Content or Context: Which Carries More Weight in Predicting Popularity of Tweets in China

Lun Zhang[1], Tai-Quan Peng[1,2], Ya-Peng Zhang[3] and Xiao-Hong Wang[3]

1. Web Mining Lab, Dept. of Media & Communication, City University of Hong Kong
2. Faculty of Humanities & Arts, Macau University of Science and Technology
3. Sina.com

Abstract

Through writing short tweets in microblogging sites, millions of users document their life, provide commentary and opinions, express deeply felt emotions, and articulate ideas. How does the content generated by a user compete with the content by others to attract limited human attention? With a framework developed from heuristic-systematic model of information processing, the study aims to uncover content factors and contextual factors that will affect the popularity of tweets. In the study, the popularity of tweets is decomposed into two dimensions: width of distribution and depth of deliberation. The data of the study are 10,000 tweets randomly drawn from a popular microblogging website in China. It is found that content and contextual factors play equally important roles in predicting width of distribution of tweets, while content factors outperform contextual ones in predicting depth of deliberation of tweets. Specifically, topics of tweets and availability of supplementary information are two important content factors in predicting popularity of tweets, while user characteristics (i.e., number of followers and user types) are important contextual factors. Our findings also suggest that re-tweeting and commenting are two distinct behaviors in the context of microblogging.

Keywords:

Microblogging, Heuritic-Systematic Model, Information Processing, Tweets Popularity

The annual conference of the World Association for Public Opinion Research, Hong Kong, June 14-16, 2012.
Correspondence should be addressed zhanglun1525@gmail.com.

Introduction

Different from traditional blogs, microblogging allows users to post short messages that are limited to no more than140 characters (known as tweets). These tweets could be distributed by forwarding the post (known as "re-tweet" function) to any other users who has declared to be the follower of the author (i.e., the users who have subscribed the updates of the author) and discussed by leaving comments by any users in the platform (known as "comments" function). Thanks to those technical features of microblogging, it has become a popular social platform for ordinary users and a significant marketing channel for organizations.

It is reported that Twitter, the most popular microblogging site around the world, has 140 million active users and more than 340 million tweets each day[1]. Microblogging users have been deluged by voluminous information available there. How will users act on millions of tweets available there? What kind of tweets will be more popular among users? This study aims to uncover the characteristics of tweets that will influence their popularity.

Conceptual Framework and Research Hypotheses

The study will draw on Heuristic-Systematic Model of information processing (HSM; Shelly Chaiken, 1980) to explain why some tweets will be more popular among users (i.e. be retweeted or commented more by users). HSM argues that people use two parallel modes, systematic and heuristic, to process information. Systematic mode is a deliberative processing mechanism, in which message- and topic-

[1] http://blog.twitter.com/2012/04/shutting-down-spammers.html (Accessed on May 7, 2012)

relevant cognition plays a critical role in forming judgments (S. Chaiken, 1987). Heuristic mode of information processing, which could co-occur with systematic mode and exert independent effects on opinion judgment (Chen & Chaiken, 1999), argues that individuals will assess the validity of the message based on a superficial assessment of extrinsic cues (e.g., characteristics of communicators), rather than the message itself (S. Chaiken, 1987).

According to HSM, users of microblogging sites will process billions of tweets available in two modes: systematic mode and heuristic mode. In systematic mode, users will make their behavioral decisions (i.e., re-tweet or comment) based on their evaluation of the tweet content *per se*, while in heuristic mode, users will make their decisions based on their assessment of contextual features of the tweet. Therefore, a conceptual framework, as shown in Figure 1, is proposed.

In the framework, popularity of tweets is decomposed into two dimensions according to specific characteristics of microblogging: depth of deliberation and width of distribution. Two sets of antecedents that will affect popularity of tweets are included in the framework: content factors of tweets and contextual factors of tweets (Suh, Hong, Pirolli, & Chi, 2010). Content factors refer to features of semantic cognition representing central cues for information processing, while contextual factors refer to characteristics of authors of tweets representing peripheral cues for information processing (Petty & Cacioppo, 1986).

