Analysis and Research on Factors Affecting Information Dissemination of Emergencies in Social Media Environment

Ling Zhang
School of Management, Center for Service Science and Engineering, Wuhan University of Science and Technology, Wuhan, 430080?China

Jie Wei
School of Management, Wuhan University of Science and Technology, Wuhan, 430080?China

Lijun Zhu
Institute of Scientific and Technical Information of China, Beijing, 100038, China

Analysis and Research on Factors Affecting Information Dissemination of Emergencies in Social Media Environment

Ling Zhang[1,2] *Jie Wei*[1] *Lijun Zhu*[3]
1（School of Management, Wuhan University of Science and Technology, Wuhan, 430080，China）
2（Center for Service Science and Engineering, Wuhan University of Science and Technology, Wuhan ,430080, China)
3（Institute of Scientific and Technical Information of China, Beijing, 100038, China）

Abstract: We examine whether emotional expression (joy, anger, sadness, fear, disgust) occurring in social media content is associated with a user's information sharing behavior. Our research is in the context of Hurricane Irma and Tweets associated with that event. We find that negative emotions play an important role in the communication of information, among which, fear has the most significant effect. Meanwhile, the initial stages of the information life cycle have the most prominent influence on the information dissemination.

Keywords: Emergency; Social Media; Information Dissemination; Emotional evolution; Semi-supervised

1. INTRODUCTION

Social media has experienced tremendous growth in user base and has an impact on the public discourse and communication in society. Using social media, such as Twitter and Facebook, users post many messages, including their opinions and feelings as well as facts. One of the most successful social media services, Twitter, allows users to post tweets, which are short messages with a limit of 140 characters. As a result, Twitter is an ideal platform for users to spread information where the original tweet is propagated to a new set of audiences, namely, the "followers" of the tweeter. By retweeting, users may not only share information but also engage a certain audience or publicly agree or disagree with someone [1]. Compared with traditional communication platforms, the communication and dissemination of information and emotions in Twitter is more active and influential [2]. In addition, because of the low threshold, low cost, and instant and concise information distribution method used by twitter, people are more willing to use it as the preferred channel for personal and emotional communication. When emotions promote large-scale emotional resonance and discourse coordination among netizens, the imbalance of social emotions may be aggravated. The accumulation of negative emotions may be the hotbed of network events and social conflicts. Therefore, in public health emergencies such as natural disasters, it is of great practical significance to identify emotions in online social media and to analyze the differences in the transmission of different emotions at various stages of network evolution.

Factors affecting word-of-mouth message dissemination in social media have been extensively analyzed [3-4]. For instance, it has been shown that tweets with features such as URLs, hashtags, and emotional words are more likely to be retweeted than those without these features [5]. Each user in the online social media could be a social censor. A large number of tweets convey complicated signals about the users and the real-world events they are experiencing, of which the sentiments are an essential part. Therefore, emotion states of the users play a key role in understanding the user behaviors in social media, whether from an individual or group perspective [6-8].

The paper is organized as follows. The next section provides a literature review on microblogging information dissemination with an emphasis on factors, particularly regarding text content and time. In the subsequent section, the methodology is laid out. Followed by the results of our study. Finally, the paper concludes with a discussion of our results, research, and practical implications, as well as limitations and potential future work.

