ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

Microblogging Content Propagation – Hybrid Propagation Models and Analysis

N. Baggyalakshmi[1], Dr. A. Kavitha[2] and Dr. A. Marimuthu[3]

Computer Science Department, Kongunadu Arts and Science College, Coimbatore[1]

Assistant Professor in Computer Science, Kongunadu Arts and Science College, Coimbatore[2]

Assistant Professor in Computer Science, Government Arts and Science College, Coimbatore[3]

Abstract: Micro-blogging is a type of social networking deal that has develop permeating in Network 2.0 era. Micro-blogs agrees bloggers to interchange data, deliberate concepts, and stake capabilities with groups or even guests with analogous safeties. Due to the growth of situates similar Facebook, Twitter, and Weibo, administrative statistics is extra and further habitually encountered in a social context: even stories published by mainstream media sites are often encountered by users after having been mutual by others. Obviously, this social environment can impact how data is construed and re-shared. In recent times, there has been an extreme pact of attention in questioning inherent configurations in posts on microblogs such as Facebook, Twitter. While many works consume a well-known topic exhibiting technique, we instead suggest to apply a Hybrid Propagation Model and Hybrid Propagation Analysis in Microblogging. In propagation model use Affinity, Modeling and Visualizing Information Propagation. In analysis side virality and susceptibility type of techniques used. The Microblogging propagation model system shows the propagation paths and social graphs, influence scores, timelines, and geographical information among people for the user-given terms. Propagation analysis, based on this framework, it develop a numerical factorization model and another probabilistic factorization variant. The work also develops an efficient algorithm for the models' parameters learning.

Keywords: Micro-blogging content propagation, Hybrid propagation models, Hybrid propagation analysis.

I. INTRODUCTION

The microblogging service Twitter allows users to broadcast short messages, tweets, to their followers. Millions of users have enthusiastically embraced Twitter, using the 140character limit to express opinion, describe experiences, and spread ideas and information. The resulting flood of data can potentially be mined to discover the "buzz" about products, people, and events, discover emerging trends, and facilitate real-time information search. One of the key challenges that need to be solved is to identify related tweets that are about the same topic.With the rapid growth of social network services and applications such as Facebook, Twitter and Weibo, research on social networks and social media is becoming a hot area. One example is social advertising [1], which utilizes user's relationships, interests and published data to target social advertisement to potential users. Microblogs, also called micro posts, allow users to exchange small elements of content such as short sentences, individual images, or video links. Micro bloggers post about topics ranging from the simple, such as "what I'm doing now," to the thematic, such as "sports cars".The update cycles of other online data sources, such as web pages and blogs, are measured in days, weeks or even months. In contrast, microblogs are updated every few minutes or even seconds, so they can be regarded as real-time services. Therefore, micro-blog data is a perfect resource for studying the dynamic nature of information, in particular how information is propagated and distributed in a large social network. In recent years, a number of micro-blogging services have provided APIs that enable developers to extract the content of messages and obtain information about the social connections among users. As a result, some interesting analyses of microblogging have been published. For example, Sun et al. [5] used data provided by Facebook to determine the correlations between different communities; and Kwak et al. [6] used data from Twitter to construct the relationships among people for advanced analysis of social networks.Content propagates among microblogging users through their follow links, from followers to followers. The former are the senders, and the latter are known as the receivers. A receiver may adopt the content exposed to her based on a number of factors, namely the: (a) virality of the sender, (b) susceptibility of the receiver, (c) virality of the content topic and (d) strength of relationships between sender and receiver. Topic virality refers to the tendency

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

of a topic in getting propagated. Since microblogging has been shown rather an information source than a social networking service. We assume in this paper that most relationships among users in a microblogging site are casual and identical in strength. We therefore focus on modeling the user and content factors that drive content factors that drive content propagation. The modeling of the virality and susceptibility factors has many important applications. In advertisement and marketing, companies may hire viral users to propagate positive content about their products, or to the advertisement with viral content so as to maximize their reach. Similarly, politicians may leverage on viral users to disseminate their messages widely or to conduct campaigning. Also one may detect events by tracking those mentioned by non-susceptible users and detect rumors based on susceptible users interactions with the content.In this proposed techniques, use Hybrid Propagation Model and Hybrid Propagation Analysis in Microblogging. In propagation model use Affinity, Modeling and Visualizing Information Propagation. In analysis side virality and susceptibility type of techniques used. Here after, discuss about different propagation models and analysis in below chapters.

