Transactions on Replication Research

Conceptual Replication DOI: 10.17705/1atrr.00079 ISSN 2473-3458

Emotions and Information Diffusion on Social Media: A Replication in the Context of Political Communication on Twitter

Linus Hagemann
University of Potsdam
mail@linushagemann.de

Olga Abramova
University of Potsdam
olga.abramova@uni-potsdam.de

Abstract:

This paper presents a methodological and conceptual replication of Stieglitz and Dang-Xuan's (2013) investigation of the role of sentiment in information-sharing behavior on social media. Whereas Stieglitz and Dang-Xuan (2013) focused on Twitter communication prior to the state parliament elections in the German states Baden-Wurttemberg, Rheinland-Pfalz, and Berlin in 2011, we test their theoretical propositions in the context of the state parliament elections in Saxony-Anhalt (Germany) 2021. We confirm the positive link between sentiment in a political Twitter message and its number of retweets in a methodological replication. In a conceptual replication, where sentiment was assessed with the alternative dictionary-based tool LIWC, the sentiment was negatively associated with the retweet volume. In line with the original study, the strength of association between sentiment and retweet time lag insignificantly differs between tweets with negative sentiment and tweets with positive sentiment. We also found that the number of an author's followers was an essential determinant of sharing behavior. However, two hypotheses supported in the original study did not hold for our sample. Precisely, the total amount of sentiments was insignificantly linked to the time lag to the first retweet. Finally, in our data, we do not observe that the association between the overall sentiment and retweet quantity is stronger for tweets with negative sentiment than for those with positive sentiment.

Keywords: Twitter, Information Diffusion, Sentiment, Elections.

The manuscript was received 02/28/2022 and was with the authors 5 months for 2 revisions.

1 Introduction

1.1 Motivation for Replication

Enabling information diffusion (i.e., transmission from one actor to another, at unprecedented scale and pace), social media has become indispensable in today's public discourse, for example, during elections (Mallipeddi et al., 2021; Hagemann & Abramova, 2022), protests (Valenzuela, 2013), crises events like terrorist attacks (Stieglitz et al., 2017; Fischer-Preßler et al., 2019) and disease outbreaks (Abramova et al., 2022). Granting access to a multimillion-person audience worldwide, platforms like Facebook, Twitter, Instagram, and, lately, TikTok represent a unique environment used to spread the news and, subsequently, to persuade and influence a person's opinions, attitudes, motivations, and, eventually, behavior. As psychology literature suggests, persuasion leans on the emotions (a.k.a. sentiment) conveyed in a message and can become a powerful tool for management, marketing, politics, or personal use. Multiple empirical examinations verified correlations between content sentiment and the sharing of news articles (Berger & Milkman, 2012; Heimbach & Hinz, 2016), the probability of reaching the funding goal of crowdfunding projects (Faralli et al., 2021), or patients' support in online health communities (OHC) (Chen et al., 2020).

Due to its polarizing and controversial nature as well as long-term and nationwide implications, politics and online political communication attracted the wide attention of scholars and practicians (Perloff, 2021). Traditional political communication, initiated and handled by a narrow circle of politicians and journalists, has fundamentally changed in the age of social media. In these new settings, which empower "average Joes" with a voice to express, challenge, support, or reject (Stier et al., 2018), Twitter serves as a "digital mirror" reflecting the offline community's mood and intentions (Tumasjan et al., 2010; Perloff, 2021).

1.2 Background

Our study offers a methodological and conceptual replication of the paper of Stieglitz and Dang-Xuan (2013), who were among the first to investigate whether the sentiment of political Twitter messages is associated with information diffusion assessed by retweet quantity and retweet speed. Sampling state parliament elections in three German states in 2011, Stieglitz and Dang-Xuan (2013) concluded that emotionally charged content is shared more frequently and swifter than neutral content.

Since then, Twitter's audience has skyrocketed from 68 million monthly active users as of Q1 2011 (Statista.com, 2019) to 211 million daily active users as of Q3 2021 (Businessofapps.com, 2021). The communication density grew from 100 million tweets per day as of January 30, 2011, to over 835 million tweets per day as of November 14, 2021 (Internetlivestats.com, 2021). Remarkable is the rapidity of information diffusion: based on tweets around the 2016 EU referendum ("Brexit") and the 2016 US presidential elections, Gorodnichenko et al. (2021) report that information diffusion is largely complete within 1–2 hours. Twitter's expansion and the importance of speed for extensive audience engagement, as shown, for example, for collective movements, as noted by Trottier and Fuchs (2014), justify the need to reexamine the earlier claimed insights and serve as the basis for the current replication.

Whereas Stieglitz and Dang-Xuan (2013) focused on communication on Twitter prior to the state parliament elections in the German states of Baden-Württemberg, Rheinland-Pfalz (south-western regions), and Berlin, we test their theoretical propositions in the context of state parliament elections in the German state Sachsen-Anhalt (Saxony-Anhalt, central region). Since the original study by Stieglitz and Dang-Xuan, there have been multiple shifts in the political landscape of Germany. The German Pirate Party, one of the most relevant parties in the original study, no longer holds any seats in state-level or federal parliaments. Instead, the right-wing party Alternative für Deutschland (AfD), founded in 2013, was part of all 16 state-level parliaments and part of the federal parliament at the time this research was conducted. However, as our replication is only focused on the sentiment of political tweets and their relationship to tweet metrics, we believe this change should not alter the outcome of our replication.

Moreover, the original paper used the German version of SentiStrength for automatic sentiment analysis. We also used the German version of SentiStrength, thus offering methodological replication as a contribution. Additionally, we employ the Linguistic Inquiry and Word Count (LIWC) software to check the potential sensitivity of results to the used textual analysis tool, targeting conceptual replication merits.

Based on the original study, the link between sentiment and information sharing on Twitter is tested with the following hypotheses:

Hypothesis 1: The larger the total amount of sentiment (positive or negative) a political Twitter message exhibits, the more often it is retweeted.

