International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 13(2021) pp. 286-299

Submitted: 30 April, 2021; Accepted: 26 May, 2021; Publish: 8 Sep., 2021

Modeling Information Credibility in Time-Specific Online Social Media based on Structural and Attribute Properties

Md. Habibur Rahman[1], **Tabia Tanzin Prama**[2] **and Md Musfique Anwar**[3]

[1]Jahangirnagar University, Computer Science and Engineering Department,
Savar, Dhaka-1342, Bangladesh
habibur.stu2015@juniv.edu

[2]Jahangirnagar University, Computer Science and Engineering Department,
Savar, Dhaka-1342, Bangladesh
prama.stu2017@juniv.edu

[3]Jahangirnagar University, Computer Science and Engineering Department,
Savar, Dhaka-1342, Bangladesh
manwar@juniv.edu

Abstract: **Nowadays, millions of people actively participate in online social networks (OSNs) everyday to express and share their ideas, opinions, interests, feelings about various events, governmental policies, current economics, societal debates etc. Thus, these numerous user generated contents diffuse very rapidly in OSNs through different peers in various formats such as tweets, images, videos etc. Therefore, it is very crucial to extract credible information from these mass volume of social user-generated contents. Existing research works on measuring trustworthiness of any content in OSNs are typically deliberated on contents generated by the social users. However, such methodologies overlook the prospective temporality of users' interests as well as paid less attention to the structure of the social network. In this work, we propose an approach to measure the level of credibility of a piece of information in OSNs based on the social users' temporal topical interests and the structural properties of the underlying social network. The effectiveness of the proposed methodology is justified using three benchmark datasets.**

Keywords: Online social networks; User-generated content; Information credibility; Structural property.

I. Introduction

Since the beginning of Information Technology, communication among the general people has received a major breakthrough over time. In this modern era, we are heavily influenced in our daily life by one of the greatest inventions of recent times called Online Social Networks (OSNs). The reason people are so much indulged in using these networks are manifold but rapid messaging, interactions with peers digitally (with photos, reacts), video chat service etc. features can be mentioned at first place. Apart from these special features people also engage themselves heavily into these networks in order to receive news about various topics like politics, social affairs, national issues, sports etc because of hardly having the time to sit before traditional news medium like TV or newspaper. We have got a good amount of social media platform over the years. Twitter is one such network which got immense popularity as a micro-blogging platform that allows its users to write anything on any topic but no more than 280 characters (previously it was 140 characters). Any post is known as *tweets* on Twitter which may contain topic related URLs, audio, video, image or plain texts etc. Users can connect with each other using *follower-followee* relationships. Users also have the privilege to spread or share any content with their peers by the *retweet* feature. Restricting users to a no more than 280 characters post at a time may seem cumbersome, In fact, it paves the way for the users to put a second thought before posting anything and inspires people to provide succinct content all across the network. This is one of the reasons why Twitter is so popular. At the time of special or popular events, people express their thoughts, opinions etc. by retweeting (share) the content mentioning some special words related to the topic with # symbol like #blacklivesmatter or #MeToo known as *hashtags*. Finding anything on social media is like searching a needle in a haystack. So, a hashtag is used as a norm so that users can track all the updates and contents related to a specific incident or topic. Twitter generates a large pool of information at every moment dynamically and information takes very little time to propagate over the network. This makes it an ideal platform to create any sudden mass gathering or civil unrest in the shortest possible time like a wildfire. Twitter is one of the mediums that was heavily responsible for the mass

uprising movement called "Arab Spring"[13] in the middle-east countries. It also played a pivotal role in the 2009-10 Iranian[14] after election protest. So, the impact of Twitter as a social media or micro blogging service can not be denied at any point. Sometimes rumors or invalid information have been propagated which leads to social instability. In 2016, the US presidential election was heavily influenced by the "fake news" ecosystem established on Twitter. A study[15] performed on almost 10 million tweets from 700,000 accounts by the *Knight Foundation* states that more than 600 fraud news and conspiracy news agencies were associated with more than 6.6 million tweets. So it is necessary to validate the credibility of the information that is being shared or tweeted. But this doesn't always signify Twitter as a fake or unauthorized medium of content propagation. Things also happen in other ways. During the Fukushima[12] Daiichi power plant accident(2011) in Japan, a large number of tweets (almost 70%) under the #fukushima tag were found to be highly credible even though the tweets were synthesis derivative (tweets having third party information) in nature. Another interesting fact is that anonymity usually carries a negative impact on the information propagation. But it turned out that during this event, Japanese users who didn't have their location information publicly in their account actually shared less low-credible news than non-Japanese users. Tweets form the basis of social interactions in Twitter where a user is kept updated of the tweets of someone she is following. Each tweet is first represented as a tuple of raw data: *<user id, tweet content, time stamp>*. The *user id* is the Twitter identification number, the *tweet content* is the text of the tweet and the *time stamp* is the time when the tweet has been published on Twitter. In terms of social connectivity, Twitter allows a user to follow any number of other users without requiring approval. The following user becomes a *follower* and will receive all the tweets from the followed user. Twitter facilitates real-time propagation of information to a large group of users [7, 29, 34]. This makes it an ideal environment for the dissemination of breaking-news directly from the news source and/or geographical location of events.

The goal of this research work is to measure the credibility or trustworthiness of a tweet in Twitter from two different perspective, namely, from user level and topic level, respectively. Authors in [7, 27] defined credibility as the believability of information. Usually, the users in social media try to measure the credibility or trustworthiness of any content based on its source, accuracy, timeliness, objectivity etc. Earlier research works emphasized more on the properties of the contents such as presence of URLs, sentiment words, hashtags, total number of words etc. They also focused on the users' profile such as name, number of connections, locations etc. [1, 2]. However, the structural properties of OSNs didn't get much attention on these approaches. Usually, credibility or trustworthiness of an information or topic T_x is considered as the total number of times T_x is mentioned. We observe that most of the times, a topic T_x is discussed heavily in a group of well connected users. So, in order to measure the credibility of a topic, we emphasize more to a topic which is discussed by the number of unconnected (*unrelated*) users. Again, most of the existing research works ignore the temporal topical activeness or interests as we observe that users' degree of topical activeness vary widely over time.

We define two types of *credibility* in the context of a target topic of interest:

Definition 2. *Topic-Level Credibility*: The expected trustworthiness a particular topic considering it's impact in the social network and how the users show their interests on that topic.

Definition 1. *Tweet-Level Credibility*: A degree of believability that can be assigned to a tweet about a target topic, i.e.: an indication that the tweet contains believable information [5].

The contributions of our proposed model are as follows:

- Emphasize the structure of the underlying social network along with the features of the social media contents.
- Consider the impact of the number of the unconnected (i.e. unrelated) users for measuring the credibility of a topic.
- We perform experiments on three real datasets to show the effectiveness of the proposed model.

This paper is an extended version of a conference paper that appeared as [30]. The new contributions to this journal version are summarized as follows.

1. We emphasize more on users' recent activities by applying time-based forgetting factor on social stream. The reason behind this is that not all the actions performed by the users carry equal importance. Again, users' interests change over time widely.
2. We perform experiments on two different Twitter[1] datasets. The length of each tweet are very short (at most 140 characters) and often noisy. So, in order to improve the quality of data and the performance of the subsequent steps, we apply some advance pre-processing steps (such as filter spam words, convert slang words into their standard forms etc.) on the original tweets using natural language processing (NLP) tools (details are mentioned in Section 4).
3. In this extension work, we also incorporate hashtags along with the topics generated by the topic modeling approach (such as Twitter-LDA model) to observe their effects on information credibility.
4. Our proposed extension model is also capable to list candidate spammers at different time intervals for some given query topics.
5. To address all the above issues, we modified the existing algorithmic framework (proposed in conference paper version [30]) which is explained in more details in Section 4.
6. New experiments are conducted on different Twitter datatset as well as other social network like Flicker dataset.

[1] https://twitter.com/

The rest of the paper is organized as follows. Section 2 describes the relevant works on information credibity in OSNs. We introduce some relevant terms and the problem statement in Section 3. Our proposed approach for measuring topic and tweet/tag related credibility is described in Section 4. Section 5 covers the experimental evaluations on three real datasets along with discussions. Finally, we conclude the paper in Section 6.

II. Related Work

Earlier research works mainly concentrating on creating feature based models of credibility rely on user surveys. Castillo et al. [7] proposed a model that first checks whether a topic is newsworthy and then samples a collection of 10 tweets about that topic. Later, it surveys whether the set of tweets is credible or not. Mendoza et al. [28] also evaluates trust in news dissemination on Twitter, focusing on the Chilean earthquake of 2010.

Some research works also focused on the temporal contents of the social posts. Leskovec et al. [3] proposed a model that used the temporal properties of "memes" which are shared across the blogosphere. Kwak et al. found that headline news or persistent news were the most trending topics in Twitter [1]. Lianwei et al. [11] proposed a method based on adversarial networks and multi-task learning to capture differential credibility features for information credibility evaluation. Tao Ma et al. performed meta-analysis of perceived credibility concerns for user-generated-online-health information [32]. Daniel et al. [36] proposed a supervised learning approach where they considered various categorical features like sentimental words, length of the post, number of followers, number of tweets, URLs, hashtags etc. The focus of their study was to build an automatic model to measure the credibility level of an information.

