answer the first research question, we obtained the average intensity of each of the eight basic emotions for each event (see Table 1). Moreover, we computed the bivariate correlation between each pair of emotions.

RQ2: Which messaging behavior do users exhibit during positive, negative, and polarizing events?

For RQ2, we studied how users react to the three types of events (positive, negative, polarizing) in terms of platform-specific user actions. Thus, for Twitter we consider the number of re-tweets, number of likes, tweeting rate, tweeting count per user, and one-to-one communication. For Facebook, we study the number of replies to a comment, like count, daily time rate, and number of comments per user. And for YouTube, we examine the number of replies and likes to a comment, as well as a daily time rate, and the number of comments per user.

RQ3: Are there differences in the messaging behavior when users are faced with messages that convey expected emotions and those with a shifted emotional valence?

For RQ3, we study how users respond to the emotions conveyed in messages. In particular, we contrast the behavior towards the expected emotions and the shifted emotions and provide a time-series analysis of each.

4 Results

In this section, we first show the intensities of emotions expressed in each OSN during positive, negative, and polarizing events (Section 4.1). We then examine the user behaviour as a reaction to emotionally-charged messages and show a time-series analysis of the shifted emotions with respect to the expected emotions (Section 4.2).

4.1 Emotion intensity during positive, negative, and polarizing events

Our analysis shows that OSN users express emotions with a similar intensity over Twitter, Facebook, and YouTube upon encountering polarizing, positive, and negative events (see Figure 2, 3, and 4 where negative emotions are colored red (anger, sadness, disgust, fear), positive green (joy, trust), and conditional (i.e., context-dependent) emotions are colored yellow (surprise, anticipation)).

Furthermore, for each category (positive, negative, polarizing) Figure 5 shows the respective difference between the *expected* and the *shifted* emotions. Thus, in positive events, the negative emotion score (shifted) is subtracted from the positive emotion score (expected). In negative events, the positive emotion score (shifted) is subtracted from the negative emotion score (expected). Moreover, since there is no expected emotion for polarizing events, we chose to subtract the positive emotion score from the negative emotion score.

To mitigate bias in the results which may emerge due to the length of a message (i.e., tweets are restricted to 140 characters[10], while Facebook and YouTube posts can be considerably longer), we present the scores of each emotion averaged over the sentence count. Finally, to show the relative presence of each emotion in the data-set, we divide the averaged emotion scores e (based on the sentence count S) with the message count in the data-set (N)

$$\frac{\sum_{i=1}^{n} \frac{e_i}{S_i}}{N}.$$

We found that messages sent during polarizing events exhibited no tendency of a particular group of emotions to greatly dominate over the other, as compared to the positive and negative events. As shown in Figure 2, OSN users expressed positive and negative emotions with a similar intensity. For polarizing events, Figure 5 further shows that the relative difference between the scores assigned to negative and positive emotions only exhibit a low difference (0.02 for Twitter, 0.08 for Facebook, and 0.04 and YouTube). These results were expected to a certain degree, as users tend to either approve/support or disapprove/oppose a topic of interest during polarizing events (e.g., political campaigning).

With respect to OSN-related differences in emotional intensities during polarizing events, we found that our Facebook data-set contained 39% emotionally neutral messages, while YouTube and Twitter messages were more emotionally charged (24% and 21% emotionally neutral messages, respectively). These platform related differences are depicted in Figure 2.

In contrast, and as shown in Figure 3, positive events exhibited a higher intensity of positive emotions (joy, trust) as compared to negative emotions (anger, fear, disgust, sadness). In fact, in positive events the differences between the intensities of positive and negative emotions are considerably higher (0.70 for Twitter, 0.40 for Facebook, and 0.26 for YouTube) as compared to the data-sets for polarizing and negative events.

Interestingly, when comparing the intensities of specific emotions communicated over the three OSN platforms, we found that a single tweets carries on average a more intense positive emotion, when compared to messages sent via

[10] Note that the increased limit of 280 characters that has been introduced by Twitter in November 2017 was not in effect during our data-extraction period.

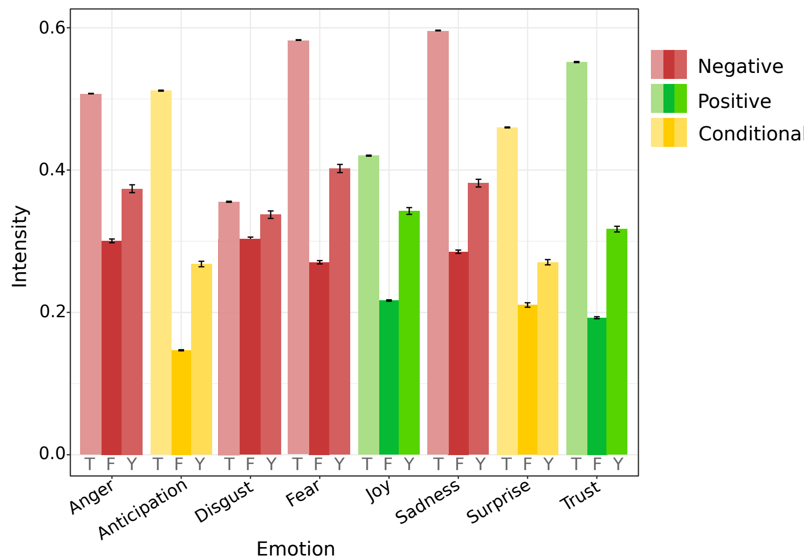

Fig. 2. Emotions expressed during polarizing events on Twitter (T), Facebook (F), and YouTube (Y).

