Framework for Rumors Detection in Social Media

Rehana Moin, *Zahoor-ur-Rehman, Khalid Mahmood
Department of Computer Science COMSATS University
Islamabad, Attock Campus, Pakistan

Mohammad Eid Alzahrani, Muhammad Qaiser Saleem
College of Computer Science and Information Technology,
Al Baha University, Al Baha, Saudi Arabia

***Abstract*—The development of social networks has led the public in general to find easy accessibility for communication with respect to rapid communication to each other at any time. Such services provide the quick transmission of information which is its positive side but its negative side needs to be kept in mind thereby misinformation can spread. Nowadays, in this era of digitalization, the validation of such information has become a real challenge, due to lack of information authentication method. In this paper, we design a framework for the rumors detection from the Facebook events data, which is based on inquiry comments. The proposed Inquiry Comments Detection Model (ICDM) identifies inquiry comments utilizing a rule-based approach which entails regular expressions to categorize the sentences as an inquiry into those starting with an intransitive verb (like is, am, was, will, would and so on) and also those sentences ending with a question mark. We set the threshold value to compare with the ratio of Inquiry to English comments and identify the rumors. We verified the proposed ICDM on labeled data, collected from snopes.com. Our experiments revealed that the proposed method achieved considerably well in comparison to the existing machine learning techniques. The proposed ICDM approach attained better results of 89% precision, 77% recall, and 82% *F*-measure. We are of the opinion that our experimental findings of this study will be useful for the worldwide adoption.**

Keywords*—Social networks; rumors; inquiry comments; question identification

I. Introduction

A rumor is an unverified claim about any event, transmitting from person to person. It may refer to an incident, object or problem of public concern. It may prove to be a social destructive phenomenon in any human culture. Usually, the social media rapidly transmits the unverified statements that may be harmful for anybody. Nowadays, social networks like Twitter and Facebook are more popular with regards to acquiring and propagating information. On social networks everybody is free to obtain and share information, anywhere at any time [1].

Besides, it has been reported that these social sites are capable to spread rumors [2]. In general, a rumor refers to the information that lacks source and its truthfulness. Ordinarily, it is generated in an emergency situation, leading to anxiety, disruption of social activities; thus, reducing the government credibility, even endangering the national security, for instance, on March 2011, after Japan Earthquake followed by tsunami and nuclear disaster. A rumor was propagated by microblog platforms, advising use of iodized salt for protection of people by nuclear radiation. Consequently, the public in general rushed to markets to buy salt, which was totally untrue and unnecessary practice. In the future, to avoid such unfruitful happenings, at the earliest, rumor detection is essential.

Earlier, much work has been done on rumor detection using the Twitter. We did work on Facebook to address the problem of rumor detection. We selected Facebook reason being the most popular social network. In Oct 2012, Facebook was having one billion users per month. Cameron Marlow, one of the research scientists, considered Facebook as world's most powerful instrument for studying human society [3].

A framework diagram is developed for rumors detection, starting from Facebook data collection, preprocessing of data, extraction of English text, apply TopicRank to obtain keyphrases and based on those keyphrases (topics) extract the event data and detect assertion to filter assertive event posts and finally detecting the inquiry comments on assertive posts using our proposed ICDM approach . We used labeled data from snopes.com to check the validity of our proposed ICDM approach and to make comparison with machine learning techniques.

We aim to tackle the rumors detection problem using inquiry comments identification through ICDM approach. This comprises two steps. In the first phase, we identify questionable statements named as "inquiry comments". We adopt both machine learning supervised approach like classifiers to detect questions and rule-based method to detect question marks, 5W1H words and regular expressions [4] which utilizes patterns to filter inquiries. In the second phase, we extract inquiry comments asking question about the event. We define the threshold to identify the rumors and test our ICDM model using labeled data from snopes.com.

Consequently, following research questions are formulated:

- How English text is separated from different languages?
- How to develop rumors detection framework that can correctly identify the rumors?
- How can we verify our proposed ICDM (Inquiry Comments Detection Model)?

Remaining part of this paper is organized as follows: Section II provides the related work; in Section III, methodology is presented; in Section IV, the results are presented: Section V concludes the whole work and addresses the future research directions of this study.

II. Related Work

This research related work aims to explore the role of Social Media in real-world emergencies, news diffusion, and rumor detection approaches.