2. RELATED WORK

Factors affecting “retweet ability” of tweets have been analyzed in previous works. This was done by using content-related features such as, topics[9],URL（uniform resource locator）, and hashtag inclusion[3-4]. Hong et al [5]. have shown that tweet topics determined by topic modeling, which is a widely used natural language processing technique, and the number of followers of the tweet publisher are useful features for predicting the volume of retweets. As part of human communication, social media content usually conveys information about the author's emotional state, his or her judgment or evaluation of a person or topic, or anticipated emotional communication (that is, the sender wishes to have an emotional impact on the recipient) [10], commonly known as "sentiment". Hansen et al. [11] analyzed the relation between sentiment contained in a tweet and “retweet ability”. An analysis of approximately 560,000 tweets showed negative tweets have higher “retweet ability” than the positive tweets, while the opposite is true for non-news related tweets. Stieglitz et al. [12] have analyzed about 170,000 tweets related to political elections in Germany and revealed that the number of negative and positive tweets forwarded was higher than that of neutral tweets. However, few studies have identified emotions as another potential driver of information dissemination in social media environments, in particular, the user’s information sharing behavior. Physiological arousal has been shown to be a driver of information sharing [13,14]. Content that evokes high-arousal, or activating, positive or negative emotions is more viral. Content that evokes low-arousal, or deactivating, emotions is less viral. Festiger[15]deems that any form of cognitive dissonance can cause anxiety, so individuals will have the desire to reduce or eliminate this disorder. Cognitive imbalance theory refers to the fact that individuals recognize the contradiction between their attitudes and behaviors, that is, people tend to find reasons for their actions, such as free justification and self-persuasion. Therefore, this study will focus on the impact of emotional factors in the dissemination of emergency information.

The evolution process of emergencies often has a certain life cycle. If the dynamic change of public sentiment with the life cycle of an event is analyzed we can accurately capture the evolution process of public opinion information with the life cycle transmission. The famous division of the life cycle by scholars includes: in 1976, Turner [16] who described the development of the disaster model and divided the evolution process of the disaster into six stages: 1) the beginning point, 2) the incubation period, 3) the rush period, 4) the outbreak period, 5) the rescue period, and 6) the social adjustment period. In 1986, Steven, introduced the life cycle theory into crisis management and put forward the four-stage theory of crisis communication: potential stage, outbreak stage, spread stage and resolution stage [17]. Based on previous studies, this paper divides hurricane events into six stages: the initial stage, the outbreak stage, the first recession stage, the second growth stage, the second recession stage and the recovery stage.

3. METHODOLOGY

3.1 Data collection

Aiming at Hurricane Irma in the Caribbean on September 6, 2017, this study uses NodeXL to crawl data from September 6, 2017 to September 20, 2017 on Twitter with the keyword "Hurricane Irma". After removing non-English, duplicate and event-independent data from the collected microblog data, 238769 pieces of data were obtained, and then 183148 pieces of data were obtained after removing the data with less than 1 forwarding volume. At the same time, in the data preprocessing part, special symbols are filtered out for the remaining data, such as "/@@xxx", "#xxx", hyperlinks, and the like, which do not contain actual topic meaning and emotional information. The paper uses CRF algorithm [18] (conditional random field algorithm) to divide words in the text and further processes related expressions. The emotional annotation section is combined with psychologist Plutchik's multi-dimensional emotional model [19].

3.2 Methods for inferring tweet emotion

The basic idea of semi-supervised learning [20] is to use a small number of marked samples and a large number of unmarked samples to conduct effective emotional classification. Support Vector Machines (SVM) is proposed by Vapnik[20] on the basis of statistical theory and then based on structural risk minimization theory, which is often used as a supervised learning method for two classification. And the study has shown that SVM also has a good classification effect on multi-classification of emotions [21,22]. Therefore, this paper adopts the semi-supervised learning method of SVM to classify emotions in microblog texts. Its algorithm steps are shown in table 1 and its calculated results are shown in table 2, with an average accuracy of 83.8%. Experiments show that this method has a good effect in the classification of micro-blog emotions.

Table1. Semi-supervised emotion classification algorithm based on SVM

Input:Annotated dataset L containing emotion types(joy,anger,sadness,fear,disgust), unlabeled dataset U

Output:New annotated sample set and model C

1):Using TF-IDF to Extract Features.

2):Randomly extract 20% labeled data set R from L and 80% training set L'= L-R for training model.

3):While(Training threshold not reached)

4):Using L' to train SVM Classifier C.

5):Use classifier C to predict unlabeled data U.We can get the classified emotion set Si (i = 0,1,2,3,4), where i is the category label of emotion.