Some recent works consider to apply machine learning methods to their appraoches. Iftene et al. [16] suggested a neural network based model to find the credibility of tweets and users in real time. Along with the neural network model they also integrated sentiment analysis. For each individual tweet, they observed various parameters like retweet numbers, favourites number, creation date, words number, relevant words ratio, characters number etc. On the other hand, for each user they considered features like 40 most recent tweets, location, description, geolocation, verification, account creation date, number of followers etc. Hassan et al. [17] came up with an automatic text mining approach to detect credible events on Twitter. They used a dataset of popular twitter events which was manually labeled by credibility ratings to build up a topic-term matrix. The resultant number of topic documents generated by the matrix was used as the feature for a supervised classifier later. It turned out that decision tree classifier outperformed other similar approaches by almost 82%. Alrubaian et al. [25] proposed an automated classification analysis system for the assessment of information credibility on Twitter where they discussed four components- a reputation-based model, a credibility assessment engine, a user expertise component, and a feature-ranking algorithm. In the measurement of the score, the features they extracted for the previously mentioned models are a user's sentiment history, how popular he/she is, the number of retweets or replies a tweet has, number of static or animated emoticons in a tweet, the number of duplication(retweeting same thing more than once) by a user etc. Gupta et al. [18] developed a web based system "TweetCred" to evaluate the credibility of a tweet in real-time. They incorporated more than 45 features such as number of unique characters, existence of stock symbol, colon symbol, happy smiley and sad smiley, number of seconds since the tweet, source of tweet (mobile or web), tweet contains 'via', tweet contains geo-coordinates, ratio of likes and dislikes in terms of a YouTube video. Considering these features a semi-supervised ranking algorithm is used to rate a user's tweet in the scale of 1(low) to 7(high). Xia et al. [19] performed a study focusing on the emergency situation. They suggested the idea of using a Twitter Monitor online system. The whole system is composed of two phases. At first, identifying the emergency situation using the monitor model and K-means algorithm. Secondly, determining the credibility with the help of Bayesian Network. Whether user is verified, author's registered age, time of the tweets been cited, total number of tweets in a day, time interval of last 2 tweets, total time spent in Twitter etc are the attributes that have been taken into account in this approach under four categorical features- Author-based, Topic-based, Content-based and Diffusion based. Malith et al. [20] suggested a user reputation-based model using a K-means clustering algorithm to determine the source credibility of Twitter content. They also integrated additional two methods namely- sentiment analysis and news category analysis. Introduction of the agreement score of a particular user is one of the noteworthy contributions of this paper. Thandar et al. [21] figured out a way to measure the credibility of the opinions expressed in Twitter based on the user's background knowledge. Firstly, polarity of the tweets are classified using the support vector machine algorithm. Then features like user's bio, topic related ratio of a user, user's tweet-retweet ratio for a given topic, List feature etc. are examined to measure the expert score. Sabbeh et al. [22] proposed a machine learning based methodology to determine the credibility especially for the Arabic news on Twitter. Various classification algorithms such as Naive Bayes, Support Vector Machine and Decision Tree were applied to four main modules. The properties they observed for this model are whether a user is verified or not in Twitter, user's location near the event, user contains URL in the bio or not, activeness of the account, polarity of user's comment, content has image or URL, real user name etc. Canini et al. [31] introduced a potential algorithm for automatically measuring the trustworthiness of contents based on a topic on Twitter. The main focus was to combine the topic-based content with network-based structure. They identified the users associated with the searched topic first and then developed a ranked list for those relevant users for a topic which helped them identify potential users to follow. The comparison of the rankings provided by the proposed algorithm and a commercial site showed the efficacy of the system.

However, the common aspect of the above methods is that they paid less attention to both the structural properties of the social network as well as the temporal effect of users' topical interests. This research investigates how all these parameters (temporal topical interests, structural properties etc.) can be

employed in order to improve the quality in measuring the credibility information in OSNs.

III. Problem Statement

We now define some basic concepts before formally introducing our problem.

Online Social Graph: An online social graph is denoted as *G = (N, E, $\mathcal{T}$)*, where *N* is the set of users (representing nodes in G), *E* is the set of links between the nodes (for example, the friendship relations in Facebook, the follower, followee relations in Twitter) and $\mathcal{T} = \{T_1, T_2..., T_m\}$ is the set of topics associated with nodes in N.

Topic: A topic is a distribution over words, i.e., it contains most representative words for that topic [29, 35]. For example, *genetics* topic has words like chromosome, DNA, gene, mutation etc. about genetics.

Action in Twitter: An action refers to an activity in Twitter that a user n_i performs at a time point t_x such as posting a tweet or reply to an existing tweet about certain Topic T_j. This activity is recorded as a tuple like $\langle n_i, T_j, t_x \rangle$.

Time-Based Forgetting Factor: Generally, users' recent activities carry more importance compare with past activities. Hence, we emphasize greater importance to user's most recent activities by a measure called *recency score*, denoted by μ. Its value is computed by an exponential time-decay function in Equation 14, which assigns lower importance to user's older activities, as they are less likely to match the user's current interest. The parameter a is an external factor to control the speed of decay and $age_{\langle n_i, T_j, t_x \rangle}$ denotes the amount of time passed since the activity occurred.

$$\mu_{\langle n_i, T_j, t_x \rangle} = exp(-a \times age_{\langle n_i, T_j, t_x \rangle}) \tag{1}$$

Activity Stream: The collection of all the actions performed by social users in known as activity stream which is denoted as S.

Topical Expertise Score: For each social user $n_i \in N$, we compute her topical expertise score (denoted by σ) towards a given topic T_j, using Equations 4 and 5. There are two factors related to topical expertise of a user n_i. The first factor $f_1(n_i, T_j)$ is the user n_i's interest towards a topic T_j.

$$f_1(n_i, T_j) = \frac{\sum \mu_{\langle n_i, T_j, t_x \rangle}}{|ACTS(n_i, *)|} \tag{2}$$

where, $ACTS(n_i, *)$ denotes the set of all actions on any topic(s) performed by user n_i.

The second factor $f_2(n_i, T_j)$ is the participation of user n_i related to T_j in comparison to the most active participant user related to T_j in the social network G.

$$f_2(n_i, T_j) = \frac{\sum \mu_{\langle n_i, T_j, t_x \rangle}}{max_{u_z \in U^Q} |ACTS(u_z, T_j|} \tag{3}$$

Then, the topical expertise score (denoted as σ) of n_i related to T_j is

$$\Gamma_{(n_i, T_j)} = (\alpha \times f_1(n_i, T_j)) + (1 - \alpha) \times f_2(n_i, T_j) \tag{4}$$

$$\sigma_{(n_i, T_j)} = \frac{\Gamma_{(n_i, T_j)}}{max_{u_z \in N} \{\Gamma_{(n_i, T_j)}\}} \tag{5}$$

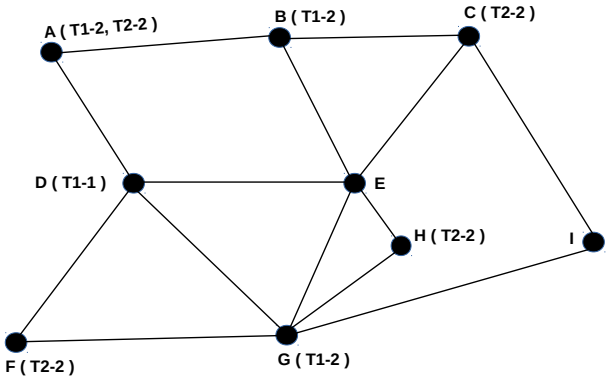

Figure. 1: Sample social graph with users' activities on topic $T1$ and $T2$

The parameter $\alpha \in [0, 1]$ in Equation 4 balances the two factors of $f_1(n_i, T_j)$ and $f_2(n_i, T_j)$.

We measure the trustworthiness of a topic considering the structural properties of G. Our goal is to give more emphasize for trustworthiness on topics that are discussed heavily by unconnected nodes using Equation 6 and 7.

$$f_{uncon}(T_x) = \sum_{n_i \varepsilon U} C(n_i, T_x) \sum_{n_k \in U \setminus n_i \cup N_i} C(n_k, T_x) \tag{6}$$

where $C(n_i, T_x)$ denotes the number of mentions in the form of (n_i, T_x) and N_i indicates neighbors of user n_i. So $f_{uncon}(T_x)$ counts the total number of mentions of T_x by the number of unconnected pairs which captures the notion of widespread interest in G and thus reflects the trustworthiness of a topic T_x.

$$\lambda_{uncon}(T_x) = \frac{f_{uncon}(T_x)}{max_{T_z \in \mathcal{T}} \{f_{uncon}(T_z)\}} \tag{7}$$

The value of $\lambda_{uncon}(T_x)$ indicates the *trustworthiness* score of topic T_x. Similarly, we measured the trustworthiness score from the perspective of connected neighbors (as indicated by $\lambda_{con}(T_x)$) using Equation 8 and 9.

$$f_{con}(T_x) = \sum_{n_i \varepsilon U} C(n_i, T_x) \sum_{n_k \in N_i} C(n_k, T_x) \tag{8}$$

$$\lambda_{con}(T_x) = \frac{f_{con}(T_x)}{max_{T_z \in \mathcal{T}} \{f_{con}(T_z)\}} \tag{9}$$

Then, for each tweet in testing phase,

$$\pi_{uncon}(t_k) = \beta \times \sigma_{(n_i, T_j)} + (1 - \beta) \times \lambda_{uncon}(T_x) \tag{10}$$

$$\pi_{con}(t_k) = \beta \times \sigma_{(n_i, T_j)} + (1 - \beta) \times \lambda_{con}(T_x) \tag{11}$$

where, $\beta \in [1, 10]$ in Equation 10 and 13 is a weighting factor that balances the topical expertise score ($\sigma_{(n_i, T_j)}$) and trustworthiness score of topic T_x.

