



Detecting Influential Users in a Trending Topic Community Using Link Analysis Approach

Myat Mon Oo^{1*} May Thu Lwin¹

¹*University of Computer Studies, Mandalay, Myanmar*
* Corresponding author's Email: myatmonoo@ucsm.edu.mm

Abstract: Social media has become a popular platform to post or share personal information, opinion, photos, videos, etc. Detection of influential users is a significant problem in information diffusion or propagation. Previous researches find influential users based on follower or retweet relationships and centrality measurement approaches. In this paper, influential users are detected by network topology that was obtained from communication relationships among users, link analysis approach, and user's profile features. The proposed approach aims to detect trending topic influencers. Firstly, communication relationships namely retweet, mention, and reply between users in a trending topic are extracted and a trending topic graph is constructed. Secondly, influential users are detected using a link analysis approach combined with the ability of users' profile features. The performance of the system can be proved to compare with influencer detection methods. The experimental result shows that the trending topic influencers can detect using the interaction relationships and user's features.

Keywords: Trending topic, Social network, Influential users, Twitter, Network graph.

1. Introduction

Social media is encountered with significant growth because of the increased volume of content and the number of users. Some of the most popular social media are Google+, Myspace, Instagram, Facebook, Twitter, etc. The large volume of information from social media attracted many researchers, scientists, and companies to obtain significant information [1]. Information on social media can serve many purposes such as rumors, the spread of unwanted content, hindering negative behaviors, marketing advertisement with low cost, etc. There are many research areas concerned with social media such as Event detection, Public health monitoring and surveillance, Community structure, Network measures, Information cascade, Influence, and Homophile, Recommendation, Sentiment analysis, Social spammer detection, etc. In this research, influencers on a trending topic are detected on Twitter.

Twitter has become a popular platform in which users can share information. Users can post statuses

with the limitation of 280 characters, which are called statuses or tweets. The tweet contains plain text, URLs, images, mention to other users (the symbol "@"), and hashtags (a keyword or a phrase used to describe a topic or a theme by placing the symbol "#" in front of them). When a user x is following user y , user y is a followee of user x and user x is a user y ' follower [2]. There is a feature namely reply, a tweet of user can be replied or commented through one of his tweets. There are two powerful features namely sharing or retweet and favorite or like. A retweet is simply reposted of another user's tweet and allows people to spread discussion easier. Favorites are used when users like a tweet.

The rise of social media leads to emerging a temporal set of users who are participated in a topic [3]. A word or a hashtag that is being discussed more than others with a high frequency is called a trending topic [4]. In this research work, Twitter users' activities are characterized, and then influential users associated with a trending topic were detected by using a link analysis approach.

Influence takes place when opinions, behaviors, or emotions of a person are changed by others. Understanding social influence in Online Social Networks (OSNs) is vital and can result in many appealing applications, for example, disease spread, viral marketing, searching, expertise recommendation, information propagation, etc.[5]. Detecting social influence is still allowing open since that has been proposed because of social influence lacks a general definition [6].

A social network is a network composed of actors and the relationships between these actors [7]. Most of the previous researches detected influential users using follow relationships to provide the popularity of the users. However, follow relationships contain active and inactive users who do not provide the dissemination of information. Recently, the researchers considered the usefulness of retweet or mention relationships. In this paper, we used retweet, mention, and reply relationships to create a trending topic network.

Actors denoted as Twitter user accounts and relationships denoted as a reply, mention, and retweet. Every actor in the network can be connected to one or more other actors to form eventually a structure. By studying social networks, the expected benefit can be achieved with the research questions like finding influential users.

Influential users can accelerate spreading and hindering the messages by the power of fans in the microblog networks [33]. If influential people are predicted, unwanted information will be hindered. And a company wants to introduce their product at a low cost to the public, the influential person can promote their product in a short time. This paper aims to detect influential users in a common interesting topic by analyzing activities and communication relationships of users. To achieve this aim we emphasized the following objectives:

- to extract the retweet, mention and reply relationships between Twitters' users in a trending topic
- to build the trending topic community graph using relationships
- to calculate users' personalization weight using profile features
- to find influential users in a trending topic community using link analysis approach

The main contribution of this paper is to detect the influential users on a trending topic by integrating network topology from the communication relationship among users and publishing and interaction ability of users. The next Section 2 describes related work, and Section 3 discusses the process of data collection, preprocessing, and the

methods we used to identify influencers. Section 4 explains experimental results related to network structure and analysis of influence. Section 5 includes the conclusion of the paper and future work regarding the study.