the other two platforms. However, when observing the shifted emotions in positive events (i.e., negative emotions related to positive events), our results reveal that YouTube users tend to express (on average) more intense negative emotions as compared to Facebook and Twitter users. This difference is particularly evident in Figure 5, where the difference between positive and negative emotional intensities on YouTube is only 0.26, compared to 0.70 on Twitter and 0.40 on Facebook.

Figure 4 shows emotional intensities communicated during negative events. As expected, negative events showed a comparatively higher intensity of negative (expected) emotions on Twitter. However,we also found a considerable presence of positive emotions (see Figure 4) and only a low difference between the intensities of negative and positive emotions (see Figure 5). In contrast to Twitter, emotions in messages related to negative events communicated on YouTube and Facebook are even predominantly positive on average (see Figure 5, the difference between negative and positive emotions is -0.05 on YouTube and -0.04 on Facebook, i.e., the shifted emotion is slightly dominant over the expected emotion).

Next, we examine whether different emotions belonging to the same emotional valence are communicated jointly in a single message. To this end we performed a bivariate correlation analysis for each pair of emotions (e.g., anger with disgust, anger with joy). Our results show a high Spearman's ρ coefficient between *disgust* and *anger* (ρ=0.81) as well as *sadness* and *fear* (ρ=0.87) on

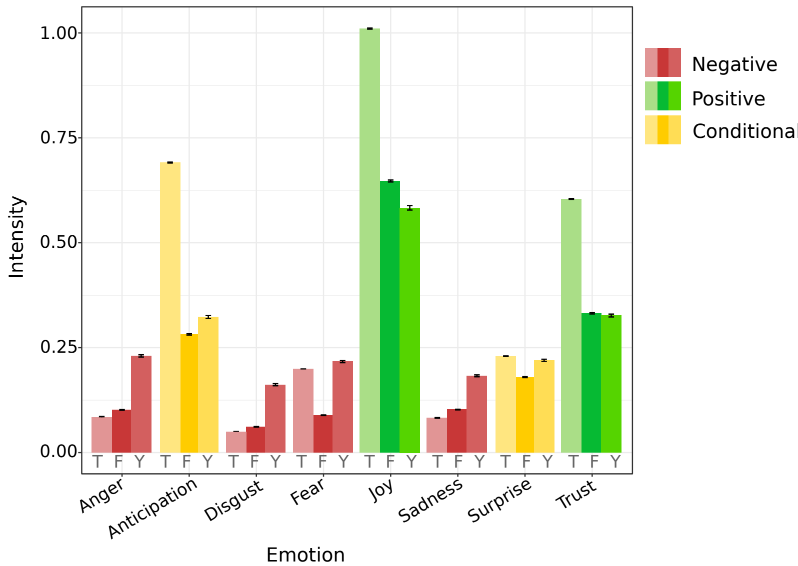

Fig. 3. Emotions expressed during positive events on Twitter (T), Facebook (F), and YouTube (Y).

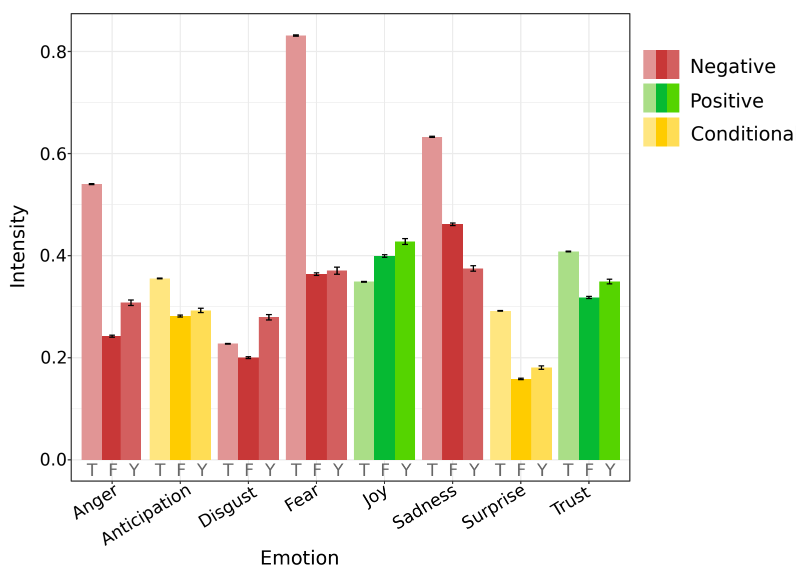

Fig. 4. Emotions expressed during negative events on Twitter (T), Facebook (F), and YouTube (Y).