Let us explain the concept of measuring the trustworthiness of a topic from the perspective of connected and unconnected (unrelated) neighbors with an example as shown

in Figure 1. The sample social graph contains 9 users and two topics of $T1$ and $T2$. The topic and the numbers (in a bracket) present after the name of a node indicate the node's actions related to a specific topic. For example, user A performed two actions on both topic $T1$ and $T2$, whereas, user D performed one action on topic $T1$. Now, if we want to measure the trustworthiness of topic $T1$ from connected neighbors perspective, we see that users A has two neighbors (B and D), D has two neighbors (A and G), B and G has one neighbor of A and D, respectively. So, the trustworthiness score of topic $T1$ (according to Equation 8) is: 16 $((2 \times (2+1)) + (1 \times (2+2)) + (2 \times 2) + (2 \times 1))$. Again, from unconnected neighbors perspective, A and B, both have one unrelated neighbor i.e. G. Similarly, G has two unconnected neighbors (A and B) and D has one unrelated neighbor i.e. B. So, the trustworthiness score of $T1$ from unconnected neighbors perspective (according to Equation 6) is 18 $((2\times2)+(2\times2)+(2\times(2+2))+(1\times2))$. Similarly, the trustworthiness score of topic $T2$ from connected neighbors perspective is 0 as no user, who are active on $T2$ has connected neighbor. Again, the trustworthiness score of $T2$ from unconnected neighbors perspective is 48 $((2\times(2+2+2))+(2\times(2+2+2))+(2\times(2+2+2))+(2\times(2+2+2)))$. We can see that topic $T2$ got more emphasize from unconnected neighbors perspective as all the active users on $T2$ are unrelated and this actually indicates high credibility of $T2$.

Our proposed model also lists possible spammers in the social network. For finding such a list for a certain topic, we consider two factors: i) Those who performed more activities (such as posting tweets etc.) than the average number of activities performed by all the users and ii) the collection of keywords in all those activities are mostly similar and limited in numbers. Equations 12 and 13 formulates the above mentioned two factors.

$$|ACTS(n_i, T_j)| = AVG_{ACTS(n_i,T_j)} + \rho \qquad (12)$$

where, $ACTS(n_i, T_j)$ indicates the number of actions performed by user n_i on topic T_j, $AVG_{ACTS(n_i,T_j)}$ denotes the average number of actions performed by all the users and ρ indicates the percentage of actions compare with $AVG_{ACTS(n_i,T_j)}$.

$$W_{|ACTS(n_i,T_j)|} = |ACTS(n_i, T_j)| \times b \qquad (13)$$

where, $W_{|ACTS(n_i,T_j)|}$ indicates the number of keywords in all the activities related to T_j performed by n_i. The parameter b controls threshold in the keyword numbers to identify possible spammer candidate.

IV. Modeling Topic Specific Tweet Credibility

There are three major stages in our proposed model to measure the credibility of a testing tweet as presented in Figure 2. Firstly, the pre-processing is performed to remove irrelevant data from the social stream. Secondly, we apply topic modeling method on the cleaned data to identify the latent topics. Finally, we develop and apply our proposed algorithm on the processed social activity streams to measure the credibility level of each testing tweet.

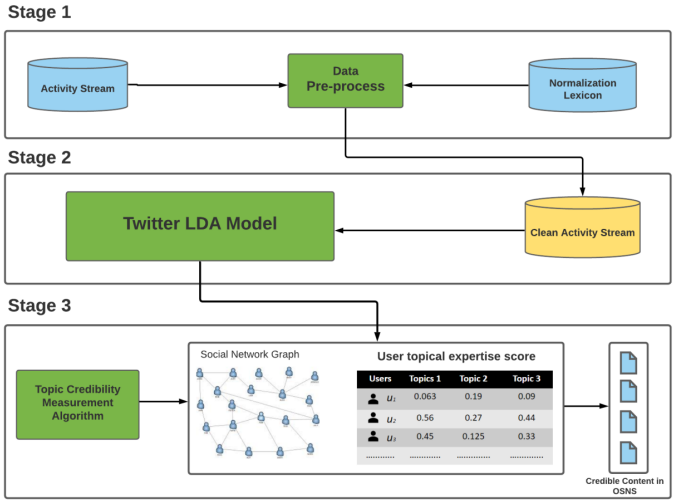

Figure. 2: Overview of the proposed system architecture

A. *Data Pre-processing for Topic Detection*

In this section, we present the first stage of our proposed approach which is the pre-processing of the social activity streams.

1) Filtering Spam/Noisy Tweets.

Twitter users often publish tweets which contain many unrelated words known as spam/noisy tweets. As a result, we need to filter these noisy tweets from the dataset to perform better analysis. We first check the structure of the tweets, and filter out tweets that have less than 4 text tokens as these kind of tweets usually do not carry any meaningful topic-like content. Next, we remove spam words from the tweets, using a "black list"[2] of words to filter noisy tweets.

2) Twitter Slang Words and Abbreviations Conversion.

Many tweets often contain grammatically incorrect sentence structures with misspellings, and words with non-standard forms such as informal abbreviations, phonetic substitutions. Table IV-A.2 shows example of Twitter slang words. We performed normalisation of the tweets through direct substitution of lexical variants with their standard forms with a normalisation lexicon proposed by Han et al. [33] and Internet slang dictionary[3].

English Words	Slang Words
tomorrow	toommorrow, twomorow, tomorow, tomorrowww, tmrw
network	netwrk, ntwork, networ, netwerk, netwrks, netowrk
took	toook, tok, tookk, tokk, tookk
good	gooda, goodi, ggoood, gud, gooodd, goooddd
weekend	wknd, weekknd, weknd, wekend

Table 1: Example of slang words in Twitter

[2]https://github.com/splorp/wordpress-comment-blacklist/blob/master/blacklist.txt

[3]https://www.noslang.com/

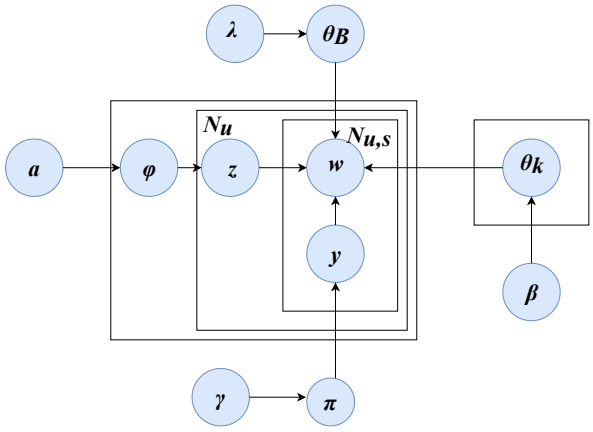

Figure. 3: Graphical Representation of Twitter-LDA Model

3) Linguistic Processing of the Tweets.

In order to improve the quality of our tweet corpus and the performance of the subsequent steps, we need to clean the tweets set using a linguistic process and use of a normalisation lexicon. In the linguistic processing phase, at first all the tweets are converted into lower case. Then we apply WordNet lemmatiser from NLTK [23] to lemmatise the contents of the tweets. Then, we removed stop words, punctuations, numbers, URLs. We also remove the keyword RT(which is used for retweet). The remaining words are converted into a seed word (stemming word) for example: plays, playing, etc.->play by using Lucene 4.9.0 Java API[4].

B. Detection of Topics from Social Activity Streams

Most of the times, the topic of a tweet is not mentioned explicitly by the Twitter user. The users often use hashtags to mention the topics of the tweets. For example, the different hashtags that are used to mention the election held in USA in 2012 are #USElection, #election2012, #Obama etc. Use of hashtag is optional and there is no specific standard rules of using hashtag to mention a topic. As a result, it is difficult for any system to correctly extract topics from hashtags.

1) Topic Discovery based on Twitter-LDA Model.

Topic modelling is the task of identifying which underlying concepts are discussed within a collection of documents, and determining which topics each document is addressing. Conventional topic models such as Latent Dirichlet allocation (LDA) [6] and probabilistic Latent Semantic Analysis (pLSA) [9] learn topics from document-level word cooccurrences by modeling each document as a mixture of topics, whose inference suffers from the sparsity of word cooccurrence patterns in short texts. Hence, we apply topic modeling approach such as Twitter Latent Dirichlet Allocation (T-LDA) [4], a variant of Latent Dirichlet allocation (LDA) [6], to detect the latent topics from the tweets.

The formulation of T-LDA as shown in Figure. 3 is as follows:

- Topical interest of each user is indicated by ϕ_i which represents the distribution over N topics.

[4]https://lucene.apache.org

Topic 1	Topic 2	Topic 3	Topic 4
court	Father	business	twitter
democracy	wishes	product	message
protest	happy	money	followers
election	dad	credit	photo
rescue	man	benefit	search

Table 2: Sample Word Topic Distribution in T-LDA Model

- The background word distribution (θ_B) and topic word distribution (θ_K) are analyzed to imply each word by N number of topics.

- A word can be used to indicate a background word or a topic word determined by the latent value of y.

Table IV-B.1 shows sample word topic distribution for the tweets from CRAWL dataset (we will discuss more in the Experimental section) using Twitter-LDA model. We can see that Topic 1 focuses on *politics*, Topic 2 is related with *daily/social life*, Topic 3 concentrates on *business and marketing* while Topic 4 focuses on *social media*.