2. Related work

At first, a twitter account that has a large number of followers is identified as an influential user [8]. Nowadays, researchers used user content and network structure to identify influential users. According to the investigation of previous researches, influence measurement can be classified as Local Measure, Shortest Path Measure, Iterative Calculation Measure, Coreness Measure, Machine-Learning Algorithms, and others [33]. In Local Measure, the number of links connecting to a user is denoted as the degree of that user. In an undirected network, a user with high degree centrality is identified as an authorized user for information dissemination. In the directed network, in-degree centrality mostly mentions the user's popularity, and out-degree centrality assumes the social ability of a user. In Shortest Path Measure, Closeness centrality, Betweenness centrality, and Katz algorithms can find the importance and influence of nodes in the network. In Iterative Calculation Measure, PageRank like algorithms and Eigenvector centrality were used to calculate user influence or discover a vital node in a graph. This measurement considers not only the score of influence of connected users but also the direct links among users. The coreness measure support that the position of users is vital than direct connection in reachability and influence diffusion. To predict influential users, machine-learning algorithms are applied. Otherwise, researchers have introduced hybrid methods to detect influencers.

P. E. Nalwoga Lutu [9] argued that only follow relationships cannot access useful information such as influencers because of a lack of intensity of interaction between users. They proved that mention relationships get more information about influencers compared with the following relationships. However, only mention relationships do not provide the intensity of interactions between users.

A. K. Bhowmick, M. Gueuning, J. C. Delvenne, R. Lambiotte, and B. Mitra [10] argued existing methods mainly depended on the network structure only for the influential user detection. They proposed an algorithm considering both the temporal retweet relationships and local structural information to find the influential users set in the network. However, only retweet relationships cannot provide a network structure.

M. W. Al-Nabki, E. Fidalgo, E. Alegre, and L. Fernández-Robles [11] proposed a ToRank algorithm for ranking and detecting influential hidden services in the Tor network. They used graph structure and detected influential nodes by considering their degree centrality as initial weight and combining their followers' weight and followee's weight.

Z. Z. Alp and Ş. G. Ögüdücü [12] identified future influencers on the topic on Twitter using topic-specific features and user-specific features. They focused on the idea of sharing information using a retweet relationship between users. In our work, communications between users were extracted from retweet, reply, and mention relationships from a tweet. Z. Jianqiang, G. Xiaolin, and T. Feng [13] proposed UIRank (User Influence Rank) algorithm to find the influential users in the following network in Sina Weibo. They take into account both tweet influence and network influence for each user. Tweet influence is measured by the number of comments and retweets count, and network influence is measured by degree, betweenness, and closeness centrality of each user. However, they cannot categorize influencers for each topic. In our work, we emphasized to detect the most influencers in a topic.

K. Almgren and J. Lee [14] proposed content-based influence measurement (CIM) that considers both context and content to predict influential users in the social network. They showed that followers have different interaction strengths and content alone does not predict influential users. However, they ignored that calculation of the centrality takes time quite slow, their approach only worked on the network that has a significant number of short loops. P. I. B. and A. A. Y. E. E. L. V. S. Zamyatina [15] proposed a method to develop the influence index in an expertise area that provides steady results on the data stream. They used the communication efficiency measurement method because of strong performance. However, they left to express the sentiment response of users.

F. Erlandsson, P. Bródka, A. Borg, and H. Johnson [16] proposed association rule learning to detect the relationship between users using the ECLAT Algorithm. Their approach estimated the users who influence other users in new groups. There is no significant difference ranking result compared with PageRank and Degree. However, their system cannot pull out rules for the biggest pages in their dataset.

N. A. B, T. F. Gharib, M. Hamdy, and Y. Afify [17] proposed a new model (Influence Ranking Model (IRM)) to find information spreaders based on K-shell decomposition. Graphical approaches, non-graphical approaches, and hybrid approaches can

detect influential users. Their model consists of Influence Weight Calculation Phase, Node Pruning Phase, and Node Ranking Phase. Their proposed IRM significantly suppresses weighted k-shell decomposition methodology on uniquely ranking SN nodes.

S. Peng, G. Wang, and S. Yu [18] proposed a voting algorithm to identify the influential nodes in the mobile social network studying the communication behaviors of mobile users, the influence score is computed. A. E. Cano, S. Mazumdar, and F. Ciravegna [19] identified topical influencers with a retweet relation. They proposed Topic-Entity PageRank to find influencers.

3. Methodology

In this part, we present the methods we used, collecting data, preprocessing, and how to detect influential users in a trending topic. Fig. 1 shows the structure of the proposed influential detection system.

3.1 PyQuery

Python search engine (PyQuery) enables Python developers to search for Python Packages and Modules (code) and encourage them to take benefit of an important software development practice, code reuse [20]. PyQuery facilitates the search process and improves the query results. PyQuery allows users to perform a generic keyword search without limitations in the input.

