1-2021

Effects of Sentiments on the Morphing of Falsehoods and Correction Messages on Social Media

Kelvin K. King
The University of Texas Rio Grande Valley

Bin Wang
The University of Texas Rio Grande Valley

Diego Escobari
The University of Texas Rio Grande Valley, diego.escobari@utrgv.edu

Recommended Citation

King, Kelvin, Bin Wang, and Diego Escobari. 2021. "Effects of Sentiments on the Morphing of Falsehoods and Correction Messages on Social Media." In Proceedings of the 54th Hawaii International Conference on System Sciences 2021, 6563. Manoa, HI. https://doi.org/10.24251/HICSS.2021.789.

Effects of Sentiments on the Morphing of Falsehoods and Correction Messages on Social Media

Kelvin K. King
University of Texas, Rio Grande Valley
kelvin.king@utrgv.edu

Bin Wang
University of Texas, Rio Grande Valley
bin.wang@utrgv.edu

Diego Escobari
University of Texas, Rio Grande Valley
diego.escobari@utrgv.edu

Abstract

The burgeoning literature on fake news reveals that the emotional context of the message is a major factor that drives its diffusion on social media. However, studies have largely missed a major aspect of the diffusion process, which is the morphing of the textual contents themselves during this process. Our study first visually illustrates through hazard functions that, while falsehoods morphs aggressively at the initial stages, correction messages morph more aggressively in the long run. In addition, we leverage on cosine distance and econometric modeling to empirically investigate how sentiment affects the morphing of fake news and their correction messages. We find that positive sentiments, emotionally charged messages and correction messages positively affect the morphing of messages during the diffusion process. Our results also show that, as time goes by, the impacts of sentiments on morphing change.

1. Introduction

Deception in social media, in the form of "fake news" engineered as a deliberate campaign to wage war and influence user perception, has received much attention from both academia and industry alike [6]. The research foci have been on content verification through reactive measures such as presentation and source credibility and their ratings [31, 43] and through proactive methods such as detection [34, 49] and correction messages [27, 45]. Little research has examined how messages mutate or morph during the diffusion process through changes in their contents. The current research fills this gap in the literature that has investigated the diffusion of falsehoods and correction news as a static communication process. Our study differs significantly from previous research [24, 31, 52] in several ways. First, we focus on the evolution of tweets over time rather than combining them into a corpus. To the best of our knowledge, this is the first study examining how fake news and their correction messages morph over time. Second, through the use of repeated event hazard functions, we visually illustrate the difference in the morphing rates of falsehoods and correction messages. Third, we employ a unique panel data set on 14 verified falsehoods and corrections topics that circulated on Twitter during Hurricane Harvey in 2017 to investigate the role sentiment plays in the morphing process. We attempt to answer the following research questions:

1. How do false and correction messages morph on social media?
2. What are the effects of sentiments on the morphing of false news and correction messages? How do these effects change over time?

Because emotive components are important factors in the virality of messages on social media [44], we examine how they may be a major factor in the morphing of not just falsehoods but also correction messages. Using cosine distance to measure the morphing of the messages, our empirical analysis shows that emotive components affect the morphing of both falsehoods and correction messages and positive emotions are more influential as compared with negative and neutral ones. We also show that emotionally charged messages with both strong positive or negative emotions morph more than neutral ones irrespective of the period. Our study also shows that the impacts of emotions on morphing change over time.

2. Background Literature

2.1. Fake News

Following prior research, we define fake news or falsehoods as "*any news item mimicking legitimate news and designed to mislead*" [1, 33]. Due to the wide spread of false news on social media, practitioners and academics have proposed several approaches to combat fake news. Using a cognitive reflection test (CRT), a

URI: https://hdl.handle.net/10125/71409
978-0-9981331-4-0

recent study finds that users rely on their analytical thinking to assess the veracity of headlines, irrespective of the consistencies or inconsistencies between the stories and their political ideologies [47]. In particular, users' susceptibility to fake news is influenced by "lazy-thinking" rather than partisan bias. In addition, users with "reflexive open mindedness" have the propensity to fall for fake news stories [48]. Other studies have shown that as rumors propagate rapidly through social media, tools such as rumor combating sites and tools are quite effective in creating awareness and slowing its spread [15].

Information systems (IS) researchers have proposed leveraging "source ratings" in the fight against misinformation [32]. Empirical evidence shows that the use of source rating and news presentation are effective in the fight against fake news propagation [28, 31]. In contrast, although flagging fake news triggers increased cognitive activity and stimulates users to extend more time in considering news headlines, it ultimately has no effect on the users' beliefs or judgment about the veracity of the news item [40]. Similarly, in the event that falsehoods are tagged, untagged headlines even in cases of falsehoods are automatically assumed to be more accurate and are given more consideration for sharing on social media [46].

2.2. Information Morphing

We define *information morphing* on social media as the change in textual contents from its original message over time. This can be achieved by adding, subtracting and substituting characters in the text [14]. Despite an abundance of research on information and rumor diffusion, the focus of the extant literature has been on the diffusion rate and its contributors with no study done on the morphing of the messages during the diffusion process [16, 30]. There have been very few studies, if any, that have attempted to understand how news evolve over time. For example, Friggeri et al. [24] found that rumors do not particular die out on Facebook but persist in low frequencies and come back after a while. Furthermore, using political tweets, a recent study analyzed the average change in corpus of false and real tweets when they resurfaced and found that falsehoods change at an average of 0.5 when they are reintroduced, while real news was not investigated because they were not observed to return [52]. Scholars argue that false news gains its strength through repetition [23]. Experiments showed that messages get distorted as they flow through a channel [57].

