

A study of tweet veracity to separate rumours from counter-rumours

Alton Y.K. Chua and Banerjee Snehasish

Abstract

Rumors are known to propagate easily through computer-mediated communication channels such as Twitter. Their outbreak is often followed by the spread of counter-rumors, which are messages that debunk rumors. The probability of a tweet to be a counter-rumor is referred as tweet veracity in this paper. Since both rumors and counter-rumors are expected to contain claims, the two might not be easily distinguishable. If Internet users fail to separate rumors from counter-rumors, the latter will not serve its purpose. Hence, this paper investigates the extent to which tweet veracity could be predicted by content as well as contributors' profile. The investigation focuses on the death hoax case of Singapore's first Prime Minister Lee Kuan Yew on Twitter. A total of 1,000 tweets (500 rumors + 500 counter-rumors) are analyzed using binomial logistic regression. Results indicate that tweet veracity could be predicted by clarity, proper nouns, visual cues, references to credible sources as well as contributors' duration of membership, and number of followers. The significance of these findings are highlighted.

Keywords

Counter-rumor; information quality; information veracity; online rumor; Twitter; user-generated content.

Introduction

Rumors are unverified yet instrumentally relevant messages that propagate in situations of ambiguity, danger, or potential threat. Despite being unverified, they help people make sense of uncertain situations, and manage risks [2], [10].

Rumors often spread on the Internet through computer-mediated communication (CMC) channels such as Twitter, Facebook or WhatsApp. In fact, CMC enables rumors to propagate faster and geographically farther than ever before [12]. As rumors ripple through the Internet, they become menacing to the online communities by fueling misunderstanding and panic.

The outbreak of rumors is often followed by the spread of counter-rumors. For the purpose of this paper, counter-rumors are defined as messages that debunk rumors. They are generally created by authoritative sources to correct the misunderstanding caused by rumors. They seek to lower the perceived accuracy of rumors, thereby making people less likely to be misinformed [26].

Scholars hitherto have generally investigated rumors and counter-rumors in isolation of each other. For example, the literature abounds with works on how rumors spread via CMC channels such as Facebook [11], [22]. Some prior works have also shed light on the impact of counter-rumors on users of CMC channels such as Twitter [26].

However, relatively little is known about how rumors differ from counter-rumors. Given that both rumors and counter-rumors are expected to contain claims, the two might not be easily distinguishable. If Internet users are unable to pick up any differences between rumors and counter-rumors, the primary purpose of the latter is defeated.

Hence, the objective of this paper is to investigate differences between rumors and counter-rumors on Twitter. The probability of a tweet to be a counter-rumor is referred as tweet veracity. Informed by prior works [10], [25], [29], this paper specifically hypothesizes that tweet veracity could be predicted in terms of intra-textual content, para-textual content as well as contributor profile. While intra-textual content includes properties that are derivable from the use of words within (intra) the text, para-textual content entails properties that are identifiable from around (para) the words used in the text [32]. Contributor profile encompasses properties that summarize tweet contributors' standing in the community [24]. To test the hypotheses, this paper studies the death hoax case of Singapore's first Prime Minister Lee Kuan Yew on Twitter.

The rest of the paper proceeds as follows. The next section reviews the literature. This is followed by a description of data collection, coding and analysis. The results are presented next, followed by the discussion. The paper concludes with remarks on its significance, limitations, and directions for future research.

Literature Review

Related works have shown that in situations that are marred with ambiguity and uncertainty, Internet users commonly engage in speculation with one another via CMC channels [25]. Users' collective attempts to make sense of uncertain situations set the stage for the outbreak of rumors on the Internet [2], [10].

An emerging stream of works in the rumor literature suggests that believability of rumors is possible to be lowered by exposing users to counter-rumors. In particular, counter-rumors refer to messages that debunk rumors [26], [29].

However, the extant literature sheds little light on the phrasing of counter-rumors. This paper argues that a counter-rumor could debunk a given rumor in at least two different ways. One, it could state the rumor, and follow it up with a refuting statement. Two, the counter-rumor could simply put forth its claim without explicitly specifying the rumor that it seeks to debunk.

The latter could be more viable on a CMC channel such as Twitter that imposes a length restriction of 140 characters. Given the little room to maneuver, a counter-rumor tweet might not always have the luxury to state the rumor explicitly before using a refuting statement. In that case, both the rumor and the counter-rumor would appear as tweets that contain claims. Such claims are not always easy to verify in the absence of official information [13]. As a result, users might mistake a counter-rumor tweet as a rumor tweet, especially if they come across the former before the latter.

This conundrum highlights the need to understand the factors that could be correlated to tweet veracity. In this paper, tweet veracity is defined as the probability of a tweet to be a counter-rumor. Relatively little scholarly efforts have been trained hitherto to identify the antecedents of tweet veracity. To extend the literature, this paper proposes a conceptual framework, which is shown in Figure 1, to predict tweet veracity. Informed by related works, the framework hypothesizes that tweet veracity could be predicted based on intra-textual content (e.g., [9], [10], [34]), para-textual content (e.g., [7], [30], [34]), and contributor profile (e.g., [5], [17], [29]). The hypotheses are summarized in Table 1.

