

Does Negatively Toned Language Use on Social Media Lead to Attitude Polarization?

Jürgen Buder^a, Lisa Rabl^a, Markus Feiks^b, Mandy Badermann^b, and Guido Zurstiege^b

Leibniz-Institut für Wissensmedien^a, University of Tübingen^b

Author Note

Correspondence concerning this paper should be addressed to Jürgen Buder, Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076 Tübingen, Germany. E-Mail: j.buder@iwm-tuebingen.de

Abstract

Prior research has indicated that both attitudinal homogeneity of communication networks (“echo chambers”) and attitudinal heterogeneity of communication networks (“adversarial debates”) can lead to attitude polarization. The present paper argues that communication in both echo chambers and adversarial debates is dominated by network negativity, a negative valence in the tone of discussions which might be associated with attitude polarization. Combining methods from sentiment analysis and social network analysis, more than 4 million tweets on two controversial topics (Brexit, Trump) were analyzed to investigate the occurrence of network negativity and its association with two proxies of attitude polarization (extremity and ambivalence). Results indicate that negativity in users’ own tweets was most strongly related to polarization, whereas negativity among users’ friends, or consonance of sentiments between users and friends had less impact on polarization. The findings are related to literatures on negativity bias, optimal distinctiveness theory, and intergroup contact theory.

Keywords: echo chamber, attitudes, attitude strength, social media, sentiment analysis, negativity bias

Does Negatively Toned Language Use on Social Media Lead to Attitude Polarization?

In 1993, the Internet pioneer Howard Rheingold likened the Usenet, an early type of social media for group discussions, as a “giant coffee-house with a thousand rooms” (Rheingold, 1993, p. 130). A similar spirit of optimism over the potential benefits of connecting to people across the globe characterized societal and scientific discourse about the Internet in general and social media in particular (Delli Carpini et al., 2004; Donath, 2007; McKenna & Bargh, 1998). However, over the last decade public and scientific discourse about the Internet and social media has shifted dramatically towards the negative end of evaluative space. The ability to communicate across the globe is no longer seen as something that unites people, but is often regarded as a hotbed for spreading misinformation, fake news, and hate speech (Ferrara et al., 2016; Pennycook et al., 2018; Zollo et al., 2015). Key among the concerns is the assumption that the current (social) media landscape may give rise to polarization where average attitudes of users are becoming increasingly distant from a reference point (e.g., the neutral scale mid-point, an earlier and more moderate average attitude, or the average attitude of a different group). Many scholars have argued that polarization can lead to a fragmentation of society into isolated pockets of opinion (e.g., Abramowitz & Saunders, 2008; Adamic & Glance, 2005; Taber & Lodge, 2006), and might be a factor in the rise of extremist parties, ideologies, and policies that can be observed in the US and Europe. Consequently, many social scientists have investigated potential antecedents that might lead to polarization in offline and online contexts.

Extant theories in this field have postulated attitude polarization to be a function of the content of discussion – either the quality of arguments being exchanged with others, or the extent to which a given content is congenial with one’s worldview. Using data from Twitter, the present paper offers a new account to describe factors that might lead to attitude polarization in social media. In particular, it investigates in how far the tone or sentiment of

discussion – rather than the content – may fuel polarization. It is argued that network negativity, i.e. negative sentiments in a person's network – which includes both one's own utterances on social media, and the utterances of the accounts one is following – can predict polarization.

Antecedents of Polarization

Within psychology, research on polarization had its origins in the discovery of the group polarization phenomenon (Myers & Lamm, 1976). In the corresponding experimental paradigm, individual attitudes of group members are measured both before and after discussing an issue with others. Typically, it shows in such settings that the average post-discussion attitude in a group has moved away from the neutral scale mid-point – in other words, the average attitude in the group has become more extreme. Early theories on group polarization have argued that group members process discussion arguments in a rational manner so that groups will ultimately converge on the attitude position that has the highest number of sound arguments in its favor (Vinokur & Burnstein, 1974). However, more recent work on polarization is heavily influenced by the observation that individuals or groups rarely engage in such kind of rational and unbiased information processing of argument quality (Kunda, 1990). Consequently, it has been proposed that people have a strong preference to attend to arguments that support their existing worldview (Hart et al., 2009; Iyengar & Hahn, 2009; Knobloch-Westerwick, 2014), for engaging with people who are similar to oneself (McPherson et al., 2001), and for conforming to information that originates from one's in-group (Mackie, 1986). Rather than being rational and unbiased processors of arguments, people tend to seek and weigh information more heavily when this information is congenial to one's worldview and/or congenial to the views of normatively significant others in one's in-group. This view of humans as biased information processors has led to the formulation of the echo chamber hypothesis (Sunstein, 2007). According to this hypothesis, people's tendency to

prefer congenial information and disregard uncongenial information leads to the creation of attitudinally homogeneous networks (“echo chambers”) where group members mutually validate and perpetuate their worldviews, becoming immune to other viewpoints, and isolate themselves from potentially threatening out-groups, leading to polarization or even radicalization. Moreover, Sunstein (2007) and others (Pariser, 2011) have argued that the current, high-choice media landscape can further fuel these dynamics.

The question whether the homogeneous informational environments of echo chambers will lead to polarization has been addressed by a number of scholars using a wide variety of approaches ranging from experimental work (e.g., Garrett, 2009; Taber & Lodge, 2006), over survey studies (e.g., Binder et al, 2009; Stroud, 2010), field studies using actual social media data (e.g., Barberá et al., 2015; Conover et al., 2011; Del Vicario et al., 2016) to computer simulations (e.g., Gescke et al., 2019; Mäs & Flache, 2013). While many studies found support for the hypothesis according to which a homogeneous communication network of like-minded interaction partners is associated with higher attitude polarization, the echo chamber hypothesis has also had its detractors. A number of studies have reported that it is very difficult to completely block out uncongenial information from one’s communication networks (Mutz & Mondak, 2006), that people do not avoid uncongenial information (Conover et al., 2011; Fletcher & Nielsen, 2017; Webster & Ksiazek, 2012), that some people even quite explicitly seek out for cross-cutting dialogue on the Internet (Buttliere & Buder, 2017; Stromer-Galley, 2003), and that the assumed complicity of Internet algorithms in the creation of echo chambers might be overstated (Nechushtai & Lewis, 2019). In sum, this line of work suggests that actual networks of social media users are often more heterogeneous than the echo chamber hypothesis proposes.