6):Use the classifier C to predict the reserved 20% annotation data R

7):Combine the unlabeled data with the highest confidence in the predictions in U and their corresponding emotion types, and then form a new set L'

8):The label data with the lowest prediction confidence in the merge R constitutes a new set

9):End While

10):Terminate iteration

Table 2. Emotional classification results

Emotional categories	Accuracy	Recall rate	F-score
joy	0.83	0.84	0.84
anger	0.83	0.89	0.86
sadness	0.86	0.80	0.83
fear	0.86	0.82	0.84
disgust	0.81	0.80	0.82
Average accuracy		0.838	

3.3 Measuring Method of Information Dissemination

In this paper, the number of forwards per tweet is used as a measure of the amount of information dissemination, and the number of tweets less than 1 is removed. Each tweet has its corresponding tweet time and emotional category, and the average daily forwarding volume is counted as the time interval to divide the network evolution stage.

This study uses regression analysis to investigate the influence of emotion and life cycle on information transmission, and other relevant text content factors, including URL, hashtag and followers, are also taken into account, whose main influencing factors are shown in table 3. The statistical data of tweet description are shown in table 4. Since the standard deviation of forwarding amount is much higher than its mean, it is necessary to adjust and analyze excessive dispersion. Therefore, this study uses stata15.0 for negative binomial regression to analyze the influence of emotion and life cycle on information transmission, and the model is shown in (1) and (2). In addition, two-way analysis of variance is used to test whether emotions and life cycles have an interactive effect on information dissemination

$$log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \, log(x_3) + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \quad (1)$$

$$log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \, log(x_3) + \beta_4 \, x_9 + \beta_5 \, x_{10} + \beta_6 \, x_{11} + \beta_7 \, x_{12} + \beta_8 x_{13} \quad (2)$$

Table3. Variables used in regression analysis

Variable	Description	Sym
Forwarding volume	Number of retweets	y
URL	Categorical variable for whether tweet includes URL	x_1
hashtag	Number of hashtags	x_2
follower	Number of followers	x_3
anger	Categorical variable, showing that tweet is anger	x_4
sadness	Categorical variable, showing that tweet is sadness	x_5
fear	Categorical variable, showing that tweet is fear	x_6
disgust	Categorical variable, showing that tweet is disgust	x_7
the first stage	the initial stage	x_8
the second stage	the outbreak stage	x_9
the third stage	the first recession stage	x_{10}
the fourth stage	the second growth stage	x_{11}
the fifth stage	the recovery stage	x_{12}

Table4. Statistics for collected tweet dataset

Variable	minimum value	Maximum value	Mean	SD
Forwarding volume	1	167240	323.43	3037.138
hashtag	0	16	0.97	1.470
URL	0	1	0.446	0.68
follower	0	999495	12991.2	66577.75

4. FINDINGS

4.1 Joy messages spread widely

In order to study the relationship between an emotional microblog and its amount of spreading, this paper calculates the average forwarding amount for tweets in each emotional category. The study found that when the average number of tweets forwarded by each emotional category was counted, the number of tweets forwarded by joy emotional micro-blog was larger, just like other negative emotional micro-blogs. In this study, considering the curiosity of researchers in this project, some of the tweets with joy emotional micro-blog ranked first were selected and found that some of them were humorous as shown in Table 5. Nahon and Hemsley[23] believed that the information dissemination content of social media has the following categories: humor, novelty, production quality, emotional influence, resonance and interest, while Bakshy and Hofman[24] et al. also found that more interesting content is more likely to be disseminated. Therefore, according to the research results, humor can also be used as one of the elements to promote the dissemination of information content.