As a microblogging site, Twitter depends on a directed friendship or followership even though reciprocity is not required [37]. Retweeting, which is basically reposting an original post, can introduce the content to a new audience and such retweeted messages can usually be modified so that they lose any reference to the original and can even be posted to a different social network [14]. This propels tweets to go even further without the knowledge of the original tweeter as they reach a wider audience [37]. This implies that the morphing of Twitter messages can take any form, as Twitter allows modification to whatever extent that suits the re-tweeter's need. In the current study, we use cosine distance, which equals 1 - cosine similarity, to measure how dissimilar a tweet is to the original tweet. The morphing of a message is the inverse of the similarity between the original tweet and subsequent tweets.

Research shows that rebuttals and corrections at times can be very effective in addressing misinformation on social media by reducing the credibility of the refuted content [27]. Study further shows that message or rumor-correcting tweets are more propagated or spread more than the rumors themselves [17]. This is very important as it shows the power of rebuttals, coupled with the fact that such rebuttals that are retweeted can be altered and modified. This study thus allows us to have a better understanding of the mechanics on how false news morphs over time and the role rebuttals play in the evolution framework. Our study is quite different from previously mentioned studies and places relatively less emphasis on the generality of the spread of the underlying phenomenon. In addition, these previous studies tend to treat the mutability of misinformation as a corpus thereby losing valuable information. On the other hand, our study takes an alternative perspective, which views misinformation as verifiable false news that are mutable and robust as they diffuse. For a message to morph it first needs to be shared or propagated. However, a message that propagates does not have to morph. We explore this idea using a fixed effects model with multiple time series levels while controlling for the word counts and variability.

2.3. Sentiment

With the rise of Twitter as one of the most archetypal social media platforms for user-generated content, researchers in IS and beyond have since relied on Twitter sentiments for inferring user beliefs and perception [36]. These studies have ranged from the use of microblogs on unidirectional platforms such as Twitter which leads to asymmetrical connections [55] to bidirectional platforms such as Facebook [54]. These studies have revealed the importance of sentiments, an affective or emotional state affecting a user's judgment of a topic. These studies illuminate how sentiments can be inferred from textual contents [13] and applied in understanding user behavior and their reactive

tendencies to information sharing [8]. However, results on the impacts of sentiments on user behavior on social media have been contradictory. For example, though a recent study showed that emotionally charged political messages are tweeted more [55], other studies have alluded to the efficacy of mostly negative valence over positive ones in influencing virality, especially when it comes to news [25]. Some of the reasons alluded to this is the moderating effects of novelty or the newness of the news stories [29, 58].

3. Research Hypotheses

A recent study revealed that positive news are more likely to go viral than negative or neutral ones, even after controlling for novelty or usefulness [10]. This is due to the strong emotions elicited by the positive news and hence the retweeting behavior. For quite some time researchers have argued that rumors and falsehoods are infamously effective in causing disruptions due to their ability to cause reactions from their highly emotional contents [8, 11]. These reactions may be manifested in several ways, including the modification of news item in order to synch with the user's current affective state. Due to character limitations from Twitter, users are known to perform several of the following modifications: shortening tweets through deleting, preserving and adapting tweets for their own purposes and the use of authorship and attribution [14]. The use of these methods leads to changes in the original content but not necessarily the context. Hence, we argue that there exists a positive relationship between sentiment and morphing, similar to the positive relationship between sentiment and virality, due to the strong reactions elicited by the positive sentiment. We therefore hypothesize:

H1a: *An original tweet's sentiment is positively associated with its morphing*.

Emotionally charged messages influence reactivity in receivers as compared to neutral ones [55]. This may be because they influence the affective components in the brain and induce reactions without the user extending their cognitive process. Studies have since tried to show that those affective components trigger a peripheral thought process [3, 34, 44] but not their cognitive process, and this may lead to irrational negative behaviors [34]. A study using electroencephalogram [35] showed that emotional words influence high amount of brain responses as compared to neutral ones. In general, we argue that emotive tweets will cause users to react and change the contents of tweets before sharing in order to synchronize and personalize their own feelings as compared to neutral tweets. Hence, we have:

H1b: *An original tweet with more positive or negative sentiment is positively associated with its morphing.*

Bad news, emotions or events have long held sway over those that were inherently good, as a general principle across a broad range of psychological phenomena [7]. Fake news and correction news can be categorized as good and bad. Although fake news stories have been shown to be more viral and influential in sharing behavior as compared to real news [58], studies comparing the propagation of false news and correction are limited. The novelty of such fake news stories entice users on social networks to take ownership of them in order to increase their social media standing [29]. A study has shown that when a user takes possession of such a tweet they are more likely to engage in authorship attrition and/or the preservation and adaptation of the original message [14]. This adaptation is what leads users to shortening or deleting part of the tweets and adapting them to their own purpose and writing styles. When this happens, the similarities between the original tweet and the retweets will change. In comparison, real news stories lack the elements of novelty seen in fake news stories [24, 58] to warrant such zealous modifications. Nor are they known to cause such reactivity. However, we argue that correction news is very different both in tone and intensity from real news as they rebut falsehoods and usually do so in the strongest possible terms. We posit that the strength of correction messages lies in their strongly worded context and how they counter falsehoods. When arguing against a topic, one is usually expected to imply the topic in question and modify the argument against. While real news does not contain novel information, correction messages which is "counter-fake news" may contain more novel information as to efficiently rebut the argument in question. This means that correction messages may not only stimulate more interest but also has the potential to be more modified more than false news. As a result, we hypothesize:

H2*: An original tweet's veracity is positively associated with its morphing.*

By analyzing news articles in the New York Times, a recent study revealed that positive affections highly influence virality [9]. This may be as a result of people's decision making being geared towards maintaining a sense of positivity as they go about their everyday tasks [22]. As a result, individuals are more likely to maintain and even increase a positive status quo when modifying a positive text. Such modifications may include improving on a positive tweet to include jokes and emoticons which may increase morphing and positivity. We argue that, with each tweet, each user over time may upend the positivity of the previous tweet. Thus, as time goes by the morphing and sentiment increases over time.

On the other hand, the modification of tweets with negative sentiments may not be sustainable as time goes by possibly due to the loss of newness or surprise value. We therefore hypothesize:

H3a: *The positive association between sentiment and morphing gets stronger over time.*

Considering emotionally charged tweets are expected to influence virality more than neutral tweets [55], we expect that group emotional contagion may in fact assist in the transfer of moods and emotions [5]. This means that if there are no emotions or the emotional valence of the tweet is neutral, it may not receive much attention and as such may not be retweeted more. We argue that this behavior could be akin to herd mentality, such that emotive messages causes the sentiments (positive or negative) to be transferred and snowball over time. Just as the original message may convey such emotions, positive and negative emotions will be transferred to the recipient and their modifications would then be a direct reflection of their emotional state. The user's modification of the tweet whether positive or negative can then be easily seen from the modification of the text. And as time goes by and more users receive the tweet, the emotions are transferred to and from and expressed by the modification of the textual contents. Thus over time, reactivity and emotion will influence several modifications. Moreover, we argue that, as time goes by and the novelty in a tweet decays, neutral tweets will quickly lose traction and be modified less. In contrast, an emotive tweet (positive or negative) will more likely withstand the test of time due to the emotion contained in the message and continue to increasingly morph.

H3b: *The positive association between positive or negative sentiments and morphing gets stronger over time.*

As correction news morphs more than fake news due to the desire to confront the "fakeness" of a news article, we argue that it is more likely to also morph more as time goes on. Although studies have shown that fake news in general may diffuse faster in a short amount of time than other news [58], we posit that correction news are more emotive and aggressive in their response in debunking falsehoods. This reaction will give way to a more aggressive morphing behavior as time goes by. Also, considering that falsehoods must first be introduced in the nomology for correction messages to even exist, we argue that the mechanisms underlying correction messages may be playing "catchup" and as such need to increase their morphing behavior over time. We also argue that although falsehoods will initially be introduced and therefore morph faster initially, correction messages will eventually morph faster as time goes by till falsehoods are extinguished. We therefore foresee that over time due to the aggressive stances employed in rebutting falsehoods, correction messages may increase morphing behavior at both the short term and the long run more than falsehoods. We argue that as times goes by morphing may increase more for correction news than false news.

H4*: The positive association between veracity and morphing gets stronger over time.*

4. Sample and Methodology

4.1. Sample

Hurricane Harvey was a Category 4 hurricane that made landfall along the Texas coast on August 25, 2017. This hurricane displaced more than thirty thousand residents and caused over one hundred and ninety billion dollars of damage. Considering its high social and economic costs and the fact that rumors have predominantly been observed during crisis events, Hurricane Harvey is an ideal event that can serve as a natural setting for our study on the morphing of fake news and their correction messages.

We investigate the morphing of tweets by first identifying and collecting all tweets for each day from Hurricane Harvey's formation on August 17 through September 27, 2017 through Twitter. We only retained verifiable false and correction tweets based on the Federal Emergency Management Agency's (FEMA) rumor control page and three fact checking websites including Factcheck.org, Snopes.com, Truth or fiction. We obtained 28 original tweets with 14 fake tweets and 14 correction tweets as a result. Next, we collected all the retweets of these topics for a 5-week period. We obtained a total of 150,907 tweets and retweets for our first-step exploratory analysis on the morphing hazard rates of falsehoods and correction messages.

Next, we leveraged SpaCY and the natural language processing libraries in Python to calculate the sentiments of the tweets as SpaCy provides a fast and accurate syntactic analysis following an approach by [33]. We marked up words in our corpus as corresponding to a part of speech using its meaning and its association with related words in the sentence and calculated the polarity and subjective scores for each sentence. The standardized polarity score is the raw sentiment orientation of the textual content, which ranges from 1 to 99.99 for positive sentiment, 0 for neutral, and -1 to -99.99 for negative sentiment. Since our dependent variable is the change in characters of a tweet, we controlled for word count and used time in hours as an exogenous variable in order to reduce endogeneity. We obtained a total of 133,319 verified tweets and retweets for our second-step empirical analyses on the factors affecting the morphing of falsehoods and correction messages.