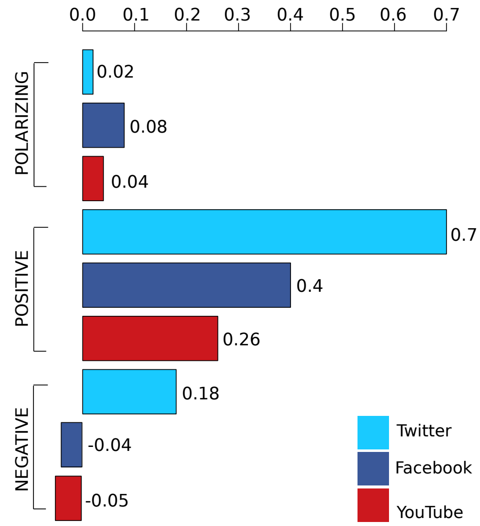

Fig. 5. Difference between the emotional intensity of expected and shifted emotions.

YouTube, between *sadness* and *anger* (ρ=0.92), as well as *joy* and *trust* (ρ=0.86) on Facebook, and between *sadness* and *anger* (ρ=0.70) on Twitter.

During positive events, *anger* and *disgust* were highly correlated (ρ=0.83) on YouTube, while the same holds for *sadness* and *fear* (ρ=0.89) on Facebook. Negative events exhibited a high correlation between *disgust* and *anger* (ρ=0.83), as well as *fear* and *sadness* (ρ=0.82) on YouTube, a high correlation between *fear* and *anger* (ρ=0.83) as well as *fear* and *sadness* (ρ=0.81) on Facebook, and a high correlation between *sadness* and *fear* (ρ=0.71) on Twitter.

Based on the aforementioned results, we conclude that emotions belonging to the same emotional valence tend to be be communicated together in a single OSN message. This observation is particularly evident in our Twitter data-set (see Figure 6a), where users are limited to 140 characters only, i.e., Twitter users only have limited space express their emotions and opinions. We also observed that when users are allowed to post longer messages, there is a higher correlation between positive and negative emotions (e.g., posts that convey joy also convey anger) (see Figures 6b and 6c).

However, it is worth mentioning that emotions belonging to two different categories in terms of emotional valence (positive vs. negative) are weakly or at most moderately correlated as compared to emotions belonging to the same emotional valence (e.g., joy and trust; anger and disgust). In a similar way, we found that different negative emotions are only weakly or moderately correlated during positive and polarizing events.

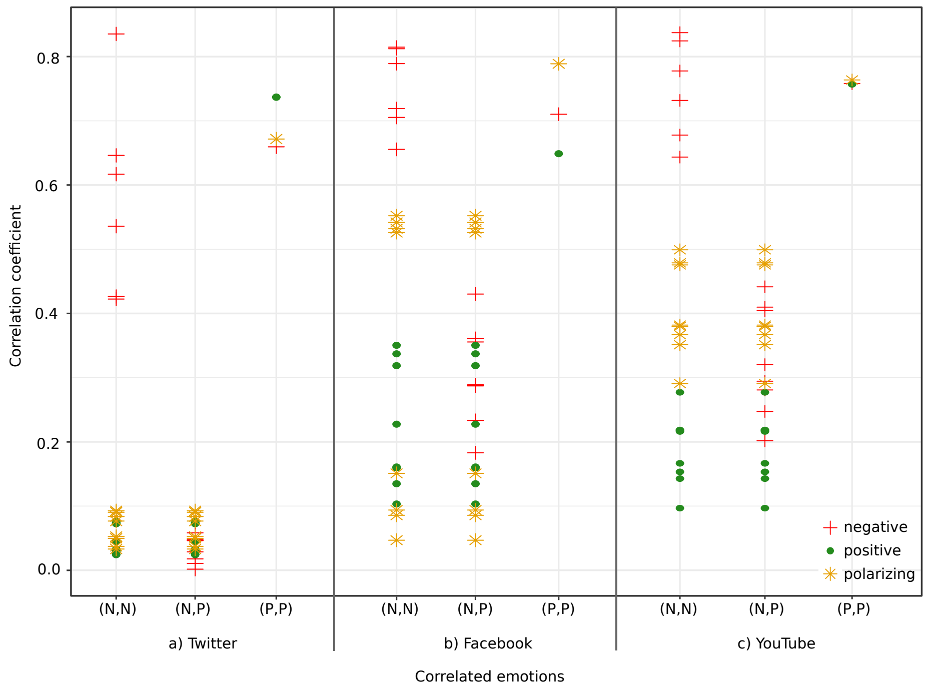

Fig. 6. Correlation between pairs of negative emotions (N,N), negative and positive emotions (N,P), and pairs of positive emotions (P,P) on Twitter, Facebook, and YouTube during negative (red cross), positive (green dot), and polarizing (yellow star) events.