A. Role of Social Media in Real World Emergencies

The use of social media in emergencies and crisis has gone up many folds in recent years [5], by involving reports from the eyewitnesses. Furthermore, the use of social media in actual emergencies has also been studied. These studies have shown the importance of Social media for breaking news, information gathering and coordinating in different situations, including emergencies, protests [6], natural disasters like earthquake, floods, hurricanes and forest fires. During natural and un-natural emergencies, the social media has a significant role in both transmitting information to the situation affected people and getting live reports from eyewitnesses. The local infrastructure if intact can provide the situation information to the public in much faster rate on their mobile services,rather than using the traditional news media.

B. News Diffusion through Social Media

The social media is a medium for conversation besides a source of news for the public. Mostly the current topics on the social media are news related therefore, they can be used to detect breaking news. Breaking news reporting enables the public getting to know the current information through eyewitnesses [7]. Researchers aim to use this feature of social networks to develop the tools for latest news-gathering [8] and to report the current situation, analyzing the user-generated content (UGC) [9] and discovered the potential of social networks to give rise to citizen journalism as well as verification of reports posted in the social media platforms [10].

C. Rumor Detection Approaches

Nowadays, the discovery of social network services has led to the public to spread rumors at fastest rate. Castillo et al. [11] studied the Twitter reports of the public during 2010 Chile earthquake thus, analyzing user's behavior on microblogging platforms to reach credibility of such information and examined the retweets system to analyze the rumors propagation pattern on the Twitter. Qazvinian et al. [12] classified the rumors related tweets using the matching regular expression with the keyword query.

Zhao et al. [4] approach based on the assumption that rumors will provoke tweets from the user's inquiring about their reliability; it implies that such tweets are possibly rumor having a number of enquiring tweets. The author prepared a list of five regular expressions (such as "is (that | this | it) true") that are useful to identify the inquiry tweets.

There has been found little work on automatic rumor detection regardless of the extensive study to examine the rumors in social media and developing techniques to tackle this problem [13], [14]. A set of predefined rumors (e.g. Obama is Muslim) is fed to a classifier, which classifies new tweets as being linked to the predefined rumors or not (e.g. I think Obama is not Muslim would be about the rumor, while Obama was talking to a group of Muslims wouldn't).

Tolosi et al. [15] use feature analysis on different events datasets and found it difficult to distinguish between rumors and non-rumors as features change dramatically across events. Zubiaga et al. [16] resolved these discoveries at the tweet platform, by showing that generalizability can be attained by leveraging context of the events. In [17] studied the use of a crowdsourcing platform to detect rumors and non-rumors in social network.

III. Methodology

The framework diagram for rumors detection is presented in Fig. 1.

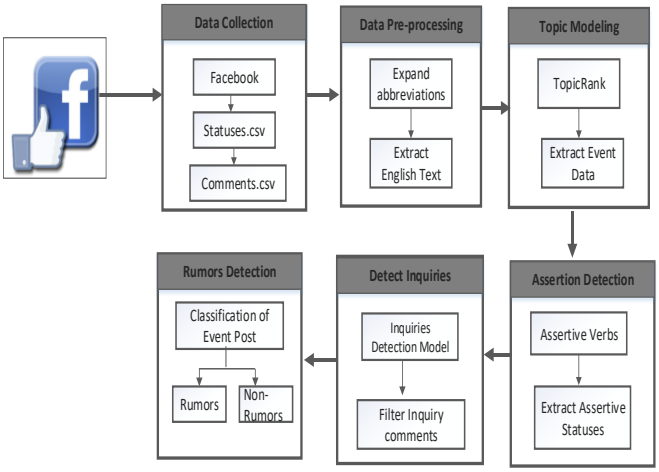

Fig. 1. Framework diagram for rumors detection.

A. Data Extraction

The Facebook data can be easily accessed and publicly available. Fig. 2 presents the flow diagram of data collection.