Figure 1. The proposed conceptual framework to distinguish between rumors and counter-rumors.

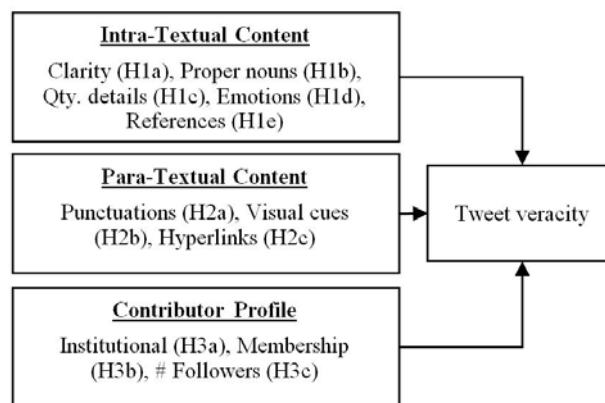


Table 1. Summary of the hypotheses to be tested

	Hypotheses
Intra-Textual Content	H1a: Clarity is positively correlated to tweet veracity.
	H1b: The presence of proper nouns is positively correlated to tweet veracity.
	H1c: The presence of quantitative details is positively correlated to tweet veracity.
	H1d: The presence of emotions is negatively correlated to tweet veracity.
	H1e: The presence of references to credible sources is positively correlated to tweet veracity.
Para-Textual Content	H2a: The presence of punctuations is positively correlated to tweet veracity.
	H2b: The presence of visual cues is positively correlated to tweet veracity.
	H2c: The presence of hyperlinks is positively correlated to tweet veracity.
Contributor Profile	H3a: The presence of institutional contributors is positively correlated to tweet veracity.
	H3b: Contributors' membership duration is positively correlated to tweet veracity.
	H3c: Contributors' follower count is positively correlated to tweet veracity.

Intra-Textual Content

Intra-textual content includes properties that are derivable from the use of words within (intra) tweets [32]. There are at least five such properties that could help distinguish between rumors and counter-rumors on Twitter. These are clarity, proper nouns, quantitative details, emotions, and references to credible sources.

Clarity is a measure of the interpretive definiteness or dubiousness expressed in a message [25]. Lack of clarity has long known to catalyze the spread of rumors [18], [33]. Rumors cause panic mainly due to their dubious nature [10]. In contrast, counter-rumors are expected to debunk rumors explicitly. Clearly-articulated counter-rumors are likely to inspire confidence, making them effective in debunking rumors [29]. Hence, the following is hypothesized:

H1a: Clarity is positively correlated to tweet veracity.

Noun is one of the most content-rich parts of speech in English language [31]. In particular, the use of proper nouns in a message enhances its definiteness. Conversely, the lack of proper nouns in a message often contributes to its dubiousness. Given that rumors are less definitive than counter-rumors, the former could contain fewer proper nouns [10]. Hence, the following is hypothesized:

H1b: The presence of proper nouns is positively correlated to tweet veracity.

Quantitative details include numerals (e.g., 2, 100), numbers (e.g., second, hundred), or quantifiers (e.g., few, much). Quantitative details are known to enhance the level of specificity in a message. Hence, they are generally inconspicuous in rumors [6], [9]. In contrast, counter-rumors could contain quantitative details in order to affirm their veracity. Prior works have shown that the presence of numbers such as date and percentage is a crucial indicator of information veracity [34]. Hence, the following is hypothesized:

H1c: The presence of quantitative details is positively correlated to tweet veracity.

Rumors generally highlight either positive or negative outcomes. These are referred as wish rumors, and dread rumors respectively [20]. In addition, rumors often contain words that express anxiety [25]. The emotive nature of rumors mainly contributes to making them sensational in nature. However, counter-rumors are expected to debunk rumors with facts by maintaining an objective tone. They may not necessarily be emotionally charged. Hence, the following is hypothesized:

H1d: The presence of emotions is negatively correlated to tweet veracity.

Messages that make references to sources, especially credible ones such as the mainstream media, are generally perceived as being compelling and persuasive [2]. Rumors are rarely corroborated with references to credible sources [25]. However, the veracity of counter-rumors could be enhanced by citing such sources. In fact,

counter-rumors could be mostly issued by authoritative sources to correct the misunderstanding caused by rumors. Hence, the following is hypothesized:

H1e: The presence of references to credible sources is positively correlated to tweet veracity.

Para-Textual Content

Para-textual content entails properties that are identifiable from around (para) the words used in tweets [32]. There are at least three such properties that could help distinguish between rumors and counter-rumors on Twitter. These are punctuations, visual cues, hyperlinks.

Punctuations are para-textual expressions that serve to attract readers' attention. Specifically, in online settings, punctuations in the form of emoticons and ellipses are often used strategically to make messages stand out [15]. Rumors that arise out of people's collective sense-making efforts may not be written in order to stand out. They could be inherently eye-catching due to their emotive and sensational nature. However, counter-rumors are expected to be deliberately written in ways so that they can grab the eyeballs. After all, if they do not become the cynosure, they would not serve their purpose of debunking rumors. Conceivably, conspicuous messages are critical to raise public awareness [16]. Hence, the following is hypothesized:

H2a: The presence of punctuations is positively correlated to tweet veracity.

