

# Detecting Misinformation in Twitter using Text Analytics

**M. Sunitha**

PG Scholar

Department of Information Technology  
Rajalakshmi Engineering College  
Chennai, India

[sunitha.m.2014.mese@rajalakshmi.edu.in](mailto:sunitha.m.2014.mese@rajalakshmi.edu.in)

**R. Anto Arockia Rosaline**

Assistant Professor

Department of Information Technology  
Rajalakshmi Engineering College  
Chennai, India

[antoarockiarosaline.r@rajalakshmi.edu.in](mailto:antoarockiarosaline.r@rajalakshmi.edu.in)

**Abstract**— The Online Social networking sites has transformed the mode of communication between people and offered them to interact steadfast. Social networking sites enable the users to build, exchange the post, images and videos in an effective environment. Twitter is one such social networking site that enlarged in last few years. It is an online service which enable the user to tweet and retweet. It is commonly used to share realistic content such as news, events and credible information. These online services prompts or encourage the incredible post to propagate, which in turn lead to recognize the misinformation. This paper explores the detection of misinformation in twitter using credible level of the post and also to discern the source for incredible post. The proposed methodology allot the monitors to delineate about the post retain the credible level of the post. Each post are compared with the associated press tweets to identify the misinformation. This provides a quality of credible information to the user.

**Keywords:** *Twitter, Text Mining, Text analytics, Credibility, Misinformation, Detection.*

## I. INTRODUCTION

Twitter is an online service network that permits its users to post and view the 140 character content called “tweets”. It has been playing a prominent role in connecting people and assist users to learn about current events occurring worldwide. A name, term or subject that stated at significant rate are called ‘Trending Topic’. These trending topics is the best feature in twitter to find the users interest.

Trending topics became more familiar among the users because it prompts the post in a real time environment. It assist the user to know “What is happening?” in the world and what are the opinion of the event among the users. In twitter only authorized users can view and publish their tweets, whereas unauthorized user cannot post the tweets. These trending are created by the teenage to propagate the happening without interruption. Users of twitter can also tag their short message using “Hashtags”, with an anterior of ‘#’ symbol. The short post named “tweets” that are posted on network are not credible every time.

Twitter furnish a platform for the erroneous tweet to elevate or spike among the users. It holds more than 70 million tweets per day. To point the type of tweets it has been labeled as ‘Conversation, News and Misinformation.

The labeled tweets may vary in content but scatter over the network in a fraction of second. The rumor, misinformation, false news or fake content projected over the twitter reduce the credibility on those short messages. This form content indicates an anecdote or tale which spreads in twitter to make the people believe. The propagation of credible or incredible content need to stem before it impact users. The most familiar tweet in twitter which created a sensational moment is death of ‘Osama bin laden’. The fake news or incredible information plays a significant role during the sensational moments. Another famous rumor tweet spiked in twitter is about Boston Marathon bomb blast. The following are some tweets posted during that event.



Fig 1. Tweet posted during Boston Marathon.

In 2013, Boston marathon tweets spiked in twitter within 48 hours. The post which furnish an environment for propagation of misinformation as to be detected earlier. To provide a credible or a minimal misinformation spread the stemming method is undertaken to find that post. The technology which assist the researchers to spot the post within credible content is Text Mining and Text Analytics. As text mining emerged from data mining it comprises text categorization, entity extraction and clustering of text. It also cover the lexical analyzer in information retrieval technique. The quality of information is extracted using text mining.

Text analytics involves in the process of parsing the content and tagging the text. Sentiment analysis is carried out text analytics. The roles of text analytics in assistance with an online social network can be: analyzing the reviews to discover the sentiment or opinion of the user, classification of the content based on the keywords. To

provide a credible information text analytics is used in this paper. There are multiple and hybrid technique for discover the false post. This paper handles the trustee level of the user account. The trustee are the users or followers of an individual who outline the content they viewed on the twitter. Based on the outline of trustee the trustee level is maintained to detect incredible tweet or post. The prime goal is to detect the source for incredible tweet and make the online social service to provide nourished post about the events occurring over the world.

## II. TRUSTEE SELECTION METHOD

Twitter can be represented as directed graph to follow the path of the tweet propagating in the form of retweet. The credibility of the tweet is identified by collating the post with associated press. The associated press are the news accounts. Not all the post are compared. The keyword classifier discovers the query post which in parallel compared to associated press for clarifying the content credible level. Each time the credible level is reduced.

The trustee are individuals who are engaged to outline the credible of content. Trustee can be categorized the post into two forms: positive and negative. The credible level of the content will decrease simultaneously as the negative rate increases. When the credible level of a post crosses the threshold.