Hypothesis 2: The larger the total amount of sentiment (positive or negative) a political Twitter message exhibits, the shorter the time lag to the first retweet.

Hypothesis 3: The associations between sentiment and (a) retweet quantity as well as (b) retweet time lag are stronger for tweets with negative sentiment than for those with positive sentiment.

2 Method

Figure 1 summarizes the data collection and analytical procedures we followed and contrasts them with Stieglitz and Dang-Xuan (2013).

	Stieglitz & Dang-Xuan (2013)	Our study
Data acquisition basis	State parliament elections in the German states • Baden-Wurttemberg & Rheinland-Pfalz (21-27.03.2011) • State parliament election in Berlin (29.08.-25.09.2011)	State parliament elections in the German state • Saxony-Anhalt (09.05-20.06.2021)
Keywords	• Names of the six most important German parties **or** • Name of the front-runner of one of these parties in the respective election	• Tweet made by the official account of one of the most relevant political parties in Saxony-Anhalt (AfD, Bündnis90/Die Grünen, CDU, FDP, Die LINKE, SPD) **or** • by the leading nominee of those parties **or** • contained the name of at least one of the leading nominees **or** • contained the name (or abbreviation) or the official hashtag of the party for this election • **and** tweet contained at least one of several used hashtags to identify tweets regarding this state-level election (as opposed to other German elections later in the same year)
Collection tool	• Self-developed Java-based crawler that uses the Twitter “Search API”	• Twitter API for Academic Research and custom python script
Data pre-processing	• Removal of duplicates and non-German tweets • Removal of irrelevant tweets (e.g. advertising)	• not applicable
Sentiment analysis	• SentiStrength (German version)	• SentiStrength (German version) → method. replication • LIWC (German version) → conceptual replication
Regression analysis	• H1, H3a: negative binomial regression model • H2, H3b: OLS	• H1, H3a: negative binomial regression model • H2, H3b: OLS

Figure 1. Overview of data collection and analytical procedures for Stieglitz and Dang-Xuan (2013) and our paper

2.1 Data

We collected a sample of tweets around the state parliament election in Saxony-Anhalt, Germany, on June 06, 2021. Our sample comprises a period of six weeks, from May 09 to June 20, 2021. Thus, the data collection period consists of four weeks prior to the election and two weeks after Election Day on June 06.

We used the Twitter API for Academic Research (Twitter.com, 2021) to collect all tweets that fulfilled the following criteria: The tweet was made by the official account of one of the most relevant political parties in Saxony-Anhalt (AfD, Bündnis90/Die Grünen, CDU, FDP, Die LINKE, SPD) or by the leading nominee of those parties, or contained the name of at least one of the leading nominees or the name (or abbreviation) or the official hashtag of the party for this election. Additionally, only tweets with at least one hashtag as an explicit marker for the Saxony-Anhalt state-level election (#ltw2021 OR #ltw21 OR #ltwlsa21 OR #ltwlsa2021 OR #ltw21 OR #ltw2021 OR #sachsenanhalt OR landtagswahl OR "sachsen-anhalt" or "sachsen anhalt") were included in the sample. The complete query can be found in Appendix A.

The rationale behind these criteria lies in additional German elections conducted later in 2021, precisely, state-level elections in Berlin and federal elections. At the time of data collection, the campaigning and discussions for these elections had already begun. For the sake of replication, we kept the focus on a specific state-level election as Stieglitz and Dang-Xuan (2013) did. Such a strategy might have resulted in a slightly smaller sample size, as the discussions of the state-level election might have been overshadowed by the federal election.

In their data preprocessing, Stieglitz and Dang-Xuan (2013) report the manual elimination of irrelevant tweets, namely: (1) advertising tweets based on typical (German) keywords that signify ads and (2) tweets in languages other than German by applying different language detection tools. In our case, according to the documentation for the Twitter API for Academic Research (Twitter.com, 2021), both issues (1) and (2) are addressed with the search parameters we specified (see Appendix A).

After removing duplicate tweets (based on the ID of the tweets), we used a Python script to collect all retweets of each tweet in our sample, again using the official Twitter API, since the API does not offer the possibility to retrieve retweets of a tweet directly. This step was necessary to calculate the time difference between the time a tweet was originally posted and its first retweet. Our script contained a check so that in the final sample, only tweets where we could indeed retrieve all retweets (the number of found retweets equaled the retweet count returned by the original API call) remained. As a result, 9,848 observations were left for analysis.

Although this sample size is smaller than the sample used by Stieglitz and Dang-Xuan (2013), it is large enough to conduct a meaningful analysis. To recognize small effect sizes (e.g., $f^2 = 0.02$) in a linear regression model with five predictors (H2, H3b), a sample size of 543 would suffice. Thus, the more crucial factor here is that we deem our sample to properly reflect the population of German Twitter users tweeting about politics.

2.2 Measures and Procedures

2.2.1 Sentiment Analysis

Sentiment analysis means the extraction of the emotional tone of a message. Although both manual (e.g., judgments of human coders) and automatic approaches are legitimate, the latter is often preferred when working with big data sets. The estimates (or judgments) vary and can be binary (e.g., positive/negative), trinary (e.g., positive/negative/neutral), scale-based (e.g., -5 for strongly negative / +5 for strongly positive or faceted (e.g., joy (0-100), trust (0-100), sadness (0-100) (Thelwall et al., 2017)). At this point, our analytical procedure bifurcates: We extract sentiment from tweets using: (1) SentiStrength (Thelwall et al., 2017) - the same tool as Stieglitz and Dang-Xuan (2013) – thus repeating the original methodological procedure; (2) LIWC (Pennebaker et al., 2015) – an alternative tool – thus verifying the results conceptually.