4.2. Cosine Distance

An efficient way of measuring the similarities or differences in data and documents with textual contents such as tweets is the use of clustering techniques [12]. Agglomerative clustering is a type of hierarchical clustering method used in data mining that begins at some point and repeatedly combines two or more suitable clusters [12]. Cosine similarity is an agglomerative clustering technique that calculates the similarities or differences between textual contents and has been used intensively in face detection [41] and Web clustering [56]. It is an effective means for cataloging and documenting large corpuses of documents [20, 53]. The cosine similarity between vectors X and Y $Cosine(X; Y) = X*Y/(|X|*|Y|)$, where ||X|| and ||Y|| are the Euclidean norms of X and Y, respectively.

We define morphing as the change of characters in a tweet that does not change the original meaning of a tweet. The cosine distance is a term-based similarity measure and equals 1-cosine similarity. It considers the distance between two documents and is commonly used in natural language processing. It applies to the vector representation of documents, and the cosine distance vectorizes the text by converting them into numerical data [26]. We calculated the cosine distance of the word vectors based on their dissimilarity to measure morphing. This method is also used to divide the data into various groups based on object similarity or differences [12]. It is able to show the distances between corpuses of tweets that are in a multidimensional term vector space which is defined by the cosine of the angles [52]. The cosine distance metrics for the tweets begin when the initial tweet is assigned a numerical value of 0 and then its cosine distance is compared with subsequent tweets and assigned values based on distances between the tweets. The initial value assigned is a comparative between the initial tweet on itself and should show no differences and is assigned a value of zero. This approach measures the differences based on distances with a tweet. The larger the cosine distance, the more different a tweet is from the original tweet.

4.3. Exploratory Survival Analysis

Survival analysis analyzes the occurrence of an event as a failure process starting from a certain point in time and the factors associated with the occurrence of the event [21, 39, 42]. It relies on the expected duration of time until one or more events occur. Survival analysis has been applied in IS to study behavioral patterns such as the diffusion of technologies [4, 38, 51]. Treating the mutation or morphing of a tweet on the same topic with the same veracity as an event, we analyzed the hazard functions for the morphing of both falsehoods and correction messages. Because mutation can occur multiple times for the same original tweet, we treated the mutation of falsehoods and correction messages as independent recurring event where the characters change in an original (first) tweet over time [19].

The Andersen-Gill (AG) model, an extension of the Cox proportional hazards model, is the most frequently used model to examine the occurrence of recurrent events [2]. It relates the intensity function of event recurrences to the covariates multiplicatively and treats each subject as a multi event with independent increments which has a common baseline hazard function for all recurring events. The (AG) is appropriate for our analysis because it assumes that each tweet and retweet is independent and does not rely explicitly on previous events before they occur. The hazard function $\lambda_{ik(t)}$ for the k^{th} event of the i^{th} subject is denoted as $\lambda_{ik}(t) = \lambda_0(t)e^{X_{ik}\beta}$.

We assume that morphing occurs as a result of the message contacts between users of the network per topic. During the diffusion process, a 1 means that there was a change in the original tweet or mutation, while a 0 means the observation was censored and was not observed to morph during the period of the analysis.

We analyzed our tweet data of 150,907 observations and present the Nelson cumulative hazard functions for falsehoods and correction messages using the AG model Figure 1. The Nelson cumulative hazard function for recurring events represents the expected number of events for a unit that has been observed for the given amount of time. The results indicate that although falsehoods had a slightly higher initial morphing rate, correction messages morphed faster than falsehoods after the first 60 hours. This might be as a result of competition between both falsehoods and correction. Lastly, falsehoods also morphed 20 hours longer than correction messages as no events were observed after about 680 hours for correction messages.

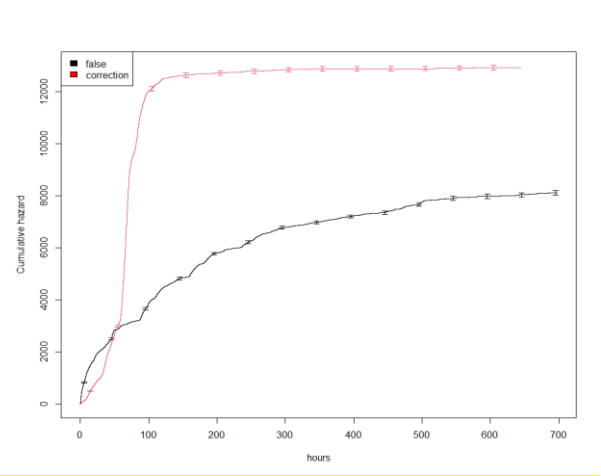

Figure 1. Nelson cumulative hazard functions for false and correction tweets.

4.4. Empirical Analyses

4.4.1. Variable Definition

We summarize our variable definitions in Table 1. In addition to our dependent and independent variables, we also included two control variables including the word count and variation to control for the length of the tweet and the morphing history on the morphing of a tweet at time t. We performed both the Breusch Pagan and the White's test for heteroskedasticity and used the White heteroscedastic-consistent robust estimates. Table 2 summarizes the sample descriptive statistics.