Figure 1 about here

Content Factors Predicting Popularity of Tweets

In systematic mode of information processing, individuals' decisions toward messages are affected by contents implied in the message (Stone & Hoyt, 1974). Four content-related factors are included in the study: *topics of tweets* (Bandari, Asur, & Huberman, 2012), *length of tweets, affective degree of tweets* (Bandari, et al., 2012), and *availability of supplementary information*. .

Micro-blogging site, due to its main feature of diverse content generated by users in the platform, spots a niche in providing various type of information, such as news, socialization information, utility information, entertainment information, and information of self-assurance (Wang & Huberman, 2011). Wang et al. (2012) investigated the popularity of four types of topic (i.e., entertainment, technology, offbeat, and lifestyle) on Digg. With a topic modeling approach, Zhao et al (2011) found that world events and travel information as the most two popular topics in Twitter, followed by the topics of technic-science message, sports, arts, family & life, health, business and education. By introducing the categorization of news topics to Twitter, Bandar et al. (2012) found that the most popular tweets belong to the category of technology, health, fun stuff, and programming. Thus, it is hypothesized that:

H1: Tweets with different content foci will vary in their popularity.

People engage in message- or issue-relevant thinking depending on the level of informativeness. The length of the tweet is a basic indicator of the message informativeness. Individuals would be more responsive to longer tweets. However, the average length of a single tweet is about 14 words or 78 characters (Go, Bhayani, & Huang, 2009), which is less informative compared with the content in other social media (Ehrlich & Shami, 2010). Users of microblogging sites are allowed to insert a URL in their tweets, which may direct audiences to external webpages for more information. Thus the availability of supplementary information could enrich the informativeness of tweets, which may further improve the popularity of tweets. Thus, it is hypothesized that:

H2: The longer the tweet is, the more popular the tweet will be.

H3: The availability of supplementary information in tweets will increase the popularity of tweets.

Tweets in microblogging sites are expressed in a more sentimental way (Bollen, Pepe, & Mao, 2009). It has been found that individuals are more responsive to emotion-inducing information (Lord & Putrevu, 1993). Bandari et al. (2012) found that messages written in a more emotional manner could receive more attention in Twitter. Thus, it is hypothesized that:

H4: Affective degree of tweets will positively affect the popularity of tweets.

Context Features Predicting Popularity of Tweets

In heuristic mode of information processing, characteristics of information source have been identified as indispensable cues that will affect information judgment (S. Chaiken, 1987). Messages send from highly credible sources would be much easier to be acted upon. Four author-related features will be included as contextual factors in the study: *users' degree of activeness*, *users' self-disclosure degree*, *users' experience*, and *users' authoritativeness* (Kwak, Lee, Park, & Moon, 2010).

Castillo et al. (2011) identified that the number of tweets that the author has posted will exert influences on evaluation of the credibility of tweets. However, in the circumstance of micro-blogging site, over-active users, who publish too many tweets on microblogging sites, may be regarded as spammers by others. Thus, it is hypothesized that:

H5a: There would be an inverse-U shape effect of degree of activeness on popularity of tweets.

According to previous studies in spammer detection, the number of followers that the author has is found to positively moderate the relationship between number of tweets and tweets popularity. The number of followers and the posting frequency are used as effective indicators to detect spammers in Twitter (Benevenuto, Magno, Rodrigues, & Almeida, 2010). Users who have large number of followers and have actively posted tweets in the past are identified as influential users. Thus,

H5b: The number of followers positively moderates the effect of users' degree of activeness on popularity of tweets.

Users' self-disclosure degree, their authoritativeness, and their using experience are three other contextual factors that contribute to the source credibility. The degree of online self-disclosure, defined as the amount of message about self that an individual reveals online (Gibbs, Ellison, & Heino, 2006), reduces the uncertainty among users, fosters online closeness, and consequently increases the degree of information credibility. Tweets published by official and reputable source are considered more trustful and valuable (Castillo, et al., 2011). On the contrary, tweets by ordinary persons lacking expertise in certain domains are usually not trustful. Thus, given a cue that the author is verified as elite member (e.g., organization, media, celebrity, or school), the message would be more likely to be acted upon (S. Chaiken, 1987). Moreover, users with longer using history tend to spread more credible information (Castillo, et al., 2011). Hence, it is hypothesized that:

H6: Users' self-disclosure degree will positively affect popularity of tweets.