II. LITERATURE REVIEWS

A. Topic Modeling

Probabilistic topic models such as LDA were introduced by [6]. [10] Presented the Author-Recipient-Topic (ART) model to learn the distribution specific to author-recipient pairs. [11] Proposed a supervised learning approach to categorize links and quantify influence of web pages. Neither work considered information propagation. The supervised learning approach requires a training data set that is a link-labeled and link weighted graph. Our work does not require such training data because it works directly on the microblog messages published by users.

B. Influence Maximization

Influence maximization proposed in [2] aims to identify a set of seed users who could influence the most number of other users in a social network. Two popular influence propagation models are Independent Cascade Model and Linear Threshold Model. These models assume influence probability based on simple heuristics, such as uniform probability or probability proportional to the degree of a node. Moreover, this problem does not have a target message nor consider the topics for a link. Most previous works focused on improving the efficiency of greedy algorithms [15], [16], such as the CELF optimization based on the sub modularity of incremental influences.

C. Model-based Information Diffusion

Richardson and Domingos [11] proposed a probabilistic method for extracting information from a knowledge-sharing network and put forward a hypothesis about the most effective individuals for viral marketing. Subsequently, Kempe et al. [12] proposed a model to maximize the influence of a social network. First, they showed that finding the most influential people is an NP-hard problem. Then, they proposed two models, the Linear Threshold Model and the Independent Cascade Model, and used them to simulate information propagation in a social network.

D. Information Propagation on Real Data

With the increasing availability of social network data in recent years, researchers have applied different models to analyze the data. Cha et al. [15] exploited Flickr data to construct the relationships between photos and the photographers. They also tried to determine how widely information can be spread and what role word of mouth plays in such a network. Sun et al. [5] investigated the propagation phenomenon of Facebook's News Feed, and created a social network based on users and fans of Facebook for analysis.

III. CONTENT PROPAGATION METHODOLOGIES

This framework apply a Hybrid Propagation Model and Hybrid Propagation Analysis in Microblogging. In propagation model use Affinity, Modeling and Visualizing Information Propagation. In analysis side virality and susceptibility type of techniques used. The Microblogging propagation model system shows the propagation paths and social graphs, influence scores, timelines, and geographical information among people for the user-given terms. Propagation analysis,

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

based on this framework, it develop a numerical factorization model and another probabilistic factorization variant. The work also develop an efficient algorithm for the models' parameters learning.

3.1 Hybrid Propagation Model

In propagation model use Affinity, Modeling and Visualizing Information Propagation.

i. Affinity Propagation

Affinity Propagation [4] is a clustering algorithm that identifies a set of exemplar points that are representative of all the points in the data set. The exemplars emerge as messages are passed between data points, with each point assigned to an exemplar. AP attempts to find the exemplar set which maximizes the net similarity, or the overall sum of similarities between all exemplars and their data points.

In this paper, we describe AP in terms of a factor graph [7] on binary variables, as recently introduced by Frey [5]. The model is comprised of a square matrix of binary variables, along with a set of factor nodes imposed on each row and column in the matrix.
A graphical model for affinity propagation is depicted in Figure 1, described in terms of a factor graph. In a log-form, the global objective function, which measures how good the present configuration (a set of exemplars and cluster assignments) is, can be written as a summation of all local factors.

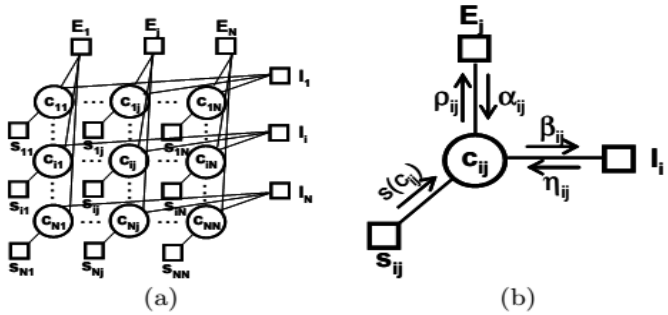

Fig 1 Binary variable model for Affinity Propagation proposed:

(a) a matrix of binary hidden variables (circles) and their factors (boxes); (b) incoming and outgoing messages of a hidden variable node from/to its associated factor nodes.
In our setting, we treat each microblog post or tweet as a data point, and we wish to identify clusters of similar tweets. The similarity between tweets i and j, or S (cij), is determined from the textual similarity. In particularly, we simply use word frequencies of the tweets (weighted using TF-IDF scheme) to compute cosine similarities between them. Tweets' words are normalized as follows: words are stemmed and lowercased, and all non-word characters are discarded.

ii. Modeling and Visualizing Information Propagation

We propose an information propagation model to measure the ability of users to propagate ideas via micro-blogs. Our measure focuses on three factors: the number of people influenced, the speed of propagation, and the geographic range of propagation. We also define a loose criterion and a rigid criterion to bind the level of quantification. We propose a method for measuring the level of spread of a query term in a micro-blog. The method allows us to quantify the influence of query terms in micro-blogs and rank their popularity.

We provide a visualization framework with an online search-based service that implements the proposed methods for demonstration purposes. Given a concept, our system finds the top micro-bloggers that have disseminated the concept in the network, and then displays different kinds of propagation values for users.

System Overview
The system framework is shown in Figure 2 First, the user inputs the query term and the time period of interest. Given the input constraints, the system identifies a set of corresponding posts and their replies, as well as the time stamps on them. Then, based on the above information, two kinds of inference trees are constructed for each blogger: a rigid lower-bound influence (LBI) tree and a loose upper-bound influence (UBI) tree. For each tree, it is possible to produce three kinds of propagation values: (1) the number of individuals influenced by the propagated information, (2) the speed of the propagation, and (3) the geographic distance of the propagated information.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

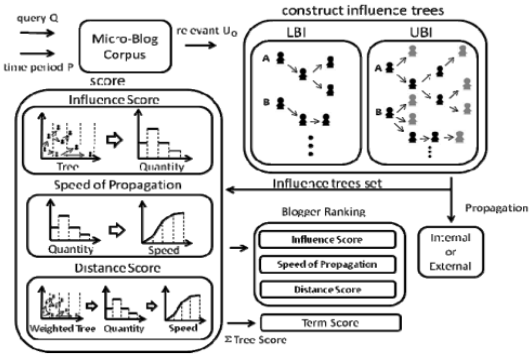

Fig 2 System Overview

Tree Construction

For a given user query Q, we construct a set of influence trees to model the propagation among users in a micro-blog platform. Before discussing the tree construction, we define some key terms.

Definition 1.Micro-blog Corpus: A micro-blog corpus C is comprised of three kinds of entities: users, posts, replies of posts. The corpus C contains a set of users U. Each user has a set of posts and each post might contain a set of reply messages. All posts and replies are time stamped.

Definition 2.Relevant Micro-Bloggers: Given a query Q and a specific time period P, we define relevant micro-bloggers with respect to Q as a set of bloggers UQ={u1,...,un} that satisfy the following requirements: (1) ui has a set of posts Ai=<ai,1, ai,2, ..., ai,m> (k>1) containing the query term Q; and (2) the time stamp of each ai k €Ai is within the specified period P. We also associate each ui €UQ with a time stamp of the first message that ui posts about Q.

Quantify Propagation

Scale of Propagation. To quantify the scale of the propagation, we can simply count the total number of people in the corresponding influence tree. The higher the number, the greater will be the scale of the propagation, as shown by the two-dimensional diagram. The horizontal axis represents the sequence of time stamps {t1,t2,…,tm}, and the vertical axis represents the number of people influenced during the specified time period.

Speed of Propagation. Besides the amount of propagation, we are interested in the speed of propagation. A person who is capable of affecting a large number of people in a short time is considered as a strong candidate for disseminating information. Based on the constructed influence trees and the time stamp of each entity, we propose a method for estimating the propagation speed of a message sent by the person at the root node.

Distance of Propagation, Geographic information enables us to observe whether a message is disseminated globally or locally. In this section, we propose a method for measuring the propagation in terms of the geographic distance. An intuitive way to determine the propagation distance is to calculate the total distance from the root user to the people he/she influences. This can be achieved by a micro-blogging service as long as it provides location information, such as city and country of the users.

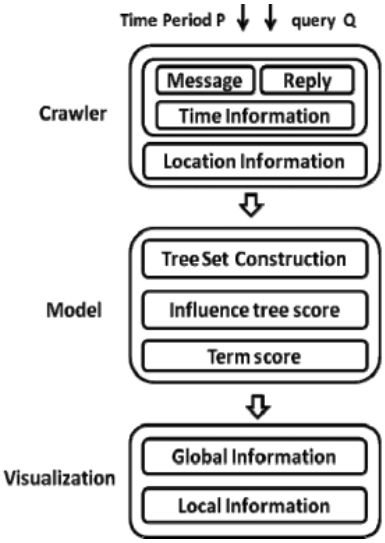

Fig 3 Working of scheme

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

Our system utilizes a topic term and the indicated time period to crawl the related posts and replies as well as the relevant users of Plurk. We crawl four kinds of content: 1) relevant posts; 2) the replies to each post; 3) the time stamps of the posts and the replies; and 4) the geographic information provided by the relevant users. Second, we use the crawled data to construct propagation trees for each relevant user. Then, we use the proposed measures to quantify the propagation capability of each relevant user. Finally, we display the information propagated by top-ranked users. We show the top five micro bloggers for each measure.

3.2 Hybrid Propagation Analysis

i. Virality and Susceptibility Analysis

Conduct an empirical analysis of content propagation on a large dataset collected from Twitter. The methodology used to derive content propagation behavior and topics will be presented. The study will show that virality and susceptibility contributing to content propagation should be modeled at topic level. Use retweet to define propagation in the remaining part of this section.

That is, each original tweet m is considered as a content item, and we say user v is exposed to m if (a) v follows m's author, and (b) v receives and reads m. Lastly, m is said to be propagated from its author u to v if (i) v follows u and (ii) v retweets m. We do not consider in this work the subsequent retweets of m by v's followers and by followers of the followers, since: (1) only less than 5% of retweets are subsequent retweets , and (2), as aforementioned, Twitter.

Methodology -Both content propagation and content topics are usually not observable when the microblogging data are crawled. We have therefore devise the methodological steps to infer them as described below. Determining user-tweet exposure - In Twitter, the latest tweets posted by a user's followers always appear at the top of her timeline. Hence, many tweets may have been missed by the user who does not monitor the timeline closely, and such tweets would never be retweeted. As Twitter API does not reveals the tweets seen by users, we define a time window in which the received tweets will be read. We know that every retweet by a user v comes with a corresponding tweet m that v must have read. We first count the number of other tweets v receives within the duration from the time v receives m to the time v retweets m. Based on this count we estimate N the number of tweets a user may read on her timeline whenever she performs a retweet.

Topic discovery -We applied Twittered model to automatically identify the topics of every original tweet. This step is conducted for every time window, independently from each other's. We first re move all retweets and non-informative tweets, e.g., tweets generated by third party applications like Foursquare or Instagram. We then remove from remaining tweets all stop words, slang word, and non-English phrases.