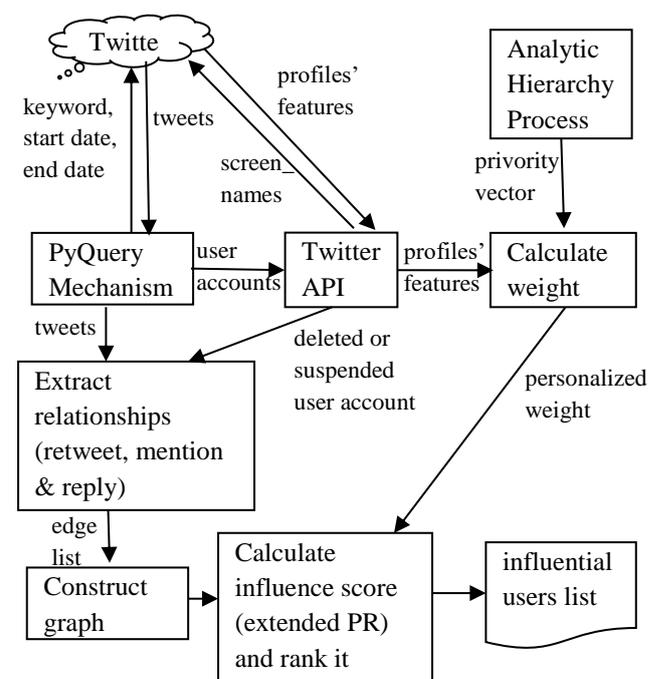


Figure. 1 Structure of proposed influential users detection system

3.2 Twitter API

Twitter allows accessing the public information of tweets and users. Twitter API¹ is a powerful tool to get data from non-protected users.

3.3 Analytic hierarchy process (AHP)

Analytic Hierarchy Process (AHP) is a powerful tool for solving decision making and can help the decision-maker to choose the best decision and to set priorities [21, 22]. A series of pairwise comparison is achieved by reducing complex decisions, and then synthesizing the results, the AHP helps to achieve both subjective and objective views of a decision.

3.4 Pagerank (PR)

The PageRank algorithm was used in the prototype of Google's search engine [23–25]. The aim is to calculate the rank or the popularity of a webpage, depended on the interrelationship of the web. PageRank is also called Link Analysis Approach because it calculates the Web Pages' score according to the Web Page's in-link and out-link. The formula is as follow:

$$PR(u) = c \sum_{v \in N(u)} \frac{PR(v)}{L(v)} \quad (1)$$

where $v \in N(u)$ denotes the set of pages v that point to page u . $L(v)$ represents page v 's outgoing links. A parameter c is used for normalization. $PR(u)$ and $PR(v)$ are page u and v ' rank scores, and initially assigned 1.

In a simple PR algorithm, there were many problems according to dangling nodes and reducible web graphs. The extended PR algorithm is

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in N(u)} \frac{PR(v)}{L(v)} \quad (2)$$

where $v \in N(u)$ denotes the set of pages v that point to page u . $PR(u)$ and $PR(v)$ are page u and v ' rank scores, and initially assigned $1/N$. N is the number of web pages. A damping factor d is usually set to 0.85. $L(v)$ is the number of outgoing links of page v .

3.5 Data collection and preprocessing

Previous research analyzed Online Social Network (OSN) graphs using partial network data [26]. Because data collection from the whole network is difficult. Full network data allows detailed

description and social structures analysis [27]. But, full network data can also be costly and difficult to collect. Twitter microblogging has been extensively used in several domains. This paper focused on data generated from Twitter using data access mechanisms. Usually, data from Twitter are crawled using Twitter API (Application Programming Interface). Many researchers, companies, and governmental institutions used the API service to extract knowledge from social media data [28].

Data acquisition from real-world Twitter data is a challenging problem in social media mining because Twitter does not allow to share tweets. In this paper, a dataset was created related to a trending topic. We used Python's jQuery (PyQuery) that allows making queries on the XML document.

Social media grows all the time and is dynamic. So, the information obtained at the collecting time is used. We collected 22037 tweets related to a trending topic namely "#rp19" from May, 1st 2019 to May 10th, 2019. "#rp19" (re: pubblica 2019) is a trending topic related to the thirteenth edition of the conference (some call it a festival) that happened on May 6, 7, and 8 at station Berlin. The number of followers, friends, statuses, and favorites of each Twitter user who contains in these tweets is collected to analyze the activities of each user using Twitter API.

After data collection, we made the following processes before constructing a graph.

- Edge list is extracted from the tweets with the reply, retweet, and mention relationships between Twitter account to create a network graph. If user A retweets user B ' post, we add a directed edge from A to B and so on. The linking relationships between Twitter accounts define as a mathematical object called a **graph**. At first, 7319 unique users (nodes) and 17281 edges are collected.
- But there were deleted or suspended user accounts, so these users were dropped.
- Isolated nodes and self-loops are removed.
- Duplicated edges are combined into a single edge.

Finally, 6065 nodes and 16234 edges were collected for a trending topic #rp19 see in Table 1.

Table 1. Number of users and edges in #rp19

	Before	After	Eliminate
Number of users	7319	6065	1254
Number of edges	17281	16234	1047

¹ <https://developer.twitter.com/en/docs>

3.6 Influential users detection

Trending topic network is defined as a directed graph, $G(V, E)$, where vertices set V represented as Twitter user accounts in a trending topic community, edges set E represented as retweet, mention, and reply relationships among users u to v . PageRank algorithm is used for reference. The extended formula is shown in Eq. (5). A_{uv} represents as follow:

$$A_{uv} = \begin{cases} 1 & \text{if user } v \text{ has a relationship to user } u \\ 0 & \text{otherwise} \end{cases}$$

$n \times n$ diagonal connectivity matrix D with the out-degree of each user is following:

$$D_{uu} = d(v_u) \quad (3)$$

M is the transition matrix of a random walk can be defined as follow:

$$M := (D^{-1}A)^T \quad (4)$$

The formula to calculate each user influence is as follow:

$$PR(u) = \alpha \times \sum_{v \in N(u)} M \times PR(v) \times W_{uv} + \frac{1-\alpha}{N} \vec{1} \quad (5)$$

where α is a teleportation parameter or damping parameter, $0 \leq \alpha < 1$ and set to 0.85 because of quickly convergence over iteration [25], $PR(u)$ and $PR(v)$ is the PageRank value of user u and v , initially assigned 1, N defined as the number of users, $\vec{1}$ is the column vector of length N containing only one, $v \in N(u)$ denotes that v is the set of users who has a relationship to user u and W_{uv} is personalization weight of user u to v . Influential user are sorted ascending based on correspondent resulted PageRank score.

Given two individual u and v are directly connected in trending topic network, u can alter the opinion of v . W_{uv} is defined as the personalization weight of user u on its neighbor nodes v . The weight is calculated based on: How many users are followed by this user? How many users are friends with this user? How many statuses are posted by this user? How many like or votes are acknowledged by this user? The weight assigned to each user is important to address the publishing and interaction ability of OSN users. The analytic hierarchy process allows the weight coefficients as numerical form.

The personalization weight W_{uv} for each user can be determined by the following Eq. (6):

$$W_{uv} = \log(a \cdot follower_u + b \cdot friends_u + c \cdot status_u + d \cdot favorite_u + 1) \quad (6)$$

where $follower_u$, $friends_u$, $status_u$, and $favorite_u$ are the number of followers, friends, status, and favorites of node u respectively, a , b , c , and d are coefficients with different weights based on the importance of the ability of followers, friends, statuses, and favorites, and they are calculated using the Analytic Hierarchy Process (AHP). These features are extracted from each Twitter account's profile who is contained in the tweets about a trending topic. Follower counts indicate the popularity of the Twitter account. If a user has the more list, the more popular that user. Friend counts indicate that how many people communicate directly to show user's social connections level. Status counts indicate how many statuses the user post to represent the user's activeness [29]. Favorite counts indicate "upvote" or "like" when a user acknowledge the post content.

All of the above user's profile features can be ranked as Table 2. We built a pairwise comparison matrix using Saaty's 1 to 9 scales [30] based on these characteristics see in Table 3. For example, if we consider that follower is moderately important than friend, the follower-friend comparison cell will fill the value 3. By reciprocally, friend-follower cell will fill the value, $1/3$. The follower-follower cell will fill 1 when follower is equally important. The follower is extremely important than favorite, the follower-favorite comparison cell will fill the value 9. By reciprocally, favorite-follower cell will fill the value, $1/9$, and so on.

Since the value of the proportion of inconsistency Consistency Ratio (CR) is 0.062, which is less than 0.10, so the judgments matrix is reasonably consistent. The priorities of the criteria such as follower, friend, status and favorite were calculated with the following steps: (1) add each column's value, (2) divide each cell by the total of the column (normalization matrix) (3) average value of each row (final priorities).

And then the resulted weighting coefficients are shown in Table 4. To counter skewness towards the very high number of followers, friends, status, and favorite, the logarithm function is used. When the log argument's value is zero, the arithmetic error occurs, so the number one is added.

Table 2. Rank the criteria

Follower	A very high degree of ability
Friend	High degree of ability
Status	An average degree of ability
Favorite	A certain degree of ability

Table 3. Comparison matrix based on users' profile features

Characteristic	Follower	Friend	Status	Favorite
Follower	1	3	7	9
Friend	1/3	1	5	7
Status	1/7	1/5	1	3
Favorite	1/9	1/7	1/3	1

Table 4. Priority vector

Weighting coefficients	Priority Vector
a	0.57
b	0.29
c	0.1
d	0.044

4. Experimental results

In this section, the experiments are presented concern with the network graph nature and the performance of the system. After the structure of the graph is described, the performance of the system measures in terms of precision, recall, and F1-measure in information retrieval.

4.1 Structure of the network graph

In this work, the python programming language is used. A graph-based approach is used to describe the network's structure. The main different parts from other's work, we detected edges with retweet, mention, and reply relationships between users on a trending topic. At first, 7319 vertices and 17281 edges are collected. When a user *A* mentions user *B* many times, the duplicated edges can occur. Duplicated edges are deleted because of *degree* may be inaccurate. So, the duplicated edges were combined into a single edge. Self-loops are removed because out-degree and in-degree of the graph can be affected. Deleted or suspended users are removed because they are not currently active. Finally, the #rp19 network contained 6065 vertices and 16234 edges.

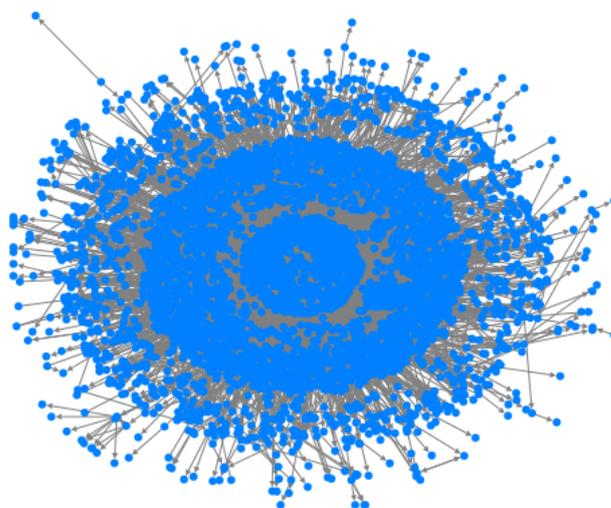


Figure. 2 A trending topic #rp19 graph with 6065 vertices 16234 edges

Fig. 2 shows the graph of a trending topic #rp19 network. Nodes represent blue color and edges represent a black arrow. The #rp19 network consists of 6065 users (nodes) and 16234 relationships (edges), with a maximum geodesic distance (diameter) of 11 and graph's density is 4.41×10^{-4} .

Fig. 3 shows the density of the number of followers, friends, statuses, and favorite of each Twitter user in the "#rp19" network.

According to Fig. 3, a very high number of followers, friends, status, and favorite, the distribution of resulted personalize weight for all users is skewness. So, personalize weights are normalized using the logarithm function.

4.2 Evaluation of the proposed method

In this section, we describe how will our proposed method meets the influential users in a trending topic. Our proposed method compared with four influencer detection approaches. As a reference for cross-validation, the methods proposed by researchers that have been used. The top *k* influential nodes are composed as a reference set where $k = 10, 20, 50, 100, 200, 500, 800, 1000, \text{ and } 1500$. The precision, recall, and F1-Measure are tested on the reference set.

Our proposed method compared with four methods of finding influential individuals: the URank [13] that calculates tweet influence and network influence; the FansRank [31] that determines user's influence based on the number of fans; the ToRank [11] that calculates nodes' influence by considering degree centrality as initial weight and then combining its followers' weight and followee' weight; the RetweetInfluence [32] that calculates user influence by retweet count and the proposed method. Top *k* sets of influential nodes

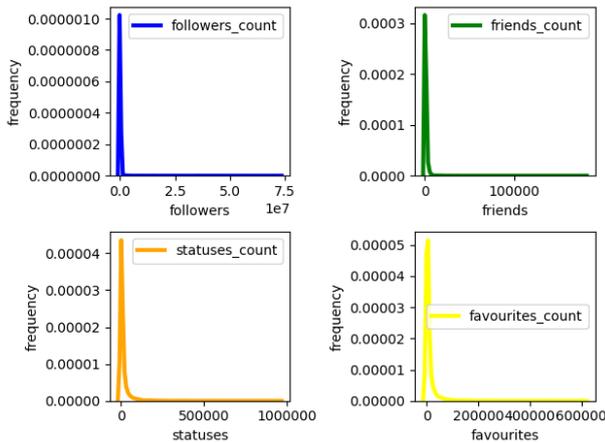


Figure. 3 Densities of the number of followers, friends, statuses, and favorites of each twitter user in the “#rp19” network

form each method are identified as $S_{propose}$, S_{fans} , S_{ret} , $S_{storank}$, and $S_{suirank}$. Let the set of the retrieved top k users of each five methods be S_{iudm} . The reference set for any two methods ($N=2$) is defined as S_2 in the following Eq. (7):

$$S_2 = (S_{propose} \cap S_{fans}) \cup (S_{propose} \cap S_{ret}) \cup (S_{propose} \cap S_{storank}) \cup (S_{propose} \cap S_{suirank}) \cup (S_{fans} \cap S_{ret}) \cup (S_{fans} \cap S_{storank}) \cup (S_{fans} \cap S_{suirank}) \cup (S_{ret} \cap S_{storank}) \cup (S_{ret} \cap S_{suirank}) \cup (S_{storank} \cap S_{suirank}) \quad (7)$$

Then the precision is defined as follow:

$$P_{iudm} = \frac{|S_{iudm} \cap S_2|}{|S_{iudm}|} \quad (8)$$

The recall is defined as follow:

$$R_{iudm} = \frac{|S_{iudm} \cap S_2|}{|S_2|} \quad (9)$$

The F1-Measure is defined as follow:

$$F1_{iudm} = \frac{2 \times P_{iudm} \times R_{iudm}}{P_{iudm} + R_{iudm}} \quad (10)$$

The reference set for any three ($N=3$) and any four ($N=4$) methods are also defined where N means the number of methods. Top- k influential user sets of each approach are gathered on a specific trending topic and precision, recall, and F1-Measure of each approach was calculated on any two ($N=2$), three ($N=3$), and four ($N=4$) reference sets. At any five reference sets, there are the same results of the precision and recall of all methods, so $N=5$ experiment is absent.

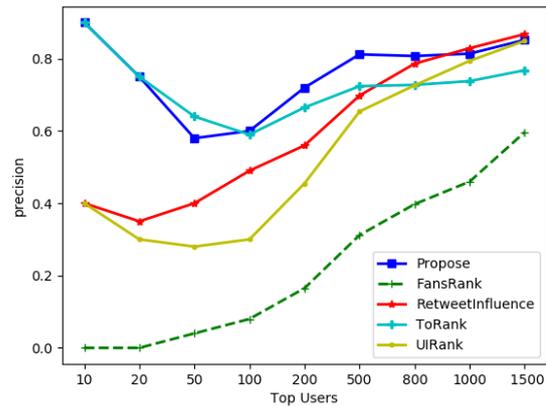


Figure. 4 Each method’s precision on a reference set $N=2$

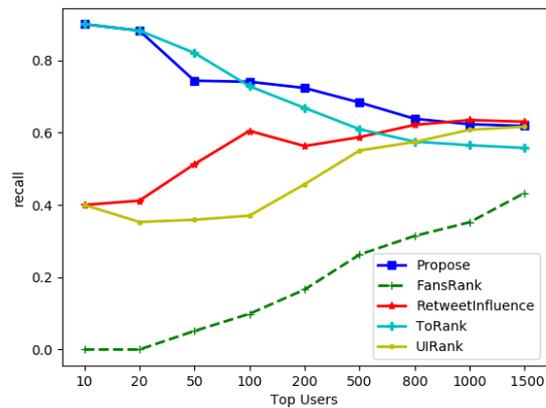


Figure. 5 Each method’s recall on a reference set $N=2$

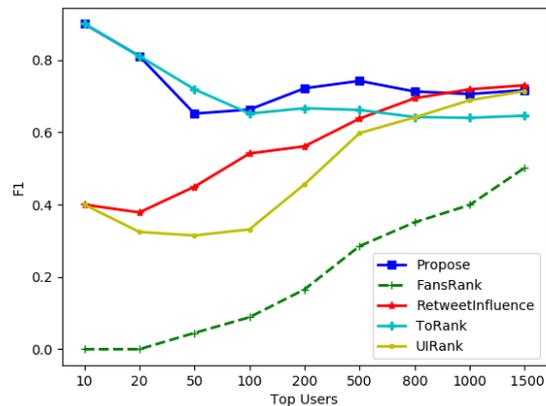


Figure. 6 Each method’s F1-measure on a reference set $N=2$

Fig. 4, 5, and 6 show precision, recall, and F1-measure of each method on a reference set of any two methods. The influential users of our proposed method and ToRank are the same up to the top 20. Between the top 20 and top 90, ToRank has high precision, recall, and F1-measure. However, from the top 90, our proposed method is higher than ToRank. RetweetInfluence, UIRank, and FasnRank are followed by our proposed method. FansRank has low

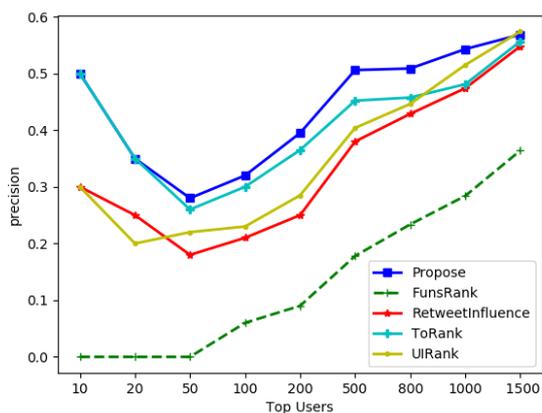


Figure. 7 Each method’s precision on a reference set N=3

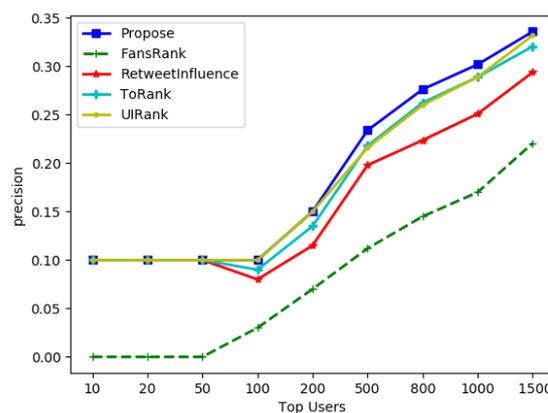


Figure. 10 Each method’s precision on a reference set N=4

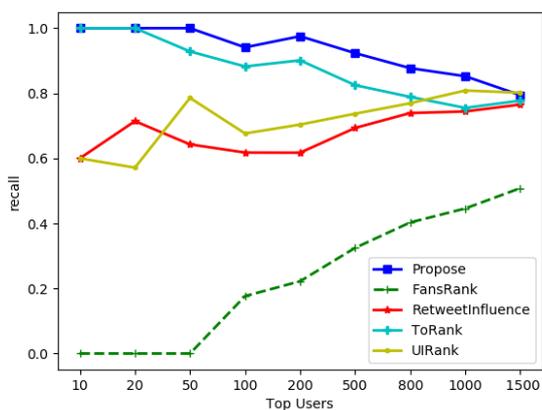


Figure. 8 Each method’s recall on a reference set N=3

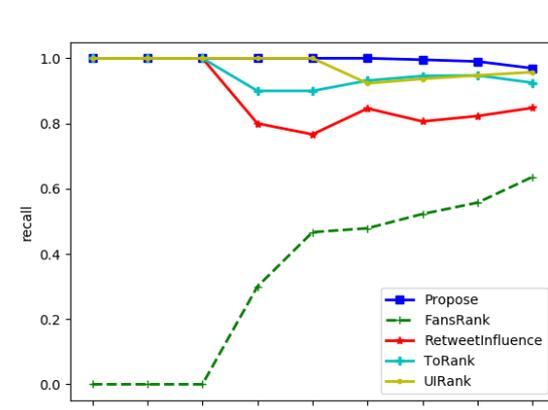


Figure. 11 Each method’s recall on a reference set N=4

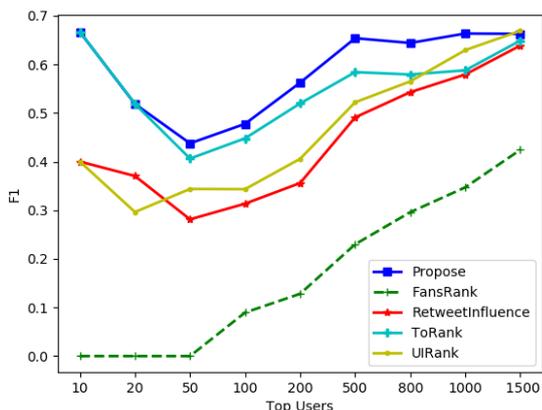


Figure. 9 Each method’s F1-measure on a reference set N=3

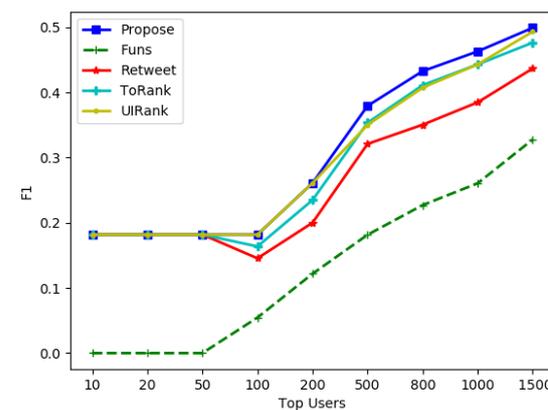


Figure. 12 Each method’s F1-measure on a reference set N=4

precision, recall, F1 which means the user who has a huge number of followers may not be influential. The method UIRank has also low precision, recall, and F1 that shows tweet influences are not enough to find influential users. The precision of our proposed method is 0.8 at the top 10 which means the probability of the influential users retrieved that are relevant to the reference set is 80%.

Fig. 7, 8, and 9 show precision, recall, and F1-measure of each method on a reference set of any three methods. Our proposed approach outperforms all of the other measurements. The proposed approach and ToRank have the same precision, recall, and F1 value up to the top 20. Between the top 20 and top 1500, our proposed approach’s precision, recall, and F1 are better than ToRank. FansRank also has the

lowest than other measurements. RetweetInfluence has also low precision that shows only the high value of tweets cannot increase user influence. The recall of our proposed method at top10 and top 20 is one which means the probability of relevant influential users in a trending topic is retrieved 100%.

Fig. 10, 11, and 12 show precision, recall, and F1-measure of each method on a reference set of any four methods. The precision, recall, and F1 of each method up to the top 50 are the same except for FunsRank. But between the top 50 and top 200, the precision, recall, and F1 of our proposed approach and ToRank are the same and higher than other measurements. From the top 200, the precision, recall, and F1-measure of our proposed method are highest than other methods. FansRank also has the lowest than other measurements. The recall of our proposed method is one that means the relevant influential users in a trending topic is retrieved.