H7: Users' experience in using microblogging sites will positively affect popularity of tweets.

H8: Users' authoritativeness will positively affect popularity of tweets.

In addition to the aforementioned content and contextual factors, two control variables are included in the study: number of followers that the users have and the time when tweets are posted. The more followers that a microblogging user has, the more potential audience tweets posted by the user will have. Tweets published at rush hour (i.e., 8 am to 11 pm) are found to be able to reach the majority of audiences (Krishnamurthy, Gill, & Arlitt, 2008), while tweets released at other time slots (i.e. 0 am to 7 am) would have lower reachability.

Method

We randomly extracted 10,000 tweets published at the end of 2011 from Sina Weibo, one of the most popular microblogging sites in China. The data includes the information about the user (e.g., registration time, location, verification as VIP users, number of followers, and number of followees), tweets content, and the time when the tweet is posted.

Measurement of Popularity of Tweets

Popularity of tweets, the dependent variable in the study, is decomposed into two dimensions: width of tweets distribution and depth of deliberation on tweets. Microblogging sites distinguished itself from social networking sites by empowering users to spread message beyond the reach of their original online social contacts via the function of "re-tweet" (Kwak, et al., 2010). Meanwhile, micro-blogging sites also allow users to interact with their contacts under the posts via the function of “comment/reply”. Hence, width of distribution is operationalized as number of times a tweet is re-tweeted and depth of deliberation as number of comments a tweet receives.

Measurement of Content Factors

Categorization of *tweets' topics* is adopted from Bandari et al (2012). These 10,000 tweets are manually coded into following nine categories: (1) personal interests information (e.g., sports, movie, pets & plants, and travel); (2) utility information (e.g., IT/technology, education, health, and job information); (3) social/political/business news; (4) social interaction; (5) personal life; (6) marketing/advertising information; (7) joke/gossips; (8) self-assurance information; and (9) system message (system notifications). The inter-coder reliability is about .85.

The *length of the tweet* is measured as the total number of words and punctuations included in tweets. The *affective degree of tweets* is measured as proportion of emotion icons (e.g., smiling face, upset face, etc) and number of modal particles in tweets. *Availability of supplementary information* is operationalized as a dichotomous variable measuring if there is an inserted URL included in a tweet or not.

Measurement of Contextual Factors

Users' degree of activeness is measured by the total number of tweets that a user has published. The dichotomous variable whether there the author has provided a text-based self-description is employed to measure users' self-disclosure degree. Users' experience is measured as the duration between the registration date and the date when the author published the tweet. Users' authoritativeness is operationalized as a dichotomous variable recording whether a user is verified as elite members (i.e, celebrities, organization, media, etc.) by the microblogging site or not.

Measurement of Control Variables

Number of followers is directed recorded in the extracted data. Tweeting time extracted from the microblogging site is recorded into a dichotomous variable with two categories: rush hour (8 am to 11pm) and non-rush hour (0 am to 7am).

Findings

Descriptive Statistics

The retweet times and number of comments received by tweets follow a heavy-tailed distribution, as shown in Figure 2. This heavy-tailed distribution suggests that most of tweets receive quite limited attention from users and only very few tweets are actively acted upon by users. In subsequent multivariate analysis, logarithmic transformations are performed on the number of re-tweet times and comment times to correct its skewness (Cohen & Cohen, 1983).

Figure 2 about here

Personal life is the most popular topic of tweets posted in the microblogging site, as shown in Figure 3 which reports the distribution of topics of 10,000 tweets included in the study. Tweets on personal life account for 45% of 10,000 tweets examined in the study, which is followed by social interaction (14%), system notification (11%), personal interests (7%), online marketing/advertising (6%), and self-assurance (5%). Interestingly, tweets on social/political/business news account for a very limited proportion (3%) of the sample. The highly skewed distribution suggests that the microblogging site is largely used as a platform for updating personal status and socialization.