4.2 User behavior

For the purposes of this paper, OSN user behavior is defined as all user actions that result in sending or forwarding a public message/comment as well as user actions related to appreciating ("liking") a public message/comment. Moreover, in our analysis we also consider the number of messages per user and the speed of message generation (i.e., the number of messages per time unit).

User behavior on Twitter: As shown in Table 2, positive events trigger the highest number of re-tweets and likes. This shows a tendency of users to prefer engaging in positive discussions rather than negative (which confirms the *Pollyanna hypothesis*, see also [8,12]). Moreover, we found that users tend to engage in a one-to-one communication (via *@username*) more frequently during positive events than during negative or polarizing ones. However, the results also indicate that users tend to send comparatively more tweets during negative events (2.86 tweets per user) than during polarizing (1.92 tweets per user) or positive (1.62 tweets per user) events. This finding corresponds to the ones discussed in [6], which suggest that emotions of a high arousal (such as anger) increase the social transmission of information. Interestingly, in our data-set polarizing events exhibited the highest tweeting rate per minute (49.42 tweets/minute).

The results of the Welch's t-test (with a 95% significance level) indicate that there is a significant difference in how users respond to tweets conveying expected emotions and those containing shifted emotions. Apart from one exception in the like count ($p>0.05$), all other tweeting behaviors that we considered exhibit a clear pattern: expected emotions receive more re-tweets. Moreover, for each of the 24 events that we analyzed, OSN users tend to send more tweets per minute that convey an expected emotion. Another interesting insight can be observed in the one-to-one communication (social sharing) of expected emotions between OSN users. During positive events, users tend to engage in a one-to-one communication by sharing predominantly positive emotions. Analogously, negative events exhibit a comparable pattern (see Table 2).

	Negative	**Positive**	**Polarizing**
Retweet count	1600.10	**5677.63**	4821.90
	$t=1.98$, $p<0.05$	$t=243.511$, $p<0.05$	
Like count	0.98	**1.49**	1.22
	$t=-0.97$, $p>0.05$	$t=0.98$, $p>0.05$	
Time rate	37.58	42.48	**49.42**
	$t=25.94$, $p<0.05$	$t=57.47$, $p<0.05$	
Tweet per user	**2.86**	1.62	1.92
	$t=13.33$, $p<0.05$	$t=26.04$, $p<0.05$	
@ count	1.02	**1.19**	1.02
	$t=25.77$, $p<0.05$	$t=69.93$, $p<0.05$	

Table 2. Twitter user behavior in positive, negative, and polarizing events (confidence interval 95%).

Figure 7a) shows the time series of tweets separated into those that carry a predominantly positive emotion and those that carry a negative emotion. In particular, there is a noticeable smaller number of tweets that convey negative (i.e., shifted) emotions during positive events. However, our data also shows that, although small in size, negative tweets occurred consistently throughout the extraction period (*mean(set difference)=36668.11, sd(set difference)=45844.64*)[11].

In contrast, we found that during negative events events a considerable number of tweets conveying positive (i.e., shifted) emotions occur (*mean(set difference)=4231.67, sd(set difference)=5603.162*). This observation was consistent over the entire extraction period (see Figure 7b). Interestingly, our data-set also revealed unexpected cases where the positive tweets (i.e., the shifted emotions) even exceed the (expected) negative tweets (the largest difference between the two subsets is 6403 tweets).

User behavior on Facebook: Table 3 shows that Facebook users also slightly prefer replying to and liking Facebook posts that have a positive emotion score,

[11] Set difference refers to the difference between the count of the expected emotions and shifted emotions, while *sd* stands for standard deviation.

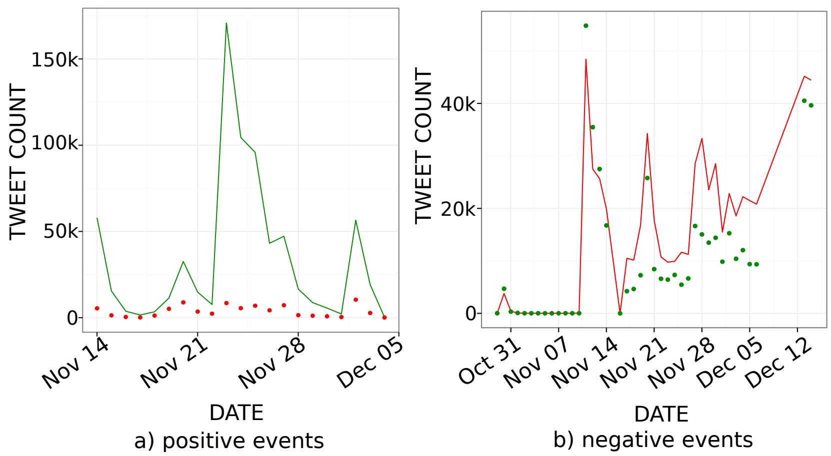

Fig. 7. Occurrences of negative messages during positive events on Twitter and vice versa. Positive messages are depicted in green and negative messages in red.

while in polarizing events we again found the highest average number of comments per unit of time.