Topics of tweets and retweets *at network level* -To compare the likelihood of getting retweeted across topics, in each time window and for each topic k, we derive the relative popularities of topic k among the set of all original tweets and the bag of retweets in the time window. The former is called generating popularity of the topic k, denoted by Gk, and the later is called propagating popularity.

Topics of tweets and retweets at individual level- In each time window, to compare the likelihood of user u getting retweeted for different topics, we compare the relative popularities of each topic k in the set of tweets posted by u, and in the bag-of retweets that u got. The former is called sender-specific generating popularity of u for topic k, while the latter one is called sender-specific propagating popularity of u for topic k.

We study user and content factors underlying content propagation in microblogging. Motivated by an empirical studying showing that different topics have different likelihood of getting propagated at both network and individual levels, we propose to model the factors to topic level. We develop V2S, a tensor factorization based framework and its associated models, to learn topic-specific user virality and susceptibility, and topic virality from content propagation data.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

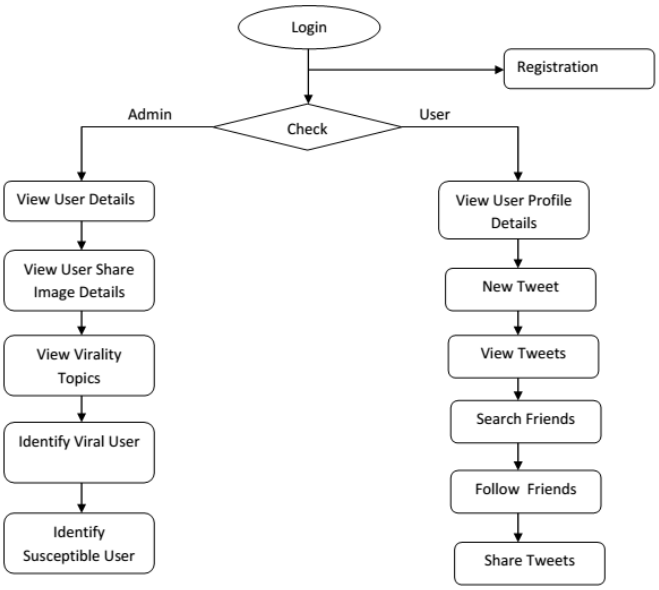

Fig 4. Data Flow Diagram

IV. EXPERIMENTAL RESULTS

Each work is implemented and simulated under various configuration parameters to know their performance measure values. Here compare proposed model to existing process.

Hybrid Propagation Models and Analysis (HPMA) with Topic-specific Behavioral Factors (TSBF).

The performance measures that are considered for evaluating the improvement of the proposed research methodologies are, "Accuracy, Precision, Recall," The comparison results of this performance metrics are illustrated and explained in the following sub sections.

Accuracy (%)

Accuracy is determined as the overall correctness of the model and is computed as the total actual classification parameters $(T_p + T_n)$ which is segregated by the sum of the classification parameters $(T_p + T_n + F_p + F_n)$. The accuracy is computed as like :

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)}$$

Where T_p - True positive, T_n-True negative, F_n-False negative, F_p-False positive

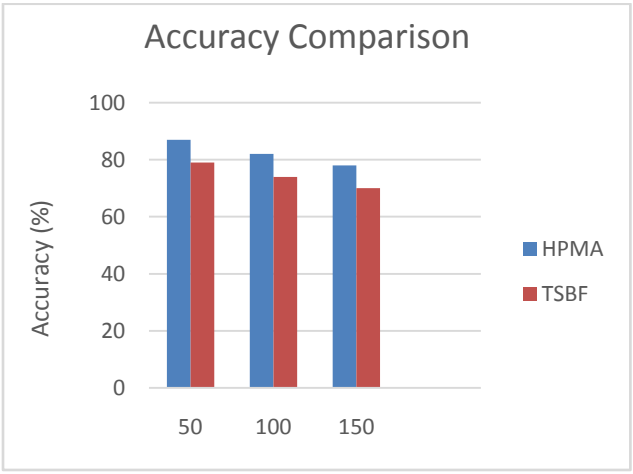

Fig 4.1 Accuracy Comparison

DOI 10.17148/IJARCCE.2018.7614

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

No of input Samples	Accuracy (%)	
	HPMA	TSBF
50	87	79
100	82	74
150	78	70

Table 4.1. Accuracy Measure

From the above Figure 1, it can be observed that the comparison metric is evaluated using existing and proposed method in terms of accuracy. For x-axis the algorithms are taken and in y-axis the accuracy value is plotted.

Precision (%)

Precision (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of positives that are correctly identified as such (e.g., the percentage of sick people who are correctly identified as having the condition).

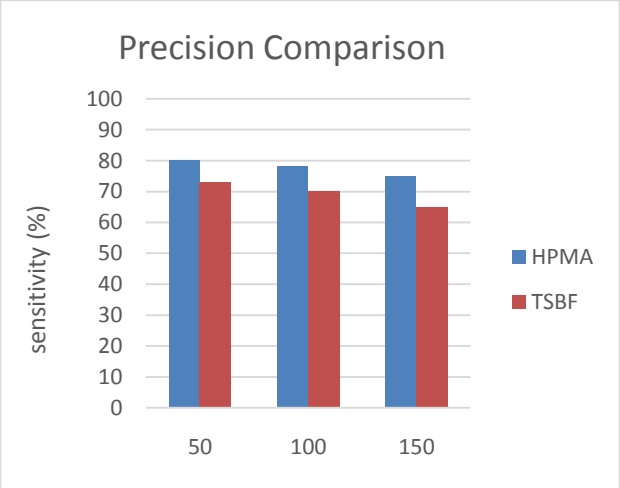

Fig 4.2 Precision Comparison

No of input Samples	Precision (%)	
	HPMA	TSBF
50	80	73
100	78	70
150	75	65

Table 4. 2. Precision Measure

In figure 4.2 Precision measure comparisons of the proposed research methodologies is given. This graph proves that the proposed research method can accurately predict the faults present in the software efficiently with improved performance.

Recall (%)

Recall (e.g., the percentage of healthy people who are correctly identified as not having the condition). Specificity relates to the test's ability to correctly detect classifier without a condition. Mathematically, this can also be written as:
The graphical representation of the recall measurement values of the proposed research methodology is given in figure 4.3.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

Fig 4.3 Recall Comparison

No of input Samples	Recall (%)	
	HPMA	TSBF
50	80	75
100	75	70
150	72	68

Table 4.3 Recall Measure

In figure 4.3 Recall measure comparisons of the proposed research methodologies is given. This graph proves that the proposed research method can accurately predict the faults present in the software efficiently with improved performance. From this comparison analysis, it can be predicted that the proposed method shows better outcome than previous techniques.

VI. CONCLUSION

This framework applied a Hybrid Propagation Model and Hybrid Propagation Analysis in Microblogging. In propagation model use Affinity, Modeling and Visualizing Information Propagation. In analysis side virality and susceptibility type of techniques used. Some people believe that online advertising could provide a profitable business model for social network services. Thus, being able to quantify and measure the propagation of information would facilitate the expansion of online advertising services in social networks. In this work, we propose a model for estimating information propagation. Present novel ways to measure the propagation speed and the geographical distances. We have also implemented an online system, called Propagation, trying to visualize information propagation in Plurk. The system displays global information about concept propagation as well as local, dynamic information that allows users to gain more insights into propagation patterns. Our system and model are for general purposes, so can easily be applied to other microblog services, such as Twitter. We believe that our system will provide researchers in other areas, such as social science, with an alternative way to collect data and conduct social-network research.In the future, we will exploit certain NLP techniques, such as opinion analysis, to analyze the content of messages and provide a deeper analysis of the propagation. Moreover, the current system is not very efficient because the query API is slow. To resolve the problem, we will investigate cloud computing based approaches to collect information through the Map Reduce framework in order to speed up the process.