The use of visual cues such as pictures in messages serves two purposes. One, visual cues help make messages conspicuous. Two, visual cues offer confidence about the veracity of messages [7], [30], [34]. Even though rumors could occasionally contain visual cues, serving these two purposes is not too imperative. In contrast, serving both of these purposes is crucial for the success of counter-rumors. Hence, the following is hypothesized:

H2b: The presence of visual cues is positively correlated to tweet veracity.

Hyperlinks in messages serve as avenues for its readers to acquire additional information about the entries with the click of a mouse. They could be especially useful on a CMC channel such as Twitter that does not offer the luxury of contributing lengthy messages [34]. Since rumors are typically characterized by lack of comprehensive information, they might not use hyperlinks liberally. On the other hand, counter-rumors might need to use hyperlinks in order to justify their debunking claims. Hence, the following is hypothesized:

H2c: The presence of hyperlinks is positively correlated to tweet veracity.

Contributor Profile

Contributor profile encompasses properties that summarize tweet contributors' standing in the community [24]. There are at least three such properties that could

help distinguish between rumors and counter-rumors. These are institutional account, membership duration, and number of followers.

In the context of information seeking, institutional authority refers to the influence exerted by cultural, political or social institutions [8], [28]. When a message is contributed by an institutional authority, it inspires greater confidence about its veracity compared with an entry submitted by any ordinary individual. Unlike rumors, counter-rumors tend to be created by institutional authorities [29]. Hence, the following is hypothesized:

H3a: The presence of institutional contributors is positively correlated to tweet veracity.

Membership duration refers to the longevity of a user in an online community. The longer a user belongs to a community, the greater will be the sense of oneness and commitment [5]. This in turn may make users cautious while sharing information. On the one hand, they would not want to share rumors. On the other hand, they would want to help others by sharing counter-rumors. Hence, the following is hypothesized:

H3b: Contributors' membership duration is positively correlated to tweet veracity.

Number of friends is known to affect the ways in which a user acts in online settings [17]. Specifically, the higher the number of friends, the greater will be the user's sense of altruism for the community [5]. On Twitter, a contributor with a large network of followers may be more motivated to act responsibly than one who has few followers. The motivation to act responsibly in turn could make users reluctant to share rumors yet willing to share counter-rumors. Hence, the following is hypothesized:

H3c: Contributors' follower count is positively correlated to tweet veracity.

Methodology

Data Collection

As indicated earlier, this paper studies the death hoax case of Singapore's first Prime Minister Lee Kuan Yew on Twitter. Lee passed away on 23 March 2015. However, on 18 March, a rumor reporting his death rippled through the CMC channel Twitter. To debunk the rumor, several news organizations posted counter-rumors [3], which were also shared on Twitter.

Tweets related to this death hoax posted on 18 March 2015 form the data for this paper. They were retrieved using the hashtags #LeeKuanYew and #LKY. The data collection process yielded 5,885 tweets altogether.

Data Coding

Three research assistants were recruited for data coding. They were Singaporeans, graduate students of Information Systems in a large public university in Singapore, conversant with the use of Twitter, and familiar with the death hoax case under investigation.

The coders conferred to randomly select tweets from the pool of 5,885 entries, and jointly label them as one of the following: rumors (0), counter-rumors (1), or non-rumors (-1). Tweets that asserted the death of Lee Kuan Yew were labeled as rumors while those that confirmed that he was alive were labeled as counter-rumors. All other tweets were labeled as non-rumors. Disagreements among the coders were resolved through discussion. This procedure of random selection followed by joint labeling of tweets continued until unanimously agreed 500 rumors, and 500 counter-rumors were identified. These 1,000 tweets were admitted for further coding and analysis.

To code these tweets based on the variables in the proposed conceptual framework, the coders followed a two-step approach. In the first step, they jointly coded 400 randomly-selected tweets (200 rumors + 200 counter-rumors) from the pool of 1,000 entries. This allowed establishing inter-coder agreement that was measured in terms of Cohen's Kappa (κ). A value above 0.60 indicated substantial non-chance agreement [21]. In the second step, the remaining 600 tweets were distributed uniformly among the coders to be coded independently. Such a two-step coding procedure is informed by the literature [19], [21], [25].

In terms of intra-textual content, tweets were measured for clarity, proper nouns, quantitative details, emotions, and references to credible sources. With respect to clarity, tweets were coded as 1 if they were unambiguous; 0 otherwise (Cohen's $\kappa = 0.81$). With respect to proper nouns, tweets were coded as 1 if they contained names of persons (apart from Lee Kuan Yew), institutions or places; 0 otherwise (Cohen's $\kappa = 0.94$). With respect to quantitative details, tweets were coded as 1 if they contained numerals, numbers or quantifiers; 0 otherwise (Cohen's $\kappa = 0.96$). With respect to emotions, tweets were coded as 1 if they contained emotive words; 0 otherwise (Cohen's $\kappa = 0.71$). With respect to references to credible sources, tweets were coded as 1 if they contained names of news organizations; 0 otherwise (Cohen's $\kappa = 0.82$).