Table 1. Variables and definitions

Variable	Definition
Dependent Variable	
Morphing	A value between 0 and 99.99 that equals 100 times the cosine distance between two tweets.
Independent Variables	
Veracity	1 if the original tweet is a verified true or correction tweet, and 0 if verified false.
Sentiment	The raw score of the sentiment of the tweet from -99.99 to 99.99
Time	The number of hours that had elapsed since the original tweet on the same topic.
Control Variables	
Word Count	The number of words in a tweet.
Variation	The average cosine distance from the second tweet to the last tweet on the same topic with the same veracity.

Table 2. Sample descriptive statistics (N=133,319)

Variable	Mean	Std. Dev.	Min	Max
Morphing	4.873	1.446	0	29.292
Sentiment	-0.468	3.585	-20	18.75
Veracity	0.505	0.500	0	1
Time	72.960	59.895	0	637
Word count	18.767	4.873	1	111
Variation	4.755	0.476	1.079	5.931

4.4.2. Model Specification

Equations 1 and 2 specify our empirical model to examine the morphing an original tweet $X_{i,0}$ to $X_{i,t}$ at time t.

$$\text{Morphing}(X_0; X_t) = \beta_0 + \beta_1 \text{Sentiment}_i + \beta_2 \text{Veracity}_i + \beta_3 * t + \beta_4 \text{Sentiment}_i * t + \beta_5 \text{Veracity}_i * t + \beta_6 \text{WordCount}_i + \beta_7 \text{Variation}_{i,(2,t-1)} + \varepsilon_{i,t}, \text{ and} \quad (1)$$

$$\text{Morphing}(X_0; X_t) = \beta'_0 + \beta'_1 |\text{Sentiment}_i| + \beta'_2 \text{Veracity}_i + \beta'_3 * t + \beta'_4 |\text{Sentiment}_i| * t + \beta'_5 \text{Veracity}_i * t + \beta'_6 \text{WordCount}_i + \beta'_7 \text{Variation}_{i,(2,t-1)} + \varepsilon_{i,t}. \quad (2)$$

Table 3 summarizes the results of our empirical analyses. All our independent variables had variance inflation factors less than 4 with a mean value of 2.17.

Table 3. Results of robust model during Hurricane Harvey (N=133,319)

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1.506*** (0.078)	-0.408*** (0.090)	-0.036 (0.080)	-0.512*** (0.095)	-0.134* (0.081)
Sentiment		0.016*** (0.001)		0.028*** (0.002)	
|Sentiment|			0.007*** (0.001)		-0.009*** (0.002)
Veracity		0.909*** (0.008)	0.887*** (0.008)	0.936*** (0.017)	1.033*** (0.016)
Time	0.003*** (0.0001)	0.004*** (0.0001)	0.004*** (0.0001)	0.004*** (0.0001)	0.004*** (0.0001)
Word Count	0.021*** (0.001)	0.049*** (0.001)	0.047*** (0.001)	0.049*** (0.001)	0.047*** (0.001)
Variation	0.576*** (0.015)	0.766*** (0.017)	0.695*** (0.015)	0.790*** (0.018)	0.714*** (0.015)
Sentiment*time				-0.0002*** (0.00001)	
|Sentiment|*time					0.0002*** (0.00001)
Veracity*time				-0.0005** (0.0002)	-0.002*** (0.0002)
RMSE	1.409	1.341	0.140	1.340	1.341
R-Squared	0.051	0.141	1.341	0.142	0.141

Notes: RMSE: Root mean square error. *p<0.10; ** p<0.05; ***p<0.01.

Our first model depicts the morphing (100*cosine distance) of a tweet as a function of our control variables. This is our baseline model which has time and two control variables: the word count and variation. We find that the intercept and all variables were significant at the 0.01 level.

In Model 2, we added our two of our independent variables: sentiment and veracity. The coefficients for veracity and sentiment were both positive and significant at the 0.01 level. Thus, H1a and H2 are supported.

In Model 3, we added veracity and the absolute value of sentiment to capture the magnitude of sentiment no matter whether it is positive or negative. The coefficients for both veracity and the absolute value of sentiment were all positive and significant at the 0.01 level. Thus, H1b and H2 are supported.

In Models 4 and 5, we added the interaction terms between time and veracity, sentiment, and the absolute value of sentiment. The coefficient estimates for the interaction term between veracity and time were negative and significant in both models. Thus, H4 was not supported. As time goes by, the morphing rate difference between correction messages and fake news decreased. The coefficient for the interaction term between sentiment and time was negative and significant at the .01 level in Model 4. Hence, H3a was not supported. This suggests that as time goes by, the positive impact of sentiment on morphing reduces. In Model 5, after adding the interaction term between the absolute value of sentiment and time, we noticed that the coefficient of |sentiment| became negative and significant and the coefficient of |sentiment|*time was positive and significant. These results showed that even though overall the impact of |sentiment| on morphing was positive (Model 3), the impact was not static over time. Early on, |sentiment| was negatively related to morphing. As time went by, the negative impact started to reduce and became positive after about four days. As a result, H3b was not supported.

5. Discussion

5.1. Theoretical Contributions

This study makes several contributions to literature. First, we provided a visualization of the hazard rates of morphing for both fake news and their correction messages on Twitter using survival analysis. Our results show that correction messages morph more aggressively than falsehoods.