REFERENCES

[1] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, “Everyone’s an influencer: quantifying influence on twitter,” in WSDM,2011.
[2] S. A. Macskassy and M. Michelson, “Why do people retweet? antihomophily wins the day!” in ICWSM, 2011.
[3] Z. Liu, L. Liu, and H. Li, “Determinants of information retweetingin microblogging,” Internet Research, 2012.
[4] S. Stieglitz and L. Dang-Xuan, “Political communication and influence through microblogging–an empirical analysis of sentiment intwitter messages and retweet behavior,” in HICSS, 2012.
[5] T.-A. Hoang, W. W. Cohen, E.-P. Lim, D. Pierce, and D. P.Redlawsk, “Politics, sharing and emotion in microblogs,” inASONAM, 2013.
[6] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, “Want to be retweeted?large scale analytics on factors impacting retweet in twitter network,” in SocialCom, 2010.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 6, June 2018

[7] J. A. Berger and K. L. Milkman, "What makes online contentviral?" Journal of Marketing Research, 2012.
[8] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of socialnetworks in information diffusion," in WWW, 2012.
[9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a socialnetwork or a news media?" in WWW, 2010.
[10] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitterpower: Tweets as electronic word of mouth," JASIST, 2009.
[11] Z. Zhou, R. Bandari, J. Kong, H. Qian, and V. Roychowdhury,"Information resonance on twitter: watching iran," in SOMA,2010.
[12] J. H. Parmelee and S. L. Bichard, Politics and the Twitter revolution:How tweets influence the relationship between political leaders and thepublic. Lexington Books, 2011.
[13] P. Achananuparp, E.-P. Lim, J. Jiang, and T.-A. Hoang, "Who isretweeting the tweeters? modeling, originating, and promotingbehaviors in the twitter network," ACM TMIS, 2012.
[14] C. Castillo, M. Mendoza, and B. Poblete, "Information credibilityon twitter," in WWW, 2011.
[15] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonalves, A. Flammini,and F. Menczer, "Detecting and tracking political abuse in socialmedia," in ICWSM, 2011.
[16] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in WWW, 2004.
[17] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topicsensitive influential twitterers," in WSDM, 2010.
[18] L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang, "Mining topic-levelinfluence in heterogeneous networks," in CIKM, 2010.
[19] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," inICWSM, 2010.
[20] G. Szabo and B. A. Huberman, "Predicting the popularity of onlinecontent," Comm. ACM, August 2010.
[21] D. Romero, W. Galuba, S. Asur, and B. Huberman, "Influence andpassivity in social media," 2011.
[22] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who shouldshare what?: item-level social influence prediction for users andposts ranking," in SIGIR, 2011.
[23] J. L. Iribarren and E. Moro, "Affinity paths and informationdiffusion in social networks," Social networks, 2011.
[24] R. Bandari, S. Asur, and B. A. Huberman, "The pulse of news insocial media: Forecasting popularity." in ICWSM, 2012.
[25] T.-A. Hoang and E.-P. Lim, "Virality and susceptibility in information diffusions," in ICWSM, 2012.
[26] S. Aral and D. Walker, "Identifying influential and susceptiblemembers of social networks," Science, 2012.
[27] L. Weng, F. Menczer, and Y.-Y. Ahn, "Virality prediction andcommunity structure in social networks," Scientific reports, 2013.
[28] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who sayswhat to whom on twitter," in WWW, 2011.
[29] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonc¸alves, A. Flammini, and F. Menczer, "Political polarization on twitter," inICWSM, 2011.
[30] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microbloggingduring two natural hazards events: what twitter may contributeto situational awareness," in CHI, 2010.