In terms of para-textual content, tweets were measured for punctuations, visual cues, and hyperlinks. With respect to punctuations, tweets were coded as 1 if they used emoticons, ellipses, repeated question marks or exclamation marks; 0 otherwise (Cohen's $\kappa = 0.72$). With respect to visual cues, tweets were coded as 1 if they contained multimedia content such as pictures, animations or videos; 0 otherwise (Cohen's $\kappa = 0.76$). With respect to hyperlinks, tweets were coded as 1 if they contained links to websites; 0 otherwise (Cohen's $\kappa = 1.00$).

In terms of contributor profile, tweets were measured for institutional account, membership duration, and number of followers. With respect to institutional account,

tweets were coded as 1 if they originated from institutional sources; 0 otherwise (Cohen's $\kappa = 1.00$). Contributors' membership duration on Twitter was calculated in years, while the number of followers was readily available from the dataset.

Data Analysis

Binary logistic regression was used for analysis. This is because the dependent variable tweet veracity was dichotomous—coded as 1 for counter-rumors, and 0 for rumors. The number of tweets, and the number of retweets made by tweet contributors were added as control variables.

There were 11 independent variables. These included the five measures corresponding to tweets' intra-textual content, the three measures corresponding to tweets' para-textual content, and the three measures corresponding to contributor profile. Only two of these independent variables, namely, membership duration, and number of followers, were measured on a continuous scale. The remaining nine independent variables were dichotomous in nature. Due to the presence of dichotomous independent variables, other statistical procedures such as discriminant analysis—also suited to deal with dichotomous dependent variables but cannot handle dichotomous independent variables—was not employed in this paper [27].

Prior to the analysis, the dataset was checked for the problem of multicollinearity. The values of variance inflation factor for all the control variables, and the independent variables were less than the recommended threshold 10. This confirmed that the dataset was free from multicollinearity. The highest variance inflation factor value was 2.44 corresponding to visual cues.

Results

As indicated earlier, the dataset admitted for analysis comprised a corpus of 500 rumors, and a corpus of 500 counter-rumors. Table 2 presents the descriptive statistics of the dataset corresponding to the dichotomous variables. Rumors appeared relatively richer in intra-textual content properties such as emotions, and para-textual content properties such as punctuations compared with counter-rumors. Conversely, counter-rumors seemed richer in intra-textual content properties such as proper nouns, and para-textual content properties such as visual cues.

Table 2. Descriptive statistics for the dichotomous variables

Variables	Corpus (N = 500 each)	Coded as 1 (%)
Clarity	Rumors	4.60
	Counter-rumors	89.40
Proper nouns	Rumors	62.90
	Counter-rumors	98.10
Qty. details	Rumors	7.50
	Counter-rumors	4.80
Emotions	Rumors	62.90
	Counter-rumors	15.10
References	Rumors	2.90
	Counter-rumors	60.90
Punctuations	Rumors	13.50
	Counter-rumors	3.70
Visual Cues	Rumors	3.50
	Counter-rumors	33.10
Hyperlinks	Rumors	28.20
	Counter-rumors	38.70
Institutional	Rumors	4.80
	Counter-rumors	1.40

With respect to contributor profile, users who posted rumors had relatively shorter membership duration in years ($M = 4.26$, $SD = 1.66$) compared with that of users who posted counter-rumors ($M = 4.88$, $SD = 1.67$). With respect to number of followers, users who posted rumors ($M = 724.10$, $SD = 2856.13$) had fewer followers on Twitter compared with those who submitted counter-rumors ($M = 28766.80$, $SD = 36347.02$).

Table 3 presents the odds ratios—denoted by $\text{Exp}(\beta)$ —for the 11 independent variables obtained from the binomial logistic regression analysis. Variables that were found to be statistically significant ($p < 0.05$) are italicized and marked with asterisks based on their levels of the significance value.

The Omnibus test confirmed a significant model performance ($\chi^2 = 895.09$, $p < 0.001$, $-2 \log \text{likelihood} = 421.88$, prediction accuracy = 92.10%, pseudo- $R^2 = 61.00\%$). The binomial logistic regression model accounted for 61% of variance in the dependent variable tweet veracity. However, the control variables—number of tweets, and number of retweets—exhibited non-significant relations with the dependent variable.

Table 3. Results of the binomial logistic regression analysis

	Variables	Exp(β)	Sig. (p)
Intra-Textual Content	<i>Clarity</i>	69.58***	< 0.001
	<i>Proper nouns</i>	4.70***	< 0.001
	Qty. details	0.87	> 0.05
	<i>Emotions</i>	0.45**	< 0.01
	<i>References</i>	6.99***	< 0.001
Para-Textual Content	Punctuations	1.21	> 0.05
	<i>Visual cues</i>	3.02**	< 0.01
	Hyperlinks	0.70	> 0.05
Contributor Profile	Institutional	0.61	> 0.05
	<i>Membership</i>	1.15*	< 0.05
	<i># Followers</i>	1.19*	< 0.05

In terms of tweets' intra-textual content, significant relations were identified for four properties. These included clarity, proper nouns, emotions, and references to credible sources.

