Aspects of Rumor Spreading on a Microblog Network

Sejeong Kwon[1], Meeyoung Cha[1], Kyomin Jung[2], Wei Chen[3], and Yajun Wang[3]

[1] Korea Advanced Institue of Science and Technology, Republic of Korea
{gsj1029,meeyoungcha}@kaist.ac.kr
[2] Seoul National University, Republic of Korea
kjung@snu.ac.kr
[3] Microsoft Research Asia, China
{weic,yajunw}@microsoft.com

Abstract. Rumors have been studied for several decades in social and psychological fields, where most studies were theory-driven and relied on surveys due to difficulties in gathering data. Rumor research is now gaining new perspectives, because online social media enable researchers to examine closely various kinds of information dissemination on the Internet. In this paper, we review social psychology literature on rumors and try to identify the key differences in the dissemination of rumors and non-rumors. The insights from this study can shed light on improving automatic classification of rumors and better comprehending rumor theories in online social media.

Keywords: Rumor, Social Media, Diffusion Structure, Linguistic Properties

1 Introduction

A rumor is defined as an unverified explanation of an event at the time of circulation [16]. Nwokocha et al. says that the essence of rumors is in their ambiguity [13], where ambiguity of evidence makes rumors spread more widely. Another study about rumors says that a cognitive mechanism exists in the way people tend to modify a message they heard in the past [8]. Definitions of rumors vary in research [14]. A piece of information can be considered either *verified* or *unverified*, based on the judgments made at the time of circulation. The latter, a piece of information that cannot be verified at the time of circulation (i.e., unverified), is commonly considered to be a rumor in social psychology fields. In this paper, we rigorously divide the latter further into three types: *true*, *false*, and *unknown*, based on the judgments made after the time of circulation. The first type, true, describes when a piece of information that was unverified during circulation is officially confirmed as true after some time. This could be interpreted as information leakage, marketing, or prediction with enough reliable evidence. The other two types, false and unknown, which later in time are confirmed as false or remain unverified respectively, are what we define as rumors. Based on this definition, we built a rigorous set of ground truth data on rumors by recruiting four coders to manually annotate a large amount of social media data and identify rumors.

We test numerous theories and beliefs about rumor propagation in a social network. For instance, Alison says that people spread rumors to feel superior, to feel like part of the group, to get attention, or out of anger, boredom, envy, or unhappiness [17]. Others hypothesize that rumors are dominated by certain sentiments and polarities [18, 20]. These studies, which are based on surveys, bring interesting insights into the characteristics of rumor spreading.

The growth of online social media has made propagation of informative and creative content, as well as rumors, spam, and misinformation more prevalent. In order to handle the spread of potentially harmful information, researchers have investigated the problem of detecting unusual behaviors such as misbehaving users [6] and spammers [9]. Similarly, our main goal is to identify the patterns of spreading that are unique to rumors. In doing so, we also try to explain how the findings from social media research are related to well-known social and psychological theories on rumors.

We bridge theory and practice in this work and characterize the key properties of rumor spreading based on human-annotated data. We use near-complete data from Twitter and examine real rumor spreading cases in this network. We start by reviewing the social psychology literature on the theories and ideas related to rumors which we will then test one by one.

2 Theories on Rumor Spreading

Examining how a rumor spreads has been challenging, because the researcher had to be at the right place at the right time. Since this was nearly impossible prior to the use of social media data, previous studies on social and psychological aspects of rumors have mainly been theory-driven and have relied on a small amount of manually collected anecdotal evidence. We summarize four main hypotheses from the literature for an in-depth investigation in this paper.

Rumor spreaders and the direction of information flow
Besides its ambiguity, another essential characteristic of a rumor is its influence [13]. A rumor has the power to arouse people's interest; therefore, people gossip or spread rumors to get attention. This means that rumors are one of the ways that people gain influence over friends. However, highly influential individuals, who do not want to put their reputations at risk, will not likely initiate conversations on rumors because rumors have low information credibility [4, 20]. As a reasonable proxy of measuring a user's influence, we consider the time the user has been on Twitter (i.e., registration) and the user's number of followers (i.e., degree) in this paper. The first hypothesis we test is,
H1: Rumor spreaders are likely new based on registration time and has fewer followers; thus, rumors more likely disseminate from low-degree users to high-degree users.