Table 5. Some of the top-ranked tweets on joy's emotional forwarding volume

Tweet	Senti	Numberof
Buying iPhone X in Europe? why not add a free weekend trip to NYC .	joy	38869
We have Thick Rihanna, Thick Beyonce and now Thickye West. 2017 is officially the year of getting thick .	joy	35045
Thats scrappy doo, scoobs nephew you uncultured swine.	joy	33925
TRUMP: Sorry about that, here's a corporate tax cut ， I still have no house.	joy	26579
Hurricane Irma, you're doing amazing sweetie.	joy	24368

4.2 Life cycle division

As shown in figure 3, the hot degree of the event is mainly maintained from September 10, 2017 to September 20, 2017, the data of Hurricane lrma from September 6, 2017 to September 9, 2017 are very small, so the Alma hurricane event is divided into six stages: initial stage (September 11 -September 10), the outbreak stage (September 12), the first recession stage (September 13 to September 14), the second growth stage (September. 15), a second recession stage(September 16), the recovery stage (September 17 to September 20).

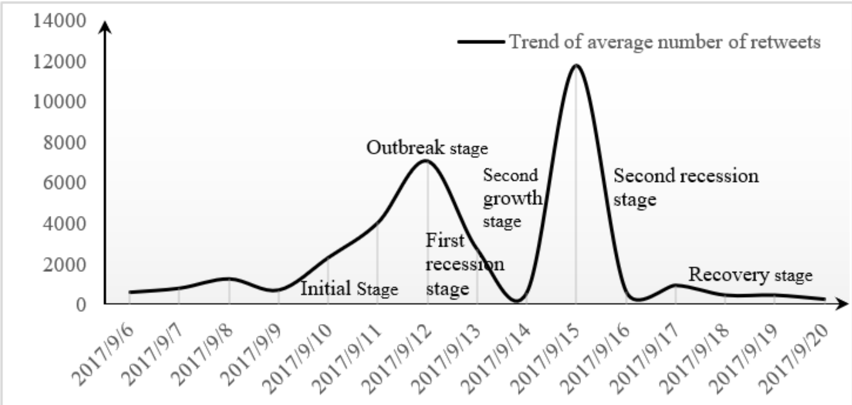

Figure 3. Average number of retweets of --Time Series Diagram

4.3 Emotion, time and Retweet Behavior

This paper first uses the negative binomial regression to study the relationship between emotional microblog and the amount of communication, and the regression equation is (1). Since the negative binomial regression is used, the coefficient of the explanatory variable is needed to be exponentially transformed. Table 6 shows the results of the analysis, the values of β and e^{β} in the table are used to measure the impact of explanatory variables on the interpreted variables, and e^{β} is an index for estimating the impact of explanatory variables on the interpreted variables, which means that when all other explanatory variables are controlled unchanged, adding each unit to explanatory variables is expected to cause e^{β} times transfer. Since R^2 increases with the number of variables or observations, R^2 cannot fully represent the goodness of fit of the regression equation. The regression coefficient indicates that anger, sadness, fear and disgust all increase the number of the transfer of tweets, among which fear is the most frequent (β=0.330, p< 0.01), anger (β=0.150, p< 0.01), disgust (β=0.149, p<0,01) are consistent with the previous results, and sadness (0.061, 0.05) is the weakest. In addition, fear regression coefficient shows that the amount of transfer of tweet with fear emotion is 39.1% higher than that of joy emotion, and the effect of URL is similar to that of other emotions, which indicates that negative emotions are the main driving factors of text transmission.

Secondly, this paper also uses negative binomial regression to study the relationship between life cycle and propagation quantity, such as equation (2). The influence of different explanatory variables on the transfer amount of tweets is shown in table 6. According to the analysis results, the effect of the initial stage (β=0.280, p<0.01) and the explosive stage on the transfer amount of tweets is more significant than that of the other four stages (β=0.294, p<0.01), and the results are consistent with 4.2. In addition, the regression coefficient in the outbreak stage indicates that the transfer amount in this stage are 34.2% higher than those in the recovery stage. Looking at other control variables, it can be found that URL (β=0.266,p<0.01),hashtag (β=0.069, p<0.01) and logfollower (β=0.690, p<0.01)all significantly affect the forwarding amount, especially logfollower.