Clarity in tweets was positively correlated to tweet veracity (Exp(β) = 69.58). Clear tweets were more likely to be counter-rumors compared with unclear ones, thus lending support to H1a.

The presence of proper nouns was positively correlated to tweet veracity (Exp(β) = 4.70). Tweets with proper nouns were more likely to be counter-rumors compared with those containing no proper nouns, thus lending support to H1b.

The presence of emotions was negatively correlated to tweet veracity (Exp(β) = 0.45). Emotive tweets were less likely to be counter-rumors compared with non-emotive ones, thus lending support to H1d.

The presence of references to credible sources was positively correlated to tweet veracity (Exp(β) = 6.99). Tweets with such references were more likely to be counter-rumors compared with those that did not refer to credible sources, thus lending support to H1e.

The presence of quantitative details in tweets was however not significantly correlated to the dependent variable. Hence, the hypothesis H1c was rejected.

In terms of tweets' para-textual content, significant relation was identified for only one property, namely, visual cues. Specifically, the presence of visual cues was positively correlated to tweet veracity (Exp(β) = 3.02). Tweets rich in visual cues were more likely to be counter-rumors compared with those that were merely textual, thus lending support to H2b.

The use of punctuations, and the presence of hyperlinks in tweets were however not significantly correlated to the dependent variable. For these reasons, the hypotheses H2a and H2c could not be supported.

In terms of contributor profile, significant relations were found for two properties. These included duration of membership, and number of followers.

Contributors' duration of membership was positively correlated to tweet veracity ($\text{Exp}(\beta) = 1.15$). A contributor with a long membership duration on Twitter was more likely to post counter-rumors compared with one who had been in the community for a short period of time, thus lending support to H3b.

Number of followers was positively correlated to tweet veracity ($\text{Exp}(\beta) = 1.19$). A contributor with a large number of followers on Twitter was more likely to post counter-rumors compared with one who was seldom followed, thus lending support to H3c.

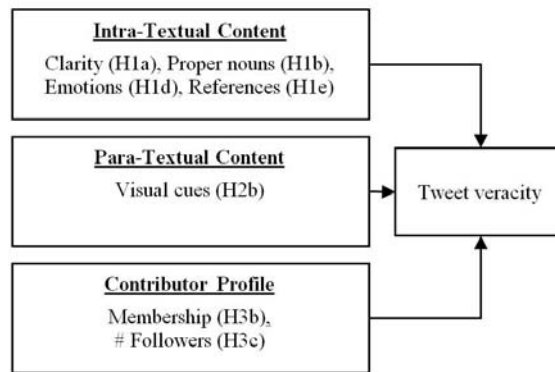
Institutional nature of Twitter accounts was however not significantly correlated to the dependent variable. Hence, the hypothesis H3a was rejected.

Table 4 summarizes the results of the hypotheses testing. The proposed conceptual framework as revised after empirical examination is depicted in Figure 2.

Table 4. Results of the hypotheses testing

	Hypotheses	Results
Intra-Textual Content	H1a: Clarity is positively correlated to tweet veracity.	Supported
	H1b: The presence of proper nouns is positively correlated to tweet veracity.	Supported
	H1c: The presence of quantitative details is positively correlated to tweet veracity.	Rejected
	H1d: The presence of emotions is negatively correlated to tweet veracity.	Supported
	H1e: The presence of references to credible sources is positively correlated to tweet veracity.	Supported
Para-Textual Content	H2a: The presence of punctuations is positively correlated to tweet veracity.	Rejected
	H2b: The presence of visual cues is positively correlated to tweet veracity.	Supported
	H2c: The presence of hyperlinks is positively correlated to tweet veracity.	Rejected
Contributor Profile	H3a: The presence of institutional contributors is positively correlated to tweet veracity.	Rejected
	H3b: Contributors' membership duration is positively correlated to tweet veracity.	Supported
	H3c: Contributors' follower count is positively correlated to tweet veracity.	Supported

Figure 2. The conceptual framework as revised after empirical examination.



Discussion

Prior works have shown that rumors on the Internet can be distinguished from non-rumors [25]. Other works have demonstrated the possibility to separate true rumors from false ones on the Internet [34]. To dovetail the rumor literature, this paper found that online rumors are distinguishable from counter-rumors, which refer to messages that debunk rumors. Specifically, in the context of Twitter, intra-textual content appeared more helpful to distinguish between rumors and counter-rumors—based on prediction strength as indicated by the odds ratios in Table 3—compared with either para-textual content or contributor profile. Nonetheless, intra-textual content, para-textual content and contributor profile cumulatively accounted for more than 60% of variance in tweet veracity.

Overall, three major findings could be gleaned from the results. One, rumors were more emotive than counter-rumors. The presence of emotions was negatively correlated to tweet veracity ($\text{Exp}(\beta) = 0.45$). Prior works provide evidence that rumors could be rich in emotion words in order to appear sensational [20], [25]. This paper confirms the emotive nature of rumors. For example, an emotive rumor tweet stated, “So sad at the passing of my idol #LeeKuanYew.” The lack of emotional expressions in counter-rumors offers a possible reason why they tend to be less sensational in nature. An example of such a counter-rumor tweet is as follows: “#LeeKuanYew #StillAlive.”

Two, counter-rumors were more compelling compared with rumors insofar as veracity is concerned. In particular, traits such as clarity as well as use of proper nouns, visual cues and pictures are known to be crucial predictors of information veracity [2], [9], [10], [34]. Unlike rumors, counter-rumors were not only clearer ($\text{Exp}(\beta) = 69.58$), but also richer in proper nouns such as “Lee Hsien Loong [Lee Kuan Yew’s son]” ($\text{Exp}(\beta) = 4.70$), visual cues such as pictures ($\text{Exp}(\beta) = 3.02$), and references to credible sources such as “CNN” ($\text{Exp}(\beta) = 6.99$). Hence, this new finding extends the literature by suggesting that counter-rumors on CMC channels such as Twitter tend to be informative, factual, and are often corroborated with a wide range of evidences. Hence, they could be compelling and newsworthy enough to inspire confidence among the online community.

In contrast, rumors were generally ambiguous. For example, an ambiguous rumor tweet stated, “He could be having his last few breath as I am writing this . . .” The inherent speculative nature of rumors prevented them from being triangulated with evidences in the form of proper nouns, visual cues, or references to credible sources. This is why rumors generally appear more spurious and conjured vis-à-vis counter-rumors [25].

Three, users with a long duration of membership, and a large number of followers were likely to post counter-rumors. As shown in Table 3, both duration of membership ($\text{Exp}(\beta) = 1.15$) and number of followers ($\text{Exp}(\beta) = 1.19$) were positively correlated to tweet veracity. This could be explained in light of the literature on online altruism and commitment [5], [17], [23]. With longer membership durations and larger networks on Twitter, users perhaps felt a greater sense of oneness with the community, thereby resulting in feelings of altruism and commitment. This could be a reason why such users were reluctant to share rumors yet willing to share counter-rumors. It seems Internet users could be made cautious in their information sharing behaviors by piquing their altruistic attitudes.

That said, this finding should be viewed by keeping in mind the idiosyncrasy of the case under investigation—the death hoax case of Lee Kuan Yew, Singapore’s political patriarch. Singapore has been governed by a dominant ruling party since its independence. It ranks lowly in terms of press freedom, and information flow is tightly regulated by the government [4], [14]. Anti-government sentiments are generally dealt with austerity [1]. In such a setting, users with longer membership duration or larger networks were fortunately more willing to share counter-rumors. This trend might not be replicated in cases that are associated with liberal societies. In fact, under low political pressure, long-serving users with large networks might be audacious enough to share anti-government rumors, or even rejoice at the death of political figures.

Conclusion

This paper investigated differences between rumors and counter-rumors, messages that debunk rumors, on Twitter. It proposed a conceptual framework which argued that tweet veracity could be predicted by their intra-textual content, para-textual content as well as contributor profile. The framework was empirically validated by studying the death hoax case of Singapore’s first Prime Minister Lee Kuan Yew on Twitter.

In terms of tweets’ intra-textual content, rumors and counter-rumors were distinguishable based on their clarity, proper nouns, emotions, and references to credible sources. In terms of tweets’ para-textual content, rumors could be separated from counter-rumors based on the presence of visual cues. In terms of contributor profile, it was possible to predict tweet veracity based on contributors’ duration of membership, and number of followers.

The significance of this paper is two-fold. One, it dovetails existing rumor literature by proposing a conceptual framework that allows predicting tweet veracity. The

framework was found to yield generally promising results when empirically validated. Specifically, it could predict whether a tweet was a rumor, or a counter-rumor with an accuracy of above 90%. More than 60% of variance in tweet veracity could be accounted for.

Two, the revised conceptual framework after empirical validation (Figure 2) can be used by software companies to develop classifiers to predict tweet veracity. Efforts can be made to automate the manual coding process as much as possible. The results of such classifiers can be used to recommend counter-rumors to users, and weed out rumors. This could be a modest step toward curbing the spread of online rumors via CMC channels.

A few limitations in this paper need to be acknowledged. For one, it is limited by its choice of data sources. It analyzed data related to a single case. Besides, data were collected only from Twitter. Hence, caution is advocated in generalizing its findings to other cases for data retrieved from other CMC channels such as Facebook. Moreover, as with any research that draws data from CMC channels, this paper failed to uncover differences in users' motivations to post rumors and counter-rumors on Twitter. It derived its findings solely based on the analyzed tweets. The link between contributors' motivation and tweet veracity still remains unknown.

This paper identifies three directions for future research. One, interested scholars could build on the findings of this paper to predict information veracity in a variety of contexts such as celebrity gossip, and disaster hoax. Two, user studies could be conducted to examine the extent to which CMC users' perceptions of rumors differ from those of counter-rumors. Three, information and communication scholars could explore ways to phrase counter-rumors so that they are maximally effective in debunking rumors. This paper hopes that such lines of future inquiry will minimize the likelihood of Internet users from being deceived by rumors in the long run.

Acknowledgements

This work was supported by the Ministry of Education Research Grant AcRF Tier 2 (MOE2014-T2-2-020). The authors would like to thank Eddy Effendi, Wang Dezhou, and Yong Kim Fook for their assistance in data coding and analysis.

References

- [1] Au-Yong, R. 2015. Fresh police report lodged against Amos Yee. The Straits Times (17 April 2015). Retrieved from <http://www.straitstimes.com/singapore/courts-crime/fresh-police-report-lodged-against-amos-yee>
- [2] Bordia, P., DiFonzo, N., Haines, R., and Chaseling, E. 2005. Rumors denials as persuasive messages: Effects of personal relevance, source, and message characteristics. *Journal of Applied Social Psychology*. 35, 6 (Jun. 2005), 1301-1331. DOI=<http://dx.doi.org/10.1111/j.1559-1816.2005.tb02172.x>.
- [3] Chan, F., and Sim, W. 2015. Police looking into hoax website that falsely announced death of Mr Lee Kuan Yew. The Straits Times (19 March 2015).

- Retrieved from <http://www.straitstimes.com/singapore/police-looking-into-hoax-website-that-falsely-announced-death-of-mr-lee-kuan-yew>
- [4] Chang, T. K. 1999. Reporting public opinion in Singapore: Journalistic practices and policy implications. *The Harvard International Journal of Press/Politics*. 4, 1 (Jan. 1999), 11-28. DOI=<http://dx.doi.org/10.1177/1081180X99004001003>.
- [5] Cho, H., Chen, M., and Chung, S. 2010. Testing an integrative theoretical model of knowledge-sharing behavior in the context of Wikipedia. *Journal of the American Society for Information Science and Technology*. 61, 6 (Jun. 2010), 1198-1212. DOI=<http://dx.doi.org/10.1002/asi.21316>.
- [6] Chua, A. Y. K., and Banerjee, S. 2016. Linguistic predictors of rumor veracity on the Internet. In *Proceedings of the IAENG International MultiConference of Engineers and Computer Scientists: International Conference on Internet Computing and Web Services (Hong Kong, March 16 – 18, 2016)*. IAENG, Hong Kong, 387-391. Retrieved from http://www.iaeng.org/publication/IMECS2016/IMECS2016_pp387-391.pdf
- [7] Chua, A. Y. K., Banerjee, S., Guan, A. H., Xian, L. J., and Peng, P. 2016. Intention to trust and share health-related online rumors: Studying the role of risk propensity. In *Proceedings of the IEEE SAI Computing Conference (London, UK, July 13 – 15, 2016)*. IEEE, New York, NY, 1136-1139. DOI=<http://dx.doi.org/10.1109/SAI.2016.7556120>.
- [8] Collier, C. W. 1992. Intellectual authority and institutional authority. *Inquiry*. 35, 2 (1992), 145-181. DOI=<http://dx.doi.org/10.1080/00201749208602286>.
- [9] DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. 2003. Cues to deception. *Psychological Bulletin*, 129, 1 (Jan. 2003), 74-118. DOI=<http://dx.doi.org/10.1037/0033-2909.129.1.74>.
- [10] DiFonzo, N., and Bordia, P. 2007. *Rumor Psychology: Social and Organizational Approaches*. American Psychological Association, Washington, WA.
- [11] Doerr, B., Fouz, M., and Friedrich, T. 2011. Social networks spread rumors in sublogarithmic time. In *Proceedings of the Annual ACM Symposium on Theory of Computing (San Jose, California, USA, June 6 – 8, 2011)*. ACM, New York, NY, 21-30. DOI=<http://dx.doi.org/10.1145/1993636.1993640>.
- [12] Doerr, B., Fouz, M., and Friedrich, T. 2012. Why rumors spread so quickly in social networks. *Communications of the ACM*. 55, 6 (Jun. 2012), 70-75. DOI=<http://dx.doi.org/10.1145/2184319.2184338>.
- [13] Edy, J. A., and Risley-Baird, E. E. 2016. Rumor communities: The social dimensions of Internet political misperceptions. *Social Science Quarterly*. 97, 3 (Jun. 2016), 588-602. DOI=<http://dx.doi.org/10.1111/ssqu.12309>.
- [14] George, C. 2007. Consolidating authoritarian rule: Calibrated coercion in Singapore. *The Pacific Review*. 20, 2 (May. 2007), 127-145. DOI=<http://dx.doi.org/10.1080/09512740701306782>.
- [15] Glazer, C. 2002. Playing nice with others: The communication of emotion in an online classroom. In *Proceedings of the Annual Distance Education Conference (Austin, Texas, USA, January 2002)*. Retrieved from http://www.scholarlypursuits.com/dec_comm.pdf
- [16] Harrison, J. L. 2016. Trident Health boasts advanced care for women's breast health. *Summerville Journal Scene* (26 October 2016). Retrieved from