SentiStrength assigns each text a positive sentiment score from 1 to 5 and a negative one from -1 to -5 and is dictionary-based, where words are assigned different scores. The initial dictionary was derived in part from the LIWC dictionary. The scoring procedure is as follows: Initially, the scores for each word are given. A sentence's total score comprises the highest positive and negative scores for its constituent words. For a multi-sentence tweet, the highest scores from any sentence are taken, and adjustments are made for non-sentiment terms with a score of 1 (no positivity) or -1 (no negativity). For example, "Yesterday it was horrible [-4] and nasty [-3] outside but the evening was lovely [2]. Rainbow today is fantastic [3]" gives us first sentence scores of 2 and -4. The second sentence scores 3 and -1 (no negativity), and the overall metrics for the tweet are 3 (maximum positive) and -4 (maximum negative).

To capture the degree of emotionality, like Stieglitz and Dang-Xuan (2013), we computed the normalized variable $sentiment = (positive - negative) - 2$, which prevents mutual canceling out of positive and negative scores and lets the variable range from [0, 8] (Table 1).

SentiStrength was developed with the peculiarities of online platforms in mind by optimizing the term weights used in the dictionary based on a MySpace Corpus and considering that spelling and punctuation are used less correctly than in formal publications (Thelwall et al., 2011). The version of SentiStrength used in the original study is the same one used for this research (i.e., there have been no changes to the SentiStrength algorithm or its underlying data/dictionaries).

Contrary to SentiStrength, which produces output purely for positive and negative emotions, LIWC assigns scores for a multitude of factors, such as Time Orientation, Perceptual Processes, and others (Meier et al., 2018). We use LIWC's output for positive emotions and negative emotions. The scores are calculated by processing each word in a given input text and increasing the counters for each category the processed word belongs to. The categories are represented in the form of different dictionaries. For example, finding the word “happy” in a text would increase the counter for positive emotions by one. The final score per category is the percentage of words of each category contained in the input text. For example, “happy” would increase the score for positive emotion in a ten-word sentence by 0.1. Since the dictionaries for positive and negative emotions are mutually exclusive, both scores cannot add up to a sum larger than 1, and both scores are always guaranteed not to be negative. Thus, both scores cannot cancel each other out, and the total sum of sentiment detected with LIWC ranges from [0,100]. Unlike SentiStrength, LIWC has not been developed especially for short-messaging / micro-blogging sites like Twitter. However, past research has shown that it is a suitable tool for analyzing Twitter messages, even focusing on political content. For example, LIWC was used by Stieglitz and Dang-Xuan (2012) to find a positive relationship between the number of words indicating affective dimensions in a tweet and its retweet rate. Research by Tumasjan et al. (2010) used LIWC to map tweet contents to political sentiment.

2.2.2 Regression Analysis

The variables (Table 1) and estimation methods repeat the work of Stieglitz and Dang-Xuan (2013). Following the original paper, to test H1 and H3a, which imply count data for the dependent variable, we used the negative binomial regression model, assuming that the dependent variable follows the negative binomial distribution. Indeed, the data in our sample is overdispersed (with $\chi^2(1)$rt_no >> 1,000 and p-value << 0.0001), and the usage of a negative-binomial model instead of Poisson regression is justified.

Specifically, the regression equation for H1 is:

$$(1)\quad log(E(rt\_no \mid *)) = \beta 0 + \beta 1\, sentiment + \beta 2\, hashtag + \beta 3\, url + \beta 4\, log(follower) + \beta 5\, log(activity)$$

where E(*rt_no* |*) is the expectation of *rt_no* conditional on the set of the independent variables (Stieglitz & Dang-Xuan, 2013, p. 228).

For H3a, the negative sentiment dummy variable (*negative*) and the interaction term (*sentiment* × *negative*) were added:

$$(2)\quad log(E(rt\_no \mid *)) = \beta 0 + \beta 1\, sentiment + \beta 2\, negative + \beta 3\, (sentiment \times negative) + \beta 4\, hashtag + \beta 5\, url + \beta 6\, log(follower) + \beta 7\, log(activity).$$

Table 1. Variables’ definitions and measurements

Variable	Description	Analytical Method/Source
	Dependent variables	
rt_no	The number of retweets	Stieglitz and Dang-Xuan (2013, p. 227)
rt_timelag	Time lag between the tweet and the first retweet (in minutes).	Stieglitz and Dang-Xuan (2013, p. 227)
	Independent variables	
sentiment_ss	Total amount of sentiment calculated with SentiStrength (i.e., the sum of positive/negative emotions detected). Ranges [0, 8], with 0 as the least emotional and 8 being very emotional.	Sentiment analysis with SentiStrength, Stieglitz and Dang-Xuan (2013, p. 227)
sentiment_liwc	Total amount of sentiment calculated with LIWC (i.e., the sum of the negative and positive sentiment detected). Ranges [0,100], with 0 as the least emotional and 100 being very emotional.	Sentiment analysis with LIWC

Table 1. Variables' definitions and measurements

negative	Whether or not the negative sentiment detected by SS/LIWC is stronger than the positive sentiment detected. Either 1 (negative sentiment is dominating) or 0 (negative is not dominating).	Sentiment analysis with SentiStrength/LIWC, Stieglitz and Dang-Xuan (2013, p. 228)
sentiment x negative	Multiplicative interaction term between negative and sentiment. Thus, ranging from [0,8] for SS, and from [0,100] for LIWC.	Sentiment analysis with SentiStrength/LIWC, Stieglitz and Dang-Xuan (2013, p. 228)
Control variables		
hashtag	The number of hashtags a tweet contained.	Stieglitz and Dang-Xuan (2013, p. 228)
url (dummy)	Dummy (binary) variable for whether a URL was included in the tweet.	Stieglitz and Dang-Xuan (2013, p. 228)
follower	Number of followers of a user (tweet's author).	Stieglitz and Dang-Xuan (2013, p. 228)
activity	Number of tweets the user has posted during the sample period.	Stieglitz and Dang-Xuan (2013, p. 228)

H2 and H3b involve the numeric dependent variable *rt_timelag*. The regression was estimated with ordinary least squares (OLS). To account for non-normality, we log-transformed the dependent variables before employing OLS regression. Given the large sample size of more than 3,500 observations, the validity of parametric tests lies in the central limit theorem (Lumley et. al., 2002).