The post which spiked with misinformation will be detached from the online social network. The paper is concentrated on how to detect the propagation of post with minimum credible level and how to stem the source for post. Later the stemmed source will be detached from the network for a period of time. This paper will furnish the networking environment with more credible information to their users.

This paper is standardized as follows: Section III discuss on literature survey, Section IV provides the framework, , Section V experimental results, Section VI Implication and Section VII References.

## III. LITERATURE SURVEY

### A. Propagation of Misinformation

Aditi Gupta, Hemank Lamba et al. [1] worked on online social media which plays a vital role during real world events, especially during the crisis events. Malicious content is posted online during such events, which can result in damage, chaos and monetary loss in the offline world. They highlight the role of twitter in two major crisis events: Hurricane Sandy and Boston Marathon Bombings in spreading fake content about the events. They performed a characterization analysis, to understand the temporal, social reputation and influence patterns for the spread of such fake

information. Their results indicate that automated techniques can be used to distinguish the characteristics of fake information on twitter.

Mengdie Hu, John Stasko et al. [2] presented a study of microblogs in breaking and spreading news by analyzing how Osama Bin Laden's death leaked through twitter. They identified three groups of "elite users" who played key roles at different stages of the news circle. Certainty analysis shows that the people who broke the news were able to convince many twitter users. They analysed on user attention and links suggested on mass media. Those results confirms twitter's rising potential in news reporting and their method could be applied to study other cases of social media breaking news.

Fang Jin, Wei Wang et al. [3] describe a quantitative analysis of tweets during the Ebola crisis reveals that lies, half-truths, and rumors can spread just like true news. From the gathered tweets, they the removed stop-words for processing and constructs a word clouds for specific days. The SEIZ model is well suited for studying the rumor propagation, as it captures distinctions in how people respond to ideas. They define the notion of a response ratio that provides a relative measure of the population. They stated that the received news from real time social platforms need to be distinguished.

Balaji Vasan Srinivasan, Khushi Gupta et al. [4] aim is to solve the problem of stemming the flow of a negative campaign in a network by observing the parts of the network. Given a negative campaign and information about the status of its spread through a few candidate nodes, proposed algorithm estimates the information flow in the network and based on this evaluated flow, finds a set of nodes which would be instrumental in stemming the information flow. The proposed algorithm is tested on real-world networks and its functionalities are compared against existing works.

### B. Impact of Rumor in Twitter

Paul Resnick, Zhe Zhao et al. [5] proposed the Rumor Lens, a suite of interactive tools that are designed to help journalists identify new rumors on Twitter and assess the audiences that rumor and correction tweets have reached. The tools make efficient use of human labor to assess whether a rumor's content is interesting enough to warrant further exploration, to label tweets as spreading, correcting, or unrelated to the rumor, and to analyze the rumor visually.

Rumor Lens is a promising tool for combining human effort with computation behind the scenes to systematically detect new rumors in Twitter, retrieve almost all the tweets related to them, and interactively analyze how many people tweeted about them or were exposed to the rumor or a correction. It depends on human labor. The amount required,

however, is small enough that it seems plausible that journalists and enthusiasts about particular topics might voluntarily provide the labors. The paper is concentrated on detecting the post that propagates with minimum credible level and how to stem the source for post. Later the stemmed source will be detached from the network for a period of time.

Kate Starbird, Jim Maddock et al. [6] stated an exploratory research examines three rumors, later demonstrated to be false, that spreads over twitter in the aftermath of the bombings. Their findings suggest that corrections to the misinformation emerge are compared with the propagation of the misinformation. The similarities and differences they observe in the patterns of the misinformation and corrections contained within the stream over the days that followed the attacks suggest directions for possible research strategies to automatically detect misinformation.

### C. Information Credibility

Aditi Gupta and Ponnurangam Kumaraguru [7] described about the twitter has evolved from being a conversation or opinion sharing channel among friends into a platform to share and disseminate information about present events. Events in the real world build a tweets on twitter. Not all content posted on Twitter is trustworthy or useful in providing information about the event. Using regression analysis, they identified the important content and sourced based features, which can forecast the credibility of information in a tweet.

Prominent content based features were number of unique characters, promising words, pronouns, and emoticons in a tweet, and user based features like the number of followers and length of username. They adopted a supervised machine learning and relevance feedback approach using the above features, to rank tweets according to their credibility score. The performance of our ranking algorithm significantly enhanced when we applied re-ranking strategy. Results show that extraction of credible information from twitter can be automated with high confidence.

Barbara Poblete, Marcelo Mendoza et al. [8] analyzed the information credibility of news propagated through twitter, a popular microblogging service. This paper focuses on automatic methods for assessing the credibility of a given set of tweets. Specifically, they analyzed the microblog postings related to “Trending” topics, and classify them as credible or not credible, based on features extracted from them.

They used features from users posting and re-posting (“re-tweeting”) behaviour, from the text of the posts, and from citations to external sources. They evaluated a significant number of human assessments about the

credibility of items on a recent sample of twitter postings. Results shows that there are measurable differences in the way messages propagate, that can be used to categorize them automatically as credible or incredible, with precision and recall in the range of 70% to 80%.

John O’Donovan, Sujoy Sikdar et al. [9] argued for a methodology for making such studies more useful to the research community. First, the underlying ground truth values of credibility must be reliable. The specific constructs used to define credibility must be carefully defined. Secondly, the underlying network context must be quantified and documented. Newsworthiness of tweets can be a useful frame for specific applications, but it is not necessary for achieving reliable credibility ground truth measurements. Using content based method constructs classes of credible and not credible messages.

### D. Rumor Detection

Eunoo Seo, Prasant Mohapatra et al [10] described an approach for finding the rumor source and assessing the likelihood that a piece of information is in fact a rumor, in the absence of data provenance information. That model the social network as a directed graph, where vertices represent individuals and directed edges represent information flow (e.g., who follows whom on Twitter). A number of monitor nodes are injected into the network whose job is to report data they receive. This algorithm identifies rumors and their sources by observing which of the monitors received the given piece of information and which did not. And shown that, with a sufficient number of monitor nodes, it is possible to recognize most rumors and their sources with high accuracy.

### E. Epidemiology Approach

Segun M.Akinwumi [11] he carried out the work innovated mathematical epidemiology concepts to create a model used to identify critical social issues. A compartmental Susceptible Exposed Infected model was developed to investigate the flow of issues in social media. The basic reproduction number  $R_0$ , the number of secondary cases resulting from the introduction of an index case into an otherwise uninfected population, was derived for the model. Any social issue with  $R_0$  greater than one was defined to be a critical issue.

Fang Jin, Edward Dougherty et al. [12] described a Characterizing information diffusion on social platforms like Twitter enables us to understand the properties of underlying media and model communication patterns. As Twitter gains in popularity, it has also become a venue to broadcast rumors and misinformation. They used an epidemiological models to characterize information cascades in twitter resulting from

both news and rumors. Specifically, they used the SEIZ enhanced epidemic model that explicitly recognizes skeptic to characterize eight events across the world and spanning a range of event types. They demonstrate that our approach is accurate at capturing diffusion in these events. This approach can be fruitfully combined with other strategies that use content modelling and graph theoretic features to detect (and possibly disrupt) rumors.

#### IV. FRAME WORK

The mainstream framework, which is depicted and explained below. It describes how to detect the misinformation and detach the source from OSN. This framework is categorized into three: Creation, Propagation and Detection.

In Creation, an online social network is created with all the basic features of communication networks. This creation segment allows people to create their own account with the facilities to change the profile picture and background images. They can invite distinctions friends and accept their friend request. As an account is created the trustee level of the account is assigned. The people can post and view others post in the wall. To outline the post, the trustees can be added. In Propagation, the post shared by the user are walled in communication network. In Detection, each post is collated with associated press news. All the post are not misinformation. The post which reduces in credible will be detached for a period of time.

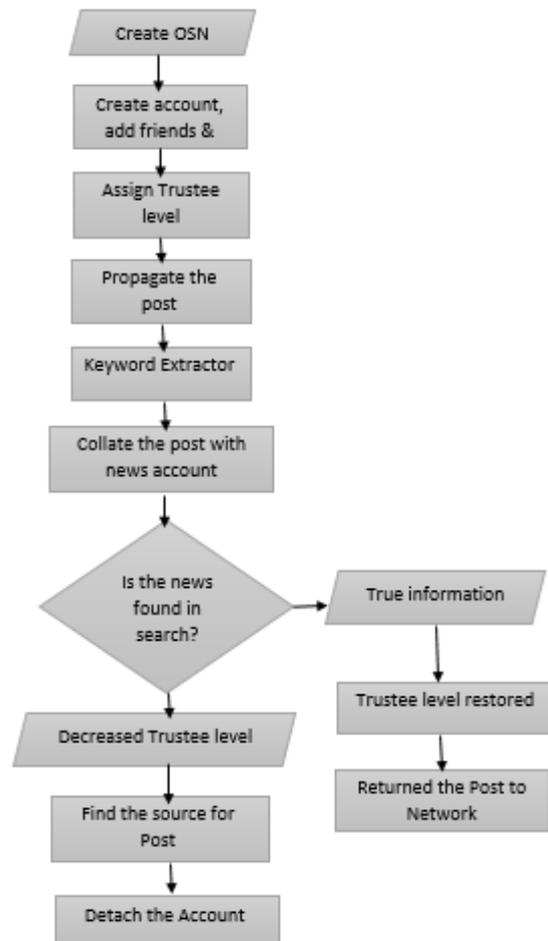
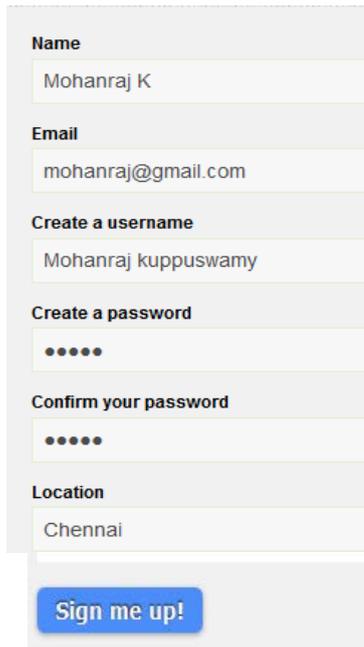