Skeptics and participation
Psychological theories describe how people react to a given rumor. When a person hears about a rumor, he will first doubt the meaning and rely on his knowledge [8]. He will then check with factual sources to verify the rumor [3]. This process of doubt ends when he gathers enough evidence, at which points he either accepts the rumor and propagates it further or disapproves it and expresses negating comments. Solove [19] says that reputation gives people a strong incentive to conform to social norms. Because

rumors spread without strong evidence, rumor receivers may simply neglect the message, incurring low infection rate and often terminating the propagation process. The low credibility of rumors and the doubts incurred by the rumor's audience will result in a different writing style in rumor conversations compared to non-rumors.
H2: Rumors contain more words related to skepticism and doubts such as negation and speculation *and are less successful as conversation topics.*

Sentimental difference
Now we examine what kinds of rumors have been studied in social psychology. A classical study was done by Knapp [12], where he gathered a large collection of World War II rumors printed in the Boston Heralds Rumor Clinic column and categorized them into several types: pipe-dream (or wish-fulfillment), bogie (or fear), and wedge-driving (or aggression). The same approach was adopted in a study of 966 rumors from the Iraq War [11], giving insights into the societal attitudes and motivations of rumor spreaders. Wish-fulfillment rumors are fantasies about the world in which all desires are fulfilled [1]. Such rumors contain positive emotions like satisfaction and happiness. On the other hand, there is a general lay belief that rumors are dominated by negative sentiment and polarity [20].
H3: Rumors contain several characteristic sentiments (e.g., anger) compared to other types of information.

Social relationships and communication
While unverified information like rumors are often neglected and have low infection rates, this does not mean all rumors are short-lived. In contrast, certain rumors have been reported to be alive for a long period of time. What are the dissemination channels for those successful rumors? Could portals and prominent websites play a role (as they often do for other viral content)? We could not confirm this since the popularity of even the most famous rumor websites like `snopes.com` and `networkworld.com` was far lower than mass media websites and portals according to `Alexa.com`. This means that the primary channel of rumor dissemination is not through websites but through other means. The word-of-mouth of individual users can be one alternative mean, in which case rumor spreaders will attribute their source to social relations like friend, mate and family. Based on this assumption, we hypothesize that a large portion of rumors spread from person to person. Knapp's theory also supports this [12].
H4: Rumors will more likely contain words related to social relationships (e.g., family, mate) and actions like hearing.

3 Methods

We use data crawled from Twitter as explained in previous work [5]. The dataset contains profile information for 54 million users, 1.9 billion follow links between them, and the 1.7 billion public tweets posted from March 2006, when Twitter was launched, through August 2009. The link information is based on a snapshot of the network in August 2009. The complete set of users, links, and tweets provides us a unique opportunity to study user behaviors surrounding real information diffusion.

Collecting events and annotation

Given a data set of tweets, we need to collect real rumor cases that circulated on Twitter. We rigorously define a rumor as follows: (i) a statement that was unverified at the time of circulation and (ii) either remains unverified or is verified to be false after some time (i.e., at the time of this study).

Table 1. Representative rumor and non-rumor cases and their tweet data summary

Topic	Spreaders (Audience)	Tweets (Mentions)	Description (Regular Expression)
	Example tweet		
		Rumor	
Bigfoot	462	1006	The dead body of bigfoot is found
	(1731926)	(40)	(bigfoot & (corpse | (dead body))
	“Bigfoot Trackers Say They’ve Got a Body, I Say They Don’t”		
AdCall	325	719	Call a specific number to avoid advertisement
	(780300)	(151)	(888-382-1222)
	“Tired of telemarketers? call 888-382-1222 from the phone you want registered”		
ObamaAnti	119	135	Obama is muslim and antichrist
	(780300)	(19)	(obama & (muslim | antichrist)) ”
	“Obama may reach out to world’s Muslims on first international trip as president.”		
Swineflu	21896	26290	Don’t eat pork killed by swine flu
	(5300366)	(7710)	(swine flu & pork)
	“swine flu...don’t eat pork it’s disgusting”		
		Non-rumor	
Dell	1581	1909	Dell enters into smartphone market
	(1814798)	(389)	(dell & smartphone & market)
	“Would you buy a Dell smartphone? Seems you’ll soon have the chance.”		
Iphone3G	16056	31003	iphone3G is launched and its review
	(433215)	(4454)	(iphone3g)
	“got Iphone 3G and it is amazing”		
Havard	219	448	A black Harvard professor is arrested at his house
	(603911)	(111)	((harvard & arrest) | (henry louis & arrest))
	“Arrest of Harvard prof H.L. last week in his own home by cops ”		
Summize	2054	969	Twitter buys an IT company (twitter & buy & summize)
	(4367672)	(285)	(twitter & buy & summize)
	“Twitter buying summize is BRILLIANT. I bet it powers the home screen.”		