More criticism against the echo chamber hypothesis can be raised on the basis of findings which have shown that congenial information consumption is not the only predictor of attitude polarization; it seems to be the case that exposure to uncongenial information

might also lead to polarization. For instance, some experimental work has reported that participants who expose themselves to uncongenial information do not engage in such behavior in order to develop a moderate viewpoint, but in order to generate refutational thought, which by itself could even bolster or polarize existing attitudes (Edwards & Smith, 1996; Taber & Lodge, 2006). These experimental findings were confirmed in a survey study (Lee et al., 2014) which showed that being in a heterogeneous social media network actually was a strong predictor of opinion polarization for those users who frequently participated in political discussions. More recently, a field experiment also showed that users who were paid to create a more heterogeneous social media network by following an ideologically uncongenial chatbot for a month were becoming more polarized in their original attitudes (Bail et al., 2018).

However, these interesting findings pose an even more interesting puzzle. On the one hand, the literature has established fairly good evidence of a relationship between network *homogeneity* and polarization, in line with the echo chamber hypothesis. On the other hand, recent studies have established a relationship between network *heterogeneity* and polarization (Bail et al., 2018; Lee et al., 2014). How could it be that both talking to like-minded others and engaging in cross-talk with disagreeing others leads to polarization? How could echo chambers and their exact opposite (adversarial debates) lead to such similar results?

Network Negativity

The present paper offers the proposal that it might not only be the attitudinal composition (homogeneity/heterogeneity) of one's network that drives attitude polarization, but also the use of language within one's network. Language use is regarded as a marker for various psychological phenomena (Pennebaker et al., 2003), with emotional tone of language being one of the key variables. For instance, recent research has established that emotional content is more likely to go viral than non-emotional content, and this might be a driving

factor in the spread of misinformation (Brady et al., 2019). In particular, negativity in language use has been discussed as a facet in recent accounts of polarization (Hills, 2019; Yarchi et al., in press). Based on these considerations, we propose that the use of negatively toned language – both the language that a person expresses, and the language that a person is exposed to in one’s network – can lead to attitudinal polarization. This assumption is based on three findings in the social scientific literature. First, psychological research has consistently shown that negatively valenced information captures attention better, is weighted more in people’s evaluations, and is often remembered better than neutral or positive information (a so-called negativity bias; Baumeister et al., 2001). Therefore, negativity in one’s social network could also impact the strength of attitudes that a person holds, providing an avenue to attitude polarization. Second, research on journalism has established that negativity of events has very high news value (bad news sells better than good news; Galtung & Ruge, 1965). There is some evidence that users on social media like Twitter are guided by similar principles of news dissemination (e.g., Rudat & Buder, 2015), and this would nicely explain the penchant of social media contributors to focus on negative sentiments. Third, some studies on social media language use have reported that negativity outweighs positivity, particularly for controversial topics (Chmiel et al., 2011; Zollo et al., 2015). Most importantly, there is some evidence that negativity might be the prevalent sentiment in both attitudinally homogeneous environments (echo chambers) and attitudinally heterogeneous environments (adversarial debates). The best empirical evidence for such an assumption is coming from a study which used a technique called sentiment analysis to compare the discussion tone of users on Facebook sites that either deal with scientific topics or with conspiracy theories (Zollo et al., 2015). Within the conspiracy theory sub-community (which arguably meets many criteria for an echo chamber), posts with a negative sentiment (54%) clearly outweighed posts with a positive tone (11%), suggesting that talk within an echo chamber is characterized by negativity. The same study also investigated the sentiment on an adversarial debate site

where conspiracy theorists met with users of scientific groups, and again, negative posts (60%) clearly outweighed positive ones (5%). In other words, negativity was the norm for discussions in both homogeneous and heterogeneous groups. Other studies provide converging evidence that negativity could be the norm both among like-minded and other-minded discussants. For instance, within the homogeneous environment of partisan blogs it was found that liberal politicians were mentioned more often by conservatives and conservative politicians were mentioned more often by liberals (Adamic & Glance, 2005). While this study did not employ sentiment analysis it is easy to imagine that mentioning a liberal politician in a conservative environment is accompanied by tonal negativity. As for heterogeneous environments of adversarial debates, one study investigated the sentiment of comments on the BBC discussion forum (Chmiel et al., 2011) and did not only find a much higher occurrence of negative than positive sentiment, but also reported that negativity in comments was associated with greater length of discussion threads. In sum, there is ample evidence suggesting that network negativity is the glue that binds together both congenial and adversarial discussions on social media.

These findings on the importance of language, sentiment, and particularly negativity informed the conceptual framework that we used for our own investigation (see Figure 1 for an overview). The framework is fashioned after the model of interpersonal situations (Kelley et al., 2003) which argues that an interpersonal outcome can be driven by oneself (actor control), by others (partner controls), or by the alignment of actor and partners (joint control). Applied to the present context, we propose that the relationship between network negativity (control) and polarization (outcome) could play out on three different levels. First, attitude polarization of a person could be determined by the person's own contributions to discourse (actor control). In fact, psychological research has provided evidence that polarization can result from one's own activity, such as repeated attitude expression (Brauer et al., 1995). Second, attitude polarization of a user could be determined by the contributions that other

people in one's communication network make (partner control). The idea that polarization is mainly driven by others in one's network is consistent with identity-based accounts of group polarization and radicalization (Mackie, 1986; Smith et al., 2020), and also with research on digital emotion contagion (Goldenberg & Gross, 2020). Third, polarization of a user could be determined by the degree to which the sentiment of one's own contributions and the sentiment of others in a network are matched (joint control). In this case, polarization would ensue when both users and their social environments exhibit negativity. Such a relationship would mimic the proposed mechanics underlying the emergence of polarization within echo chambers (Sunstein, 2007).