Table 6 . Negative binomial regression results for Number of retweets

Dependent variable	Number of retweets（y）					
	（1）			（2）		
Independent variables	Coeff	SE	e^{β}	Coef	SE	e^{β}
URL	0.329	0.	1.39	0.26	0.02	1.30
hashtag	0.052	0.	1.05	0.06	0.00	1.07
logfollower	0.481	0.	1.61	0.69	0.04	1.99
anger	0.150	0.	1.16			
disgust	0.149	0.	1.16			
fear	0.330	0.	1.39			
sadness	0.061	0.	1.06			
Initial stage				0.28	0.03	1.32
Outbreak stage				0.29	0.04	1.34
The first recession stage				0.07	0.04	1.07
The second growth stage				0.11	0.04	1.12
The second recession stage				0.04	0.05	1.04
Constant	1.720	0.		1.59	0.00	
Pseudo R^2	0.024			0.03		
Observations	183148					

***** Significant at 1% level; ** significant at 5% level; * significant at 10% level**

5. DISSCUSSION

Several insights emerge from the results of this study. First, we found that the affective dimensions (anger, sadness, fear, disgust) of political hurricane messages are indeed significantly associated with retweet behavior in terms of retweet quantity, in the way that negative emotionally charged tweets are more likely to be disseminated compared to joy ones, for example, fear tweets were retweeted 39.1% more often than joy tweets. To some extent, it conforms to the theory of cognitive deviation[19]. More specifically, sentiment (fear, anger) in social media content might also induce arousal-related effects (e.g., attention and physiological arousal) that affect sharing behavior in social media communication. Meanwhile, sadness will trigger a low level of physiological arousal, consistent with previous results[17-18].However, our results also show that the effects of joy in a tweet on its virality are weak. This contradicts the results with Berger and Milkman. A possible cause of the difference between our study and previous studies might be the context environment of the subject audience. Our study revealed a interesting finding. Some tweets with high forwarding amount of joy emotion and found that these contents are humorous.Positive news with optimism and humour is more communicable in disaster. Second, our results show that the effect of the initial stage and the explosive stage on the amount of transfer of tweets is more significant than that of the other four stages. Therefore, it is important for relevant government departments to timely deal with tweets of negative emotions in emergencies, especially at the beginning and outbreak stages of the life cycle.

According to the research results of this paper, in the process of public opinion management of catastrophic emergencies, relevant government departments should first timely channel negative emotions, especially fear, anger, disgust and disgust, which have a great impact on microblogs, to effectively guide the sound development of public opinions and avoid the occurrence of mass incidents. Secondly, starting from the communication subject, when microblog public opinions enter the initial and outbreak stages. At the same time, microblog public opinion can quickly enter the stage of positive emotion accumulation, and then strengthen the overall grasp of public opinion by creating a large number of positive messages to create a harmonious and upward public opinion environment.

Of course, there are still some limitations in this paper. For example, with the increase of cultural communication, network information of multiple languages influences and integrates with each other, the existing work mainly focuses on a single language, so the corpus resources and results collected in the emotional analysis of a single language cannot be directly used in a multi-language environment. In addition, when different media and different forms of emotional information are "fused" together, "qualitative change" will occur. Therefore, in the follow-up study, relevant data from multiple channels should be collected and integrated for specific events, so as to more comprehensively reflect the trend of public emotions and better serve the early warning and emergency management of emergencies.

6.CONCLUSION

In this study, hurricane lrma is taken as an example to explore the factors influencing the information dissemination of emergencies in the social media environment. We investigated the relation between the sentiment of a tweet and life cycle and its virality in terms of dissemination volume. We used the number of retweets as measures of tweet virality. The results showed that negative emotion tweets spread more widely than positive emotion tweets, and that the dissemination volume of fear tweets was 39.1% higher than that of joy tweets. We also conducted life cycle analysis of tweet dissemination and discovered that in the initial stage and the outbreak stage of an emergency, the effect on the transfer amount of tweets is more prominent than that of the other four stages. Moreover, the regression coefficient in the outbreak stage indicates that the transfer amount in this stage are 34.2% higher than those in the recovery stage.