C. Topical Credibility Measurement Algorithm

We develop an algorithmic framework to measure the credibility of the topics and the testing tweets.

Algorithm 1 `Information Credibility Measurement Algorithm`

Input: $G = (N, E, \mathcal{T}), S, \alpha, \beta, \rho, a, b$

Output: Credibility scores at topic ($\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$) and tweet/action ($\pi_{uncon}(t_k)$ and $\pi_{con}(t_k)$) level and list of candidate spammers Φ_{T_x}

1: **for** each action $\langle n_i, T_j, t_x \rangle \in S$ **do**
2: compute $\mu_{\langle n_i, T_j, t_x \rangle}$
3: **for** each $n_i \in N$ **do**
4: **for** each topic $T_j \in \mathcal{T}$ **do**
5: compute $\sigma_{(n_i, T_j)}$
6: **if** $\sigma_{(n_i, T_j)} \geq 0.25$ **then**
7: $P_{T_j}.add(n_i)$
8: **for** each action $T_x \in \mathcal{T}$ **do**
9: compute $\lambda_{uncon}(T_x)$
10: compute $\lambda_{con}(T_x)$
11: compute $AVG_{ACTS(n_i, T_j)}$
12: list Φ_{T_x}
13: **for** each activity $t_k \in \Phi_t$ **do**
14: compute $\pi_{uncon}(t_k)$
15: compute $\pi_{con}(t_k)$
16: Output $\lambda_{uncon}(T_x)$, $\lambda_{con}(T_x)$, $\pi_{uncon}(t_k)$, $\pi_{con}(t_k)$ and Φ_{T_x}

Algorithm overview. The algorithm, called `Information Credibility Measurement Algorithm`, measures the trustworthiness of the topics discussed in OSNs as well as the credibility of a particular tweet. It first computes recency scores ($\mu_{\langle n_i, T_j, t_x \rangle}$) of each actions performed by the social users (line 1-2). Then, it computes the topical expertise score ($\sigma_{(n_i, T_j)}$) for each user in the social graph G for each specific topic T_j (line 3-5) and added those users whose ($\sigma_{(n_i, T_j)} > 0.25$) in a priority list P_{T_j} of a topic T_j (line 6-7).

Next, it measures the trustworthiness of each topic T_x from the perspective of connected and unconnected neighbors as well as the average number of actions for each T_x (line 8-11). It also lists the candidate spammers for each T_x (line 12). Similarly, the algorithm computes the credibility level of each testing activity t_k from the perspective of connected and unconnected neighbors (line 13-15). Finally, it outputs the credibility scores in both topic and action level of all the testing actions (line 16).

V. Experimental Results

All experiments are performed on an Intel(R) Core(TM) i7-6600U 2.6 GHz Windows 7 PC with 8 GB RAM. We conducted the experiment on two Twitter datasets and one Flickr dataset. In all cases, we set the value of α to 0.5 .

A. Experimental Datasets

Twitter datasets. We conduct our experiments on two Twitter datasets: a small dataset CRAWL [2] and one large scale datasets - SNAP[5]. In the CRAWL dataset, we consider the user tweets from February, 2012 to April, 2012. SNAP contains 467 million Twitter posts from 20 million users from June 1, 2009 to December 31, 2009. We choose connected 1,00,000 users and consider their tweets from June 11, 2009 to July 12, 2009. We apply *Twitter-LDA (T-LDA)* [4] model to extract 50 latent topics in CRAWL and 30 topics in SNAP dataset.

Flickr dataset. Flickr[6] is a popular platform for image and video hosting service where social users can upload photos of nature, portraits, festivals, landscape, architecture and so on. Users have the option to categorize these images by adding one or more tags with those images. It is not necessary for users in Flickr to create an account to see existing photos and videos, but registration to an account is mandatory to upload their own photos. Registered users can add another Flickr user into their contact lists (which is considered as social connection/link between the users). We choose 50 most frequent tags of that are associated in the photos. Table 3 shows the statistics of our experimental data.

Dataset	# of Nodes	# of Edges	# of Topics	# of Activities
CRAWL	125	34,376	10	2,500
SNAP	3900	1,65,963	15	10,000
Flickr	312	4,260	8	3,386

Table 3: Statistics of the dataset used in the experiment

B. Information Credibility in CRAWl Dataset

Below we present both topic and tweet level trustworthiness for CRAWL dataset.

1) Result on Topic Level Trustworthiness.

Table V-B.1 shows the trustworthiness scores (i.e. $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$) for the 10 topics that are used in the experiment. In this work, we consider a topic has different trustworthiness level from connected and unconnected neighbors perspective if the difference between the scores of $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$ is greater than 0.15 . We can see that some of the topics such as *home loan, media, aggression, corporate* and *family* have similar trustworthiness scores both from connected and unconnected neighbors perspective whereas, the topics such as *environment, daily life, entertainment, politics* and *travel* (marked as yellow color) have different trustworthiness scores considering the effect of with and without direct neighbors. We can see that half of the topics discussed by the social users have different level of credibility from different point of views.

Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$
Home Loan	0.41	0.39	Corporate	0.71	0.60
Media	0.65	0.57	Entertainment	1.00	0.81
Environment	0.62	0.45	Family	0.65	0.57
Aggression	0.69	0.76	Politics	0.84	1.00
Daily Life	0.93	0.78	Travel	0.69	0.31

Table 4: Statistics of Trustworthiness Scores of the Topics in CRAWL dataset

2) Result on Tweet Level Credibility

We choose 15 tweets to measure their credibility level both from connected and unconnected neighbors perspective (i.e. $\pi_{uncon}(t_k)$ and $\pi_{con}(t_k)$) and consider two different values of the weighting parameter β (balances the topical expertise score i.e. $\sigma_{(n_i,T_j)}$ and trustworthiness score of topic T_x) of $\beta = 3$ and $\beta = 4$. Table V-B.2 shows the performance of the proposed algorithm in measuring tweets level credibility in CRAWL dataset. We asked two annotators two level the credibility of each of the 15 testing tweet from a score between 1 to 10 (indicated in 3rd column of Table V-B.2). We can see that, most of the cases, credibility scores (bold values) of the tweets considering unrelated (i.e. unconnected) neighbors matched closely to the credibility scores defined by the annotators. Therefore, considering unconnected neighbors to measure the credibility level of the tweets can be used to capture the notion of trustworthiness. In some cases (tweet number 5 and 13), there are significant gaps between the annotated credibility scores and the $\pi_{uncon}(t_k)$. We find that such cases usually happen for a *retweet* as it doesn't indicate the actual topical interest of the user who *retweet* a tweet posted by other social user. Again, in few cases (tweet number 14), the gaps between the annotated credibility scores to the $\pi_{uncon}(t_k)$ and the $\pi_{con}(t_k)$ (marked in green color) are sometimes large. That is due to the user's (who posted or retweeted) low level of topical interest on a topic which is highly popular among other users.

C. Information Credibility in SNAP Dataset

Below we present both topic and tweet level trustworthiness for SNAP dataset.

1) Result on Topic Level Credibility.

Table V-C.1 shows the trustworthiness scores (i.e. $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$) for the 14 topics that are used in the experiment. Similar to CRAWL, we consider a topic has different trustworthiness

[5]http://snap.stanford.edu/data/twitter7.html

[6]https://www.flickr.com/explore

#	Tweet	Topic	Annot. Credibility	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	$\sigma_{(n_i,T_x)}$	$\pi_{uncon}(t_k)$ ($\beta=3$)	$\pi_{con}(t_k)$ ($\beta=3$)	$\pi_{uncon}(t_k)$ ($\beta=4$)	$\pi_{con}(t_k)$ ($\beta=4$)
1	"My God have mercy on my enemies, because I w...	Politics	6.5	0.84	1.00	0.45 (0.14)	7.20 **(6.29)**	8.34 (7.43)	**6.81** (5.60)	7.79 (6.57)
2	puppies and rainbows make people happy, softwa...	Entertainment	8.0	1.00	0.81	0.07 (0.33)	7.21 **(7.99)**	5.91 (6.69)	6.28 **(7.33)**	5.16 (6.23)
3	it's starting to get really hard for me to not...	Corporate	5.0	0.71	0.60	0.37 (0.09)	6.06 **(5.20)**	5.33 (4.48)	**5.73** (4.59)	5.10 (3.97)
4	"@ThrillMeDead Brazilian Steakhouse or Sushi?"...	Daily Life	7.0	0.93	0.78	0.04 (0.11)	6.63 **(6.84)**	5.56 (5.76)	5.74 **(6.02)**	4.83 (5.09)
5	Loved meeting! RT @Cmtalcott: Meeting w @equal...	Daily Life	5.0	0.93	0.78	0.26(0.67)	7.29 (8.51)	**6.21** (7.43)	6.63 (8.24)	**5.70** (7.32)
6	Obama Releases $147M in Aid to Palestinians ht...	Media	5.0	0.65	0.57	0.02 (0.07)	**5.14** (4.71)	4.64 (4.22)	**4.71** (4.13)	4.28 (3.71)
7	My movie review of #TheRaven may be slightly o...	Environment	6.0	0.63	0.45	0.05 (0.06)	**4.53 (4.56)**	3.33 (3.36)	**3.95 (3.99)**	2.92 (2.97)
8	RT @TheOnion: "I'll teach you the hot licks an...	Family	5.5	0.64	0.57	0.05 (0.12)	**4.67 (4.88)**	4.18 (4.39)	**4.08 (4.36)**	3.65 (3.94)
9	I linked earlier to a @NYTimes piece about thi...	Travel	7.0	0.69	0.31	0.60 (0.22)	**6.71** (5.56)	3.96 (2.81)	**6.61** (5.08)	4.26 (2.73)
10	With Rep Winkler flipping to yes on the eve of...	Media	5.5	0.65	0.57	0.16 (0.13)	**5.01** (4.89)	4.50 (4.39)	**4.52 (4.37)**	4.09 (3.94)
11	RT @Moparpalooza: Here is the flyer for @Mopar...	Aggression	5.0	0.69	0.76	0.31 (0.18)	**5.79 (5.38)**	6.30 (5.89)	**5.42** (4.87)	5.86 (5.31)
12	So I'll probably just spend the next few days ...	Aggression	5.0	0.69	0.76	0.16 (0.18)	**5.31 (5.38)**	5.83 (5.89)	4.79 (4.87)	**5.23 (5.31)**
13	RT @exposeliberals: Obama TSA forces targets 7...	Politics	8.0	0.84	1.00	0.22 (0.21)	6.61 (6.51)	**7.65 (7.64)**	5.89 (5.88)	**6.86 (6.86)**
14	Getting ready to deliver my keynote speech to ...	Corporate	8.0	0.71	0.60	0.10 (0.18)	**5.24 (5.48)**	4.51 (4.75)	**4.64 (4.95)**	4.02 (4.33)
15	RT @eli_rubel: Speak with enthusiasm. Otherwis...	Politics	7.0	0.84	1.00	0.08 (0.43)	6.09 **(7.15)**	**7.23** (8.29)	5.34 (6.74)	6.31 (7.71)