In particular, the results of the Welch's t-test indicate a significant difference in the effects of expected vs. shifted emotions in each data-set (see Table 3). For negative events, we found that users tend to reply and comment predominantly on negative posts and also send more messages that covey negative emotions per day, as compared to positive posts (replies: t=2.39, $p<0.05$; comments: t=1.19, $p<0.05$; time rate: t=2.69; $p<0.05$). For positive events, we found one statistically significant results for the comment rate per user (t=18.32; $p<0.05$), which indicates that users tend to comment more on positive posts during positive events rather then on negative posts.

	Negative	**Positive**	**Polarizing**
Replies count	0.17	**0.25**	0.09
	t=2.39, p<0.05	t=0.83, p>0.05	
Like count	2.78	**2.89**	1.79
	t=1.93, p>0.05	t=0.61, p>0.05	
Time rate (daily)	505.93	713.21	**4012.23**
	t=2.69, p<0.05	t=1.04, p>0.05	
Comment per user	1.19	1.31	**2.94**
	t=31.29, p<0.05	t=18.32, p<0.05	

Table 3. Facebook user behavior in positive, negative, and polarizing events (confidence interval 95%).

By observing the time-series plots in Figure 8b), we can see that the temporal patterns of expected and shifted emotions during negative events resemble those we found on Twitter. In particular, a considerable number of messages conveying positive emotions are sent during negative events. The positive (shifted) emotions even dominate the negative emotions at certain dates (see the green dots in Figure 8b).

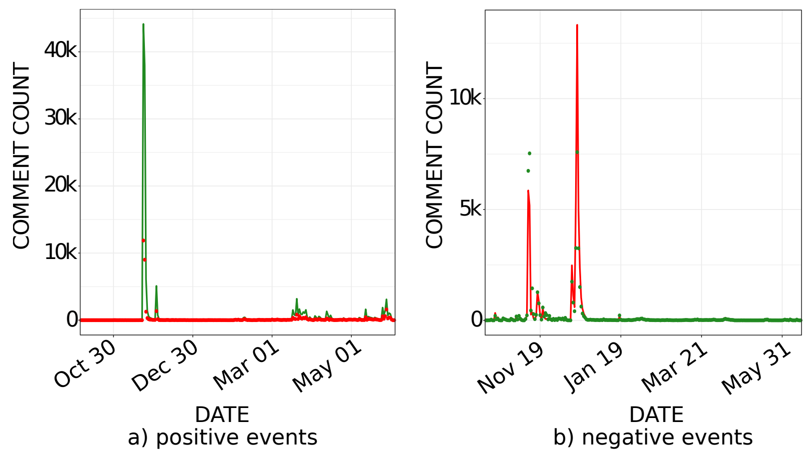

Fig. 8. Occurrences of negative messages during positive events on Facebook and vice versa. Positive messages are depicted in green and negative messages in red.

With respect to the temporal patterns of negative messages sent during positive events, Figure 8a) shows that positive emotions dominate over the negative ones throughout the entire data-extraction period. Again, this observation is analogous to the temporal patterns observed on Twitter (see Figure 7a).

User behavior on YouTube: Similarly to Facebook and Twitter, YouTube users also prefer to "like" comments on YouTube videos relating to positive events (see Table 4), as compared to comments on YouTube videos relating to polarizing or negative events.

However, unlike Facebook and Twitter users, YouTube users exhibit higher reply counts to comments on YouTube videos that are depicting a polarizing event (e.g., political campaigning, such as TV debates). Moreover, we also observed that YouTube users exhibit the highest rate of comments per time unit for videos on positive events, while Facebook users exhibited this behavior for polarizing events and Twitter users for negative events.

Analogously to the results for Facebook and Twitter, Table 4 shows that YouTube users tend to comply with the base mood of an event by replying more to negative messages during negative events ($t=3.55$, $p<0.05$), and positive messages during positive events ($t=4.16$, $p<0.05$). In the same way, YouTube users also tend to "like" positive messages during positive events ($t=2.13$, $p<0.05$) and send more negative comments per time unit during negative events ($t=2.15$, $p<0.05$).

	Negative	Positive	Polarizing
Replies count	0.69	0.54	**0.99**
	t=3.55, $p<0.05$	t=4.16, $p<0.05$	
Like count	4.59	**7.24**	5.3
	t=-0.36, $p>0.05$	t=2.13, $p<0.05$	
Time rate (daily)	958.94	**4097.81**	1782.85
	t=2.15, $p<0.05$	t=1.052, $p>0.05$	
Comment per user	1.81	1.67	**2.34**
	t=0.33, $p>0.05$	t=0.43, $p>0.05$	

Table 4. YouTube user behavior in positive, negative, and polarizing events (confidence interval 95%).