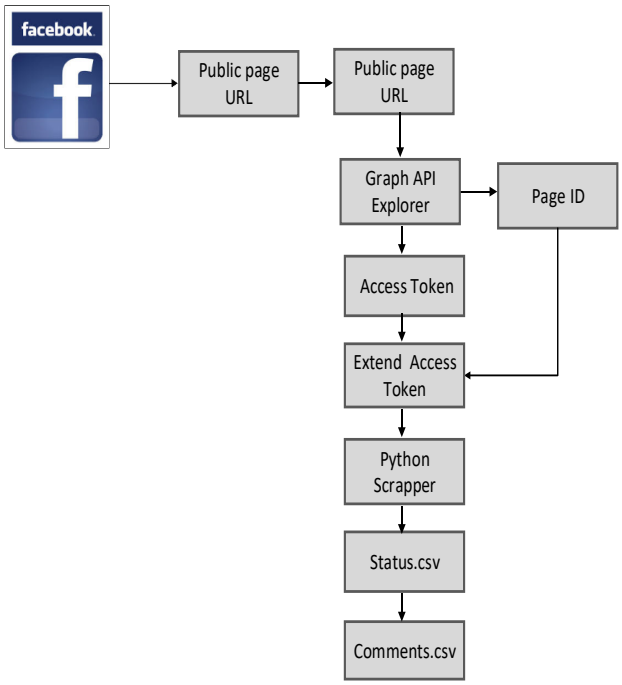

Fig. 2. Flow diagram of data collection.

Facebook data is collected using Python scraper and Facebook Graph API Explorer is used to get access token and page id of Facebook public pages. Dataset of the particular events (such as PIA flight crash PK-661 in Havelian and Pakistan Elections 2013) has been collected for the interval of one month. We have scraped all the posts along with the comments its related metadata from the Facebook news pages. The dataset is noisy and un-structured which needs pre-processing.

B. Data pre-processing

1) Data Transformation

Data pre-processing remove the un-necessary data because it degrades the system performance by making it difficult to classify the raw data. The Facebook data is pre-processed as follows:

- Expand the abbreviations.
- Extract the English text.
- Facebook content containing headlines with URLs and video links. All these links are removed as our concern is only to process the plain English text.
- Removal of such status messages having no comments

2) Expanding Abbreviations and Extraction of English Text

Public comments on the Facebook contain typos, misspelled words, unstructured and informal text. We designed an approach to tackle the typos by expanding the abbreviations to correct the short form of words and recognize the English text using dictionary-based approach. We created an abbreviation list of most commonly used abbreviations. We compared each word in a comment with the abbreviation list to extract full form of word and replaced it in the actual dataset. To separate English text, a dictionary-based approach is applied in which each word in the comment is checked in the English dictionary and extracted the English text in a separate file. For each comment, number of English words are counted and divided by the total number of words present in the comment to get its weight as presented in (1).

$$S_w = \frac{\text{Eng}_\text{w}}{\text{total}_\text{w}} * 100 \qquad (1)$$

Where, S_w indicates sentence weight, Eng_w represents number of English words in a comment and total_w represents total number of words in a comment. To detect English, threshold value is set as 80%. We check text if it contains 80% of English words then retain the row and if, no, delete the row and do not process it. Therefore, the comments with weight equal or above 80% are classified as English whereas those comments having weight below 80% are classified as belonging to non-English text. Repeat till file ends. After the completion of the whole process it generates file that contain only English Text. The flow of replacement of abbreviation with actual word is shown in Fig. 3.

Facebook data contains many abbreviations and short form of words having varying writing styles. We cannot apply text mining techniques directly to get better and acceptable results. The data need to be transformed into standard format before applying any text mining techniques. The abbreviation replacement can improved the performance of the system significantly.

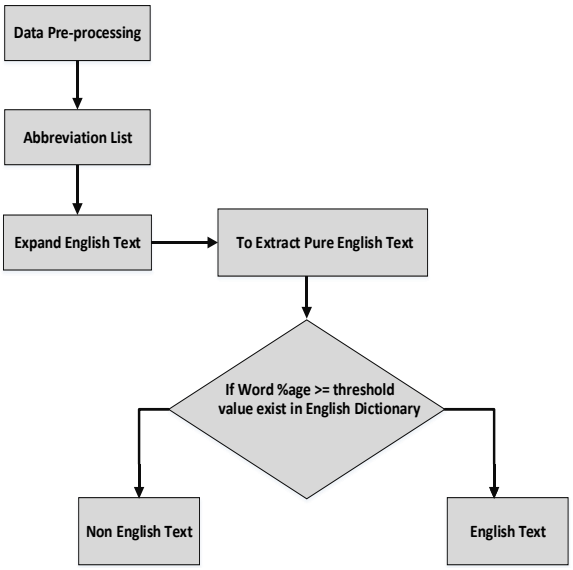

Fig. 3. Expanding abbreviations and extracting English text.