- http://www.journalscene.com/news/trident-health-boasts-advanced-care-for-women-s-breast-health/article_40ef3214-9480-11e6-a310-1fdd88a6912c.html
- [17] Huberman, B. A., Romero, D. M., and Wu, F. 2008. Social networks that matter: Twitter under the microscope. Retrieved from <http://ssrn.com/abstract=1313405>
- [18] Kendra, J. M., and Wachtendorf, T. 2003. Elements of resilience after the world trade center disaster: Reconstituting New York city's emergency operations centre. *Disasters*. 27, 1 (Mar. 2003), 37-53. DOI=<http://dx.doi.org/10.1111/1467-7717.00218>.
- [19] Krippendorff, K. 1980. *Content Analysis: An Introduction of its Methodology*. Sage, Thousand Oaks, CA.
- [20] Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. 2013. Prominent features of rumor propagation in online social media. In *Proceedings of the IEEE International Conference on Data Mining (Dallas, Texas, USA, December 7 – 10, 2013)*. IEEE, New York, NY, 1103-1108. DOI=<http://dx.doi.org/10.1109/ICDM.2013.61>.
- [21] Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*. 33, 1 (Mar. 1977), 159-174. DOI=<http://dx.doi.org/10.2307/2529310>.
- [22] Liu, D., and Chen, X. 2011. Rumor propagation in online social networks like Twitter: A simulation study. In *Proceedings of the IEEE International Conference on Multimedia Information Networking and Security (Shanghai, China, November 4 – 6, 2011)*. IEEE, New York, NY, 278-282. DOI=<http://dx.doi.org/10.1109/MINES.2011.109>.
- [23] Ma, W. W., and Chan, A. 2014. Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. *Computers in Human Behavior*. 39 (Oct. 2014), 51-58. DOI=<http://dx.doi.org/10.1016/j.chb.2014.06.015>.
- [24] Matsumura, N., Miura, A., Komori, M., and Hiraishi, K. 2016. Media and sentiments in the Great East Japan Earthquake Related Tweets—Social media as 'Meta Media'. In *Proceedings of the IEEE International Conference on Semantic Computing (Laguna Hills, California, USA, February 4 – 6, 2016)*. IEEE, New York, NY, 465-470. DOI=<http://dx.doi.org/10.1109/ICSC.2016.84>.
- [25] Oh, O., Agrawal, M., and Rao, H. R. 2013. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*. 37, 2 (Jun. 2013), 407-426. Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3100&context=misq>
- [26] Ozturk, P., Li, H., and Sakamoto, Y. 2015. Combating rumor spread on social media: The effectiveness of refutation and warning. In *Proceedings of the IEEE Hawaii International Conference on System Sciences (Hawaii, USA, January 5 – 8, 2015)*. IEEE, New York, NY, 2406-2414. DOI=<http://dx.doi.org/10.1109/HICSS.2015.288>.
- [27] Press, S. J., and Wilson, S. 1978. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*. 73, 364 (Dec. 1978), 699-705. DOI=<http://dx.doi.org/10.2307/2286261>.

- [28] Rieh, S. Y. 2002. Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology*. 53, 2 (2002), 145-161. DOI=<http://dx.doi.org/10.1002/asi.10017>.
- [29] Tanaka, Y., Sakamoto, Y., and Matsuka, T. 2013. Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. In *Proceedings of the IEEE Hawaii International Conference on System Sciences (Hawaii, USA, January 7 – 10, 2013)*. IEEE, New York, NY, 649-658. DOI=<http://dx.doi.org/10.1109/HICSS.2013.557>.
- [30] Underwood, R. L., and Klein, N. M. 2002. Packaging as brand communication: Effects of product pictures on consumer responses to the package and brand. *Journal of Marketing Theory and Practice*. 10, 4 (2002), 58-68. DOI=<http://dx.doi.org/10.1080/10696679.2002.11501926>.
- [31] Yih, W. T., Goodman, J., Vanderwende, L., and Suzuki, H. 2007. Multi-document summarization by maximizing informative content-words. In *Proceedings of the International Joint Conference on Artificial Intelligence (Hyderabad, India, January 6 – 12, 2007)*, 1776-1782. Retrieved from <http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-287>
- [32] Yulianti, N. K. D., and Hum, S. S. M. 2016. The perfection of interpreting by non-linguist interpreters. *International Research Journal of Management, IT & Social Sciences*. 3, 3 (Mar. 2016), 34-41. DOI=<http://dx.doi.org/10.21744/irjmis.v3i3.90>.
- [33] Zhang, Y. 2016. An analysis of rumor propagation based on event ambiguity. In *Proceedings of the Springer International Conference on Management Science and Engineering Management (Baku, Azerbaijan, August 30 – September 4, 2016)*. Springer, Singapore, 1607-1618. DOI=http://dx.doi.org/10.1007/978-981-10-1837-4_130.
- [34] Zhang, Z., Zhang, Z., and Li, H. 2015. Predictors of the authenticity of Internet health rumours. *Health Information & Libraries Journal*. 32, 3 (Aug. 2015), 195-205. DOI=<http://dx.doi.org/10.1111/hir.12115>.