In order to understand the diffusion characteristics of rumors, we first had to identify real rumor cases from the Twitter data. For this, we searched lists of popular events from three websites: `snopes.com`, `urbanlegends.about.com`, and `networkworld.com`. Once target rumors were identified, we further identified a set of keywords describing each target rumor by consulting these websites and informed individuals in order to extract relevant tweets. We focused on a period of 90 days starting from a key date; this either corresponds to the date when the event occurred or the date when the event was widely reported in the traditional mass media (e.g., TV and newspapers). These rumors span political, health, urban legend, and celebrity topics. For a control group, we also searched a list of popular events from various media and websites. These non-rumor events are about political controversies, IT product launches, and movie releases.

We first identified 125 topics of interest, out of which 68 were rumors and 57 were non-rumors. To ensure that all rumors and non-rumors are valid, we recruited four well-trained human coders and asked them to classify each topic as either rumor or non-rumor. For each topic, we provided four randomly chosen tweets and a list of URLs on the topic to the annotators. We tested the annotators' agreement level and found an intraclass correlation coefficient (ICC) of 0.992. This indicates that the human coders' annotations were highly reliable. Table 1 lists examples of rumors and non-rumors, respectively. In this study, we further limited our data to only those topics that contained at least 60 tweets and as a result retained 102 topics (47 rumors and 55 non-rumors).

Variables

In Section 2, variables related to the hypotheses can be divided into three categories: personal, topological and linguistic. In case of personal characteristics, we define Age and $Follower$. Both are proxies of user influence. For each topic, Age is defined as the average time between user registration and the key date of the topic as described above. $Follower$ is an average number of followers.

For topological characteristics, we first define *friendship network* and *diffusion set*. *Friendship network* is defined as a subgraph of the original follower-followee graph induced by those users who posted at least one related tweet and follow links among them. From the friendship network, we define diffusion set as a set of ordered pairs, $D = \{e_1, e_2, \dots\}$, where each element in D represents a type of information flow from one user to another. We say information flows from user A (source) to user B (target), if and only if (1) B follows A on Twitter and (2) B posts about a given topic only after A did so. Then, we represent this information flow as an ordered pair, (A, B). If a target has multiple potential sources (e.g., $(s_1, t), (s_2, t) \dots, (s_n, t)$), we pick only the source of the most recent tweets the ordered set. Thus, a target cannot have multiple sources in this work.

Next, we introduce two measures from the diffusion set; $Flow$ and $Singleton$. $Flow$, the proportion of information flow from low-degree user to high-degree user, is defined as follows where $t(e)$, $s(e)$, and ind represent target, source of a given e and number of followers of a given node in the Twitter network, respectively.

$$Flow = \frac{|\{e \in D | ind(t(e)) > ind(s(e))\}|}{|D|}$$

$Singleton$ represents the proportion of users who posted about the topic without influencing others, i.e., having none of their followers reply or talk about the topic. If rumors are not successful conversation topics, $Singleton$ will be higher for rumors than non-rumors. We formulate $Singleton$ as follows where s_i, t_i and V are source and target of a given element, e_i, in D and set of nodes (i.e., users) in the friendship network, respectively.

$$Singleton = \frac{|V \setminus \bigcup_{\forall e_i \in D} \{s_i, t_i \in e_i\}|}{|V|}$$

In addition to topological aspects, we investigate linguistic characteristics of rumor spreading by utilizing a widely used sentiment analysis tool. LIWC (Linguistic Inquiry and Word Count) has been used for text analysis of psychological and behavioral dimensions [15]. Empirical results demonstrate that it can detect meanings in a wide variety of experimental settings, including attention focus, emotionality, social relationships, thinking styles, and individual differences [21].[4] Since the tool requires some minimum amount of text as input (e.g., 50 words), we group all the tweets belonging to a single topic as an input and collectively measured the score of sentiment (e.g., anger, sad) and linguistic (e.g., negate) categories.

Table 2. Variables related to hypotheses. In the "Expectation" column, we list whether rumors or non-rumors are expected to have a higher value. In case of the Linguistic features, "Definition" column lists words related to a given symbol.