Table 5: Performance of the proposed algorithm in measuring tweets level credibility in CRAWL dataset (in all cases, $\alpha = 0.5$, the values within brackets indicate the trustworthiness scores without considering temporal factor i.e. according to the method proposed in [30])

Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$
Death	0.21	0.19	Rivalry	0.23	0.20
Random	0.69	0.73	Entertainment	0.77	0.75
Iran	0.96	0.99	Music	0.98	0.99
Persian	1.0	1.0	Internet	0.23	.20
Daily Life	0.26	0.20	Compassion	0.32	0.22
Spanish	0.33	0.22	Social Network	0.36	0.23
Corporate	0.42	0.25	Technology	0.46	0.25

Table 6: Statistics of Trustworthiness Scores of the Topics in SNAP dataset

level from connected and unconnected neighbors perspective if the difference between the scores of $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$ is greater than 0.15 . We can see most of the topics (except *corporate* and *technology*, marked as yellow color) have similar trustworthiness scores both from connected and unconnected neighbors perspective. The reason is that the average number of connections (neighbors) among the users are very less in SNAP (average connections is around 42) compare with that of CRAWL (average connections is around 275). As a result, the impact of connected neighbors have very little contribution in relation with trustworthiness of the topics.

2) Result on Tweet Level Credibility.

Similar to CRAWL dataset, we choose 15 different tweets in SNAP dataset to measure their credibility level both from connected and unconnected neighbors perspective (i.e. $\pi_{uncon}(t_k)$ and $\pi_{con}(t_k)$). Table V-C.2 shows the performance of the proposed algorithm in measuring tweets level credibility in SNAP dataset. We can see that the credibility scores (bold values) of the tweets considering unrelated (i.e. unconnected) neighbors and related (i.e. connected) neighbors are all-most same in most of the cases. The reason is that two major events (Iran election and death of popular pop star Michael Jackson) were occurred during the time period that is considered in the experiment. As a result, most of the users' paid much attention on these events and tried to post meaningful tweets relevant to the events. Again, in some cases (tweet number 1, 2, 7, 8 and 13), the gaps between the annotated credibility scores to the $\pi_{uncon}(t_k)$ and the $\pi_{con}(t_k)$ (marked in green color) are sometimes large. The reason is that these users' previously didn't have much attention on those topics. Due to occurrence of some (and possibly sudden) events, many users started posting tweets on topics that they were not shown much interest earlier.

D. Information Credibility in Flickr Dataset

Below we present both topic and image level trustworthiness for Flickr dataset.

1) Result on Topic Level Credibility.

Table V-D.1 shows the trustworthiness scores (i.e. $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$) for the 8 topics that are used in the experiment. In this work, we consider a topic has different trustworthiness level from connected and unconnected neighbors perspective if the difference between the scores of $\lambda_{uncon}(T_x)$ and $\lambda_{con}(T_x)$ is greater than 0.15 . We can see that some of the topics such as *technology, entertainment, aggression* and *architecture* have similar trustworthiness scores both from connected and unconnected neighbors perspective whereas, the topics such as *nature, daily life* and *family* (marked as yellow color) have different trustworthiness scores from connected and unconnected neighbors perspective. Lastly, we see that for topic *travel* (marked as green color), although the difference between the trustworthiness scores from different perspective is more than 0.15 but we consider the trustworthiness in both cases are similar as both the scores are above 80%.

Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	Topic	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$
Technology	0.46	0.4	Daily Life	0.38	0.60
Entertainment	0.27	0.40	Family	0.26	0.60
Nature	1.0	0.60	Aggression	0.75	1.0
Travel	0.81	1.0	Architecture	0.57	0.6

Table 8: Statistics of Trustworthiness Scores of the Topics for Flickr Dataset

2) Result on Image Level Credibility.

Similar to the Twitter datasets, we choose 15 different actions in Flickr dataset to measure their credibility level both from connected and unconnected neighbors perspective (i.e. $\pi_{uncon}(t_k)$ and $\pi_{con}(t_k)$). Table V-D.2 shows the performance of the proposed methodology for Flickr dataset. We can see that the credibility scores (bold values) of the tweets considering unrelated (i.e. unconnected) neighbors are better compare with the scores considering related (i.e. connected) neighbors in some cases (activity number 2, 6-10, 12 and 14). In other cases, performance of the proposed model from the perspective of connected neighbors perform

#	Tweet	Topic	Annot. Credibility	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	$\sigma_{(n_i,T_x)}$	$\pi_{uncon}(t_k)$ ($\beta=3$)	$\pi_{con}(t_k)$ ($\beta=3$)	$\pi_{uncon}(t_k)$ ($\beta=4$)	$\pi_{con}(t_k)$ ($\beta=4$)
1	rt @muguide essential reading: michael rowe's column re: holocaust museum murder...	Corporate	6.5	0.42	0.25	0.21(0.22)	**3.61 (3.63)**	2.42 (2.44)	**3.40 (3.42)**	2.38 (2.41)
2	rt @karoli: michael moore: "insurance companies are the halliburtons of the health...	Daily Life	6.0	0.27	0.21	0.22 (0.61)	**2.53 (3.68)**	2.12 (3.26)	**2.49 (4.02)**	2.13 (3.66)
3	rt @hotnew: #iremember michael jackson being black. (lol! i was going to say it...	Random	7.5	0.69	0.73	0.56 (0.58)	**6.55 (6.61)**	**6.82 (6.88)**	**6.42 (6.50)**	**6.65 (6.73)**
4	rt@davidmihm: rt @mdjensen: just got interviewed by the wall street journal for tweetbeep.com!...	Persian	8.5	1.0	1.0	0.39 (0.43)	**8.17 (8.31)**	**8.17 (8.31)**	7.57 (7.75)	7.57 (7.75)
5	dr. conrad murray quits medical practice #michael jackson looking ...	Random	7.0	0.69	0.73	0.45 (0.46)	**6.24 (6.25)**	**6.51 (6.52)**	6.00 (6.02)	6.23 (6.24)
6	"rt:@vaibhav: please vote for #michael jackson best song http://twtpoll.co m/5vzafj #rip mj #mj's #michael...	Music	9.0	0.96	0.99	0.34 (0.30)	**7.79 (7.67)**	**7.99 (7.86)**	7.16 (7.00)	7.33 (7.17)
7	rt @hashsocial: rt @theunseenshadow : i find it hard to ponder that when #michael jackson died on thursday...	Compassion	7.0	0.32	0.22	0.36 (0.37)	**3.38 (3.40)**	2.65 (2.67)	**3.42 (3.45)**	2.80 (2.83)
8	speechless. rt @cnnbrk: jackson's body will return 2 neverland thurs 4 public...	Death	7.5	0.21	0.19	0.23 (0.35)	**2.20 (2.55)**	2.10 (2.45)	**2.23 (2.70)**	2.14 (2.61)
9	jason rezaian's piece on female mousavi supporters 4 http:// bit.ly/8oqyi #iranelection...	Iran	9.0	0.96	0.99	0.55 (0.59)	**8.44 (8.55)**	**8.64 (8.75)**	8.03 (8.18)	8.20 (8.35)
10	rt@stopahmadi: we really need your help - the world - to make the revolution happen in iran! #iranelection...	Iran	8.5	0.96	0.99	0.43 (0.55)	**8.08 (8.44)**	**8.28 (8.64)**	7.55 (8.03)	7.72 (8.20)
11	rt @stopahmadi 'bbc persia just said "ahmadinejad is very open-minded" - traitors! #iranelection...	Persian	8.0	1.0	1.0	0.35 (0.30)	**8.06 (7.92)**	**8.06 (7.92)**	7.41 (7.22)	7.41 (7.22)
12	rt@tazahorate_ma: rt first aid info now in farsi!http://gr88.tumblr.com please rt. could save a life. #iranelection ...	Iran	8.0	0.96	0.99	0.40 (0.29)	**7.99 (7.65)**	**8.18 (7.85)**	**7.43 (6.98)**	**7.60 (7.15)**
13	#iranelection crazy shit going on post-election: http://bit.ly/zrc9f rt from #iran:...	Rivalry	6.0	0.23	0.20	0.51 (0.54)	**3.16 (3.24)**	2.95 (3.03)	**3.45 (3.56)**	3.26 (3.37)
14	@mazi: yesterday i had all the hope for #iranelection today, i have more faith in macdonalds providing...	Entertainment	7.5	0.77	0.75	0.16 (0.17)	**5.87 (5.90)**	**5.77 (5.80)**	5.27 (5.30)	5.18 (5.21)
15	rt @cody_k: mahmoud ahmadinejad... kitty thinks you're an asshole...	Iran	8.0	0.96	0.99	0.50 (0.61)	**8.27 (8.62)**	**8.47 (8.82)**	7.80 (8.27)	7.97 (8.44)

Table 7: Performance of the proposed algorithm in measuring tweets level credibility in SNAP dataset (in all cases, $\alpha = 0.5$, the values within brackets indicate the trustworthiness scores without considering temporal factor i.e. according to the method proposed in [30])

better (activity number 1, 3-5, 11, 13, 15). The reason is that the interaction among the connected users in Flickr dataset are not that much strong compare with interaction among the connected neighbors in Twitter network. As a result, the effect of connected peers don't play much role in this kind of social network.