For our YouTube data-set, Figure 9b) shows that during negative events we again found a considerable number of messages with a positive (i.e., shifted) emotion score. Similar, to our findings for Facebook and Twitter, positive messages even dominate over negative messages on certain dates.

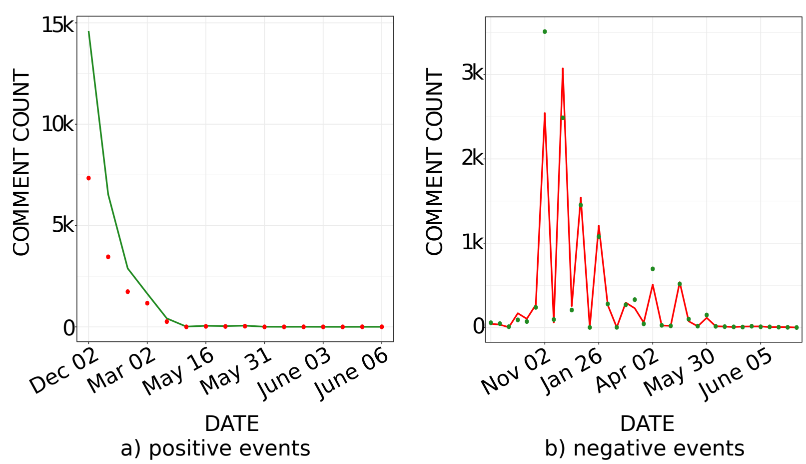

Fig. 9. Occurrences of negative messages during positive events on YouTube and vice versa. Positive messages are depicted in green and negative messages in red.

In positive events though, we again predominantly found messages conveying positive emotions (see Figure 9a).

5 Discussion

Our results bring forth interesting insights into how OSN users behave during positive, negative, and polarizing events when faced with shifted emotions. In our study, we considered OSN user behavior in terms of re-tweets and one-to-one communication (on Twitter), replies (on YouTube and Facebook), as well as likes, the number of messages per user, and the speed of message generation (time rate) (on all three OSNs). Consistent with previous work from the field of psychology, we found that to a considerable extent positive emotions also occur during negative events. An explanation for the observed phenomenon can be attributed to *social connection* [3,14] as one of the fundamental human needs. In fact, previous studies indicated that even in the times of sorrow and anxiety, people tend to eventually be supportive and positive towards one another (see, e.g., [34]).

In our data-set, we found examples of people explicitly calling for social bonding during emotionally tough events (e.g., after the 2016 earthquake in central Italy, people posted: *"please join us as we #PrayforItaly"*) and a public and explicit expression of vulnerability that triggers compassion (e.g., *"Oh dear world, I am crying tonight"*, during Aleppo bombings). Moreover, our data-set indicates that people tend to show appreciation and love for a person they care about or admire (e.g., a deceased singer, such as Leonard Cohen) or even comfort each other and send messages of hope during natural disasters (e.g., earthquake in Italy) or war (e.g., Aleppo bombing). Thus, we found empirical evidence that supports the *undoing hypothesis* [13], which states that people tend to use positive emotions as an antidote to undo the effects of negative emotions.

For Twitter, our results further indicate that expected emotions result in more re-tweets. We thereby confirm findings from [17], which suggested that people prefer sharing messages that correspond to the emotional valence of the respective event (i.e., users tend to pass along negative tweets during negative events). This might be attributed to the human tendency to conform to the situation. However, we were also able to observe a similar phenomenon on Facebook and YouTube. Beyond the mere sharing of existing messages, we also observed that users in general prefer replying to and liking messages that convey emotions which correspond to the base emotion of an event. The same holds for the message generation time rate and message per user rate on all three OSN platforms we considered in our study.

Other studies bring an additional interesting insight into the inter-personal interactions over social media, which might explain our observations of users to conform to the base emotion. According to [45], emotional messages tend to influence the emotions conveyed in other users' messages. This phenomenon, called *emotion contagion* in [45], emerges from the social connections of OSN users (or their position in the network). In this context, we observed that messages sent by "fans" (we follow an assumption that fans follow their idols on OSNs with a high probability) tend to be congruent with the messages sent by their idols. For example, a tweet posted by Pentatonix in which they announce the release of their new album triggered positive reactions from their fans. Note, however, that considering structural network properties alone might not be sufficient to study emotional contagion. For example, OSN users might form a connection (e.g., "follow") with an influential user (e.g., a politician) even though they do not actually agree with the person's ideology or point of view. Thus, the emotions passed by an influential user might also be shifted by his/her followers due to disagreement or sarcasm [15]. We leave the study of this issue for our future work.