Second, we developed an empirical model for predicting the morphing of messages on Twitter. Despite increasing interest in academia on the diffusion of fake news, to the best of our knowledge, this is the first time a research has been conducted with a high level of granularity on information morphing on social media. We identified factors such as sentiment, the absolute value of sentiment and veracity that may influence the morphing of both false and correction messages. We find that positive sentiment is associated with more morphing, which is consistent with prior research that suggests positive news influences sharing and virality [9]. During extreme events, positive news may retain some novelty and thus may cause individuals to not only share but change the textual contents before sharing. Our findings showed that certain contents that end up evoking a lower form of arousal like sadness ended up being less viral. A recent study lends credence to our findings and showed that positive emotions affects profitability and influences momentum in the financial arena [18].

We also find that in general, tweets that are emotionally charged (positive and negative) have a positive effect on morphing and are more likely to cause reactivity and content changes. This is consistent with the previous literature that showed emotionally charged messages were more likely to be shared than neutral messages [55]. A possible explanation is that emotions in general elicit the social sharing of emotions [50] and those contents may be able to induce cognitive and arousal-related effects which might compel reactivity. It is this reactivity that influences users to want to make a tweet more personal, thereby modifying the tweet to synchronize with their current affect state.

Our results show that correction news morphed more than false news. This result is inconsistent with the previous literature on virality, which showed that false news may diffuse a lot more than news that is inherently not false [58]. The difference in the findings may be because users' attempt to correct news stories with fervor such that they may keep modifying the news stories more than the competing falsehoods during extreme events. Another possibility could be that positive news or correction news may attempt to exaggerate positivity of an event already posited as bad by false news contents to sway users and lift their spirit high. This gives a sense of hope during crisis situations. For example, a recent study [52] showed that rumor resurgence often accompanied changes in textual contents and were mostly in the direction of exaggeration. Finally, users who share positive news during extreme events may want to personalize the message so that it is seen by the receivers as originating from them. That way they would be

perceived as novel disseminators of the news, and it might improve their standing in the network.

Third, we also compare the impacts of the above mentioned factors on morphing over time. Our results show that even though tweets with positive sentiments morph faster overall, the difference in morphing rate slows down over time. It shows that positively charged news are more likely to garner more changes in content initially than both neutral and negative news. As time goes by, the difference between the morphing of positive versus negative or neutral tweets starts to decrease. This can be due to the novelty and excitement generated by the positive tweets in the earlier stages of crisis situations. However at the later stages, the novelty and excitement generated by the positive sentiment may wear off [29] and the morphing slows down. However, after these types of emotion wears off, there is no longer a need to reshare with that much fervor and as such it may evolve slowly.

Furthermore, we show that even though emotionally charged tweets morphs faster than neutral ones overall, the morphing rate is not constant over time. Rather, the morphing of emotionally charged tweets is slower than neutral ones initially but accelerates as time goes by. This result in combination of the positive relationship between sentiment and morphing shows that during a crisis event such as a hurricane, the negativity present in some tweets do not elicit a strong emotional reaction among users and their desire to inject their own opinions or feelings into the modified tweets. However, emotional messages are spontaneously better remembered than neutral words [35]. This means that as time goes by, the strong sentiment in tweets would linger on and over time their impacts increase compared with neutral ones, thus leading to increased morphing.

Our results also reveal that the positive impact of veracity on morphing decreases over time. After initially seeing the correction messages, Twitter users may feel a strong urge to modify the contents and share them to generate public awareness and express their feelings. As time goes by, this urge decreases when the novelty of the news reduces, thus leading to a slowdown in the morphing rate relative to fake news.

5.2. Practical Implications

This study has several practical implications. First, our research provides not only a better understanding of the morphing behavior of false news and correction messages but also provides insights on how sentiments affect morphing on social media. Furthermore, this study can be applied to any communication process and helps us further understand the role veracity plays in the transfer of emotions on social media. Our results show that users change the textual contents of messages aggressively when the contents are emotionally charged. To minimize the aggressive nature of tweets, social media administrators need to tamper the original tweets with more neutral tweets to reduce aggression but not to reduce or dilute the true meanings of the original posts. Furthermore, our results showed that morphing increases with an increase in textual contents, therefore limiting the number of characters in those platforms can help create a safer environment devoid of toxicity.

Second, social media administrators leveraging our research can monitor and control the overflow of negative emotions that has the potential of becoming toxic over time. They can do this by limiting the duration of negative interactions on their platforms such as muting forums after an intense period of engagement. Our study shows that as time goes by, both positive and negative sentiments cause an increase in morphing. This may be used as a proxy for measuring and setting thresholds on the appropriate levels of toxicity that is allowed, and beyond this has the potential of causing disruptions in an otherwise conducive and productive environment if such negative engagements persist.

Third, social media administrators and government agencies can combat the spread of falsehoods by designing and deploying more positively charged correction messages. Considering our results showed that positively charged tweets morph more than falsehoods and correction tweets influence morphing more, government agencies and social platform administrators can design and deploy effective positively charged correction messages that morph more to counter and possibly dampen the spread and morphing of falsehoods on social media. This would ultimately increase the virality of positive news and have a ripple effect in encouraging positive emotions.

Fourth, our findings can help content creators, advertisers and marketing executives strengthen their marketing mix. Positivity when used in advertising may influence more sharing behavior and potentially impact profitability. It would also help users develop a more lasting positive view of the organization as studies have showed that people inherently like to be associated by positivity in their everyday lives [22]. Social media users may adopt this strategy and promote more real and positive news to help serve as a catalyst in spreading positive energy using their online social media presence.