Figure 3 about here

Multivariate Analysis Results

As the two measurements of popularity of tweets (i.e., re-tweet times and comment times) are correlated with each other (r = 0.3, $p < .001$), multivariate analysis of covariance (MANCOVA) is employed in the study to test research hypotheses proposed. Three blocks of variables are included in the model: (1) content factors (i.e., topic of tweets, length of tweets, affective degree of tweets, and availability of supplementary information); (2) contextual factors (i.e., users' degree of activeness, users' self-disclosure degree, users' experience, and users' authoritativeness); and (3) control variables (i.e., the tweeting time and number of followers).

In total, content factors carry more weight in predicting popularity of tweets, while contextual factors play a minor role in predicting the popularity. To assess the explanatory power of three blocks of variables included in the model, semi-partial R^2 of each block of variables are compared. As shown in Table 1, the semi-partial R^2 on re-tweet times explained by content factors (5%) is slightly greater than that by contextual factors (3%). The semi-partial R^2 on comment times explained by content factors (9%) is much greater than that by contextual factors (2%).

Table 1 about here

The multivariate main effect for all independent variables were significant (for number of followers, *Hotelling's Trace* = .07, F (2, 8708) = 280.94, p < .00l, η^2=.06; for tweeting time, *Hotelling's Trace* = .00, F (2, 8708) = 6.34, p < .00l, η^2=.00; for topic category, *Hotelling's Trace* = .06, F (16, 17414) = 33.01, p < .00l, η^2=.03; for length of tweets, *Hotelling's Trace* = .00, F (2, 8708) = 16.06, p < .00l, η^2=.00; for affective degree of tweets *Hotelling's Trace* = .00, F (2, 8708) = 9.88, p < .00l, η^2=.00); for availability of supplementary information, *Hotelling's Trace* = .05, F (2, 8708) = 203.65, p < .00l, η^2=.05; for authors' degree of activeness, *Hotelling's Trace* = .00, F (2, 8708) = 32.63, p < .00l, η^2=.01; for the interaction term of authors' degree of activeness and number of followers, *Hotelling's Trace* = .02, F (2, 8708) = 90.89, p < .00l, η^2=.02; for authors' experience, *Hotelling's Trace* = .02, F (2, 8708) = 60.22, p < .00l, η^2=.01; for authors' authoritativeness, *Hotelling's Trace* = .01, F (2, 8708) = 20.09, p < .00l, η^2=.01) except for self-disclosure (*Hotelling's Trace* = .00, F (2, 8708) = .91, p =.41). MANCOVA analysis results are summarized in Table 2.

Table 2 about here

Mixed findings are found on the impact of tweet topics on popularity of tweets, as shown in Table 2. In comparison with tweets on social/political/business news, tweets on social interaction received more re-tweets (B = .18, p < .001) and are more commented (B = .40, p < .001). This finding is consistent with what has been found in previous studies (Honey & Herring, 2009), which suggests that users of microblogging sites in China consider microblogging as a platform of social networking rather than social media.

Tweets on personal interests and jokes/gossips received more re-tweets and comments than those on social/political/business news. Tweets on personal life receive less re-tweets (B = -.07, p < .001) but more comments (B = .21, p < .001) than tweets on social/political/business news. This finding suggests that re-tweets and commentating would be two distinct behaviors in the sense that re-tweeting shows the nature of broadcasting (D. Zhao & Rosson, 2009) while replying/commentating the nature of interpersonal conversation. Tweets on utility information are found to receive significantly more comments (B = .34, p < .001) than tweets on social/political/business news. Thus, H1 is partially supported.

Longer tweets are found to be retweeted more (B = .00, p < .001) but be commented less (B = -.00, p < .001). Longer messages, thanks to their informativeness, are more beneficial for delivering the claim in advertising research (Coursaris, Sung, & Swierenga, 2010). Thus, longer tweets are able to facilitate more retweeting actions. However, the length of tweets is not conducive to garnering more comments. The negative effect of tweet length on comment times, although very weak, indicates that the commentating is a behavior focusing on the relation maintenance and interpersonal conversation. Thus, H2 is not supported.

Availability of supplementary information is found to have negative effects on both re-tweets and comments ($B_{re\text{-}tweet}$ = -.17, p < .001; $B_{comment}$ = -.42, p <.001). Compared with the tweet without URL, tweets with URL are less re-tweeted and commented. This is contradictory with our hypothesis. The presence of external sources exerts "distraction" effects on message popularity. Audiences' attention may be directed to external websites by clicking the inserted URL. Previous studies found that whether the tweet contains links among news tweets did not affect the diffusion of the tweets (Yang & Counts, 2010).

The affective degree of tweets has positive effects on re-tweets and comments ($B_{re\text{-}tweet}$ = .06, p < .001;

$B_{comment}$ = .18, p < .001). The tweets written in a more sentimental way will receive more attention from users and the effect is especially true on comments. Thus, H4 is fully supported.

Consistent with our inverse U-shape hypothesis on the impact of authors' degree of activeness on popularity of tweets, the number of tweets ($B_{re\text{-}tweet}$ = -.05, p < .001; $B_{comment}$ = -.09, p < .001) and its quadratic term ($B_{re\text{-}tweet}$ = -.03, p < .001; $B_{comment}$ = -.06, p < .001) are found to have significantly negative effects on re-tweet times and number of comments. This finding suggests that users who either over actively posted tweets or published very few tweets in the past are regarded by others as invalid users, and consequently, their tweets cannot receive much more attention. The result indicates that there is a "backfire" phenomenon—the over active users may receive punishment, rather than reward in terms of future attention. Thus H5a was supported. However, the result of the effect of users' degree of activeness is inconsistent with the findings in the study by Yang & Counts (2010) . Their study found that the number of posts published in the past had positive effect on the diffusion of the tweet. The contradiction between the current study and the study by Yang & Counts (2010) may derive from the different approaches of data collection. Their sampling frame is the interactive network of the micro-blogging site based on "@username mentioning". Moreover, only the message highlighting the topic keywords (e.g., "Iran election") using hashtag (i.e., "#") was selected. The data selection that is biased towards active users largely avoids spam users, and consequently, leads to the conclusion that is competing with the present study.

The impact of users' degree of activeness on popularity of tweets is found to be conditioned by number of followers a user has. This finding suggests that as the number of followers increases, users who frequently publish tweets are more likely to be considered as valid user. Thus the number of followers enhances the effects of author activity on the message popularity. Thus, H5b is supported.

Users' self-disclosure degree does not exert significant impact on popularity of tweets. Thus, H6 is not supported. This result suggests that users' self-disclosure degree would not affect others' evaluation of their credibility. Credibility of users of microblogging sites may be judged by other factors, rather than their personal information provided on the site.

Users' experience has no effect on re-tweeting, but positive effect on commenting (B = .00, p < .001). In line with the result that the tweets of which the topic is social interaction receive more commenting, the result also suggests that users rely on commenting function for social interaction and on re-tweeting function for information diffusion. The more experienced a user is, the more likely s/he will use micro-blogging to maintain their online relationships, and consequently, receive more replies. Thus, H7 is partially supported.

Users' authoritativeness has positive effect on re-tweeting (B = .18, p < .001) while no effect on commenting. The result indicates that users only consider the credibility when they distribute the message. Since the commentating is a conversational behavior, users may reply on the basis of user familiarity, rather than author credibility. Thus, H8 is partially supported.

In the same vein, the number of follower has significantly positive effects on both re-tweets (B = .18, p < .001) and comments (B = .17, p < .001). The result is in consistent with previous study. For example, as suggested by Cha et al (2010), the number of followers shows significant positive correlation (γ = .55) with the number of re-tweets within each author. The tweeting time is also a significant predictor of re-tweet times (B = .04, p < .01) and number of comments (B = .05, p < .05). Compared with non-rush hour, tweets published at rush time will receive more re-tweets and comments.

Conclusive Remarks

This study firstly found that tweeting time and number of followers that the author has are two indispensable variables in predicting the tweet popularity. Moreover, the tweet popularity varies across different topic category. Although the propagation of news and social issues has received extensive study (e.g., Bandari, et al., 2012; Wang & Huberman, 2011; Ye & Wu, 2010), this study surprisingly found that compared with other topic categories, social news has received very little attention.

In addition, the study also found a "backfire" phenomenon in the sense that users who over actively published tweets previously would receive less re-tweets and comments. This result would shed some lights on the spam users detection. Another surprising finding involves that the external link may exert a distraction effect on the message response. Users tend to be less responsive to the message when they are exposure to the additional information since they may have been directed to the external websites.

Finally, the finding that user experience and authors' authoritativeness show different effects on re-tweets and comment times implicates that re-tweeting and comment are two distinct behaviors. The former aims to disseminate information in which the source credibility takes important role while the latter emphasize social interaction and conversation in which the user experience may carries more weight. The difference of the two response behaviors may derive from the platform design. Firstly, the technical default of 140 characters constrains the length of response via re-tweets. Thus re-tweet cannot be efficiently used as a way of interpersonal conversation that may involve multiple rounds. In contrast, the comment function allows more engaged and extended interaction. Secondly, it is not efficient to use comments to distribute the tweets to a greater audience since the comments could only be noticed by the author him/herself. Thus, for the purpose of efficiency, users may use re-tweet to disseminate the information to a greater audience while comment to interpersonal conversation.

The model with two blocks of variables (i.e., content variables and contextual variables) only provided limited explanation power in predicting the tweets popularity. Future studies should show extensive attention on the network properties of the micro-blogging sites and develop model of social dynamics by introducing longitudinal data (e.g., Lerman & Hogg, 2010; Wang & Huberman, 2011) to predict the message popularity.

References

Bandari, R., Asur, S., & Huberman, B. A. (2012). The Pulse of News in Social Media: Forecasting Popularity. *Arxiv preprint arXiv:1202.0332*.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). *Detecting spammers on twitter*. Paper presented at the CEAS 2010 - Seventh annual Collaboration, Electronic messaging, AntiAbuse and Spam Conference Redmond, Washington, US.

Bollen, J., Pepe, A., & Mao, H. (2009). *Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.* Paper presented at the WWW2010 North Carolina, USA.

Castillo, C., Mendoza, M., & Poblete, B. (2011). *Information credibility on twitter.* Paper presented at the the 20th international conference on World wide web, Hyderabad, India.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology, 39*(5), 752-766.

Chaiken, S. (1987). The heuristic model of persuasion. In M. Zanna, J. Olson & P. Herman (Eds.), *Social influence: The Ontario sysposium Volume 5* Hillsdale, NJ: Lawrence Erlbaum Associates.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. *Dual-process theories in social psychology*, 73-96.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Coursaris, C. K., Sung, J., & Swierenga, S. J. (2010). *Effects of Message Characteristics, Age, and Gender on Perceptions of Mobile Advertising,ÄìAn Empirical Investigation among College Students.* Paper presented at the Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR), Athens, Greece.

Ehrlich, K., & Shami, N. S. (2010). Microblogging inside and outside the workplace. *Proc. ICWSM, 10*.

Gibbs, J. L., Ellison, N. B., & Heino, R. D. (2006). Self-presentation in online personals. *Communication Research, 33*(2), 152-177.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.

Honey, C., & Herring, S. C. (2009). *Beyond microblogging: Conversation and collaboration via Twitter.* Paper presented at the System Sciences, 2009. HICSS '09, HI, USA.

Krishnamurthy, B., Gill, P., & Arlitt, M. (2008). *A few chirps about twitter.* Paper presented at the the first workshop on Online social networks, Seattle, WA, USA.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is Twitter, a social network or a news media?* Paper presented at the WWW '10 Proceedings of the 19th international conference on World wide web, NC, USA.

Lerman, K., & Hogg, T. (2010). *Using a model of social dynamics to predict popularity of news.* Paper presented at the Proceedings of the 19th international conference on World wide web, NC, USA.

Lord, K. R., & Putrevu, S. (1993). Advertising and publicity: An information processing perspective. *Journal of economic psychology, 14*(1), 57-84.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, *19*(1), 123-205.

Stone, V. A., & Hoyt, J. L. (1974). The emergence of source-message orientation as a communication variable. *Communication Research*, *1*(1), 89-109.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). *Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network.* Paper presented at the Social Computing (SocialCom), 2010 IEEE Second International Conference Minneapolis, MN.

Wang, C., & Huberman, B. A. (2011). Long Trend Dynamics in Social Media. *Arxiv preprint arXiv:1109.1852*.

Wang, C., Ye, M., & Huberman, B. (2012). From User Comments to On-Line Conversations. Retrieved from http://ssrn.com/abstract=2012183

Yang, J., & Counts, S. (2010). *Predicting the speed, scale, and range of information diffusion in twitter.* Paper presented at the the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA.