6 Conclusion

In this paper, we presented a systematic study concerning the influence of emotional valence shifts on the messaging behavior of OSN users on Twitter, Facebook, and YouTube. Our study is based on a data-set including 5.6 million messages belonging to 24 real-world events. The events have been subdivided into positive, negative, and polarizing, and for each of these event categories we analyzed the intensity of Plutchik's eight basic emotions (sadness, fear, anger, disgust, joy, trust, surprise, anticipation). Thereby, our paper complements existing studies by not only considering polarizing emotion scores (positive vs. negative) but also the influence of eight basic emotions according to Plutchik's wheel of emotions. In order to study the impact of the eight emotions on user behavior in OSNs, we considered user reactions to emotionally-charged messages.

Our findings indicate that people generally prefer sharing messages that correspond to the emotional valence of the respective event. Furthermore, we conducted a time-series analysis and found a clear distinction between positive and negative events, with respect to shifted emotions. In particular, we found that positive events trigger a comparatively smaller number of negative messages. However, while negative events exhibit predominantly negative messages, they are accompanied by a surprisingly large number of positive messages. In fact, our analysis shows that in negative events positive messages may even exceed the negative ones on all three OSN platforms. To the best of our knowledge, this is the first study which found empirical evidence that supports the *undoing hypothesis* in online social networks.

In our future work, we plan to extend our analysis to studying messages written in languages other than English. In addition, we plan to investigate how sarcasm is related to shifts in the emotional valence.

Acknowledgment

Mauro Conti is supported by a Marie Curie Fellowship funded by the European Commission (agreement PCIG11-GA-2012-321980). This work is also partially supported by the EU TagItSmart! Project (agreement H2020-ICT30-2015-688061), the EU-India REACH Project (agreement ICI+/2014/342-896), by the project CNR-MOST/Taiwan 2016-17 "Verifiable Data Structure Streaming", the grant n. 2017-166478 (3696) from Cisco University Research Program Fund and Silicon Valley Community Foundation, and by the grant "Scalable IoT Management and Key security aspects in 5G systems" from Intel.

References

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment Analysis of Twitter Data. In: Proc. of the Workshop on Languages in Social Media. pp. 30–38. Association for Computational Linguistics (2011)
2. Alarid, M.: Recruitment and Radicalization: The Role of Social Media and New Technology. CCO publications (2016)
3. Baumeister, R.F., Leary, M.R.: The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin 117, 497–529 (1995)
4. Bayer, M., Sommer, W., Schacht, A.: Font size matters—emotion and attention in cortical responses to written words. PLOS ONE 7(05), 1–6 (2012)
5. BBC.com: Study: Social networks like Facebook can spread moods. (2014), `http://www.bbc.com/news/technology-26556295`
6. Berger, J.: Arousal increases social transmission of information. Psychological Science 22(7), 891—893 (2011)
7. Bessi, A., Ferrara, E.: Social bots distort the 2016 U.S. Presidential election online discussion. First Monday 21(11) (2016), `http://firstmonday.org/ojs/index.php/fm/article/view/7090`
8. Boucher, J., Osgood, C.E.: The pollyanna hypothesis. Journal of Verbal Learning and Verbal Behavior 8(1), 1 – 8 (1969)
9. Coviello, L., Sohn, Y., Kramer, A.D.I., Marlow, C., Franceschetti, M., Christakis, N.A., Fowler, J.H.: Detecting emotional contagion in massive social networks. PLOS ONE 9, 1–6 (03 2014)
10. Faraon, M., Stenberg, G., Kaipainen, M.: Political campaigning 2.0: The influence of online news and social networking sites on attitudes and behavior. eJournal of eDemocracy and Open Government 6(3), 231–247 (2014)
11. Ferrara, E., Yang, Z.: Measuring emotional contagion in social media. PLOS ONE 10(11), 1–14 (2015)
12. Ferrara, E., Yang, Z.: Quantifying the effect of sentiment on information diffusion in social media. PeerJ Computer Science 1, 1–15 (2015)
13. Fredrickson, B.L.: The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. The American Psychologist 56, 218–226 (2001)
14. Freyd, J.J.: In the wake of terrorist attack, hatred may mask fear. Analysis of Social issues and public policy 2(1), 5–8 (2002)
15. González-Ibáñez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in twitter: A closer look. In: Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2. pp. 581–586 (2011)
16. Gruzd, A., Doiron, S., Mai, P.: Is happiness contagious online? a case of twitter and the 2010 winter olympics. In: Proceedings of the 44th Hawaii International Conference on System Sciences. pp. 1–9. IEEE Computer Society, Washington, DC, USA (2011)
17. Heath, C.: Do people prefer to pass along good or bad news? valence and relevance of news as predictors of transmission propensity. Organizational behavior and human decision processes 68(2), 79–94 (1996)
18. Heimbach, I., Hinz, O.: The impact of content sentiment and emotionality on content virality. International Journal of Research in Marketing 33(3), 695 – 701 (2016)