6. Limitations and Future Research

This study has the following limitations. First, we only captured verified false and correction news during a shock event. Future research may also investigate rumors during non-crises situations to cross-validate our results. Second, we only captured the basic dimensions of sentiments, positive, negative and neutral. Future studies can examine other dimensions of sentiments, such as anger and joy, and how each of these affects morphing. Furthermore, future studies may investigate the differences in the morphing of false, real and correction messages.

7. References

[1] Allcott, H., and M. Gentzkow, "Social Media and Fake News in the 2016 Election", *The Journal of Economic Perspectives 31*(2), 2017, pp. 211–235.

[2] Andersen, P.K., and R.D. Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study", *The Annals of Statistics 10*(4), 1982, pp. 1100–1120.

[3] Angst, and Agarwal, "Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion", *MIS Quarterly 33*(2), 2009, pp. 339.

[4] B. Baesens, T. Van Gestel, M. Stepanova, D. Van den Poel, and J. Vanthienen, "Neural Network Survival Analysis for Personal Loan Data", *The Journal of the Operational Research Society 56*(9), 2005.

[5] Barsade, S.G., "The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior", *Administrative Science Quarterly 47*(4), 2002, pp. 644.

[6] Barthel, M., A. Mitchell, and J. Holcomb, "Many americans believe fake news is sowing confusion", *Pew Research Center 15*, 2016. https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

[7] Baumeister, R.F., E. Bratslavsky, C. Finkenauer, and K.D. Vohs, "Bad is Stronger than Good", *Review of General Psychology 5*(4), 2001, pp. 323–370.

[8] Bene, M., "Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014", *Information, Communication & Society 20*(4), 2017, pp. 513–529.

[9] Berger, J., and K. Milkman, "Social transmission, emotion, and the virality of online content", 2010.

[10] Berger, J., and K.L. Milkman, "What Makes Online Content Viral?", *Journal of Marketing Research 49*(2), 2012, pp. 192–205.

[11] Berger, J., and K.L. Milkman, "Emotion and Virality: What Makes Online Content Go Viral?", *GfK Marketing Intelligence Review 5*(1), 2013, pp. 18–23.

[12] Berkhin, P., "Survey of Clustering Data Mining Techniques", In *Grouping multidimensional data*. 2006, 56.

[13] Bollen, J., H. Mao, and A. Pepe, "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-economic Phenomena", *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, AAAI Press (2011), 4.

[14] Boyd, D., S. Golder, and G. Lotan, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter", *2010 43rd Hawaii International Conference on System Sciences*, IEEE (2010), 1–10.

[15] Cerf, V.G., "Information and misinformation on the internet", *Communications of the ACM 60*(1), 2016, pp. 9–9.

[16] Cheng, J.-J., Y. Liu, B. Shen, and W.-G. Yuan, "An epidemic model of rumor diffusion in online social networks", *EDP Sciences, Societ`a Italiana di Fisica, Springer-Verlag 2013*, 2013.

[17] Chua, A.Y.K., C.-Y. Tee, A. Pang, and E.-P. Lim, "The Retransmission of Rumor and Rumor Correction Messages on Twitter", *American Behavioral Scientist 61*(7), 2017, pp. 707–723.

[18] Constantinos Antoniou, John A. Doukas, and Avanidhar Subrahmanyam, "Cognitive Dissonance, Sentiment, and Momentum", *The Journal of Financial and Quantitative Analysis 48*(1), 2013, pp. 245–275.

[19] Cox, D.R., "Regression Models and Life-Tables", *Wiley for the Royal Statistical Society 34*(2), 1972, pp. 187–220.

[20] Cutting, D.R., D.R. Karger, J.O. Pedersen, and J.W. Tukey, "Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections", *15th Ann Int'l SIGIR '92*, Association of Computing Machinery (1992), 1–12.

[21] David A. Freedman, "Survival Analysis: A Primer", *The American Statistician 62*(2), 2008, pp. 110–119.

[22] Di Muro, F., and K.B. Murray, "An Arousal Regulation Explanation of Mood Effects on Consumer Choice", *Journal of Consumer Research 39*(3), 2012, pp. 574–584.

[23] DiFonzo, N., and P. Bordia, "Rumor, Gossip and Urban Legends", *Diogenes 54*(1), 2007, pp. 19–35.

[24] Friggeri, A., L. Adamic, D. Eckles, and J. Cheng, "Rumor Cascades", *Eighth International AAAI Conference on Weblogs and Social Media*, AAAI Publications (2014), 101–110.

[25] Hansen, L.K., A. Arvidsson, F.A. Nielsen, E. Colleoni, and M. Etter, "Good Friends, Bad News - Affect and Virality in Twitter", In J.J. Park, L.T. Yang and C. Lee, eds., *Future Information Technology*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, 34–43.

[26] H.Gomaa, W., and A. A. Fahmy, "A Survey of Text Similarity Approaches", *International Journal of Computer Applications 68*(13), 2013, pp. 13–18.

[27] Huang, H., "A War of (Mis)Information: The Political Effects of Rumors and Rumor Rebuttals in an Authoritarian Country", *British Journal of Political Science 47*(02), 2017, pp. 283–311.