The present research seeks to identify in how far a) the sentiment of one's own language, b) the sentiment of language use in the network of friends, and c) a match of language use between own language and friends' language will predict attitudinal polarization. To do this, more than 4 million tweets on two topics of high controversy (Trump, Brexit) were processed via sentiment analysis techniques and social network analysis techniques. In the first step of our analysis, we extracted the sentiment direction (positive or negative) from tweets in our sample to investigate whether negative sentiment occurs more frequently than positive sentiment. This would answer the question whether discourse is indeed characterized by network negativity. In the next step of the analysis, we investigated the three proposed facets of network negativity as predictor variables to see whether they influence our criterion variable of polarization. In order to assess polarization, we looked for various indicators of attitude strength (Howe & Krosnick, 2017), and identified two facets of attitude strength that a) should be linked to attitude polarization, and b) could be extracted from tweets. Thus, we arrived at two criterion variables: attitude extremity (average sentiment strength of a user's tweets; directly related to polarization) and attitude ambivalence (the simultaneous presence of positive and negative sentiments in a tweet; inversely related to polarization). Taken together, our approach resulted in the following four research questions:

RQ1: Is the sentiment of tweets characterized by negativity or positivity?

RQ2: How is the sentiment of a user's own tweets associated with polarization of the user?

RQ3: How is the sentiment of the tweets that a user encounters in the network of friends associated with polarization of the user?

RQ4: Does a match between the sentiment of a user's tweets and the sentiment of friends' tweets lead to polarization of the user?

Method

In order to address the relationship between network sentiment and polarization, the present study used cross-sectional social media data from Twitter. We chose Twitter because it can both be conceptualized as a platform for congenial debate (echo chamber) and adversarial debate (Colleoni et al., 2014). However, in order to see how network sentiment impacts polarization, we selected highly controversial and polarizing issues for our sample, thus we investigated about 4.3 million tweets about the Brexit and about Donald Trump as a basis for analysis. With a mixture of sentiment analysis techniques and techniques from social network analysis, network sentiment was determined on three levels (mean sentiment of the user's tweets, mean sentiment of friends' tweets, match between a user's and friends' sentiment). Moreover, we built on social psychological research on attitude strength (Howe & Krosnick, 2017) to extract markers of attitude polarization as outcome variables. While many, if not most components of attitude strength should be appropriate as markers of polarization, two could be directly captured by means of sentiment analysis: 1) the extremity of an attitude could be captured by the strength of sentiment in a tweet; and 2) the ambivalence of an attitude could be captured by the simultaneous presence of positive and negative sentiment in a tweet (which should be inversely related to polarization).

Data

Data from Twitter was used for this study. With a developer account, Twitter offers an access point to its user data through their application programming interface (API). There are at least two ways to collect data from Twitter: Tweets not older than 7 days could either be *searched* (REST-API) or could be filtered in real-time from a *stream* (Stream-API).

4,337,797 million tweets from 942,496 different users were collected (see Table 1). The data were primarily gathered through the Stream-API and the results were filtered using hashtags and/or usernames. For this purpose, the web service *Netlytics* (www.netlytics.org) was used and two streams were set up. The first stream was filtered for tweets that included the hashtag *#brexit*. The second stream included all available tweets that consisted of the hashtag *#trump* and/or mentioned the username *@realDonaldTrump*. We used those hashtags as they dealt with ideologically charged issues that could be linked to polarization.

Moreover, additional user data were gathered, namely the verification status of a user and a list of his or her “friends” (other Twitter accounts that the user followed). With those lists it was possible to extract those friends of a user who also contributed to the dataset (i.e., used the same hashtags), and to determine their average sentiment. For gathering the additional data, the REST-API was used together with a *Python* (3.6) script and a third-party package called “tweepy” that allows a comfortable API access.

INSERT TABLE 1 HERE

After the data were gathered, they were filtered by a set of specific predefined standards to make sure that only private accounts and non-automated communication was considered for later analysis. First, only tweets in English language were used, as our sentiment analysis algorithms (SentiStrength; Thelwall et al., 2010) were trained in English

language. Second, as we were focusing on mass polarization rather than elite polarization (Fiorina & Abrams, 2008) we wanted to exclude verified accounts, e.g. from media institutions, political parties or celebrities, so our dataset contained only tweets from non-verified accounts. Third, in order to ensure that accounts were embedded within a network structure, every user in the final dataset needed to have more than 10 friends and more than 10 followers in order to be included. In addition, users were excluded that had more than 5,000 friends and more than 10,000 followers. Fourth, a user must have sent a minimum of 3 tweets in order to be included in the final dataset, so that he or she at least could have an “interaction” with other users within the dataset. Fifth, we aimed to exclude fake accounts, thus we removed accounts that have significantly more friends than followers. After data cleaning, about 30% of the initial data was used for the analyses of sentiment direction and polarization.

Measures

Sentiment Analysis

For the sentiment analysis we used the SentiStrength algorithm (Thelwall, 2017; Thelwall et al., 2010). SentiStrength is a lexicon-based classifier that additionally uses non-lexical linguistic information and rules to detect the strength of a sentiment in short informal text. SentiStrength analyzes sentiment in customizable units of analysis (typically sentence level though we used the tweet level) by matching the words in the unit of analysis with a set of pre-defined (or pre-definable) and pre-rated lexicon of emotional words. As a result of the classification process, typically two integers are returned per unit of analysis, a positive and a negative value ranging from [1 to 5] and [-1 to -5], which stand for the highest scoring positive and highest scoring negative word in a tweet, respectively (Thelwall, 2017, p. 122). For our analyses, it was important to disentangle the direction of sentiment (an indicator of negativity/positivity) and the strength or extremity of sentiment (an indicator of polarization;

see below). Sentiment direction was coded as -1 if the absolute negative sentiment value of a tweet was larger than the positive sentiment value, was coded as +1 in the opposite case, and as 0 when the sum of sentiment values was zero. The mean sentiment of a user represents averaged mean sentiment direction values of all tweets from this user. Similarly, the mean sentiment of friends is indicated by averaged mean sentiment values of all their tweets. Finally, in order to capture sentiment similarity, we multiplied the mean sentiment direction of a given user with the mean sentiment direction of the user's friends.