Fig. 2. Framework for Misinformation Detection.



Fig 3. Trustee Social Network Login

Fig 3 describes the login page for the social network users. The registered user can login with their user id and password. In turn new user need to register for accessing use the social network.



The form contains the following fields and buttons:

- Name:** Input field with "Mohanraj K".
- Email:** Input field with "mohanraj@gmail.com".
- Create a username:** Input field with "Mohanraj kuppuswamy".
- Create a password:** Password input field with 6 dots.
- Confirm your password:** Password input field with 6 dots.
- Location:** Input field with "Chennai".
- Sign me up!** A blue button at the bottom.

Fig 4. New User Sign Up.

Fig 4 shows the new user account creation. This network allows user to create their own network for communication.



Fig 5. User Account

Fig.5 shows the user account. In which user can change their wall and profile picture. The interaction tab is available in the left top of the network site which includes notification, logout and friends list. User can post their content in the wall and read others post.

## V. EXPERIMENTAL RESULTS

This paper is an extant project. The linguistic approach of propagating post is completed and moved towards the detection phase. The trustee level is assigned and traced. Currently the project is ongoing on detection of misinformation post search in social network.

## VI. IMPLICATION

This proposed system carried out for propagation implementation. The algorithm used to propagate post is trustee system. The prime goal is to detect the

misinformation to assure user to receive true news and information.

## VII. REFERENCES

- [1] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, Anupam Joshi : “ Analyzing and Measuring the Spread of Fake Content on Twitter during High Impact Events ”, 2014.
- [2] Mengdie Hu ,Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, Kwan-Liu Ma: “ Breaking News on Twitter”, 2015.
- [3] Fang Jin, Wei Wang, Liang Zhao, Edward Dougherty, Yang Cao, Chang-Tien Lu and Naren Ramakrishnan : “ Misinformation Propagation in the Age of Twitter ”, 2014.
- [4] Balaji Vasan Srinivasan, Akshay Kuma, Shubham Gupta, Khushi Gupta : “ Stemming the flow of information in a social network ” ,2014.
- [5] Paul Resnick, Samuel Carton, Souneil Park, Yuncheng Shen, Nicole Zeffe : “ Rumor Lens: A System for Analyzing the Impact of Rumors and Corrections in Social Media ”, 2014.
- [6] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, Robert M. Mason: “ Rumors, False Flags, and Digital Vigilantes : Misinformation on Twitter after the 2013 Boston Marathon Bombing ”, 2013.
- [7]Aditi Gupta, Ponnurangam Kumaraguru: “Credibility Ranking of Tweets during High Impact Event ”,2012.
- [8] Carlos Castillo, Marcelo Mendoza , Barbara Pobelte: “ Information Credibility on Twitter”,2011.
- [9] Sujoy Sikdar, Byungkyu Kang, John O’Donovan, Tobias H’ollerer, Sibel Adal: “Understanding Information Credibility on Twitter ”, 2013.
- [10] Eunsoo Seo, Prasant Mohapatrab and Tarek Abdelzaher: “ Identifying Rumors and Their Sources in Social Network ”,2012.
- [11] Segun M. Akinwumi : “ A Mathematical Epidemiology Approach for Identifying Critical Issues in Social Media ”, 2015.
- [12] Fang Jin, Wei Wang, Liang Zhao, Edward Dougherty, Yang Cao, Chang-Tien Lu and Naren Ramakrishnan : “ Misinformation Propagation in the Age of Twitter ”, 2014.