ACKNOWLEDGEMENT

Research was sponsored in part by the National Social Science Fund Project “Study on dynamic optimization mechanism of information diffusion in social networks”, Agreement Number 15CTQ029.

REFERENCES

[1] V. M.(2008). New Media, Networking and Phatic Culture. The International Journal of Research into New Media Technologies,4(14):387-400.

[2] Boyd D M, GolderS, Lotan G. (2010).Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter: Hawaii International Conference on System Sciences,

[3] Suh B, Hong L, Pirolli P, et al. (2010)Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. International Conference on Social Computing/ IEEE International Conference on Privacy, Security, Risk and Trust. IEEE Computer Society,177-184

[4] Naveed N, Gottron T, Alhadi A C.(2011).Bad news travel fast: a content-based analysis of interestingness on Twitter. Proceedings of 11th ACM SIGKDD international conference on knowledge discovery and data mining. New York:ACM,1-7

[5] Hong L, Dan O, Davison B D.(2011). Predicting popular messages in Twitter. International Conference on World Wide Web, WWW. India: DBLP, 57-58

[6] Sheridan D P, Decker H K, Kloumann I M, et al.(2011). Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. PLOS One, 6(12):e26752.

[7] GOLDER S A, MACY M W.(2011).Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science,333(6051):1878–1881.

[8] Marchetti-Bowick M, Chambers N.(2012).Learning for microblogs with distant supervision: political forecasting with Twitter. Conference of the European Chapter of the Association for Computational Linguistics, 603–612

[9] Nagarajan M, Purohit H, Sheth A P.(2010).A Qualitative Examination of Topical Tweet and Retweet Practices. International Conference on Weblogs and Social Media. ICWSM 2010, Washington: DBLP，295-298

[10] Bollen J, Pepe A, Mao H.(2012).Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. The Sense of Touch and its Rendering.

[11] Hansen L K, Arvidsson A, Nielsen F A, et al. (2011).Good Friends, Bad News - Affect and Virality in Twitter. International Workshop on Social Computing, Network, and Services. Berlin Heidelberg: Springer,

[12] Stieglitz S, Dang-Xuan L.(2013).Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. Journal of Management Information Systems, 29(4):217-248.

[13] Berger J , Milkman K L .(2009). What Makes Online Content Viral. Journal of Marketing Research,49(8):192-205.

[14] Berger, J.(2011) Arousal Increases Social Transmission of Information. Psychological Science, 22(7):891-893.

[15] Festiger L A. A theory of cognitive dissonance.(1957). Stanford Univ. Pr., 164-165.

[16] Turner B A.(1976)The Organizational and Interorganizational Development of Disasters. Administrative Science Quarterly, 21(3):378-397.

[17] Fink S, Amacom.(2002).Crisis Management: Planning for the Inevitable. American Management Association, 4(3):875-876.

[18] Lafferty J, Mccallum A, Pereira F.(2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings of ICML, 3(2):282-289.

[19] Plutchik R. (2001).The Nature of Emotions. American Scientist, 89(4):393-409.

[20] Vapnik V, Sterin A.(1977).On structural risk minimization or overall risk in a problem of pattern recognition. Automation & Remote Control, 10(10):1495-1503.

[21] Deshmukh R R, Kirange D K.(2013). Emotion Classification of News Headlines Using SVM. Asian Journal of Computer Science & Information Technology, 2(5):104-106

[22] Chin Y H, Lin C H, Siahaan E, et al.(2013). Music emotion classification using double-layer support vector machines. 2013: 193-196.

[23] Nahon K, Hemsley J. (2013).Going Viral. UK: Polity Press Cambridge: 2013.

[24] Bakshy Eytan, Hofman, et al.(2011). Everyone's an influencer: quantifying influence on twitter.In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining.