Indonesian Journal of Electrical Engineering and Computer Science
Vol. 20, No. 2, November 2020, pp. 1037~1043
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v20.i2.pp1037-1043

Online social network relationships influenced on a retweeting

Iman K. Abbood, Saad Talib Hasson
College of Information Technology, University of Babylon, Iraq

Article Info

Article history:

Received Dec 25, 2019
Revised Mar 26, 2020
Accepted Apr 9, 2020

Keywords:

Correlation
Micro blogging
Retweet
Social network
Twitter

ABSTRACT

In this era, social media is becoming a suitable place for discovering and exchanging new updates. The ease-of-use leads to present novel breaking news to show up first on Microblogs. Twitter is one of the well-known Microblogging platforms with more than 250 million users, in which retweeting is a manageable way to share and sawing news. It is significant to foretell the retweeting and influence in a social relationship. The Correlation Coefficient formula has been used as a tool to determine the level of correlation between a user and his retweeters (followers, friends, and strangers) in social networks. Such correlation can be reached by utilizing the collected user information on Twitter with six features that have a main effect on retweet behavior. In this study, the focus is on particular friends, followers, and a retweet to be the promising source of relationships between users of social media. Experimental results based on Twitter dataset showed that the Correlation Coefficient formula can be used as a predicting model, and it is a general framework to gain better fulfillment in calculating the correlation between the user, friends, and followers in social networks, where the proposed methods refer to a relationship between the number of friends and the users who retweeting. This means that most of the friends are not retweeting friends' tweets, in other words, only 5% of the users' friends are retweeting on their friends' tweets while 95% are not retweeting.

Corresponding Author:

Saad Talib Hasson,
Iraq- University of Babylon,
College of information Technology, Iraq.
Email: saad.aljebori@gmail.com

1. INTRODUCTION

Online social networks grew to become a significant phenomenon throughout the last years [1, 2]. Where it has a vital impact on sharing, receiving and breaking news based on users' relationships [3]. In 2012, A Twitter service had worldwide popularity [4, 5], it had about 500 million users posted nearly 340 million tweets per day [6]. Twitter users' can share a tweet to reach 140 characters over their smart devices [5, 7]. They enrolled different users to mark their tweets and share emotions with their opinions in time through the interface of their websites [8].

To help scientists working in social and network fields, online network providers like Twitter have been making their data partly or fully available to the academics which has been contributing to new data-mining possibilities. The Twitter dataset has socially effected on studying many a critical problem which could not be addressed before because it serves as a source of information to many of the users [1]. The studies in [9] presented a strategy that identifies data diffusion dynamics as an important factor in the evolution in social media by using the following relationship, where he adopts a Maximum Likelihood Estimation(MLE) framework to quantify the system-wide ubiquity of various link creation strategies [9]. While the authors in [10], have been used topology-based techniques and retweet-based techniques create variety in link predictions, on the other hand, the author notes in his experiment the mixing of two techniques increased the accuracy of link production in a social network. in addition to previous studies, some authors used the relation between users and the original source of the tweet to enrich his experiment. When Twitter users write a post or repost a message, in this case, their followers can see what they are share and might decide to repost it. This leads to creating tracks that together form cascade networks [11, 12]. A user's when they are received a reposted tweets can see the track of an ancestor and the source. In the Twitter graph, a one-way directed relationship composed by following the tweets of others [13], as a result, a user may determine to follow an ancestor or original source of the tweet and receiving their future posting sprightly. These new links create shortcuts connecting users at any distance in the network [9]. In [8], the author used followee and follower features furthermore to tweet features and interactive features to predict retweet behavior. Recently, many studies benefit from a friend's graph on social networks. On social networks, like, Facebook or Twitter, users are very likely to follow influential friends in their closed society to retweet a post or "like" a photo. A good question is: how friends in a twitter network influence each other and how the influence is publishing in the social network? Answering the question is non-trivial. Actually, it is one of the challenges of the following aspects [14]. In [15] the researchers had developed an unsupervised model to guess relationship strength to automatically distinguishing weak from other strong relationships. in [16] the author investigates the retweet behavior influence by the friends in one's ego network in Microblogging network The author Sudheendra Hangal. et al. had utilized the employing weighted and directed influence edges in the social graph, that could give more effective results than a search based on binary relationships for global social search [17]. On the other hand, by combining the discovered relationship patterns in a factor graph model, the author in [18] shows in his experimental result how to derive trust relationships to expedite Alibaba's E-Commerce business from billion-scale networked data named eTrust [18].

Social influence happens when users' behaviors or (ideas, feelings) are affected by others retweet in a social network. Friends, retweets, followers .etc. represents the essential characteristic of the twitter network. Twitter is characterized by the asymmetric nature of its follower's relationships [19]. This issue is dissimilar to the friendship in a social network, like Facebook, which can only be formed with the consent of both nodes (users) and is usually understood as an undirected edge. The Twitter graphs consist of a set of nodes and edges, where a node is representing a user, while the relationship between users (connection) is represented by an edge [20]. Although in some cases follow relationships are made based on social relationships (kinship, friendships, etc.). In most instances, a person has tended to follow others based on shared interests. For example, a person X can follow Y if he is interested in sport and knows that Y is an expert in this field. Thus, it is extra accurate to talk about the Twitter graph as an interest graph instead of the social graph [21]. Briefly, the relationship in Twitter graphs can be described by the In-links and are identified as followers, while on the other side friends or 'followings' can be represented by the out-links [22].

We applied different Twitter relationships (friends, retweets, and followers) to estimate a tie of correlation in different relationships to the ratio of retweeting probability. One interesting discovery is that how the number of followers influenced the retweet probability and Identifying how user behavior predictions help in discovering influence strength.

The rest of the paper is arranged as follows. Section 2 presents a summary of present associated survey articles provides. Section 3 gives a review of the problem definition and approach. Section 4 discusses the experiment result. The last part concludes this study's results.

2. RESEARCH METHOD

In 1954, J.A. Barnes has presented the idea of social networks and defines them as the linked graph where In 1954, J.A. Barnes has presented the idea of social networks and defines them as the linked graph where nodes describe entities and edge indicate interactions between nodes [13]. Entities may be encompassing users, groups, companies, or polity agencies. The nodes build themselves as transmitters of information between users by the edges (relationships, invitations, trades, values, etc). Figure 1 shows a model of a twitter relation between nodes and edges. In this work, we used the relationship (friends, followers, retweeting) as edges and users (tweets)as nodes in the twitter network. we found a correlation among the number of (friends, followers) and the chance of producing the ratio of the (friends, followers)they are retweeting their friends' tweets after the correlation has been implemented.

Because of the unbalanced distribution of the retweet number and the retweeters of friends in a social network, the correlation between the number of friends, followers, and the users who retweet the post is used as predication estimators. Inspired from [23] the correlation coefficient formula is,

$$r = \frac{\mathrm{Cov}(x,y)}{\sqrt{s_x^2 * s_y^2}} \quad (1)$$

Where the covariance (Cov (x, y)) is calculated according to (2),

$$\mathrm{cov}(x, y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} \tag{2}$$

Where each point denotes as (x, y) pair, in this case, the x-axis represents the number of (friends, followers) it is an independent variable, while on the other side (y-axis) it is a dependent variable.While the variance (S2) of x or y is defined as,

$$s_x^2 = \frac{\sum(x-\bar{x})^2}{n-1} \tag{3}$$

$$s_y^2 = \frac{\sum(y-\bar{y})^2}{n-1} \tag{4}$$

The variance of x, y measures the variability of the (x, y) around their respective sample means $(\bar{x}, \bar{y})$, the means can be computed using (5) and (6).

$$\bar{x} = \frac{\sum x}{n} \tag{5}$$

$$\bar{y} = \frac{\sum y}{n} \tag{6}$$

Where n is the size (number of samples) in this study.

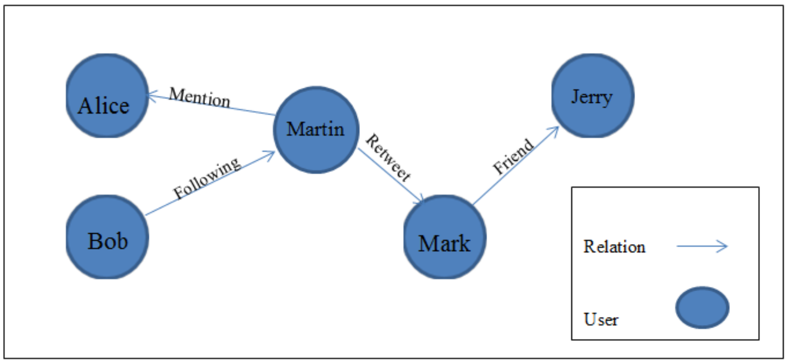


Figure 1. An Example of social network graph

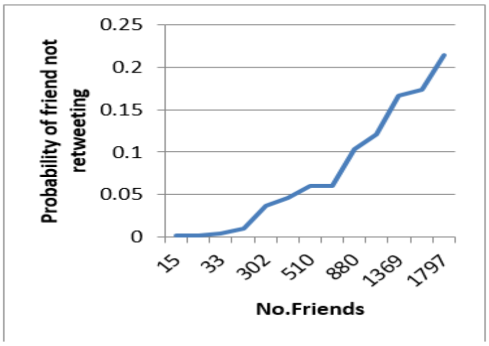
2.1. Dataset description

Twitter is a public American Microblogging service where users share status posts (called "tweets") [24]. Twitter websites interface become important sources of people's opinions and sentiments because Twitter user often writes messages about their life, share an opinion on a different kind of topics, video, picture, or express their political and religious views with his friends and followers around the world. As more and more users post about products and services they use or discuss popular events. For that, the Twitter dataset can be efficiently used for marketing or social studies. As a result, there has been a tremendous need for approaches and algorithms which can effectively process a wide variety of applications. The Twitter search API and the package "tweepy" have been used to collect data from public tweets on the twitter website.

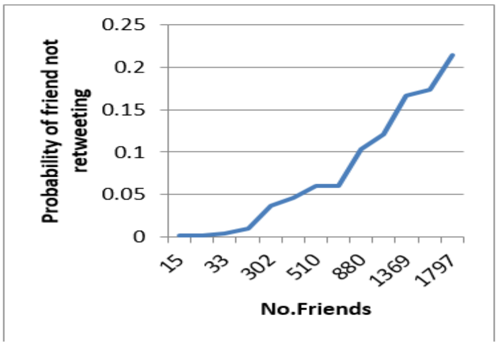
Two criteria have been added to select the tweets from a user account. First, the tweet has been written in the English language. Second, the tweet that has been retweeted was selected from the dataset of the user. We accumulated a dataset of 13 users after running the tweet downloader on the computer. To perform a pre-processing step, the information associated with those 13 users is used to extract 6495 users who retweet. Then, we have identified the (followers, friends) of each user. As the next step, build the relationship according to retweeting as (a friend and retweet, friend and not retweet, not a friend and retweet, follower and retweet, follower and not retweet, not follower and retweet). Finally, we calculate the probability of retweeting each (followers, friends, and new users). The retweeting action by new users leads to a new relation in the tweet graph. The algorithms are implemented in the Python programming language and run on an Intel Core i3 and a processor with 4 GB RAM.

2.2. Social network features

Many social network features have been used to estimate the probability of retweeting posts, like the number of user's followers, a number of following (friends), and the number of users who retweets. A social person is a user how has a good number of followers, and his tweets more likely to be retweet by his followers. These features may also indicate likely motivations for retweeting. In this research, we take a basic statistical analysis of the retweet action of users. For the retweet probability, we count the retweets of tweet posts of all the thirteen users. Figure 2 shows the probability of retweeting, where Figure 2(a) shows the probability of friend but not retweeting, while Figure 2(b) shows the probability of followers retweeting. Obviously, a number of the posted tweet in addition to a number of followers and number of friends influential on the retweet mechanism.



(a) The probability of friend not retweeting



(b) The probability of followers retweeting

Figure 2. The probability of retweeting in (a) and (b)

3. RESULTS AND DISCUSSION

The experimental results of the proposed system are described and shown in this section. The results will be analyzed with a retweet problem on twitter. A real global dataset has been prepared as an employment case study to determine the behavior of this model. A real example of retweet influence derived from the twitter dataset was implemented firstly, as shown in Table 1.

Table 1 indicates that there is a relationship between the number of friends and the users who retweeting, in other words; some of the users' friends will retweet friends tweets. It is important to estimate the ratio of the friends they are retweeting their friends' tweets. The correlation has been implemented as a tool to compute the relation between a number of friends and the number of (friends and retweet) after managing and sorting the values in Table 1 in ascending order according to the number of friends.

Table 1. The number of friends who retweet to the total number of the user's friends'

Users	No. Friend	Friend and retweet
user1	1025	47
user2	548	59
user3	33	5
user4	880	43
user5	105	25
user6	1797	69
user7	23	5
user8	1490	88
user9	302	8
user10	442	69
user11	1369	21
user12	15	2
user13	510	28

Table 2 represents the used steps to estimate the mean and variance values. The values of x and y refer to a number of (friends) and (friends and retweet) respectively.

Table 2. The variance values of both (x, y) value

Users	No. Friends	$(x-\bar{x})$	$(x-\bar{x})^2$	Friend and retweet	$(y-\bar{y})$	$(y-\bar{y})^2$
user12	15	-641.8462	411966.49	2	-34	1156
user7	23	-633.8462	401760.95	5	-31	961
user3	33	-623.8462	389184.02	5	-31	961
user5	105	-551.8462	304534.18	25	-11	121
user9	302	-354.8462	125915.79	8	-28	784
user10	442	-214.8462	46158.87	69	33	1089
user13	510	-146.8462	21563.793	28	-8	64
user2	548	-108.8462	11847.485	59	23	529
user4	880	223.15385	49797.639	43	7	49
user1	1025	368.15385	135537.25	47	11	121
user11	1369	712.15385	507163.1	21	-15	225
user8	1490	833.15385	694145.33	88	52	2704
user6	1797	1140.1538	1299950.8	68	32	1024
	$\sum x = 8539$	$\sum(x-\bar{x}) = 0$	$\sum(x-\bar{x})^2 = 4399525.7$	$\sum y = 468$	$\sum(y-\bar{y}) = 0$	$\sum(y-\bar{y})^2 = 9788$

Applying (5) and (6) to estimate the mean ($\bar{x}, \bar{y}$) of x and y based on the final row in Table 2.

$$\bar{x} = \frac{\sum x}{n} = \frac{8593}{13} = 656.8461$$

$$\bar{y} = \frac{\sum y}{n} = \frac{468}{13} = 36$$

Then the variance of x is estimated according to (3).

$$s_x^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{4399525.7}{12} = 366627.141667$$

The variance of y is estimated according to (4).

$$s_y^2 = \frac{\sum(y-\bar{y})^2}{n-1} = \frac{9877}{12} = 815.6667$$

The calculations are continuing to prognosticating the correlation by calculating the covariance by implementing (2). Table 3 shows these calculation steps for x (number of friends), and y (friends and retweet) as two variables.

Table 3. Calculation steps for x (number of friends) and y (friends and retweet) variables

Users	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
user12	-641.8462	-34	21822.76923
user7	-633.8462	-31	19649.23077
user3	-623.8462	-31	19339.23077
user5	-551.8462	-11	6070.307692
user9	-354.8462	-28	9935.692308
user10	-214.8462	33	-7089.923077
user13	-146.8462	-8	1174.769231
user2	-108.8462	23	-2503.461538
user4	223.15385	7	1562.076923
user1	368.15385	11	4049.692308
user11	712.15385	-15	-10682.30769
user8	833.15385	52	43324
user6	1140.1538	32	36484.92308
	$\sum(x-\bar{x}) = 0$	$\sum(y-\bar{y}) = 0$	$\sum(x-\bar{x})(y-\bar{y})$ =143137

The covariance is estimated based on (2) and the final row of Table 3.

$$cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} = \frac{143137}{12} = 11928.0833$$

The last step is to calculate the correlation coefficient between x and y according to (1).

$$r = \frac{Cov(x,y)}{\sqrt{s_x^2 * s_y^2}} = \frac{143137}{\sqrt{299045538.029}} = 0.689$$

This ratio represents roughly about 69% of retweeters are friends which indicates a good relationship between the number of friends and retweeting. This value used as estimation and prediction by knowing that 69 percent of the user friends will perform retweeting. Based on Table 4 results, another five important indications are suggested and tested as proposed relations. These relations are: (Friend but not retweet, Not friend and retweeting, Follower and retweet, Follower and not retweet and Not Follower and retweeting).

Friends, followers, or strange users may retweet when they follow each other as indicated in the dataset. Table 4 shows the estimated results, in addition to their correlations and percentages. Table 4 indicates a strong correlation value between friends and not retweet. This means that most of the friends are not retweeting friends' tweets, in other words, only 5% of the users' friends are retweeting on their friend's tweets (95% are not retweeting). Another strong correlation value between the followers and not retweet is indicated. This means that most of the followers are not retweeting on the user's tweets, in other words, only 13% of the user's followers are retweeting on the user's tweets (87% are not retweeting).

Table 4. The experimental result of the correlation values of different relationships between retweeters

Users	No. friend	No. Friend and retweeting	No. Friend but not retweet	No. not friend and retweeting	No. Follower and retweet	No. Follower and not retweet	No. not Follower and retweeting
user1	1025	47	978	278	130	1175	195
user2	548	59	489	403	96	915	466
user3	33	5	28	289	11	22	283
user4	880	43	837	857	103	797	797
user5	105	25	80	233	107	1449	151
user6	1797	69	1728	419	205	827	283
user7	23	5	18	203	22	1010	186
user8	1490	88	1402	475	83	1049	480
user9	302	8	294	22	15	81	15
user10	442	69	373	1078	285	644	862
user11	1369	21	1348	38	9	23	50
user12	15	2	13	974	370	1597	606
user13	510	28	482	757	55	77	736
Correlation		0.68976634	0.99937838	0.10424927	0.70180971	0.98939113	0.202210
Ratio (percentage)		0.054924464	0.945075536	0.705703244	0.133458647	0.866541353	0.456856

A very low correlation between the not friends and retweet is indicated as well as the not follower and retweeting. As a comparison, the authors in [25] showed that all users do not have an equal chance of spreading information. By studying completely the retweet chain lengths of users' on twitter, they found that the number of followers of users acts an essential role in their ability to spread information. This is coming compatible with the results in Table 4.

This study experimental results on a six correlation values showed that a number of (friends, followers) and the records of retweets is an effective approach for concluding that the enormous number of (friends, followers) in the social network not influencing on retweeting. Where the result values in Table 4 showed the ratio of friends or followers that retweeting, is very low compared with the strangers who retweeting. From this result, we can find that the characteristics of the post author have a higher influence on retweet numbers than the feature of the tweet itself.

4. CONCLUSION

In this article, we investigated and studied the impact of user retweets on a correlation between friends in the social network platform Twitter. To understand what features playing a key role in predicting an influence on a retweeting process, we examined six features that have a potential relationship with the retweetability. analyzing the six correlation values is an effective approach for concluding that the huge number of (friends, followers) in the social network not influencing retweeting. On the contrary, we found that the characteristics of a Twitter post have a greater impact on retweeting than the contents of the tweet itself. For a particular tweet of a specific user, we could apply our predictive model to predict who will retweet from his/her (friends, followers, stranger users) and who are not. This will lead to seeing how a large number of his/ her (followers, friends) will not retweet a post in Microblogs. There are other features worth studying in this work, as further work proposed to investigate how a tweet content influence on a number of a retweet, and the follower's favor forwarding which kind of tweets.

REFERENCES

[1] J. Szüle, D. Kondor, L. Dobos, I. Csabai, and G. Vattay, "Lost in the city: Revisiting Milgram's experiment in the age of social networks", *PLoS One*, vol. 9, no. 11, pp. 1-6, 2014.
[2] N. Tarmizi, S. Saee, and D. H. A. Ibrahim, "Author identification for under-resourced language Kadazandusun", *Indones. J. Electr. Eng. Comput. Sci. (IJEECS)*, vol. 17, no. 1, pp. 248-255, 2019.
[3] K. Lee, J. Mahmud, J. Chen, M. Zhou, and J. Nichols, "Who will retweet this? Automatically identifying and engaging strangers on twitter to spread information", *Int. Conf. Intell. User Interfaces, Proc. IUI*, pp. 247-256, 2014.
[4] G. Liu, C. Shi, Q. Chen, B. Wu, and J. Qi, "A Two-Phase Model for Retweet Number Prediction", pp. 781-792, 2014.
[5] M. R. Irfan, M. A. Fauzi, T. Tibyani, and N. D. Mentari, "Twitter Sentiment Analysis on 2013 Curriculum Using Ensemble Features and K-Nearest Neighbor", *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 8, no. 6, p. 5409, 2018.
[6] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist, "Understanding the Demographics of Twitter Users", *Int'l AAAI Conf. Weblogs Soc. Media*, pp. 554-557, 2011.
[7] H. F. Huo and X. M. Zhang, "Modeling the influence of Twitter in reducing and increasing the spread of influenza epidemics", *Springerplus*, vol. 5, no. 1, pp. 1-20, 2016.
[8] C. Wang, Y. Jia, J. Huang, and B. Zhou, "Retweet Prediction in Sina Weibo Based on Entity-Level Sentiment Analysis", no. Aita, pp. 343-350, 2016.
[9] L. Weng *et al.*, "The role of information diffusion in the evolution of social networks", *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1288, pp. 356-364, 2013.
[10] S. Tsugawa and K. Kito, "Retweets as a Predictor of Relationships among Users on Social Media", pp. 1-19, 2017.
[11] S. Myers and J. Leskovec, "The bursty dynamics of the twitter information network", *WWW 2014 - Proc. 23rd Int. Conf. World Wide Web*, pp. 913-923, 2014.
[12] A. Guille, "Information diffusion in online social networks", *Proc. ACM SIGMOD Int. Conf. Manag. Data*, no. April, pp. 31-35, 2013.
[13] J. Bourdeaud, "Predicting which tweets will be retweeted most," pp. 2017-2018, 2018.
[14] L. Liu, J. Tang, J. Han, and S. Yang, "Learning influence from heterogeneous social networks", no. March, pp. 511-544, 2012.
[15] R. Xiang, W. Lafayette, and W. Lafayette, "Modeling Relationship Strength in Online Social Networks", pp. 981-990, 2010.
[16] C. Xing, "Who Influenced You ? Predicting Retweet via Social Influence Locality," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 9, no. 3, pp, 1-26 ,2014.
[17] S. Hangal, D. MacLean, M. s. Lam, and J. Heer, "All friends are not equal: Using weights in social graphs to improve search", *4th SNA-KDD Work. '10*, pp. 1-7, 2010.
[18] Y. Cen *et al.*, "Trust Relationship Prediction in Alibaba E-Commerce Platform", *IEEE Trans. Knowl. Data Eng.*, vol. V, no. January, pp. 1-1, 2019.
[19] X. Zhang, M. Sun, Z. Wang, and X. Huang, "Predicting Who Will Retweet or Not in Microblogs Network", *Commun. Comput. Inf. Sci.*, vol. 568, p. v, 2015.
[20] A. Sadilek, H. Kautz, and J. P. Bigham, "Finding your friends and following them to where you are", *WSDM 2012 - Proc. 5th ACM Int. Conf. Web Search Data Min.*, pp. 723-732, 2012.
[21] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "WTF: The Who to Follow service at Twitter", *WWW 2013 - Proc. 22nd Int. Conf. World Wide Web*, pp. 505-514, 2013.
[22] L. K. Hansen, A. Arvidsson, F. F. Nielsen, E. Colleoni, and M. Etter, "Good Friends, Bad News Affect and Virality in Twitter", 2011.
[23] J. L. Rodgers, W. A. Nicewander, T. A. Statistician, and N. Feb, "Thirteen Ways to Look at the Correlation Coefficient Thirteen Ways to Look at the Correlation Coefficient", no. June 2015, 1988.
[24] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities", In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network* analysisv, pp. 56-65, 2007.
[25] C. Remy, N. Pervin, F. Toriumi, and H. Takeda, "Information diffusion on twitter: Everyone has its chance, but all chances are not equal", *Proc. - 2013 Int. Conf. Signal-Image Technol. Internet-Based Syst. SITIS 2013*, pp. 483-490, 2013.