A list of most commonly used abbreviations is presented in Table I.

TABLE I. LIST OF ABBREVIATIONS WITH EXPANSION

Abbreviations	Expansion	Abbreviations	Expansion
Y	Why	Abt	About
V	Very	Plz	Please
K	Ok	thnx	Thanks
R	Are	Sry	Sorry
&	And	Sty	Stay
Ur	Your	Dept	department
w8	Wait	Ths	This
Fi9	Fine	B/W	between

C. Topic Modelling for Keyphrase *Extraction*

Topic Rank is an unsupervised, graph-based key phrase extraction method. In Social media, event detection is a prominent research topic. TopicRank is used to discover the topics that need to be manually analyzed during post-processing to select the best topic, describing about the events. It has become a challenge to obtain the event relevant posts, since posts may have event relevant terms but describing something other than the event such as a post containing the term "earthquake" could refer to an actual earthquake or to a conference on earthquakes. Fig. 4 presents the steps involved in keyphrases extraction from a document though TopicRank. Generally in a document, one noun phrase is sufficient to convey the topic. Therefore, some candidate keyphrases are redundant to represent the topic. Existing graph-based methods (Text Rank, Single Rank, etc.) do not take that fact into account. Candidate weighting, is assigned using a random walk algorithm. N-best selection, keyphrases contains the 10 highest scored candidates as (keyphrase, score) are extracted.

D. Assertive Detection

An assertion is a forceful statement of fact or belief. The assertion detection is used to get better understanding of intention and state of mind of the users behind event posts on Facebook. For example, a post "making a statement" can help us to track the assertions being made about events and can reveal a lot about the general attitude of users about that topic.

Automatic classification of the dialogue acts is a challenging task, traditionally a number of supervised methods are used to address these challenges; however, they require substantial manual time, effort and these depend on the availability of abundant training data.

We used an unsupervised approach, for classifying assertive dialogue acts of an event. We collected the analytical verbs from online[1] resource, a total of 76 words and we used 106 assertive verbs from the published[2] (Soroush Vosoughi, 2016). By using these verbs, assertive posts are extracted and afterward we scraped comments of assertive posts and detect inquiries from them.

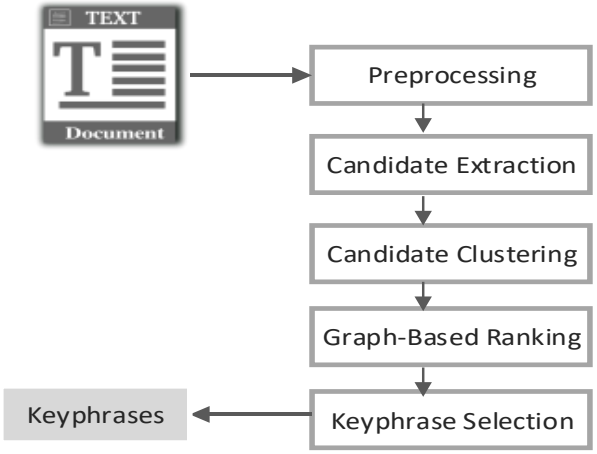

Fig. 4. Steps involved in Keyphrase Extraction through TopicRank.

E. Inquiry Comments Detection Model (ICDM)

We investigate the problem of inquiry comments detection in the textual environment; it involves detection of comments of assertive posts that contain questions (we call them "inquiry comments"). To detect inquiries from comments, a novel method is employed.

Inquiry Comments Detection Model has been implemented using Python. Algorithm employs two modules to classify sentences as inquiry comments which are useful in finding fact about some certain event propagating on the social media.

First module of our algorithm identifies a sentence as a question that satisfies following conditions:

- It ends with a question mark "?" and particularly such sentence does not contain any URL.
- It starts with an intransitive verb (like is, am, was, were, will, would, could, has, have, etc.)
- It starts with 1H-5W question words (How, What, Who, When, Why or Where).

Its second word is also checked, if it is not an intransitive verb, example: "where there is a will, there is a way." is not a question. But if it is "where are you going", the word "where" is followed by an intransitive verb, so, it is a question.

Second module of our algorithm is based on regular expression given in Table II.