For H2, the regression model looks as follows:

$$(3)\quad log(rt\_timelag) = \beta 0 + \beta 1\, sentiment + \beta 2\, hashtag + \beta 3\, url + \beta 4\, log(follower) + \beta 5 log(activity) + \varepsilon$$

For H3b, the negative sentiment dummy variable (negative) and the interaction term (*sentiment* × *negative*) were added:

$$(4)\quad \log(rt_{timelag}) = \beta 0 + \beta 1\, sentiment + \beta 2\, negative + \beta 3\, (sentiment \times negative) + \beta 4\, hashtag + \beta 5\, url + \beta 6 \log(follower) + \beta 7 \log(activity) + \varepsilon.$$

Please note that since there exist tweets that were not shared and since verification of H2 and H3b is relevant for retweeted content only, we created a subsample applying the criterion *rt_no* > 0, yielding 3,645 retweeted tweets ($N_{\text{retweeted}}$ =3,645).

We disclose our final data set as well as all code used for the collection, preprocessing, and analysis of the data.[1]

3 Results

The distribution of emotionally charged Twitter messages is presented in Appendix B (correspondence to Table 4 in Stieglitz and Dang-Xuan (2013, p. 232)). Summary statistics of variables used in regression analyses are given in Appendix C (correspondence to Table 7 in Stieglitz and Dang-Xuan (2013, p. 235)). In our sample, a tweet is retweeted about 1.54 times, and the first retweet happens 134 minutes (8046.9 seconds) after posting, on average. The average sentiment per tweet, according to SentiStrength, equals 1.04, and the number of hashtags slightly exceeds 2 (mean = 2.26). The average number of followers a

[1] Link to data, code, and results: https://github.com/linusha/twitter-saxony-anhalt-election-2021-sentiment

user has in our sample is 10,310 (with a mean of 512), which is significantly higher than in the sample of Stieglitz and Dang-Xuan (2013), who report an average of 630 followers in the Baden-Württemberg and Rheinland-Pfalz sample and 878 followers in the Berlin sample as of 2011 (date of data collection). We link the observed difference to Twitter's growth in the last ten years, as mentioned in the Introduction, Section 1, of this paper (Statista.com, 2019; Businessofapps.com, 2021).

The Correlation Matrix of Independent Variables (correspondence to Tables 8 and 9 in Stieglitz and Dang-Xuan (2013, p. 236)) is presented in Table 2. Similar to Stieglitz and Dang-Xuan (2013), variables modeled as independent have evidenced weak correlation (the highest $\rho_{follower;\,hashtag} = -0.026$), advocating that multicollinearity is not a problem in our sample.

Table 2. Correlation Matrix of Independent Variables (N = 9,848)

Variable	sentiment (Senti-Strength)	sentiment (LIWC)	hashtag	url (dummy)	follower	activity
sentiment (Senti-Strength)	1					
sentiment (LIWC)	0.395***	1				
hashtag	0.055*** [0.004,-0.01]	-0.085***	1			
url (dummy)	-0.174*** [-0.05,-0.06]	-0.144***	-0.105*** [-0.06,-0.05]	1		
follower	-0.056*** [-0.03,0.01]	-0.021**	-0.026** [-0.04,0.03]	0.102*** [0.07*,0.02]	1	
activity	-0.101*** [0.00,0.01]	-0.087***	-0.158*** [0.15*,0.06*]	0.205*** [-0.01,0.08*]	0.016* [-0.04,-0.01]	1

Notes: * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level. Values in brackets are the values from [Table 8, Table 9] in the original paper.

3.1 Methodological Replication with Sentiment Analysis done in SentiStrength

Results on retweet quantity presented in Table 3 indicate that sentiment estimated with SentiStrength is significantly associated with a higher number of retweets (b = 0.139, p < 0.001). For interpretation, we computed the exponentiated betas. As such, a one-unit increase in the overall sentiment would generate 1.149 times more retweets (exp(1.139)=1.149), or a 14.9 percent addition. The observed effect is about twice as high as the 6 percent expansion for Baden-Württemberg and Rheinland-Pfalz and more than three times higher than the 4 percent expansion for the Berlin sample reported by Stieglitz and Dang-Xuan (2013); H1 is supported.

The interaction term *sentiment* × *negative* appears to be insignificant (b = -0.025, p = 0.615), declining the speculated moderating effect of sentiment polarity on the link "sentiment – retweet quantity"; H3a is rejected. The original study is inconclusive since the interaction term was insignificant in the Baden-Württemberg and Rheinland-Pfalz sample and significant in the Berlin sample. Control variables *hashtag*, *url*, and *follower* are significantly positively associated with the retweet count, in line with Stieglitz and Dang-Xuan (2013). We also observe that user activity is significantly negatively related to retweet count (b = 0.091, p < 0.001), suggesting that simply generating more content does not yield more sharing of this content. Stieglitz and Dang-Xuan (2013) also report a significant negative effect of user activity for the Berlin sample and an insignificant effect for the Baden-Württemberg sample.

Table 3. Negative Binomial Regression Results

Dependent variable: *rt_no* (number of retweets)						
	H1			**H3a**		
Independent variables	*b*	SE	exp (b)	*b*	SE	exp (b)
sentiment	0.139***	0.018	1.149	0.190***	0.023	1.210
negative				0.139***	0.018	1.149
sentiment ×negative				0.139***	0.018	1.149

Table 3. Negative Binomial Regression Results

hashtag	0.160***	0.008	1.173	0.160***	0.008	1.173
url	0.086*	0.048	1.090	0.100**	0.048	1.104
log(follower)	0.430***	0.009	1.537	0.430***	0.009	1.537
log(activity)	-0.089***	0.016	0.915	-0.091***	0.016	0.913
constant	-3.446***	0.081	0.032	-3.456***	0.081	0.032
McFadden	0.067			0.067		
Cox and Snell	0.179			0.180		
Nagelkerke	0.188			0.190		
N observations	9,848			9,848		
Notes: *b* is the estimated coefficient, *exp(b)* is the exponentiated estimated coefficient, and SE is the estimated robust standard errors. * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level.						