[28] Indiana University, A. Kim, A.R. Dennis, and Indiana University, "Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media", *MIS Quarterly 43*(3), 2019, pp. 1025–1039.

[29] Itti, L., and P. Baldi, "Bayesian surprise attracts human attention", *Vision Research 49*(10), 2009, pp. 1295–1306.

[30] Jindal, N., B. Liu, and E.-P. Lim, "Finding unusual review patterns using unexpected rules", *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*, ACM Press (2010), 1549.

[31] Kim, A., and A.R. Dennis, "Says Who?: How News Presentation Format Influences Perceived Believability and the Engagement Level of Social Media Users", *SSRN Electronic Journal*, 2017.
[32] Kim, A., P.L. Moravec, and A.R. Dennis, "Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings", *Journal of Management Information Systems 36*(3), 2019, pp. 931–968.
[33] King, K., "The Gray Side of Fake News: A Multiclass Approach to Detecting Fake News, Real News and Everything Else in Between", *The 26th Americas Conference on Information Systems*, AIS Electronic Library (2020), 10.
[34] King, K.K., and J. Sun, "Investigating User Disclosure of sensitive information: An ELM theory", *24th Americas Conference on Information Systems*, AIS Electronic Library (2018), 12.
[35] Kissler, J., C. Herbert, P. Peyk, and M. Junghofer, "Buzzwords: Early Cortical Responses to Emotional Words During Reading", *Psychological Science 18*(6), 2007, pp. 475–480.
[36] Liang, N. (Peter), D.P. Biros, and A. Luse, "An Empirical Validation of Malicious Insider Characteristics", *Journal of Management Information Systems 33*(2), 2016, pp. 361–392.
[37] Marwick, A.E., and danah boyd, "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience", *New Media & Society 13*(1), 2011, pp. 114–133.
[38] Massimo G. Colombo, and Rocco Mosconi, "A survival model for the study of the diffusion of multiple technologies.", *Giornale degli Economisti e Annali di Economia, Nuova Serie, Anno 54*(7/9), .
[39] Michael G. Akritas, "Nonparametric Survival Analysis", *Statistical Science 19*(4), 2004, pp. 615–623.
[40] Moravec, P., R. Minas, and A.R. Dennis, "Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense at All", *SSRN Electronic Journal*, 2018.
[41] Nguyen, H.V., and L. Bai, "Cosine Similarity Metric Learning for Face Veri cation", *Asian Conference on Computer Vision-ACC 2010*, Springer, Berlin, Heidelberg (2010), 709–720.
[42] Oakes, D., "Survival Analysis", *Journal of the American Statistical Association 95*(449), 2000, pp. 282.
[43] Oh, O., K.H. Kwon, and H.R. Rao, "An exploration of social media in extreme events: Rumor theory and twitter during the Haiti earthquake 2010.", *ICIS 2010 Proceedings 231*, AIS Electronic Library (2010), 1–13.
[44] Osatuyi, B., and J. Hughes, "A Tale of Two Internet News Platforms-Real vs. Fake: An Elaboration Likelihood Model Perspective", *Proceedings of the 51st Hawaii International Conference on System Sciences*, (2018), 3986–3994.
[45] Ozturk, P., H. Li, and Y. Sakamoto, "Combating Rumor Spread on Social Media: The Effectiveness of Refutation and Warning", *2015 48th Hawaii International Conference on System Sciences*, IEEE (2015), 2406–2414.
[46] Pennycook, G., A. Bear, E.T. Collins, and D.G. Rand, "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings", *Management Science*, 2020, pp. mnsc.2019.3478.
[47] Pennycook, G., and D.G. Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning", *Cognition 188*, 2019, pp. 39–50.
[48] Pennycook, G., and D.G. Rand, "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking", *Journal of Personality 88*(2), 2020, pp. 185–200.
[49] Pérez-Rosas, V., B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News", *arXiv:1708.07104 [cs]*, 2017.
[50] Rimé, B., "Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review", *Emotion Review 1*(1), 2009, pp. 60–85.
[51] Roberts, G., O., and L. Sangalli M., "Latent diffusion models for survival analysis", *Bernoulli 16*(2), 2010, pp. 435–458.
[52] Shin, J., L. Jian, K. Driscoll, and F. Bar, "The diffusion of misinformation on social media: Temporal pattern, message, and source", *Computers in Human Behavior 83*, 2018, pp. 278–287.
[53] Steinbach, M., G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques", *6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, SIGKDD (2000), 1–20.
[54] Stieglitz, S., and L. Dang-Xuan, "Impact of diffusion of sentiment in public communication on Facebook.", *ECIS 2012 Proceedings. Paper 98*, Association for Information Systems (2012), 14.
[55] Stieglitz, S., and L. Dang-Xuan, "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior", *Journal of Management Information Systems 29*(4), 2013, pp. 217–248.
[56] Strehl, A., J. Ghosh, and R. Mooney, "Impact of Similarity Measures on Web-Page Clustering", 2000, pp. 7.
[57] Treadway, M., and M. McCloskey, "Cite Unseen: Distortions of the Allport and Postman Rumor Study in the Eyewitness Testimony Literature", *Law and Human Behavior 11*(1), 1987, pp. 19–25.
[58] Vosoughi, S., D. Roy, and S. Aral, "The spread of true and false news online", *Science 359*(6380), 2018, pp. 1146–1151.