TABLE II. PATTERNS USED TO FILTER INQUIRIES

Regular Expression	Type
wh[a]*t[?!][?1]*	Question
is (that|this|it) true	Question
Real ??|really??|unconfirmed	Question

F. Rumors Detection and Verification System

After calculating the ratio between English and Inquiry comments of a post, the ratio is compared with threshold to determine whether it is a rumor or not.

The threshold of a post (T_p) is:

$$\text{Threshold } (T_p) = \begin{cases} rumor\ if & \text{Tp} \geq 0.1 \\ non-rumor & \text{if Tp} < 0.1 \end{cases}$$

If $T_p \geq 0.1$ post is considered as a rumor.

If $T_p < 0.1$ post is non-rumor.

For reliable rumors detection and testing our proposed model, we collect a set of labeled and verified rumors/non-rumors dataset from Facebook snopes.com page. Snopes.com is a popular resource for debunking and validating rumors. It cites and aggregate trustworthy external sources (news or governmental organization) for verification of rumors, urban legends, documents Internet rumors and other questionable statements.

IV. RESULTS AND EVALUATION

A. Experimental Setup

The rule-based model has been designed for inquiry comments detection. We performed various experiments to measure the effectiveness of ICDM with traditionally designed machine learning models and to test it using labeled data.

In our experiments:

- **TP (True Positive)** are inquiry samples that are accurately classified inquiry samples.
- **TN (True Negative)** are non-inquiry samples that are accurately classified non-inquiry samples.
- **FP (False Positive)** samples are inquiry samples that are misclassified non-inquiry samples.
- **FN (False Negative)** samples are as non-inquiry

[1] http://msweinfurter.weebly.com/uploads/5/4/3/7/5437316/analytical_verbs.pdf
[2] http://soroush.mit.edu/publications/vosoughi_roy_speechact_icwsm2016.pdf

samples that are misclassified samples inquiry.

Aiming to measure the efficiency of our proposed approach**,** we have used well-known metrics: precision, recall, and *F*-measure. Precision represents the ratio of predicted positive samples that are real positives and is calculated by (2). On the other hand, recall is the ratio of true positive samples that were correctly predicted as such. Recall is calculated by equation 3. F-measure is the harmonic mean of precision and recall. F-measure is calculated by (4). Accuracy is the number of true positive and true negative samples out of total number of samples. We have calculated accuracy by using (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F} - \text{measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

B. Results

In result section, we have described the accuracy of our proposed framework of inquiry comments detection model (ICDM) for rumors detection and verification using labeled data from snopes.com.

1) Rumor Verification System

We retrieved from the Snopes.com two classifications of the rumors they have analyzed, the first is the reliability, which includes "true" and "false", but also a range of intermediate instances i.e. partly true, mixture, unproven, miscaptioned and legend. A rumor was annotated as true or false if trustworthy source snopes.com confirmed it as such. The rumor and non-rumor dataset contains 500 posts having more than 5,000 comments. The target variable labels each post to be true or false is represented by rumor or non-rumors.

TABLE III. RESULTS OF VERIFICATION SYSTEM USING ICDM AND MACHINE LEARNING METHODS

Techniques	Accuracy	Precision	Recall	F-measure
ICDM	0.70	0.89	0.77	0.82
k-NN	0.26	0.21	0.88	0.33
Naive Bayes	0.21	0.21	1.00	0.35

Table III shows that using labeled data, our approach obtained encouraging results with a precision of 0.89%, a recall of 0.77%, and *F*-measure of 0.82%.

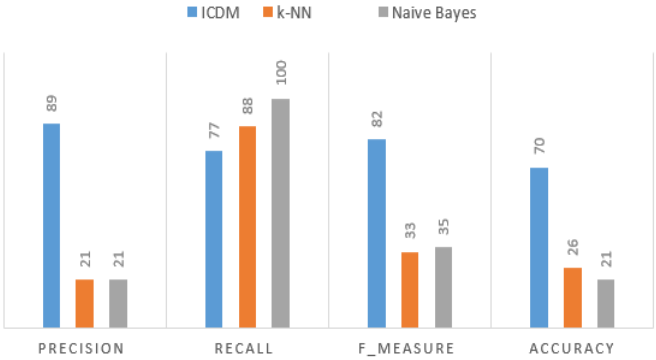

Fig. 5. Comparison between ICDM and machine learning methods.