A subset of 626 tweets was also manually coded by four native American speakers who were introduced into a slightly modified manual for coding sentiment in tweets, provided by the developers of the SentiStrength algorithm. The coders passed through three training sessions and an attendant trial coding on 125 tweets from the two data sets which was evaluated with regard to the reliability between the human coders, and it was followed by a corrective training. In order to assess agreement between human coders and SentiStrength data, Spearman correlations between mean human coding per tweet (over four human coders) and SentiStrength coding per tweet were computed. Results show significant positive correlations value of $r = .47$ for negative sentiment, and $r = .19$ for positive sentiment, respectively. While the correlation between SentiStrength and human coders for positive sentiment was relatively low (potentially reflecting the difficulty of SentiStrength to discover sarcasm; Thelwall, 2017) it provides a rather conservative test of network negativity in tweets. In other words, actual amounts of negativity might be underestimated in the given dataset. Finally, the correlation between the distance of positive and negative sentiment per tweet among human coders vs. SentiStrength was $r = .37$.

Attitude Strength

While lab and field studies on polarization measure user attitudes by asking participants about their stance towards certain political or societal issues, field studies using

actual social media data must indirectly infer attitudes. Some scholars have tried to implicitly capture attitudes from the Twitter accounts that users followed (e.g., Barberá et al., 2015), but such a procedure might be inaccurate as many other factors than political preference may influence the decision to follow a particular Twitter account. A more direct assessment of attitudinal features could be derived from the actual contributions that a user has made. Therefore, we used sentiment analysis in order to assess polarization, as this informs about the intensity or strength of attitudes. According to social psychology, attitude strength is a multifaceted construct. For instance, Howe and Krosnick (2017) presented a list of eleven different features of attitude strength. While the majority of these features cannot be estimated from social media data (e.g., accessibility, affective-cognitive consistency), we identified two features that can arguably be estimated from sentiment analysis data: attitude extremity and attitude ambivalence served as proxies for attitude polarization.

Attitude Extremity. Attitude extremity is defined as “the extent to which the attitude deviates from neutrality” (Krosnick & Petty, 1995, p. 6). For the measurement of extremity, an index was calculated for every user. To recall, SentiStrength returns two values for each tweet of a user (one for positivity, one for negativity). The extremity of a single tweet was computed by converting the two sentiment values into absolute values and keeping the largest value. The extremity of a user was then computed by taking the average of the extremity values of each of his/her tweets. By decoupling the direction of sentiment (a predictor variable; see above) from the extremity of sentiment (a criterion variable), it is ensured that these two variables are statistically not confounded.

Attitude Ambivalence. In attitude research, the construct of ambivalence was posited to account for the fact that persons can simultaneously have positive and negative evaluations of the same attitude object (Howe & Krosnick, 2017). The ability to address this dual nature of attitudes was also one of the main reasons for us to use the SentiStrength algorithm, as it captures positive and negative sentiments separately. More precisely, our *attitude*

ambivalence measure was modeled on a formula by Breckler (1994) which accounts for both the similarity and intensity of positive and negative evaluations. The formula for *Ambivalence* is

$$A = \frac{s_w + s_s}{2} - |s_w - s_s|$$

where s_w represents the weaker (smaller) absolute value of the positive and negative sentiment value of a tweet, and s_s represents the stronger (larger) absolute value of the positive and negative sentiment value of a tweet. The first term in the formula represents the intensity of sentiments, the second term captures their similarity. In order to yield the ambivalence of a user, the ambivalence scores of his/her tweets were averaged. Lower ambivalence scores are indicative of larger polarization.

Results

Sentiment Direction

Before addressing the relationship between indices of network negativity and indices of polarization, a first analysis referred to the overall sentiment direction within the datasets (RQ1). To recall, SentiStrength computes values for both the positive and the negative sentiment of a tweet. In order to compare whether average users exhibited negativity or positivity, the mean positive sentiment and the absolute value of mean negative sentiment per user were computed. A dependent t-test was carried out indicating that the mean negative sentiment ($M = 1.80$, $SD = 0.41$) of users was significantly higher than their mean positive sentiment ($M = 1.46$, $SD = 0.31$; $t(129800) = 237.46$, $p < .001$). While statistical significance in such a large sample is of little interpretive value, the effect size ($d = 0.93$) indicates a large effect (see Figure 2 for a box plot). The same analysis was also carried out for the two topics separately. Again, for both topics the mean negative sentiment of users (Brexit: $M = 1.72$, SD

= 0.24; Trump: $M = 1.83$, $SD = 0.42$;) was higher than their mean positive sentiment (Brexit: $M = 1.33$, $SD = 0.24$, $t(42054) = 179.03$, $p < .001$, $d = 1.25$; Trump: $M = 1.53$, $SD = 0.31$, $t(87748) = 171.32$, $p < .001$, $d = 0.84$). These analyses provide evidence that discourse in our sample was indeed fueled by network negativity.

INSERT FIGURE 2 HERE

Overview of Main Analyses

Multiple regression analysis was employed to address the influence of network sentiment on two indices of polarization (attitude extremity and attitude ambivalence). Results of our analyses are displayed for the total dataset, and for the two separate datasets about Brexit and Trump.

All analyses on the criterion variables of polarization included four predictor variables: the first three predictors (*Mean sentiment user*; *Mean sentiment friends*; *Mean sentiment user x friends*) referred to the variables of key interest. We also included a fourth predictor (*Friends in dataset*) which captures the number of friends who tweeted about the same topic, divided by the number of friends that a user had in total. This predictor was included to see whether polarization effects depend on the intensity with which a user's network discusses about the topics in question.

Before conducting the reported analyses, the relevant assumptions of a multiple regression were tested. The collinearity statistics (i.e., Tolerance and VIF) were all within reasonable limits and the assumption of multicollinearity was met. Residual and scatter plots showed that the assumptions of normality, linearity, and homoscedasticity were all met for the attitude strength measures of extremity and ambivalence.

Attitude Extremity

According to RQ2, it was investigated how the average sentiment direction in a user's tweets are associated with attitude extremity. Similarly, RQ3 investigated whether the average sentiment direction of friends is related to attitude extremity. Finally, RQ4 explored whether similarity in sentiment directions between user and friends leads to higher extremity of a user. Results of the regression analysis are displayed in Table 2.

INSERT TABLE 2 HERE

The amount of explained variance for the total dataset and the two isolated sets varied between 5% and 11%. With regard to RQ2, we found that users exhibited more extremity the more they expressed negativity in their own tweets. Conversely, one may say that users were less polarized the more their tweets were dominated by positivity. This lends support to the assumption that expressing one's attitudes with negatively toned language may give rise to polarization. With regard to RQ3, the data for the total dataset interestingly show that users were *less* extreme and more moderate when their friends exhibited negativity. In other words, these users were acting against the overall negativity bias in their social environment, suggesting that emotion contagion did not take place. However, this counteraction pattern has to be interpreted with caution as it was observed across the total dataset, but not in the two separate datasets. Finally, with regard to RQ4, similarity of sentiment between user and friends did have a small positive impact on extremity. When users' and friends' sentiment was matched in tone, tweets were slightly more extreme, suggesting a small tonal echo chamber effect. However, this relationship did only hold for the total dataset.

Attitude Ambivalence

RQ2 investigated how sentiment in a user's own tweets is associated with (reduced) ambivalence. Along the same lines, RQ3 assessed how the sentiment in the tweets of friends

lowers ambivalence. RQ4 examined how tonal similarity between users and friends impacts ambivalence of a user. Results of the regression analysis are displayed in Table 3.

INSERT TABLE 3 HERE

While the Brexit dataset yielded a moderate amount of explained variance (6%), overall explained variance turned out to be small, but robust for the Trump dataset (1.5%) and the total dataset (2.5%). With regard to RQ2, a small effect of user sentiment direction on attitude ambivalence was found. The more users expressed negativity in their tweets the less ambivalent, and thus the more polarized they were. Consistent with the corresponding analysis of extremity, network negativity impacted polarization via personal-individual sentiment. In contrast, the sentiment of friends did not have an impact on ambivalence (RQ3). Similarly, ambivalence was not impacted by sentiment similarity between users and friends (RQ4).

Discussion

Prior research has reported that both attitudinal homogeneity (“echo chambers”) and attitudinal heterogeneity (“adversarial debates”) in one’s communication network may give rise to attitude polarization. The present study investigated the hypothesis that the sentiment of discourse – more precisely, its network negativity – might be the common element in both the homogeneity of echo chambers and the heterogeneity of adversarial debates on discussion forums, thus becoming a single mechanism that might drive polarization. However, if network negativity drives polarization, what kind of negativity is responsible? Is it one’s own negativity, the negativity of one’s friends, or is it the tonal consonance between users and their friends? In the present study, these questions were addressed by analyzing three potential

sources of network negativity. Using a cross-sectional snapshot of more than four million tweets about Brexit and Trump, techniques from social network analysis and sentiment analysis were combined to assess the impact of negativity on polarization.

The first key finding of the study is that Twitter content on the two topics indeed was characterized by overall negativity. This finding is consistent with the extant literature on sentiments in social media (e.g., Chmiel et al., 2011; Del Vicario et al., 2016; Zollo et al., 2015). A second key finding of the present study is that negativity is linked to polarization, particularly for the attitude strength feature of attitude extremity and – to a lower extent – attitude ambivalence. To the best of our knowledge, our study is the first to demonstrate such a link in a social media setting, thus adding to a literature which shows a negativity bias in human information processing and behavior (Baumeister et al., 2001).

Our study further sought to determine the separate and combined influence of the negativity of the users and the negativity of the users' friends. The data suggest that it is users' individual negativity which is most clearly linked to polarization. The more users exhibited negativity in their own tweets, the more extreme and the less ambivalent they were. This result is in contrast to established theories of polarization which typically find an individual's social environment to be responsible for polarization effects, whether it is persuasive arguments theory (Vinokur & Burnstein, 1974), emotion contagion (Goldenberg & Gross, 2020), or the echo chamber hypothesis (Sunstein, 2007). However, the result is consistent with literature indicating that attitudes are also shaped by the private thoughts and public utterances that a person makes. For instance, one study (Brauer et al., 1995) reported that expression of one's attitude had a stronger impact on attitude polarization than exposure to arguments from congenial others, furthering the notion that it might be people's own behavior that drives polarization. However, it might be that the influence of one's environment on polarization is stronger in settings where social ties among users are potentially stronger than on Twitter, e.g. on Facebook or professional social networks.

Our investigation into the effects of the negativity in one's social environment yielded much less impact on a user's polarization. Based on the emotion contagion literature, one might have expected that negativity of a user's friends will lead to higher extremity and lower ambivalence. However, somewhat surprisingly our results tend to point into the opposite direction, suggesting that negativity in a user's social environment had a slightly depolarizing effect on attitude extremity. How could such a self-limiting mechanism of negativity be explained? One potential answer comes from optimal distinctiveness theory (Brewer, 1991). According to this theory, a person's social identity is shaped by both the motive to assimilate to one's in-group, and by the simultaneous desire to achieve distinctiveness from other in-group members. Since in-groups are typically defined by the existence of oppositional out-groups who are generally derogated (Tajfel & Turner, 1986), assimilation to the in-group norm might be responsible for a user's own negativity. In contrast, the need to achieve distinctiveness within one's in-group might lead to a tendency to counteract negativity in one's social environment.

Finally, our approach also provided us an opportunity to investigate the question whether a match between user sentiment and friends' sentiment can be associated with polarization. In other words, we examined the question whether there are tonal echo chambers that are analogous to Sunstein's (2007) attitudinal echo chambers. There have been some small effects to support the notion that tonal echo chambers may lead to polarization, as users became more extreme and less ambivalent when their friends exhibited the same sentiment direction. However, this pattern did not hold for the isolated datasets and should be interpreted with caution.

What are the implications of our findings? First and foremost, we found quite promising evidence that the sentiment of discussion matters, and that network negativity can be a precursor to attitude polarization. It is an interesting question whether the polarization effects that have been observed in many experimental studies (e.g., Taber & Lodge, 2006),

survey studies (e.g., Stroud, 2010), and field studies (e.g., Barberá et al., 2015) were also at least partially due to tonal network negativity. In any case, future research on attitude polarization within social media contexts should also consider measuring aspects of network sentiment (rather than only attitudes). Conversely, future studies that focus on network sentiment should also try to measure user attitudes, though this is quite a complicated affair in studies based on actual social media data.

Our study was built on the implicit assumption that network negativity is prevalent in both echo chambers and in cross-cutting adversarial debates. Though this assumption was made on the basis of earlier findings (e.g., Adamic & Glance, 2005; Chmiel et al., 2011; Zollo et al., 2015), our study was not designed to provide direct evidence that negativity is the norm in both types of setting (Twitter is arguably a hybrid form that caters both to exchange with like-minded and other-minded users; Colleoni et al., 2014). It would be desirable if future studies could provide more evidence for network negativity within echo chambers, but also within adversarial debates (similar to Zollo et al., 2015). For communication within echo chambers, there is strong evidence for polarization (e.g., Binder et al., 2009), but less evidence for overall negativity. In contrast, for communication in adversarial debates there is strong evidence for negativity (e.g., Chmiel et al., 2011; Zollo et al., 2015), but less empirical evidence for polarization. Nonetheless, on theoretical grounds we believe that adversarial debates can be linked to attitude polarization. After all, adversarial debates on the Internet are a form of intergroup contact (Amichai-Hamburger & McKenna, 2006), and the extant literature on this topic indicates that intergroup contact can lead to mutual understanding between groups when conditions are favorable, but can also lead to stronger fragmentation and hostility when conditions are unfavorable. Along similar lines, one could argue that it will depend on certain conditions whether adversarial debates on social media will lead to depolarization or polarization. Again, network negativity might be a key ingredient in this respect.

While the present paper provides evidence for a link between network negativity and polarization, some limitations of the study should not go unmentioned. First, we only investigated Twitter activity with regard to two topics (Brexit, Trump), and as the former President of the United States was known to be a Brexit supporter, these topics can be argued to be ideologically correlated. Therefore, it would have been better if we could have established the link between negativity and attitude polarization for a wider range of topics that cater to people with different ideological leanings.

A second limitation of the present study is that we were only investigating a single platform (Twitter). Though there is strong evidence that negativity is the norm in many social media outlets (Chmiel et al., 2011; Zollo et al., 2015), there are certainly exceptions to this rule, such as communication on fansites, or within health support groups (Wang et al., 2012). Based on our own results, we would predict that a) positivity is more frequent than negativity in such environments, and b) that positivity should be associated with reduced polarization. However, future research is needed to establish such a link.

A third limitation of our study is the use of a cross-sectional design. Though polarization can be understood and measured as both a process and a state (DiMaggio et al., 1996), it should be mentioned that the classical group polarization paradigm in psychology clearly favors repeated-measures designs. Similarly, the most compelling evidence for attitudinal echo chambers and their role in polarization comes from longitudinal survey studies (Binder et al., 2009; Stroud, 2010). A longitudinal design also would provide an opportunity to investigate causal relations among variables. Along these lines, future studies could examine whether network negativity precedes or follows from attitude polarization (or both).

Conclusion

Current research on attitude polarization is dominated by the echo chamber hypothesis (Sunstein, 2007) according to which individuals become more polarized the more they interact with attitudinally congenial others. While the notion of echo chambers has received a lot of support in experimental as well as survey-based research, there are methodological and empirical reasons for the difficulties to find unequivocal support for the echo chamber hypothesis from field studies. The methodological reasons rest on the difficulty to reliably measure an attitudinal stance from social media data, thus making it hard to establish whether a piece of information is congenial or uncongenial for a given user. For this reason, our research focused on a variable that is easier to capture with current technologies – the sentiment of messages from which a general direction (negativity/positivity) as well as attitudinal intensity (extremity/ambivalence) can be extracted. As for empirical reasons to have doubts on the echo chamber hypothesis, there is mounting evidence that both communication within and communication across echo chambers can lead to polarization. This was another reason for us to move away from attitudinal congeniality and investigate sentiment instead, particularly as negative sentiment seems to abound in both congenial and adversarial communication.

The key contribution of the present paper is to show that – irrespective of the attitudinal composition in a network – the use of language may be a precursor of polarization. In particular, it appears that large parts of online discourse are characterized by negative sentiment, and the amount of this network negativity subsequently predicts indicators of polarization like attitude extremity and reduced attitude ambivalence. The finding that network negativity might contribute to polarization and ultimately fragmentation of society might sound alarming. However, there is also a silver lining in our findings: our data suggest that it is not the negativity of one's social surrounding that drives polarization, but rather the negativity of one's own utterances. In other words, individuals have it in their own hands to actively depolarize themselves through the use of moderate language.

Acknowledgments

This research was funded by the Leibniz-WissenschaftsCampus Tübingen “Cognitive Interfaces”.

References

- Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2), 542-555. <https://doi.org/10.1017/S0022381608080493>
- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36-43). Chicago, Ill., August 21 - 25, 2005.
<https://doi.org/10.1145/1134271.1134277>
- Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The contact hypothesis reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication*, 11(3), 825–843. <https://doi.org/10.1111/j.1083-6101.2006.00037.x>
- Bail, C. A., Argyle, L. P. Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS*, 115(37), 9216-9221.
<https://doi.org/10.1073/pnas.1804840115>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531-1542. <https://doi.org/10.1177/0956797615594620>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370.
<https://doi.org/10.1037/1089-2680.5.4.323>

- Binder, A. R., Dalrymple, K. E., Brossard, D., & Scheufele, D. A. (2009). The soul of a polarized democracy: Testing theoretical linkages between talk and attitude extremity during the 2004 Presidential Election. *Communication Research*, 36(3), 315-340. <https://doi.org/10.1177/0093650209333023>
- Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10), 1802–1813. <https://doi.org/10.1037/xge0000532>
- Brauer, M., Judd, C. M., & Gliner, M. D. (1995). The effects of repeated expressions on attitude polarization during group discussions. *Journal of Personality and Social Psychology*, 68(6), 1014–1029. <https://doi.org/10.1037/0022-3514.68.6.1014>
- Breckler, S. J. (1994). A comparison of numerical indexes for measuring attitude ambivalence. *Educational and Psychological Measurement*, 54(2), 350-365. <https://doi.org/10.1177/0013164494054002009>
- Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17, 475-482. <https://doi.org/10.1177/0146167291175001>
- Buttlere, B., & Buder, J. (2017). Reading more vs. writing back: Situation affordances drive reactions to conflicting information on the internet. *Computers in Human Behavior*, 74, 330-336. <https://doi.org/10.1016/j.chb.2017.04.041>
- Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M., & Hołyst, J. A. (2011). Negative emotions boost user activity at BBC forum. *Physica A*, 390, 2936-2944. <https://doi.org/10.1016/j.physa.2011.03.040>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332. <https://doi.org/10.1111/jcom.12084>

- Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 89-96).
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific Reports*, 6:37825. <https://doi.org/10.1038/srep37825>
- Delli Carpini, M. X. D., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement. *Annual Review of Political Science*, 7, 315-344. <https://doi.org/10.1146/annurev.polisci.7.121003.091630>
- DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, 102(3), 690-755. <https://doi.org/10.1086/230995>
- Donath, J. (2007). Signals in social supernets. *Journal of Computer-Mediated Communication*, 13, 231-251. <https://doi.org/10.1111/j.1083-6101.2007.00394.x>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1), 5-24. <https://doi.org/10.1037/0022-3514.71.1.5>
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104. <https://doi.org/10.1145/2818717>
- Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11, 563-588. [doi:10.1146/annurev.polisci.11.053106.153836](https://doi.org/10.1146/annurev.polisci.11.053106.153836)
- Fletcher, R., & Nielsen, R. K. (2017). Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication*, 67(4), 476-498. <https://doi.org/10.1111/jcom.12315>

- Galtung, J., & Ruge, M. H. (1965). The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research*, 2(1), 64-90. <https://doi.org/10.1177/002234336500200104>
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2), 265-285. <https://doi.org/10.1111/j.1083-6101.2009.01440.x>
- Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology*, 58(1), 129-149. <https://doi.org/10.1111/bjso.12286>
- Goldenberg, A., & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences*, 24(4), 316-328. <https://doi.org/10.1016/j.tics.2020.01.009>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555-588. <https://dx.doi.org/10.1037/a0015701>
- Hills, T. T. (2019). The dark side of information proliferation. *Perspectives on Psychological Science*, 14(3), 323-330. <https://doi.org/10.1177/1745691618803647>
- Howe, L. C., & Krosnick, J. A. (2017). Attitude strength. *Annual Review of Psychology*, 68, 327-351. <https://doi.org/10.1146/annurev-psych-122414-033600>
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19-39. <https://doi.org/10.1111/j.1460-2466.2008.01402.x>
- Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. Cambridge University Press.

- Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. Routledge.
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1-24). Lawrence Erlbaum Associates.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, 64(4), 702-722. <https://doi.org/10.1111/jcom.12077>
- Mackie, D. M. (1986). Social identification effects in group polarization. *Journal of Personality and Social Psychology*, 50(4), 720–728. <https://doi.org/10.1037/0022-3514.50.4.720>
- Mäs, M., & Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PLoS ONE*, 8(11): e74516. <https://doi.org/10.1371/journal.pone.0074516>
- McKenna, K. Y. A., & Bargh, J. A. (1998). Coming out in the age of the Internet: Identity "demarginalization" through virtual group participation. *Journal of Personality and Social Psychology*, 75(3), 681-694. <http://dx.doi.org/10.1037/0022-3514.75.3.681>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Mutz, D. C., & Mondak, J. J. (2006). The workplace as a context for cross-cutting political discourse. *The Journal of Politics*, 68(1), 140-155. <https://doi.org/10.1111/j.1468-2508.2006.00376.x>

- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4), 602-627. <http://dx.doi.org/10.1037/0033-2909.83.4.602>
- Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90, 298-307. <https://doi.org/10.1016/j.chb.2018.07.043>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- Rheingold, H. (1993). *The virtual community: Finding connection in a computerized world*. Addison-Wesley Longman.
- Rudat, A., & Buder, J. (2015). Making retweeting social: The influence of content and context information on sharing news in Twitter. *Computers in Human Behavior*, 46, 75-84. <https://doi.org/10.1016/j.chb.2015.01.005>
- Smith, L. G. E., Blackwood, L., & Thomas, E. F. (2020). The need to refocus on the group as the site of radicalization. *Perspectives on Psychological Science*, 15(2), 327-352. <https://doi.org/10.1177/1745691619885870>
- Stromer-Galley, J. (2003). Diversity of political conversation on the Internet: Users' perspectives. *Journal of Computer-Mediated Communication*, 8(3). <https://doi.org/10.1111/j.1083-6101.2003.tb00215.x>
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556-576. <https://doi.org/10.1111/j.1460-2466.2010.01497.x>

- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Nelson-Hall.
- Thelwall, M. (2017). The heart and soul of the Web? Sentiment strength detection in the social web with SentiStrength. In J. A. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 119-134). Springer. https://doi.org/10.1007/978-3-319-43639-5_7
- Thelwall, M., Buckley, K., Paltoglou, D. C., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 6(12), 2544-2558. <https://doi.org/10.1002/asi.21416>
- Vinokur, A., & Burnstein, E. (1974). Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach. *Journal of Personality and Social Psychology*, 29(3), 305-315. <http://dx.doi.org/10.1037/h0036010>
- Wang, Y.-C., Kraut, R., & Levine, J. M. (2012). To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 833-842). Seattle, WA, February 11 - 15, 2012. <https://doi.org/10.1145/2145204.2145329>
- Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, 62(1), 39-56. <https://doi.org/10.1111/j.1460-2466.2011.01616.x>

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (in press). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*.

<https://doi.org/10.1080/10584609.2020.1785067>

Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PLoS ONE* 10(9): e0138740. <https://doi.org/10.1371/journal.pone.0138740>

Tables

Table 1

Dataset overview partitioned by search terms

	Brexit	Trump
Search term(s):	#brexit	#trump, @realDonaldTrump
Period of search	2017-11-22 to 2017-12-19	2017-11-21 to 2018-01-04
Days	27	44
n_{tweets}	1,685,560	2,652,237
n_{user}	289,012	653,484

Table 2

Regression statistics for attitude extremity

	Total dataset		Brexit		Trump	
Predictor	<i>b</i>	95% CI	<i>b</i>	95% CI	<i>b</i>	95% CI
Friends in dataset	-0.01	[-0.02, -0.01]	-0.01	[-0.02, -0.01]	0.01	[-0.00, 0.01]
Mean sentiment user	-0.07	[-0.08, -0.07]	-0.11	[-0.11, -0.10]	-0.07	[-0.08, -0.07]
Mean sentiment friends	0.04	[0.04, 0.05]	0.01	[-0.00, 0.02]	0.01	[-0.00, 0.01]
Mean sentiment user x friends	0.02	[0.00, 0.03]	0.02	[-0.01, 0.03]	0.00	[-0.01, 0.01]
R ²	.055		.105		.054	

Table 3

Regression statistics for attitude ambivalence

	Total dataset		Brexit		Trump	
Predictor	<i>B</i>	95% CI	<i>b</i>	95% CI	<i>b</i>	95% CI
Friends in dataset	0.00	[-0.00, 0.00]	-0.00	[-0.00, 0.00]	0.00	[-0.00, 0.00]
Mean sentiment user	0.02	[0.02, 0.02]	0.04	[0.03, 0.04]	0.02	[0.02, 0.02]
Mean sentiment friends	-0.00	[-0.00, 0.00]	0.00	[-0.00, 0.02]	-0.00	[-0.01, 0.01]
Mean sentiment user x friends	-0.01	[-0.02, -0.01]	-0.01	[-0.03, 0.01]	-0.00	[-0.01, 0.01]
R ²	.025		.061		.015	

Figures

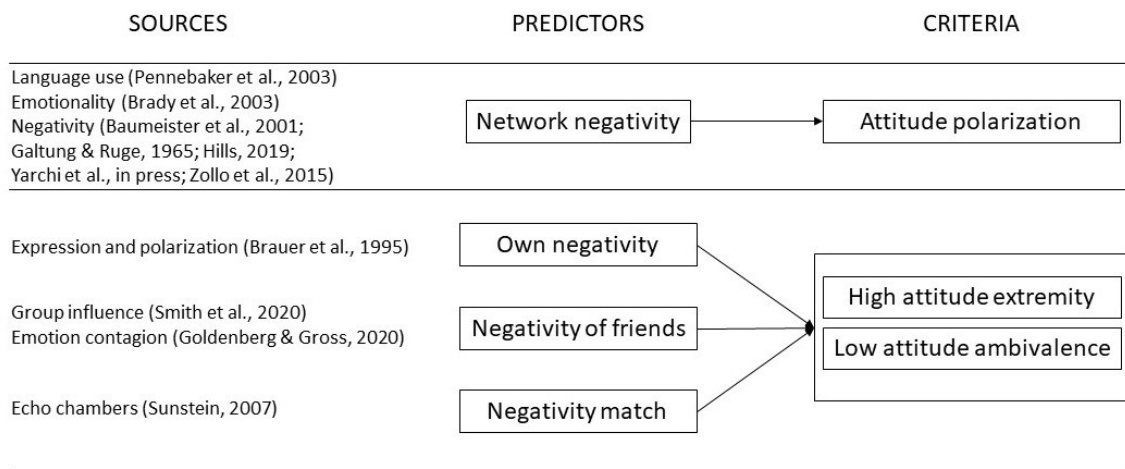


Figure 1: Conceptual framework used in this research. Left: sources for hypothesized relations; middle: predictor variables; right: criterion variables. Upper half: general framework of network negativity; lower half: detailed framework and variables.

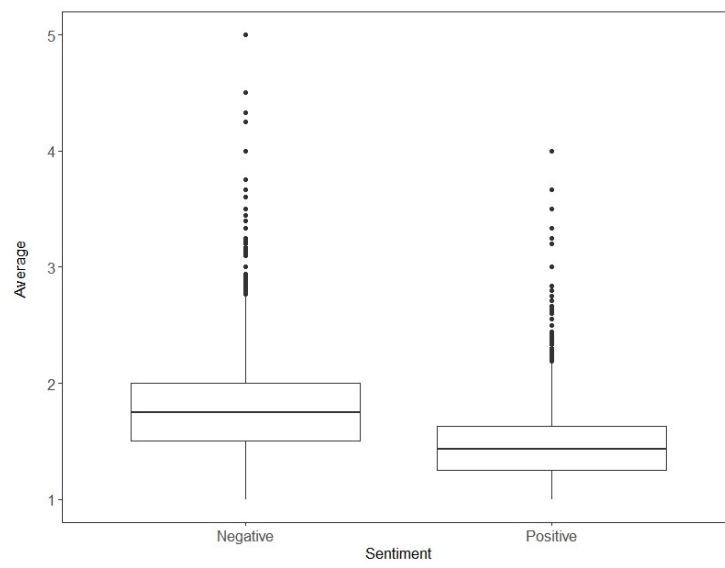


Figure 2: Box plot comparing average negativity and average positivity of users.