Again, in some cases (activity number 10 and 11), the gaps between the annotated credibility scores to the $\pi_{uncon}(t_k)$ and the $\pi_{con}(t_k)$ (marked in green color) are sometimes large. The reason is that the tags associated with images were sometimes present new technology (for example, in activity 10 which focuses on space related images) or new insights of a product (in activity number 11 that focuses new properties of iPhone) in which most of the users were not familiar.

E. Effect of α on the Social Datasets

We vary the value of α from 0 to 1 in Equation 4 to observe its effects on users' average topical expertise (σ) (Equation 5) on different topics. Figure 4, 5 and 6 presents the effect of α on users' average σ values for sample 5 topics in CRAWL, SNAP and Flickr dataset, respectively.

The parameter α balances two factors $f_1(n_i, T_j)$ and $f_2(n_i, T_j)$ that indicate user n_i's interest towards topic T_j and her participation related to T_j in comparison with the most active user in the network, respectively (Equation 2 and 3). In all cases, we observe that higher values of α which give more emphasizes on the first factor ($f_1(n_i, T_j)$) generate lower topical expertise scores (σ) in all topics and the overall differences in σ values among the topics are very less. Again, for lower values of α that focus more on the second factor ($f_2(n_i, T_j)$), we see that the differences in σ values among the topics increase a lot. In SNAP dataset, the impact of α is not much on the topics of *Iran, Iran Election* and *Random* (this topic also includes tweets related to the death of Michael Jackson). The reason is that most of the users concentrated and paid equal importance on these topics.

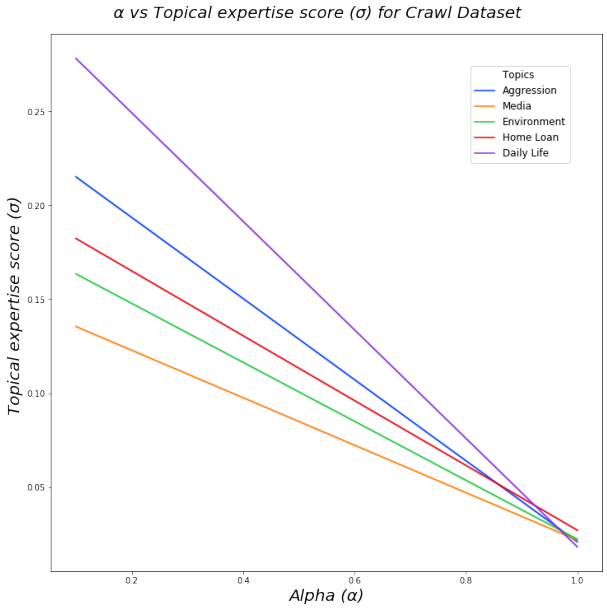

Figure. 4: Effect of α on users' avg. topical expertise score (σ) on different topics in CRAWL

#	Tags	Topic	Annot. Credibility	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	$\sigma_{(n_i,T_x)}$	$\pi_{uncon}(t_k)$ $(\beta=3)$	$\pi_{con}(t_k)$ $(\beta=3)$	$\pi_{uncon}(t_k)$ $(\beta=4)$	$\pi_{con}(t_k)$ $(\beta=4)$
1	ride bronx transportation bicycle squarer south boogie...	Daily Life	5.5	0.38	0.60	0.33	3.65	**5.19**	3.60	**4.92**
2	stockholm carnival student quarnevalen snow night winter..	Aggression	7.5	0.75	1.0	0.19	**5.82**	7.57	**6.01**	7.76
3	moscow russia canon party night film macro fun girl...	Entertainment	6.0	0.27	0.40	0.67	4.00	**4.81**	4.3	**5.08**
4	home sargent clic holiday malcolm	Family	5.0	0.26	0.60	0.25	2.60	**4.95**	2.56	**4.60**
5	photo nature tree winter wood snow flower fel...	Nature	7.5	1.00	0.6	1.00	10.00	**7.20**	10.00	**7.60**
6	squarer iphoneography app format instagram manila ...	Technology	7.0	0.46	0.4	0.67	**5.23**	4.81	**5.44**	5.08
7	stuckincustoms trey ratcliff nikon photography shoot...	Travel	6.5	0.81	1.00	0.33	**6.66**	7.99	**6.18**	7.32
8	light flower sunset family vacation los angeles catalina tree...	Nature	7.0	1.00	0.6	0.5	**8.50**	5.70	**8.00**	5.60
9	college travel northeastern china boston university hong tian sky...	Travel	8.0	0.81	1.00	0.67	**7.68**	9.01	**7.54**	8.68
10	nasa goddard space center flight sun earth satellite weather ice...	Technology	8.5	0.46	0.40	0.33	**4.21**	3.79	**4.08**	3.72
11	iphone phonegrab landscape gdzlla boston sanfrancisco...	Entertainment	8.0	0.27	0.04	0.33	2.88	**3.79**	2.94	**5.72**
12	squarer center app iphoneography instagram sutro	Technology	7.5	0.46	0.4	0.67	**5.23**	4.81	**5.44**	5.08
13	gmo free nude food salmon monsanto world organic...	Daily Life	6.0	0.38	0.60	0.33	3.65	**5.19**	3.60	**4.92**
14	stockholm street hyde nature oxford tokyo sweden ...	Nature	7.0	1.00	0.60	0.50	**8.50**	5.70	**8.00**	5.60
15	finepix spring nature olympuse quedlinburg town...	Family	4.0	0.26	0.60	0.25	2.57	**4.95**	2.56	**4.60**

Table 9: Performance of the proposed algorithm in measuring image level credibility in Flickr Dataset (in all cases, $\alpha = 0.5$, the values within brackets indicate the trustworthiness scores without considering temporal factor i.e. according to the method proposed in [30])

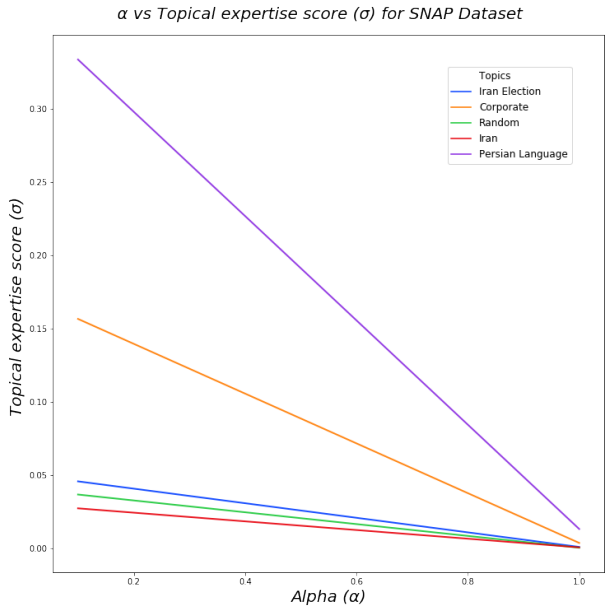

Figure. 5: Effect of α on users' avg. topical expertise score (σ) on different topics in SNAP

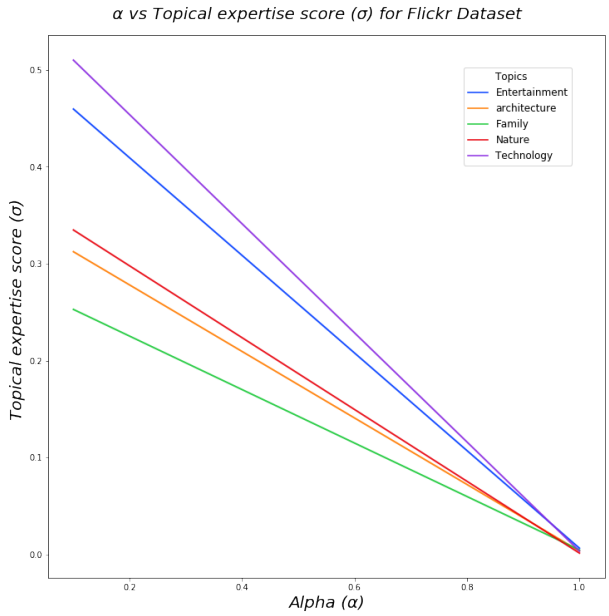

Figure. 6: Effect of α on users' avg. topical expertise score (σ) on different topics in Flickr

We also show the effect of α at action (tweet/image) level trustworthiness for all the datasets depicted in Figure 7, 8 and 9. The following three tweets in CRAWL that are tested:
T1 = puppies and rainbows make people happy, software just gets stuff done

T2 = RT @eli_rubel: Speak with enthusiasm. Otherwise its in one ear and out the other. Own it. #pdxsw

T3 = My movie review of /The Raven may be slightly over focused on how hot @johncusack is. So go read it! http://t.co/AYEKYXJj

The following three tweets from SNAP dataset are tested to see the effect of α:
T1 = rt @karoli: michael moore:"insurance companies are the halliburtons of the health...