Ye, S., & Wu, S. (2010). Measuring message propagation and social influence on Twitter. com. *Social Informatics*, 216-231.

Zhao, D., & Rosson, M. B. (2009). *How and why people Twitter: the role that micro-blogging plays in informal communication at work.* Paper presented at the ACM 2009 international conference on Supporting group work, Florida, USA.

Zhao, W., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., et al. (2011). Comparing Twitter and traditional media using topic models. *Advances in Information Retrieval*, 338-349.

Table 1 Explanatory Power of Baseline Model, Model with Content Features, Model with Contextual Features, and Full Model

	Models with Content Features	Models with Contextual Features	Models with Control Variables	Full Model
$R^2_{re\text{-}tweet}$	5%	3%	8%	16%
$R^2_{comment}$	9%	2%	6%	17%

Table 2 MANCOVA Results of Content Features and Contextual Features on Re-tweet Times and Number of Comments

	Re-tweet Times		Number of Comments	
	Coefficients	*s.e.*	Coefficients	*s.e.*
Intercept	-.69***	.04	-.77***	.06
Content Features				
Topics1 (Personal Interests = 1, Social/Political/Business News = 0)	.08*	.03	.25***	.05
Topics2 (Utility Information = 1, Social/Political/Business News = 0)	.05	.04	.34***	.06
Topics3 (Social Interaction = 1, Social/Political/Business News = 0)	.18***	.03	.40***	.05
Topics4 (Personal Life = 1, Social/Political/Business News = 0)	-.07*	.03	.21***	.05
Topics5 (Marketing/Ads = 1, Social/Political/Business News = 0)	.04	.03	.10	.05
Topics6 (Jokes/Gossips = 1, Social/Political/Business News = 0)	.24***	.04	.21***	.06
Topics7 (Self Assurance = 1, Social/Political/Business News = 0)	.07	.04	.09	.06
Topics8 (System Notification = 1, Social/Political/Business News = 0)	-.01	.03	.06	.05
Length of Tweets	.00*	.00	.00*	.00
Affective Degree of Tweets	.06*	.03	.18*	.04
Availability of Supplementary Information (Absence of short links = 0, Presence = 1)	-.17***	.01	-.41***	.02
Contextual Features				
Users' Degree of Activeness	-.05***	.01	-.09**	.01
Users' Degree of Activeness^2	-.03***	.01	-.06**	.01
N. of Followers $\times$ Users' Degree of Activeness	.09***	.01	.05*	.01
Users' Self-Disclosure Degree (Absence = 0, Presence = 1)	-.02	.01	.00	.02

Users' Experience	.00	.00	$.00^{***}$	.00
Users' Authoritativeness (Ordinary = 0, Celebrity/Organizations = 1)	$.18^{***}$	.03	.03	.05
Control Variables				
Number of Followers	$.18^{***}$	.01	$.17^{***}$	.01
Tweet Time (Rush Hour = 1, Else = 0)	$.04^{*}$	.01	$.05^{*}$	.02

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Figure 1 Conceptual Framework