Fig. 5 shows that on labeled data of 500 posts, our proposed ICDM approach obtained better results with a precision of 89%, a recall of 77%, and *F*-measure of 82% as compared to Machine learning approaches.

V. CONCLUSION

The current work explored the problem of rumors detection based on inquiry comments identification using textual content of social media especially Facebook. Data of Facebook is secure and inaccessible to access except the public pages. Therefore, we have searched for Facebook news pages and collected some event's data. The scraped data is unlabeled, huge in volume and mix of multiple languages. An expanded list of abbreviation is prepared to remove inconsistencies during the pre-processing phase. For detection of event relevant post, we used topic modeling technique such as TopicRank to discover and select best topic describing the event. Once the system is able to mark topics, assertive posts about that event are extracted and relevant comments were scraped. A rule-based approach is developed to extract inquiry comments from the assertive post comments. We verified our ICDM approach using labeled data from snopes.com and achieved better results as compared to machine learning-based approaches, with 89% precision, 77% recall, and 82% F-measure. We believed that the experimental findings from this study will be useful in real-world inquiry classification problems.

In the future, our focus will be to address multilingual content to avoid removal of other language content that reduces the data resulting in the loss of vital information to get better understanding of people judgment about the event on the social network and containing short and informal questions where 5W1H words or question mark are likely to be absent. Our focus will be to handle informal online languages to explore question characteristics and devise an automated method to detect interrogative sentences based on syntactic and lexical features.

REFERENCES

[1] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng, "Rumor Cascades," Icwsm, pp. 101–110, 2014.

[2] J. Kostka and R. Wattenhofer, "Word of Mouth : Rumor Dissemination in Social Networks," pp. 1–14.

[3] G. Goggles, "WHAT emtech MIT," no. August, 2012.

[4] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts," WWW '15 Proc. 24th Int. Conf. World Wide Web, pp. 1395–1405, 2015.

[5] S. E. Middleton, L. Middleton, and S. Modafferi, "Real-time crisis mapping of natural disasters using social media," IEEE Intell. Syst., vol. 29, no. 2, pp. 9–17, 2014.

[6] S. D. Agarwal, W. L. Bennett, C. N. Johnson, and S. Walker, "A model of crowd-enabled organization: Theory and methods for understanding the role of twitter in the occupy protests," Int. J. Commun., vol. 8, no. 1, pp. 646–672, 2014.

[7] N. Diakopoulos, M. De Choudhury, and M. Naaman, "Finding and assessing social media information sources in the context of journalism," Proc. 2012 ACM Annu. Conf. Hum. Factors Comput. Syst. - CHI '12, p. 2451, 2012.

[8] A. Zubiaga, H. Ji, and K. Knight, "Curating and contextualizing Twitter stories to assist with social newsgathering," Proc. 2013 Int. Conf. Intell. user interfaces - IUI '13, p. 213, 2013.

[9] P. Tolmie et al., "Supporting the Use of User Generated Content in Journalistic Practice," Proc. 2017 CHI Conf. Hum. Factors Comput. Syst. - CHI '17, pp. 3632–3644, 2017.

[10] J. Spangenberg, "News from the Crowd : Grassroots and Collaborative 3 . GRASSROOTS JOURNALISM - A NEW," Www, pp. 765–768, 2014.

[11] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis," Proc. First Work. Soc. Media Anal. - SOMA '10, pp. 71–79, 2010.

[12] E. Rosengren, D. R. Radev, Q. Mei, and A. Arbor, "Rumor has it : Identifying Misinformation in Microblogs," pp. 1589–1599, 2011.

[13] S. Hamidian and M. T. Diab, "Rumor Identification and Belief Investigation on Twitter," Acl, pp. 3–8, 2016.

[14] S. Hamidian and M. Diab, "Rumor Detection and Classification for Twitter Data," no. c, pp. 71–77, 2015.

[15] L. Tolos, A. Tagarev, and G. Georgiev, "An Analysis of Event-Agnostic Features for Rumour Classification in Twitter," pp. 151–158, 2012.

[16] A. Zubiaga, M. Liakata, and R. Procter, "Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media," 2016.

[17] R. McCreadie, C. Macdonald, and I. Ounis, "Crowdsourced Rumour Identification During Emergencies," Proc. 24th Int. Conf. World Wide Web - WWW '15 Companion, pp. 965–970, 2015.