T2 = rt @muguide essential reading: michael rowe's columnre: holocaust museum murder..

T3 = jason rezaian's piece on female mousavi supporters 4http:// bit.ly/8oqyi #iran-election...

The following three images from Flickr dataset are used for the purpose analysing the effect of α on credibility score:
T1 = pittsburgh art pennsylvania craft smithsonian modern renwick elder artist kid international brownsville festival learn education child summer mon abandon valley night washington dc winter bridge decay rust quilt harbor cityscape

T2 = england bath imperial event spa freshers somerset union university dramsoc olympics london abbey travel

α	T1		T2		T3	
	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)
0.1	8.51	7.02	4.89	4.33	6.02	7.14
0.2	8.35	6.87	4.86	4.30	6.13	7.26
0.3	8.20	6.71	4.82	4.26	6.24	7.36
0.4	8.04	6.55	4.79	4.23	6.36	7.48
0.5	7.88	6.39	4.76	4.20	6.47	7.59
0.6	7.73	6.23	4.733	4.17	6.59	7.71
0.7	7.57	6.08	4.70	4.14	6.70	7.82
0.8	7.41	5.92	4.66	4.10	6.82	7.94
0.9	7.26	5.92	4.63	4.07	6.93	8.05
1.0	7.09	5.77	4.60	4.04	7.05	8.17

Figure. 7: Effect of α on tweet level credibility in CRAWL

α	T1		T2		T3	
	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)
0.1	6.14	6.42	3.71	2.52	7.47	7.68
0.2	5.99	6.27	3.62	2.43	7.39	7.60
0.3	5.86	6.13	3.54	2.35	7.31	7.52
0.4	5.71	5.99	3.45	2.24	7.23	7.44
0.5	5.56	5.84	3.37	2.19	7.14	7.35
0.6	5.42	5.70	3.29	2.10	7.06	7.27
0.7	5.28	5.56	3.21	2.02	6.98	7.19
0.8	5.13	5.41	3.12	1.93	6.90	7.11
0.9	4.99	5.27	3.04	1.85	6.83	7.03
1.0	4.84	5.12	2.94	1.75	6.74	6.95

Figure. 8: Effect of α on tweet level credibility in SNAP cambridge punt asia national uppark peterield trust southsea portsmouth basketball woman uk festival kite

T3 = photo india canon photography chennai colored tamilnadu picture pics madras pic beautiful beach colorful marina beauty vibrant colour nature lighter vivid temple people canoneosdmarkii green rain portrait bird water child.

α	T1		T2		T3	
	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)	$\pi_{uncon}(t_k)$ (β= 3)	$\pi_{con}(t_k)$ (β= 3)
0.1	2.49	4.87	7.69	9.02	9.02	6.22
0.2	2.42	4.80	7.41	8.80	8.80	6.00
0.3	2.34	4.72	7.24	8.57	8.57	5.77
0.4	2.27	4.65	7.02	8.35	8.35	5.55
0.5	2.20	4.58	6.79	8.12	8.12	5.32
0.6	2.12	4.50	6.57	7.90	7.90	5.10
0.7	2.05	4.43	6.35	7.68	7.68	4.88
0.8	1.98	4.36	6.12	7.45	7.45	4.65
0.9	1.90	4.28	5.90	7.23	7.23	4.43
1.0	1.83	4.21	5.67	7.08	7.00	4.20

Figure. 9: Effect of α on image level credibility in Flickr

F. Spammers Candidate List in Twitter Datasets

We also generate possible spammers lists in both Twitter datasets according to Equation 12 and 13. Table 10 shows the candidate spammers list in CRAWL dataset for 5 topics for the time period from May 01, 2012 to May 14, 2012. The topics are *Family*, *Media*, *Environment*, *Aggression* and *Daily Life* and the average number of activities in these topics are 2.97, 1.55, 1.97, 2.35 and 2.13 respectively. We can see that all the users in the list for all the topics performed 250%-450% more actions than the average number of actions on the related topics. Again, the number of major keywords are very limited compare with the number of actions they performed. Hence, these users enlisted to the candidate spammer list.

Topic	Candidate Spammers	# of Activities	# Major Keywords
Family	2059051	Avg+336.70 %	marriag home amend vote ballot man propos state
	8795782	Avg+266.67 %	miss tourism arrest hous family wife design..
	10207502	Avg+ 233.33 %	retak countri step home occup life work..
Media	8276112	Avg+451.61 %	campaign price wife win spring home lunch
Environment	7540482	Avg+304.56 %	insomnia phone Photo laptop minut
	9283282	Avg+304.56 %	love photo post check snow.
	7496712	Avg+ 253.81 %	hear wrap legi panel host smith..
Aggression	616173	Avg+384.62 %	classroom teacher social currenc activ influenc
	5471982	Avg+384.62 %	resum network time scienc answer facebook...
Daily life	9636172	Avg+373.8 %	classroom teacher social currenc activ influenc
	3533231	Avg+ 280.37 %	demo dummi team kick friend picture...

Table 10: Candidate Spammer list in CRAWL Data set ($\alpha = 0.5, \rho = 200, b = 2$)

Table 11 shows the candidate spammers list in SNAP dataset for 3 major topics for the time period from June 16, 2009 to June 30, 2009. The topics are *Family*, *Iran* and *Entertainment* and the average number of activities in these topics are 1.5, 1.9 and 1.19, respectively. We can see that all the users in the list for all the topics performed around 420%-2300% more actions (specially, the candidate users for *Iran* topic) than the average number of actions on the related topics. Again, the number of major keywords are very limited compare with the number of actions they performed. Hence, these users enlisted to the candidate spammer list.

Topic	Candidate Spammers	# of Activities	# Major Keywords
Random	18727003	Avg + 2000.0 %	rip michael jackson family death award godfather
	24161907	Avg + 1133.33 %	micheal jackson kategosselin fashion lime light
	15571195	Avg + 1133.33 %	remember true magic michael jackson rip
	16910946	Avg + 733.33 %	remember michael jackson lovethecool legends
	2035081	Avg + 933.33 %	tribute legenfary michael music jackson news
Iran	8833302	Avg + 1000.0 %	iran rumor people election urgent neda election
	17350399	Avg + 1105.26 %	death tweeter police fight people protest report
	20745227	Avg + 1210.51 %	message croed croed danger tweet time poignant
	23277821	Avg + 1421.05 %	mousavi khamenie shoot political prisoner tortured
Entertainment	18067881	Avg + 420.62 %	knew killer ludacries samples MJ music
	20683343	Avg + 2268.90 %	tribute rap check micheal jackson here music

Table 11: Candidate Spammer list in SNAP Data set ($\alpha = 0.5, \rho = 200, b = 2$)

G. Effect of Hashtags in Tweet Level Credibility

We also incorporate the effect of hashtags to measure the credibility of a tweet. As there is no formal rules for defining hashtags, so it is not easy to define a proper mechanism for finding the trustworthiness of a tweet. We manually check the hashtags associated with a tweet and mark it as positive/negative. Next, we revise the trustworthiness score of a tweet containing hashtags as:

$$\pi_{uncon}(t_k)/\pi_{con}(t_k) = \pi_{uncon}(t_k)/\pi_{con}(t_k) + -(h \times m) \quad (14)$$

where h indicates the presence of hashtag (it can be from positive/negative perspective) and m is the number of hashtags in the tweet t_k. For each unique positive hashtags, the trustworthiness score is increased by 10%, while the score of $\pi_{uncon}(t_k)/\pi_{con}(t_k)$ is decreased by 5% for each negative hashtags.

Table V-G shows the impact of hashtags in SNAP dataset for two important events: i) Iran election and ii) death of Michael Jackson. The positive hashtags are marked in blue color and the red color is used to mark negative hashtags.