Results on retweet speed were obtained from a reduced sample (n = 3,645). Out of 9,894 tweets, 3,691 were retweeted at least once. The examination of time to first retweet (*rt_timelag*) distribution suggests a large spread of values, with a minimum of 3 seconds and a maximum of 1,272,279 sec (i.e., 21,205 min or 353 hours or 14.7 days). Bearing in mind that information diffusion is largely complete within 1-2 hours (Gorodnichenko et al., 2021) and the average time of 72 min (SD=291 min) for the Baden-Württemberg and Rheinland-Pfalz sample and 114 min (SD=695 min) for the Berlin sample (Stieglitz & Dang-Xuan 2013), along with the sensitivity of OLS estimators to outliers, we kept observations that were retweeted within 24 hours of their initial posting for further analysis.

Thus, the subsample contains 3,645 observations, and the average time to first retweet is 72 min (SD=193 min, median= 7.45 min). In our data, most tweet dissemination happens within 500 min (i.e., 8 hours 20 min), which looks plausible.

Correlation analysis of independent variables (Table 4) suggests the absence of multicollinearity in the data.

Table 4. Correlation Matrix of Independent Variables for Reduced Sample (N = 3,645)

Variable	**sentiment (Senti-Strength)**	**sentiment (LIWC)**	**hashtag**	**url (dummy)**	**follower**	**activity**
sentiment (Senti-Strength)	1					
sentiment (LIWC)	0.402***	1				
hashtag	0.047*** [0.02,-0,04]	-0.085***	1			
url (dummy)	-0.159*** [-0.05,-0,07]	-0.135***	-0.058*** [-0.04,0.03]	1		
follower	-0.094*** [-0.05,-0,01]	-0.008	-0.078*** [-0.11*,-0.01]	0.131*** [0.11*,0.06*]	1	
activity	-0.136*** [0.04,0.02]	-0.140***	0.025 [0.20**,0.12*]	0.214*** [-0.08*,0.02]	0.181*** [-0.10*,0.01]	1
Notes: * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level. Values in brackets are the values from [Table 12, Table 13] in the original paper.						

Results presented in Table 5 inform about the insignificance of sentiment (b = -0.0006, p = 0.8540) as an explanatory variable for retweet speed; H2 is rejected. This finding disagrees with Stieglitz and Dang-Xuan (2013), who report significant negative coefficients of sentiment (Baden-Württemberg and Rheinland-Pfalz sample: b = –0.05, p < 0.05 and Berlin sample: b = –0.04, p < 0.01), meaning that emotionally charged content spreads faster than a neutral one. At the same time, a closer look into Table 14 (Stieglitz & Dang-Xuan 2013, p. 240) reveals a discrepancy between the textual description and the table: for the Berlin sample, while the text claims significance at a 1% level (Stieglitz & Dang-Xuan 2013, p. 236), the table reports the significant only at a 10% level.

The interaction term *sentiment* × *negative* called to test the moderation effect of negative content appears to be insignificant (b = 0.088, p = 0.308); H3b is rejected. This finding implies that negatively charged tweets do not necessarily disseminate quicker and conforms to Stieglitz and Dang-Xuan (2013). The controls *hashtag* and *follower* are significantly negatively related to the retweet speed. Contrary to (Stieglitz & Dang-Xuan 2013), who claim the negative impact of all controls, the presence of *url* is positively related to the retweet time lag, and the user *activity* is insignificant.

Table 5. OLS Regression Results

Dependent variable: *rt_timelag* (time lag between the tweet and the first retweet)				
	H2		**H3b**	
Independent variables	*b*	SE	*b*	SE
sentiment	-0.006	0.031	-0.029	0.039
negative			-0.135	0.199
Sentiment x negative			0.088	0.086
hashtag	-0.027**	0.013	-0.027**	0.013
url	0.437***	0.078	0.432***	0.079
log(follower)	-0.164***	0.016	-0.164***	0.016
log(activity)	-0.027	0.030	-0.028	0.030
constant	3.335***	0.151	3.355***	0.152
R2 adjusted	0.036		0.035	
N observations	3,645		3,645	
Notes: b is the estimated coefficient, and SE is the estimated robust standard errors. * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level.				

3.2 Conceptual Replication with Sentiment Analysis done in LIWC

Contributing to the discussion on how best to extract sentiment using 'off-the-shelf' dictionaries – the core variable of this study – we measure the emotional intent alternatively with LIWC (Pennebaker et al., 2015) and yield conceptual replication. LIWC outputs standardized scores derived from raw frequency divided by word count, as explained in Section 2.2.

Results on retweet quantity are displayed in Table 6. Surprisingly, computed with LIWC, overall sentiment is significantly negatively related to the number of retweets (b = -0.011, p = 0.003), suggesting an aversion to emotionally charged messages. This rejects H1; furthermore, it advocates the opposite (negative) direction of the relationship between the overall sentiment and retweet volume compared to Stieglitz and Dang-Xuan (2013). Precisely, a one-unit increase in sentiment, i.e., if the fraction of emotionally charged words increases by 1 percent, would generate 1.1 percent fewer retweets.

Table 6. Negative Binomial Regression Results

Dependent variable: *rt_no* (number of retweets)						
	H1			**H3a**		
Independent variables	*b*	SE	exp (b)	*b*	SE	exp (b)
sentiment	-0.011**	0.004	0.989	-0.008**	0.004	0.992
negative				0.182*	0.109	1.199
sentiment × negative				-0.035**	0.014	0.966
hashtag	0.163**	0.008	1.176	0.162***	0.008	1.176
url	0.046	0.048	1.047	0.051	0.048	1.052

Table 6. Negative Binomial Regression Results

log(follower)	0.427***	0.009	1.533	0.426***	0.009	1.531
log(activity)	-0.111***	0.016	0.895	-0.110***	0.016	0.896
constant	-3.155***	0.080	0.043	-3.161***	0.080	0.042
McFadden	0.065			0.065		
Cox and Snell	0.174			0.175		
Nagelkerke	0.184			0.175		
N observations	9,848			9,848		
Notes: b is the estimated coefficient, exp(b) is the exponentiated estimated coefficient, and SE is the estimated robust standard errors. * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level.						

The interaction term *sentiment* × *negative* is significant (b = -0.035, p = 0.011), suggesting that the “the lower the emotionality, the higher the retweets” pattern is weaker for negative messages than for positive ones; H3a is rejected. Noteworthy, our results with LIWC indicate the opposite (negative) moderation compared to the positive moderation hypothesized and observed in the Berlin sample by (Stieglitz & Dang-Xuan 2013). Controls like *hashtag* and *follower* are positively related to the number of retweets. In contrast, *activity* is negatively linked to retweet quantity. The presence of *url* has no significant impact.

Results on retweet speed are given in Table 7 and reveal sentiment is an insignificant predictor (b = 0.001, p = 0.866); H2 is rejected. The interaction term *sentiment* × *negative* is also insignificant (b = 0.010, p = 0.667); H3b is rejected.

As for control variables, the higher the number of *hashtags* and *followers*, the lower is a tweet’s time to its first retweet. User *activity* appears to be insignificant in predicting the speed of information dissemination.

Table 7. OLS Regression Results

Dependent variable: *rt_timelag* (time lag between the tweet and the first retweet)				
	H2		**H3b**	
Independent variables	b	SE	b	SE
sentiment	0.001	0.006	-0.001	0.007
negative			0.170	0.181
Sentiment x negative			0.010	0.023
hashtag	-0.027**	0.013	-0.028**	0.013
url	0.440***	0.078	0.431***	0.079
log(follower)	-0.164***	0.016	-0.162***	0.016
log(activity)	-0.025	0.030	-0.027	0.030
constant	3.317***	0.147	3.290***	0.148
R2 adjusted	0.036		0.037	
N observations	3,645		3,645	
Notes: b is the estimated coefficient, and SE is the estimated robust standard errors. * Significant at the 10 percent level; ** significant at the 5 percent level; *** significant at the 1 percent level.				

4 Discussion

4.1 Interpretation of Results

We methodologically and conceptually replicated the study of Stieglitz and Dang-Xuan (2013) which was empirically validated in the context of political communication on Twitter. This is the first replication of the original paper to the best of our knowledge. Table 8 exhibits the comparison of findings.

Table 8. Comparison of our findings to Stieglitz and Dang-Xuan (2013)

Hypothesis	Supported in Stieglitz and Dang-Xuan (2013)?	Supported in the current study with SentiStrength?	Supported in the current study with LIWC?
H1: The larger the total amount of sentiments a political Twitter message exhibits, the more often it will be retweeted.	Yes (BW&RP sample: $p<0.01$; Berlin sample: $p<0.05$)	Yes ($b = 0.139$, $p < 0.001$)	No ($b = -0.011$, $p = 0.003$)
H2: The larger the total amount of sentiments a political Twitter message exhibits, the shorter the time lag to the first retweet will be.	Yes (BW&RP sample: $p<0.05$; Berlin sample: $p<0.1$)	No ($b = -0.0006$, $p = 0.8540$)	No ($b = 0.001$, $p = 0.866$)
H3a: The association between sentiment and retweet quantity is stronger for tweets with negative sentiment than for those with positive sentiment.	Partial (Berlin sample only) (BW&RP sample: $p>0.1$; Berlin sample: $p<0.05$)	No ($b = -0.025$, $p = 0.615$)	No ($b = -0.035$, $p = 0.011$)
H3b: The association between sentiment and retweet time lag is stronger for tweets with negative sentiment than for those with positive sentiment.	No (BW&RP sample: $p>0.1$; Berlin sample: $p>0.1$)	No ($p= 0.308$)	No ($b = 0.010$, $p = 0.667$)
Note: BW&RP - Baden-Württemberg and Rheinland-Pfalz.			

Our methodological replication confirms that emotionally charged messages are retweeted more often. Stieglitz and Dang-Xuan (2013) found significant evidence in the data scraped in 2011 for the federal German states Baden-Württemberg and Rheinland-Pfalz and Berlin, while we observe the same pattern in the online conversations taking place in 2021 around the elections in Saxony-Anhalt. Surprisingly, the conceptual replication, when sentiment was assessed with an alternative tool, LIWC, suggests the opposite pattern. Precisely, the sentiment here is negatively related to retweet count.

Further, we reject the negativity bias assumption (i.e., that the link between sentiment and retweet quantity is stronger for tweets with negative sentiment than for those with positive sentiment). Stieglitz and Dang-Xuan (2013) claim partial support and observe the significant moderation effect in the Berlin sample but not in the Baden-Württemberg and Rheinland-Pfalz sample. Both the methodological and conceptual replication reject H3. On top of that, the conceptual replication found that the link between sentiment and retweet quantity is weaker for negatively charged posts than for positively charged posts.

The most glaring difference to the original paper is proposition H2, namely that higher emotionality corresponds to a shorter time lag to the first retweet. With our data, we did not find evidence to support this statement. A possible explanation for this result could be the evolution of the Twitter community: Since 2011, Twitter has exhibited heavy expansion (Internetlivestats.com, 2021). The growth could have also led to centralization, namely the emergence of thought leaders with many followers. Thus, the large group of followers might have become the driver of information dissemination in the network. Subsequently, the importance of content emotionality would have become smaller. In other words, for those groups of followers, the original author might be more important than the content of a specific tweet, (e.g., due to past agreements with the author or the author's general fame).

Finally, our study agrees with the original paper on the insignificance of a tweet's negative sentiment in moderating the link "sentiment - retweet time lag." The control variables *hashtag* and *url* are significantly positively related to the retweets' volume, which is in line with findings from the original paper. The number of *hashtags* and *followers* contribute to quicker sharing; however, *activity* reflecting the number of posts does not matter for retweet speed.

4.2 Theoretical Contributions

Our study makes several contributions to information systems (Mallipeddi et al., 2021) and political communication literature (Perloff, 2021). First, this study adds to the discussion on whether strategic crafting of social media messages is effective for information dissemination, especially in the political domain. Similar to Stieglitz and Dang-Xuan (2013) and Mallipeddi et al. (2021), our methodological replication suggests that highly emotional content is retweeted more. However, in contrast to the original paper, our recent evidence on elections in a German federal state shows that emotionally charged messages do not necessarily spread faster. This insight points to the changes in the Twitter community due to its extreme growth, which might have influenced the content dynamics on the platform.

Next, this study provides a valuable conceptual replication of the Stieglitz and Dang-Xuan (2013) model. In contrast to SentiStrength, the sentiment assessment with another automatic tool, LIWC, yielded a different conclusion on H1, suggesting that more neutral content receives more retweets. Thus, we support the discussion of Chan et al. (2021), who compared sentiment scores of 37 "off-the-shelf" dictionaries and demonstrated how results might differ based on the dictionary choice. In line with the best practices advice, both of the tested tools are suitable and were applied for the analysis of tweets in the political context (e.g., LIWC was used to create psychological profiles of election candidates in the Federal election in Germany in 2009 (Tumasjan et al., 2010) and SentiStrength was used by the Stieglitz and Dang-Xuan (2013)). The current study passes a cautionary message for future investigations that employ a dictionary-based approach for sentiment estimation, calling for the verification of results with multiple lexicons prior to conclusions. Further research comparing LIWC and SentiStrength usage for Twitter messages could be especially interesting, as the LIWC dictionary is one part of the SentiStrength dictionary. Although outside of the scope of this replication, future research investigating questions with regard to the benefits of "social media attuned" tools like SentiStrength over off-the-shelf dictionaries like LIWC, as well as regarding the importance of more advanced algorithms over simple word-proportion-calculations could provide interesting insights. This is especially true for short messages like tweets, which might be less accurately evaluated by simple word-counting techniques. These could also benefit from further validation by exploring the manual sentiment coding of tweets. Finally, we respond to the recent call for replications studies in Information Systems to update original models and theories (Brendel et al., 2021).

4.3 Practical Implications

For practicians, our results imply the importance of followers in political conversations. A bigger base of subscribers robustly increases the post's likelihood of becoming viral if measured by the retweet volume and time to first retweet. Next, since the number of hashtags is also significantly linked to more intensive sharing and less time to first retweet, content creators might intentionally include more hashtags in their Twitter posts. As for sentiment, our methodological replication would still advise crafting emotionally charged messages for broader information dissemination. Stronger sentiment, however, cannot guarantee a high speed of distribution. Contrary to prior evidence, we did not find support for the negativity bias; thus, we claim no priority of a negatively toned message over a positive one in terms of sharing volume or rapidity.

4.4 Limitations and Future Research

Our study has limitations, opening up avenues for future research. Similar to the original paper, political conversations on Twitter in Germany were taken as a research site. Further studies might sample political discourse in other cultural settings or examine sentiment as a virality factor in another (non-political) domain (e.g., for product promotion). Moreover, submitting mixed evidence between methodological and conceptual replication, we call for further triangulation of the theoretical model with other dictionary-based approaches, beyond SentiStrength and LIWC, or more advanced classification techniques like custom machine learning or deep learning sentiment analyzers.

Further research should also investigate the importance of emotional content versus the importance of followers with regards to the retweets content accumulates, noting that our findings regarding H2 differ from the ones in the original study from 2013.

4.5 Conclusion

To sum up, our replication of the Stieglitz and Dang-Xuan (2013) study about the potential link between emotions on Twitter and information dissemination during a political event is only partially consistent with the initial findings. With the methodological replication, we also conclude that emotionally charged content (i.e., overall sentiment) is shared more but not necessarily quicker. Our conceptual replication with LIWC for the sentiment assessment exhibits the opposite: the lower the sentiment, the higher the number of retweets. Both replications are aligned with the original study on the insignificance of sentiment in explaining the time to first retweet. The number of followers and number of hashtags are consistently linked to a higher speed of information dissemination.

References

Abramova O., Batzel K., & Modesti M. (2022). Collective response to the health crisis among German Twitter users: A structural topic modeling approach. *International Journal of Information Management, 2(2).*

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192-205.

Brendel, A. B., Diederich, S., & Niederman, F. (2021). An immodest proposal–going “All in” on replication research in information systems. *European Journal of Information Systems*, 1-10.

Businessofapps.com. (2021). *Twitter revenue and usage statistics*. Retrieved from https://www.businessofapps.com/data/twitter-statistics/

Chan, C. H., Bajjalieh, J., Auvil, L., Wessler, H., Althaus, S., Welbers, K., van Atteveldt, W. & Jungblut, M. (2021). Four best practices for measuring news sentiment using ‘off-the-shelf’dictionaries: A large-scale p-hacking experiment. *Computational Communication Research*, *3*(1), 1-27.

Chen, L., Baird, A., & Straub, D. (2020). A linguistic signaling model of social support exchange in online health communities. *Decision Support Systems*, *130*, 113233.

Faralli, S., Rittinghaus, S., Samsami, N., Distante, D., & Rocha, E. (2021). Emotional intensity-based success prediction model for crowdfunded campaigns. *Information Processing & Management*, *58*(1), 102394.

Fischer-Preßler, D., Schwemmer, C., & Fischbach, K. (2019). Collective sense-making in times of crisis: Connecting terror management theory with Twitter user reactions to the Berlin terrorist attack. *Computers in Human Behavior*, *100*, 138-151.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *European Economic Review*, 103772.

Hagemann L., & Abramova O. (2022). Crafting audience engagement in social media conversations: Evidence from the U.S. 2020 presidential elections. In *Proceedings of the 55rd Hawaii International Conference on System Sciences*.

Heimbach, I., & Hinz, O. (2016). The impact of content sentiment and emotionality on content virality. *International Journal of Research in Marketing*, *33*(3), 695-701.

Internetlivestats.com. (2021). Retrieved from https://www.internetlivestats.com/twitter-statistics/#trend

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*(1), 151-169.

Mallipeddi, R. R., Janakiraman, R., Kumar, S., & Gupta, S. (2021). The effects of social media content created by human brands on engagement: Evidence from Indian general election 2014. *Information Systems Research*, *32*(1), 212-237.

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). “LIWC auf Deutsch”: The development, psychometrics, and introduction of DE-LIWC2015. Retrieved from https://osf.io/tfqzc/

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Perloff, R. M. (2021). *The dynamics of political communication: Media and politics in a digital age*. Routledge.

Statista.com. (2019) *Number of monthly active Twitter users*. Statista. Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media — Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), 217-248.

Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging — An empirical analysis of sentiment in Twitter messages and retweet behavior. In *Proceedings of the 45th Hawaii International Conference on System Sciences*.

Stieglitz, S., Mirbabaie, M., Schwenner, L., Marx, J., Lehr, J., & Brünker, F. (2017) Sensemaking and communication roles in social media crisis communication. In *Proceedings of the 13th International Conference for Wirtschaftsinformatik*.

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political Communication*, *35*(1), 50-74.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, *62*(2), 406-418.

Thelwall, M. (2017). Heart and soul: Sentiment strength detection in the social web with sentistrength. In J. A. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 119-134). Springer.

Trottier, D., & Fuchs, C. (Eds.). (2015). *Social media, politics and the state: Protests, revolutions, riots, crime and policing in the age of Facebook, Twitter and YouTube*. Routledge.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 4, No. 1).

Twitter.com (2021). *Twitter API for academic research*. Retrieved from https://developer.twitter.com/en/products/twitter-api/academic-research

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American Behavioral Scientist*, *57*(7), 920-942.

Appendix A: Keyword Query Used to Retrieve Relevant Tweets with Twitter API

(@DIE_LINKE_LSA OR @GRUENE_LSA OR @AfD_LSA OR @cdulsa OR @SPD_LSA OR @FDP_LSA OR "Lydia Hüskens" OR "Eva von Angern" OR "Oliver Kirchner" OR "Katja Pähle" OR "Reiner Haseloff" OR "Cornelia Lüddemann" OR @Eva0112 OR @Connylue OR @reinerhaseloff OR @KatjaPaehle OR @LydiaHueskens OR @O_KirchnerAfD OR ((CDU OR AfD OR FDP OR SPD OR "die grünen" OR "die linken" OR "die linke" OR "b90" OR #afd OR #fdp OR #zweitstimmegrün OR #besserdielinke OR #dielinke) (#ltw2021 OR #ltw21 OR #ltwlsa21 OR #ltwlsa2021 OR #ltw21 OR #ltw2021 OR #sachsenanhalt OR landtagswahl OR "sachen-anhalt" or "sachsen anhalt")))

Appendix B: Distribution of Emotionally Charged Twitter Messages

Table B1. Distribution of Emotionally Charged Twitter Messages

	Senti-Strength		**LIWC**	
	Top 50 retweeted users	**Total sample**	**Top 50 retweeted users**	**Total sample**
Emotionally charged tweets (sentiment > 0)	456	5607	536	6871
Positive sentiment tweets (polarity > 0)	309	3333	373	4449
Negative sentiment tweets (polarity < 0)	121	1890	106	1581
Mixed sentiment tweets (polarity = 0 & sentiment > 0)	26	384	57	841
Total	839	9848	839	9848
Note: Only tweets (and no retweets) are regarded.				

Appendix C: Summary Statistics of Variables Used in Regression Analyses

Table C1. Summary Statistics of Variables Used in Regression Analyses

Varibale	**Mean**	**Standard Deviation**
Dependent variables		
rt_no	1.54	3.95
*rt_timelag**	4320.96	11564.5
Independent variables		
sentiment (SentiStrenth)	1.04	1.17
sentiment (LIWC)	5.91	6.18
Control variables		
hashtag	2.25	2.63
url (dummy)	0.66	0.47
follower	43370.18	248868
activity	16.58	41.55
N = 9848		
Note: For rt_timelag, only tweets that have been retweeted at least once in the first 24 hours of being posted have been considered (N=3,645).		

About the Authors

Linus Hagemann was a student assistant at the Chair for Information Systems, especially Social Media and Society, at the University of Potsdam and holds a B.Sc. in IT-Systems-Engineering from Hasso-Plattner-Institute, Potsdam. His research interests lie at the intersection of computer science, big data analytics and the social sciences, esp. using large-scale behavioral and network data to investigate human activity and computer-mediated communication. His research was presented at Hawaii International Conference on System Sciences and published in Internet Research journal.

Olga Abramova is a post-doctoral researcher at the Chair for Information Systems, especially Social Media and Society, at the University of Potsdam. She studied business administration and received her Ph.D. in Information Systems at the Technical University of Darmstadt. Her research focuses on the economic, social and psychological aspects of digitalization, uncovering “the bright and dark sides” of modern ICTs. Her works have been presented at conferences such as International Conference on Information Systems, European Conference on Information Systems and Hawaii International Conference on System Sciences and published in journals such as Electronic Markets, International Journal of Information Management, Internet Research and Die Unternehmung.