Characteristic	Symbol	Definition	Expectation
H1: Rumor spreaders and the direction of information flow			
Personal	Age	Average of registration age	Non-rumor
Personal	$Follower$	Average number of followers	Non-rumor
Topological	$Flow$	Fraction of information flow from low to high degree users	Rumor
H2: Skeptics and participation			
Topological	$Singleton$	Fraction of users whose content is ignored	Rumor
Linguistic	$negate$	no, not never	Rumor
Linguistic	$cogmech$	cause, know, ought	Rumor
Linguistic	$exclusive$	but, without, exclude	Rumor
Linguistic	$insight$	think, know, consider	Rumor
Linguistic	$tentative$	may be, perhaps, guess	Rumor
H3: Sentimental difference			
Linguistic	$affect$	happy, cried, abandon	Non-rumor
Linguistic	$negemo$	hurt, ugly, nasty	Rumor
Linguistic	$anxiety$	worried, fearful, nervous	Rumor
Linguistic	$anger$	hate, kill, annoyed	Rumor
Linguistic	sad	crying, grief, sad	Rumor
Linguistic	$posemo$	love, nice,sweet	Non-rumor
H4: Social relationship and communication			
Linguistic	$social$	mate, talk, they, child	Rumor
Linguistic	$hear$	listen, hearing	Rumor

Table 2 lists variables related to the hypotheses we will test. In the, 'Characteristic' column, "Topological" and "Linguistic" mean the corresponding variables are estimated from *diffusion set* and LIWC, respectively.

4 Result

In this section, we test the significance of the variables described in Table 3 between rumors and non-rumors. Table 3 shows the result of the comparisons for each variable.

The first hypothesis, $H1$, considers three variables: Age, $Follower$ and $Flow$. These variables describe who the rumor spreaders are and how information flows. Ta-

[4] Full list available at http://www.liwc.net/descriptiontable1.php

Table 3. Extracted features and their p-values in t-test. In the "Type" column, "Non-rumor' and "Rumor" mean the feature had a higher value for non- rumors and rumors, respectively. In the "Expectation" column, we list whether rumors or non-rumors are expected to have a higher value.

Hypothesis	Symbol	Type	Expectation	p-value
$H1$	Age	None	Non-rumor	0.73
	$Follower$	None	Non-rumor	0.68
	$Flow$	Rumor	Rumor	**
$H2$	$Singleton$	Rumor	Rumor	***
	$negate$	Rumor	Rumor	***
	$cogmech$	Rumor	Rumor	*
	$exclusive$	Rumor	Rumor	***
	$insight$	None	Rumor	0.15
	$tentative$	Rumor	Rumor	**
$H3$	$affect$	Non-rumor	Non-rumor	*
	$negemo$	None	Rumor	0.92
	$anxiety$	None	Rumor	0.61
	$anger$	None	Rumor	0.85
	sad	None	Rumor	0.11
	$posemo$	Non-rumor	Non-rumor	*
$H4$	$social$	Rumor	Rumor	**
	$hear$	Rumor	Rumor	**

*P<0.05; **P<0.01; ***P<0.001

ble 3 demonstrates that a rumor flows from low-degree users to high-degree users with a statistically significantly high probability. This is in stark contrast to typical information propagation, which mostly involves flow from high-degree to low-degree users (i.e., two-step-flow of information [10]). However, we could not confirm the hypothesis that non-popular users utilize rumors to increase their influence over their friends. Rumors and non-rumors had similar registration age and number of followers.

The second hypothesis, $H2$, deals with the existence of different participation rates and writing styles in rumors. In Table 3 and Figure 1, we can see that most of the related variables show statistically significant distinct ranges of values between rumors and non-rumors, as predicted by the social and psychological theories. The high value of $Singleton$ indicates that rumor rarely initiate a conversation (i.e., no one talks about a rumor after seeing it). Other variables in $H2$ are about the speculative words that indicate doubt about the content of rumors. Our results support that process of doubt properly works to users for rumors and it induces different writing styles. The presence of a lot of negation in rumors can be attributed to people exhibiting uncertainty in their tweets. Our statistical test confirms that rumors have a clearly different writing style, providing empirical confirmation of the social and psychological theories about rumor spreading.