19. Hu, Y., Zhao, J., Wu, J.: Emoticon-based ambivalent expression: A hidden indicator for unusual behaviors in weibo. PLOS ONE 11(1), 1–14 (2016)
20. Kim, H.S., Lee, S., Cappella, J.N., Vera, L., Emery, S.: Content characteristics driving the diffusion of antismoking messages: Implications for cancer prevention in the emerging public communication environment. Journal of National cancer institute. Monographs 47, 182–187 (2013)
21. Kramer, A.D.I., Guillory, J.E., Hancock, J.T.: Experimental evidence of massive-scale emotional contagion through social networks. Proc. of the National Academy of Sciences 111(24), 8788–8790 (2014)
22. Kušen, E., Cascavilla, G., Figl, K., Conti, M., Strembeck, M.: Identifying Emotions in Social Media: Comparison of Word-emotion Lexicons. In: Proc. of the 4th International Symposium on Social Networks Analysis, Management and Security (SNAMS). IEEE (August 2017)
23. Kušen, E., Strembeck, M., Cascavilla, G., Conti, M.: On the influence of emotional valence shifts on the spread of information in social networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 321–324. ACM (2017)
24. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proc. of the 19th International Conference on World Wide Web. pp. 591–600 (2010)
25. Lin, R., Utz, S.: The emotional responses of browsing facebook: Happiness, envy, and the role of tie strength. Computers in Human Behavior 52(Supplement C), 29 – 38 (2015)
26. Mcmanus, B.: An Expert Explains How Social Media Can Lead to the 'Self-Radicalisation' of Terrorists (2015), https://www.vice.com/en_uk/article/we-asked-an-expert-how-social-media-can-help-radicalize-terrorists
27. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word-emotion association lexicon. Computational Intelligence 29(3), 436–465 (2013)
28. Myrick, J.G.: Emotion regulation, procrastination, and watching cat videos online: Who watches Internet cats, why, and to what effect? Computers in Human Behavior 52(Supplement C), 168 – 176 (2015)
29. Nabi, R.L.: Exploring the framing effects of emotion. Communication Research 30(2), 224–247 (2003)
30. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Bad news travel fast: A content-based analysis of interestingness on twitter. In: Proceedings of the 3rd International Web Science Conference. pp. 8:1–8:7. ACM (2011)
31. Nielsen, F.Å.: Afinn (2011), http://www2.imm.dtu.dk/pubdb/p.php?6010
32. Plutchik, R.: The nature of emotions. American Scientist 89(4) (2001)
33. Rordriguez, M.G., Leskovec, J., Balduzzi, D., Scholkopf, B.: Uncovering the structure and temporal dynamics of information propagation. Network Science 2(1), 26–65 (2014)
34. Savage, D.A., Torgler, B.: The emergence of emotions and religious sentiments during the September 11 disaster. Motivation and Emotion 37(3), 586–599 (Sep 2013)
35. St Louis, C., Zorlu, G.: Can Twitter predict disease outbreaks? British Medical Journal (BMJ) 344, 1–3 (2012)
36. Steinbach, M.: ISIL Online: Countering Terrorist Radicalization and Recruitment on the Internet and Social Media (2016), https://www.fbi.gov/news/testimony/isil-online-countering-terrorist-radicalization-and-recruitment-on-the-internet-and-social-media-

37. Stieglitz, S., Linh, D.X.: Emotions and information diffusion in social media- sentiment of microblogs and sharing behavior. Journal of management information systems 29(4), 217–247 (2013)
38. Stuart, S.: Why do we pay more attention to negative news than to positive news? (2015), `http://blogs.lse.ac.uk/politicsandpolicy/why-is-there-no-good-news/`
39. Suh, B., Hong, L., Pirolli, P., Chi, E.H.: Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: Proc. of the 2010 IEEE Second International Conference on Social Computing. pp. 177–184 (2010)
40. Thelwall, M., Buckley, K., Paltoglou, G., D., C., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology 61, 2544—2558 (2010)
41. Trung, D.N., Nguyen, T.T., Jung, J.J., Choi, D.: Understanding Effect of Sentiment Content Toward Information Diffusion Pattern in Online Social Networks: A Case Study on TweetScope, pp. 349–358 (2014)
42. Tsugawa, S., Ohsaki, H.: Negative messages spread rapidly and widely on social media. In: Proceedings of the ACM on Conference on Online Social Networks. pp. 151–160. ACM (2015)
43. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1079–1088 (2010)
44. Weimann, G.: Terror on Facebook, Twitter, and Youtube. The Brown Journal of World Affairs 16(2), 45–54 (2010)
45. Yang, Y., Jia, J., Wu, B., Tang, J.: Social role-aware emotion contagion in image social networks. In: Proc. of the Thirtieth AAAI Conference on Artificial Intelligence. pp. 65–71 (2016)
46. Zhang, Z., Zhang, S.Y.: How do explicitly expressed emotions influence interpersonal communication and information dissemination? A field study of emoji's effects on commenting and retweeting on a microblog platform. In: 20th Pacific Asia Conference on Information Systems. pp. 1–14 (2016)