#	Tweet	Topic	Annot. Credibility	$\lambda_{uncon}(T_x)$	$\lambda_{con}(T_x)$	$\sigma_{(n_i,T_x)}$	$\pi_{uncon}(t_k)$ ($\beta=3$)	$\pi_{con}(t_k)$ ($\beta=3$)	$\pi_{uncon}(t_k)$ ($\beta=4$)	$\pi_{con}(t_k)$ ($\beta=4$)
1	rt @hotnew: #iremember michael jackson being black. (lol! i was going to say it...	MJ	6	0.69	0.73	0.56 (0.58)	**5.9** (6.55)	6.14 (6.82)	5.78 (6.42)	**5.99** (6.65)
2	"rt:@vaibhav: please vote for #michael jackson best song http://twtpoll.co m/5vzafj #rip mj #mj's #michael...	MJ	9.0	0.96	0.99	0.34 (0.30)	**8.57** (7.79)	**8.88** (7.99)	7.88 (7.16)	**8.01** (7.33)
3	rt @hashsocial: rt @theunseenshadow : i find it hard to ponder that when #michael jackson died on thursday...	MJ	6.0	0.32	0.22	0.36 (0.37)	3.72 (3.38)	2.92 (2.65)	3.76 (3.42)	3.08 (2.80)
4	rt@stopahmadi: we really need your help - the world - to make the revolution happen in iran! #iranelection...	Iran	9	0.96	0.99	0.43 (0.55)	**8.89** (8.08)	**9.11** (8.28)	8.31 (7.55)	**8.50** (7.72)
5	rt@tazahorate_ma: rt first aid info now in farsi!http://gr88.tumblr.com please rt. could save a life. #gr88 #iranelection ...	Iran	8.5	0.96	0.99	0.40 (0.29)	**8.4** (7.99)	**8.59** (8.18)	**8.17** (7.43)	**8.36** (7.60)
6	#iranelection crazy shit going on post-election: http://bit.ly/zrc9f rt from #iran:...	Iran	5.0	0.23	0.20	0.51 (0.54)	**3.8** (3.16)	3.54 (2.95)	**4.14** (3.45)	**3.91** (3.26)

Table 12: Performance of the proposed algorithm in measuring tweets level credibility by incorporating hashtags in SNAP dataset

The value within brackets indicate the trustworthiness scores of the tweets without considering the impact of hashtags. We can see in all cases that incorporation of the effect of hashtags provide better trustworthiness as those values are much closer to the scores given by the annotators.

VI. Conclusion

Measuring information credibility in OSNs is very promising research field as because it plays vital role during emergency situations and important events. Our proposed model incorporates users' temporal topical interests as well as the network topology in order to highlight the importance of a topic and social action both from connected and unconnected neighbors perspective. Experiments on three real social network datasets depict the efficacy of the proposed methodology.

References

[1] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In WWW, pp. 591–600, 2010.

[2] P. Bogdanov, M. Busch, J. Moehlis, A. K. Singh, and B. K. Szymanski,"The social media genome: Modeling individual topic-specific behaviorin social media," in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013,pp. 236–242.

[3] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In KDD, pp. 497–506, 2009.

[4] W. X. Zhao, J. Jiang, J. Weng, J. He, E. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In Proc. ECIR, pp. 338–349, 2011.

[5] Kang, J O'Donovan, T Höllerer. Modeling topic specific credibility on twitter. In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, pp. 179–188, 2012.

[6] Blei, D. M., Ng, A. Y., Jordan, M. I.: Latent dirichlet allocation. Journal of Machine Learning Research, pp. 993–1022 (2003)

[7] C. Castillo, M. Mendoza, and B. Poblete. Information Credibility on Twitter. In Proceedings of the 20th international conference on World wide web, pp. 675–684, 2011.

[8] Sikdar, Sujoy and Kang, Byungkyu and ODonovan, John and Höllerer, Tobias and Adah, Sibel. Understanding information credibility on twitter. In International Conference on Social Computing, pp. 19 - 24, 2013.

[9] Hofmann, T. Probabilistic latent semantic indexing. In: SIGIR, pp. 50–57 (1999)

[10] W. X. Zhao, J. Jiang, J. Weng, J. He, E. P. Lim, H. Yan, and X. Li,"Comparing twitter and traditional media using topic models," in European conference on information retrieval. Springer, 2011, pp. 338–349.

[11] L Wu, Y Rao, A Nazir and H Jin. Discovering differential features: Adversarial learning for information credibility evaluation. Information Sciences (Elsevier), pp. 515-540, 2020.

[12] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, Z. Wang. Trusting Tweets: The Fukushima Disaster and Information Source Credibility on Twitter, 2012.

[13] S. Kassim, "Twitter Revolution: How the Arab Spring Was Helped By Social Media", Mic, 2012. [Online]. Available: https://www.mic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media. [Accessed: 29- Apr- 2021].

[14] A. Burns and B. Eltham, Twitter Free Iran: an Evaluation of Twitter's Role in Public Diplomacy and Information Operations in Iran's 2009 Election Crisis. Record of the Communications Policy and Research Forum, pp. 298–310, 2009.

[15] M. Hindman and V. Barash, "Disinformation 'fake news' and Influence Campaigns on Twitter", Kf-site-production.s3.amazonaws.com, 2018. [Online]. Available: https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/238/original/KF-DisinformationReport-final2.pdf. [Accessed: 30-Apr- 2021].

[16] A. Iftene, D. Gifu, A. R. Miron and M. S. Dudu. A Real-Time System for Credibility on Twitter.In Proceedings of the 12th Language Resources and Evaluation Conference,pp. 6166–6173,2020.

[17] D. Hassan. A Text Mining Approach for Evaluating Event Credibility on Twitter.In 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), pp. 1524-4547, 2018.

[18] A. Gupta, P. Kumaraguru, C. Castillo and P. Meier. TweetCred: Real-Time Credibility Assessment of Content on Twitter. In International Conference on Social Informatics, pp 228-243, 2014.

[19] X. Xia,X. Yang,C. Wu,S. Li and L. Bao. Information Credibility on Twitter in Emergency Situation. In Pacific-Asia Workshop on Intelligence and Security Informatics, pp 45-59, 2012.

[20] M. Wijesekara and G. U. Ganegoda, "Source credibility analysis on Twitter users," in Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2020, Sep. 2020, pp. 96–102, doi: 10.1109/SCSE49731.2020.9313064.

[21] M. Thandar and S. Usanavasin. Measuring Opinion Credibility in Twitter. In Recent Advances in Information and Communication Technology 2015, pp 205-214, 2015.

[22] S. F. Sabbeh and Sumayah Baatwah. Measuring Opinion Credibility in Twitter. In Recent Advances in Information and Communication Technology 2015, pp 205-214, 2015.

[23] Bird, S., Loper, E., Klein, E. Natural Language Processing with Python. Sebastopol, CA: Oâ€™Reilly Media (2009)

[24] K. R. Canini, B. Suh and P. L. Pirolli.Finding Credible Information Sources in Social Networks Based on Content and Social Structure. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Pp 1 - 8, 2015.

[25] M. Alrubaian, M. Al-Qurishi, and A. Alamri. A Credibility Analysis System for Assessing Information on Twitter.In IEEE Transactions on Dependable and Secure Computing, pp. 661 - 674, 2018.

[26] Li, Ruohan and Suh, Ayoung. Factors influencing information credibility on social media platforms: Evidence from Facebook pages. In Procedia computer science, pp. 314 - 328, 2015.

[27] B. Kang, J. O'Donovan, and T. Hollerer. Modeling topic specific credibility on twitter. In Proceedings of IUI, pp. 179–188, 2012.

[28] Mendoza, M., Poblete, B., and Castillo, C. Twitter Under Crisis: Can we trust what we RT? In 1st Workshop on Social Media Analytics (SOMA), ACM Press, pp. 71–79, (2010).

[29] Anwar, M. M., Liu, C., Li, J..: Discovering and tracking query oriented active online social groups in dynamic information network. In: WWWJ, pp. 1–36 (2018)

[30] Rahman, M.H., Prama, T. T., Anwar, M. M. Modeling Topic Specific Credibility in Twitter Based on Structural and Attribute Properties. In International Conference on Hybrid Intelligent Systems (HIS), pp. 580-589, 2020.

[31] K. R. Canini, B. Suh and P. L. Pirolli, "Finding Credible Information Sources in Social Networks Based on Content and Social Structure," 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 2011, pp. 1-8, doi: 10.1109/PASSAT/SocialCom.2011.91.

[32] TJ Ma, D Atkin. User generated content and credibility evaluation of online health information: A meta analytic study. In Telematics and Informatics, (Elsevier), pp. 472-486, 2017.

[33] Han, B., Cook, P., Baldwin, T.: Lexical normalization for social media text. In: Journal ACM Transactions on Intelligent Systems and Technology (TIST), Volume 4, Issue 1 (2013)

[34] Anwar, M. M., Liu, C., Li, J..: Uncovering Attribute-Driven Active Intimate Communities. In: ADC, pp. 109–122 (2018)

[35] Anwar, M. M., Liu, C., Li, J..: Discovering and Tracking Active Online Social Groups. In: WISE, pp. 54–69 (2017)

[36] G.A. Daniel, M. P. Takis, M. Eni, S. Markus, S. Harald, G. Peter, C. Carlos, M. Marcelo and P. Barbara. Predicting information credibility in time-sensitive social media. In Journal of Internet Research, Emerald Group Publishing Limited, 2013.

Author Biographies

Md. Habibur Rahman received his B.Sc. degree in CSE from Jahangirnagar University, Dhaka, Bangladesh in 2021. Now, pursuing Master's degree in CSE, Jahangirnagar University, Dhaka, Bangladesh. His major research areas are Machine Learning, Data Mining, Artificial intelligence, Software Engineering etc.

Tabia Tanzin Prama is currently pursuing her B.Sc. degree in CSE from Jahangirnagar University, Dhaka, Bangladesh. Her major research areas are Data Mining, Natural Language Processing, Artificial Intelligence, Machine Learning etc.

Dr. Md Musfique Anwar has received his PhD degree from the Department of Computer Science and Software Engineering, Faculty of Science, Engineering and Technology of Swinburne University of Technology, Melbourne, Australia in 2018. He has completed his M.Sc. degree from the Department of Intelligence Science and Technology, Graduate School of Informatics of Kyoto University, Japan in

2013 and B.Sc. degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2006. Since 2008, he is a faculty member having current designation Associate Professor in the Department of Computer Science and Engineering of Jahangirnagar University, Savar, Dhaka, Bangladesh. Currently his research focuses on Data Mining, Social Network Analysis, Natural Language Processing and Software Engineering.