In third hypothesis, $H3$, our purpose is to test which sentiments are more dominant in rumors compared to non-rumors. In Table 3, we can see that rumors do not necessarily contain different sentiments than non-rumors. In fact, negative sentiments like anger, sadness, and anxiety may depend on the topic rather than on information credibility. Figure 2 shows the 95% confidence intervals of sentiment variables. For instance, news about a crime and an accident show higher negative sentiments. Lower

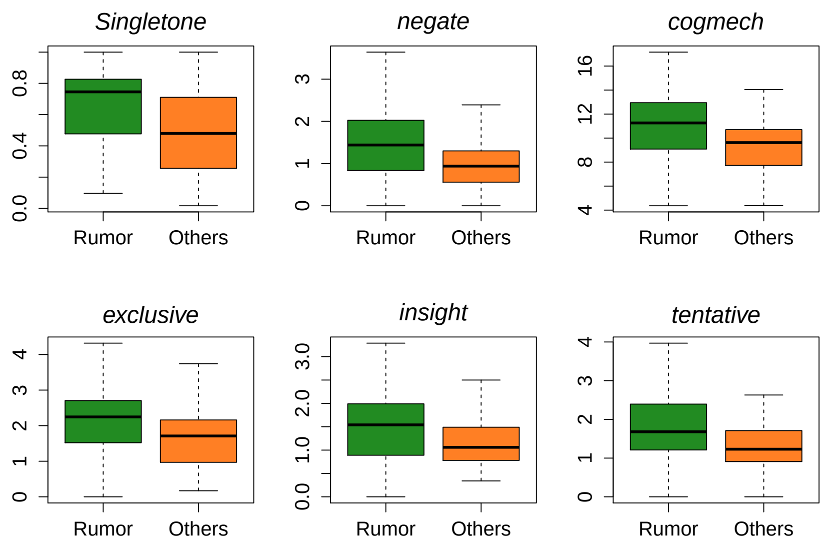

Fig. 1. 95% confidence intervals of variables on participation and writing style

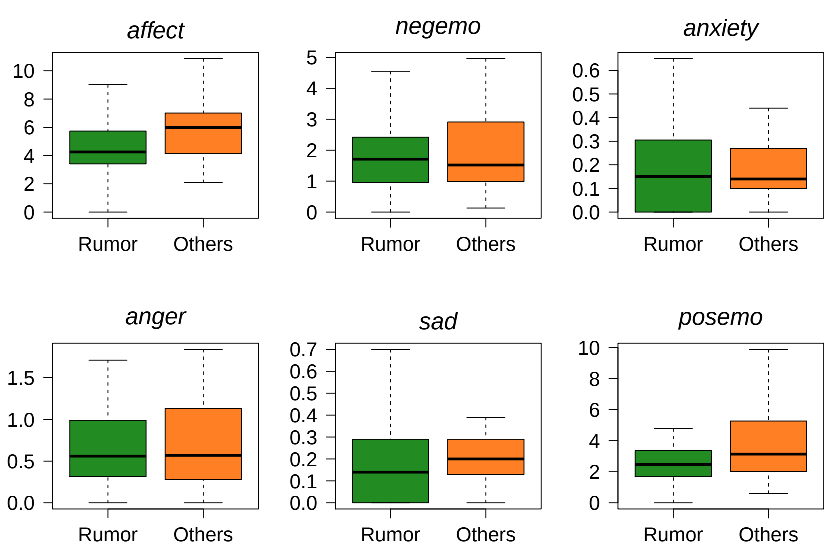

Fig. 2. 95% confidence intervals of variables on sentiments

affect score of rumor indicates that less appearance of words like happy, cried, and abandon. From this, we can infer that rumors are less likely influences human emotion. Combining these results, we conclude that unlike what a wide range of theories suggest, sentiment in rumors depends on the topic and is not statistically distinct from sentiment in non-rumors.

In Table 3, we see that rumors have a significantly higher fraction of words related to $social$ and $hear$. That is, words related to social relation like 'friend', 'buddy', and 'neighborhood' are more showed up in the rumor tweets. This would be an indicator of main propagation mechanism that rumors are more likely to be disseminated through social relation. This is quite different from other information, which originates from mass media. Thus, we conclude that there is enough evidence to support H4.

Table 4. Variable importance by Random Forest. MDA is Mean Decrease Accuracy value. Higher MDA means higher discriminative power.