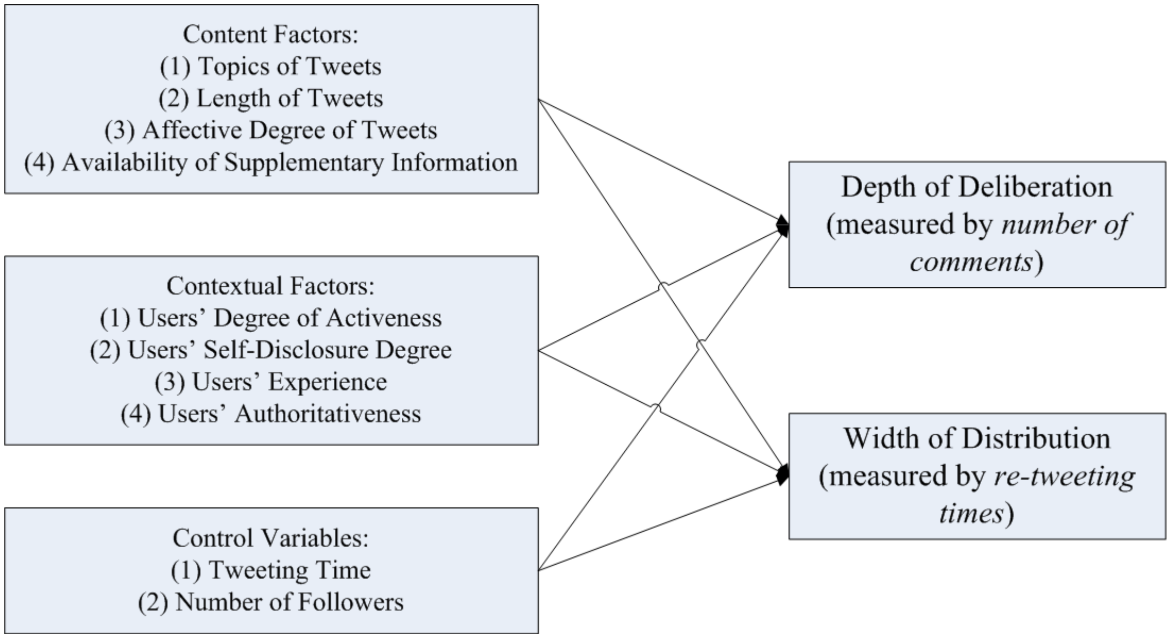

Figure 2 Distribution of Number of Retweets and Comments

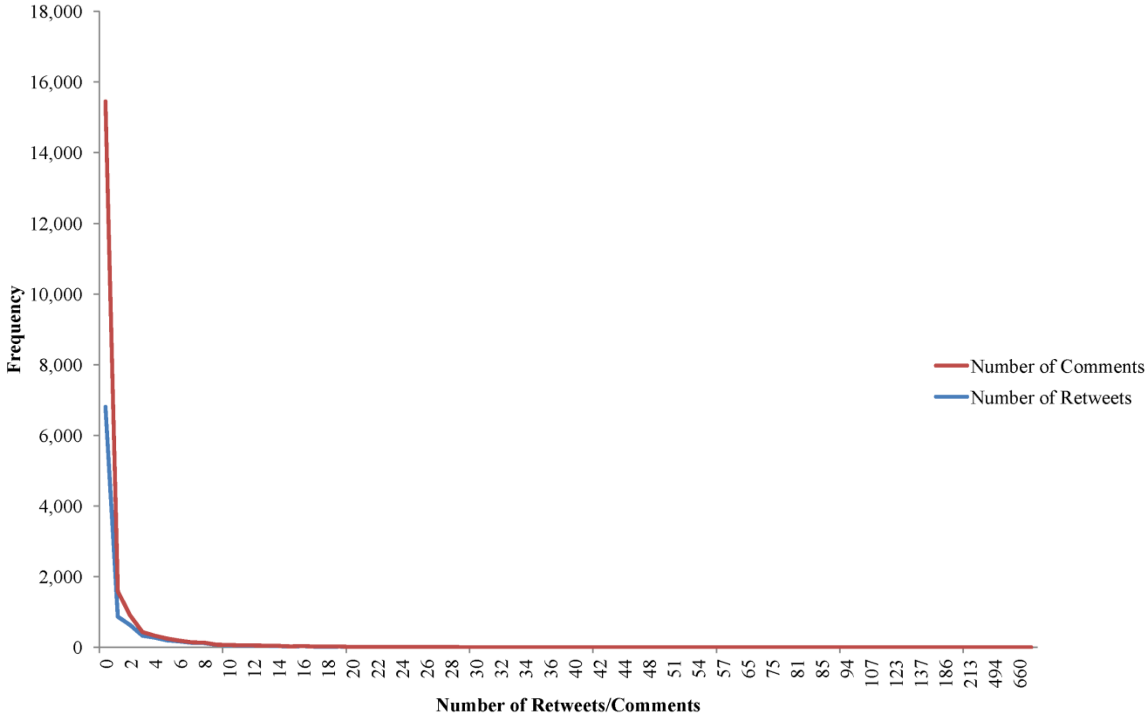

Figure 3 Distribution of Topical Categories of Tweets (N = 10,000)