Rank	Variable	Characteristic	MDA	Rank	Variable	Characteristic	MDA
1	*negate*	linguistic	18.46	10	*social*	linguistic	5.52
2	*affect*	linguistic	17.67	11	*sad*	linguistic	4.73
3	*Flow*	topological	16.16	12	*insight*	linguistic	2.84
4	*Singleton*	topological	11.63	13	*anxiety*	linguistic	2.66
5	*hear*	linguistic	10.94	14	*Follower*	personal	1.45
6	*tentat*	linguistic	10.20	15	*negemo*	linguistic	1.34
7	*excl*	linguistic	10.14	16	*Age*	personal	1.20
8	*posemo*	linguistic	9.29	17	*anger*	linguistic	0.30
9	*cogmech*	linguistic	9.10				

In addition, we estimated the discriminative power of the features. To prevent the problem of over fitting [7], we used *Random Forest*, a modified algorithm of bagging, which utilizes a large collection of de-correlated trees to measure variable importance [2]. Analysis of the discriminative power of the 17 features yields three insights. First, the personal characteristics, Age and $Follower$ (ranked 16th and 14th respectively), do not have much discriminative power. Second, sentimental differences (e.g., anger, sad) inferred from our literature reviews have no discriminative power. Third, writing style, the fraction of words related to speculation, and topological variables have the highest predictive power.

5 Discussion

Studies on rumors always have a data problem, because one must be at the right place at the right time. Thus, existing social psychology studies have been conducted only on very small-scale data (containing up to tens of users) and are mostly theory-driven. On the other hand, recent studies on rumors in online social media investigate many characteristics of rumors using large-scale data, but those are not related to theories in social and psychological studies. Our study on Twitter rumors (containing up to thousands and tens of thousands of users) serves as useful large-scale empirical data. Using the Twitter data, we examine the actual, complete diffusion instances, testing the hypotheses generated by the social and psychological literature. Hence, our tested features are more intuitive to understand a mechanism of rumor spread than others introduced in recent research and are directly applicable to real online networks for the task of classifying data as rumor or non-rumor.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (2011-0012988).

References

1. G. Allport and L. Postman. *The psychology of rumor*. Rinehart & Winston, 1947.
2. L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
3. H. Cantril. *The invasion from Mars: A study in the psychology of panic*. Transaction Pub, 1940.
4. C. Castillo, M. Mendoza, and B. Poblete. Information Credibility on Twitter. In *proc. of the International World Wide Web Conference*, 2011.
5. M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *proc. of the International AAAI Conference on Weblogs and Social Media*, 2010.
6. H. Cheng, Y.-L. Liang, X. Xing, X. Liu, R. Han, Q. Lv, and S. Mishra. Efficient Misbehaving User Detection in Online Video Chat Services. In *proc. of the ACM International Conference on Web Search and Data Mining*, 2012.
7. T. Dietterich. Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3):326–327, 1995.
8. L. Festinger, D. Cartwright, K. Barber, J. Fleischl, J. Gottsdanker, A. Keysen, and G. Leavitt. A study of a rumor: its origin and spread. *Human Relations*, 1948.
9. R. Kant, S. Sengamedu, and K. Kumar. Comment Spam Detection by Sequence Mining. In *proc. of the ACM International Conference on Web Search and Data Mining*, 2012.
10. E. Katz. The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 21(1):61–78, 1957.
11. S. Kelley. Rumors in Iraq: A guide to winning hearts and minds. *Master's Thesis, Monterey, CA: Naval Postgraduate School*, 2004.
12. R. Knapp. A psychology of rumor. *Public Opinions Quarterly*, 8:22–37, 1944.
13. N. K. Nkpa. Rumor mongering in war time. *The Journal of Social Psychology*, 96(1):27–35, 1975.
14. S. C. Pendleton. Rumor research revisited and expanded. *Language & Communication*, 18, 1998.
15. J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological Aspects of Natural Language Use: Our Words, Ourselves. *Annual Review of Psychology*, 54:547–577, 2003.
16. W. Peterson and N. Gist. Rumor and public opinion. *The American Journal of Sociology*, 57, 1951.
17. A. Poulsen. Why people gossip and how to avoid it, 2013. http://www.sowhatireallymeant.com/articles/personality-traits/gossip/.
18. R. Rosnow. Inside rumor: A personal journey. *American Psychologist*, 46:484–496, 1991.
19. D. Solove. *The future of reputation: Gossip, rumor, and privacy on the Internet*. Yale Univ Pr, 2007.
20. C. Sunstein. *On Rumours: How Falsehoods Spread, Why We Believe Them, What Can Be Done*. Penguin, 2011.
21. Y. R. Tausczik and J. W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.