According to Fig. 4, 7, and 10, the precision of all methods shows a downward trend when the increase of N, because of the nodes in the reference set decreases when N increase. When k increase, the precision of all methods is rise.

5. Conclusion

This paper emphasizes on finding influencers based on communication relationships on a trending topic. The main different parts from other's work, we detected edges with retweet, mention, and reply relationship between users on a trending topic. We created a new dataset related to a trending topic with 6065 nodes and 16234 edges. We faced duplicate edges, self-loops, and deleted or suspended user accounts. Duplicated edges are converted into single one and self-loops are removed because that can be inaccurate the degree of the graph. By removing deleted or suspended users, we cannot regard those users as influential users.

In Twitter, influence takes place not also the spread of tweets but also the interaction of users. By using this hypothesis, we proposed the influential user detection method. The number of followers, friends, statuses, and favorites from the users' profile and in-link and out-link of #rp19 network graph is used to calculate the influence score of the users. Using a real Twitter dataset, the results of the experiment show that our proposed approach can identify influential users than other related algorithms using evaluation measures such as precision, recall, and F1-measure. Only a large number of followers (FunsRank) is not enough to extract influential users on a trending topic. The influential nodes are not detected only considering

the degree centrality of nodes in a graph. In the future, the sentiment of tweets on a trending topic will take into account to measure user influence.

Conflicts of Interest

The authors declare no conflict of interest with respect to the research, authorship, and/or publication of this article.

Author Contributions

Myat Mon Oo contributed to the design and implementation of the research, the analysis of the results, and the writing of the manuscript. Dr. May Thu Lwin supervised this research.

Acknowledgments

I would like to express my very great appreciation to Dr. May Thu Lwin for her valuable and constructive suggestions during the planning and development of this research work.

References

- [1] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in Social Media Research: Past, Present and Future", *Inf. Syst. Front.*, Vol. 20, No. 3, pp. 531–558, 2018, doi: 10.1007/s10796-017-9810-y.
- [2] S. Madria and T. Hara, "Big Data Analytics of Social Networks for the Discovery of 'Following' Patterns", In: *Proc. of Big data Anal. Knowl. Discov. 17th Int. Conf. DaWaK 2015 Val. Spain, Sept. 1–4, 2015 Proc.*, Vol. 9263, pp. 123–135, 2015, doi: 10.1007/978-3-319-22729-0.
- [3] L. Recalde, D. F. Nettleton, R. Baeza-Yates, and L. Boratto, "Detection of trending topic communities: Bridging content creators and distributors", In: *Proc. of HT 2017 - Proc. 28th ACM Conf. Hypertext Soc. Media*, pp. 205–213, 2017, doi: 10.1145/3078714.3078735.
- [4] P.-M. Caleffi, "The 'hashtag': A new word or a new rule", *Skase J. Theor. Linguist.*, Vol. 12, No. 2, pp. 46–69, 2015.
- [5] J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu, "Social network user influence sense-making and dynamics prediction", *Expert Syst. Appl.*, Vol. 41, No. 11, pp. 5115–5124, 2014, doi: 10.1016/j.eswa.2014.02.038.
- [6] S. Peng, A. Yang, L. Cao, S. Yu, and D. Xie, "Social influence modeling using information theory in mobile social networks", *Inf. Sci. (Ny)*,

- Vol. 379, pp. 146–159, 2017, doi: 10.1016/j.ins.2016.08.023.
- [7] C. C. Aggarwal, “An Introduction To Social Network Data Analytics”, *Soc. Netw. Data Anal.*, pp. 1–15, 2011, doi: 10.1007/978-1-4419-8462-3.
- [8] A. Leavitt, E. Burchard, D. Fisher, S. Gilbert, W. Ecology, and P. Pub, “The Influentials: New Approaches for Analyzing Influence on Twitter”, *Web Ecol.*, Vol. 04, No. September, pp. 1–18, 2009, [Online]. Available: <http://www.webecologyproject.org/wp-content/uploads/2009/09/influence-report-final.pdf>.
- [9] P. E. Nalwoga Lutu, “Using twitter mentions and a graph database to analyse social network centrality”, In: *Proc. of 2019 6th Int. Conf. Soft Comput. Mach. Intell. ISCOMI 2019*, pp. 155–159, 2019, doi: 10.1109/ISCOMI47871.2019.9004313.
- [10] A. K. Bhowmick, M. Gueuning, J. C. Delvenne, R. Lambiotte, and B. Mitra, “Temporal Sequence of Retweets Help to Detect Influential Nodes in Social Networks”, *IEEE Trans. Comput. Soc. Syst.*, Vol. 6, No. 3, pp. 441–455, 2019, doi: 10.1109/TCSS.2019.2907553.
- [11] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and L. Fernández-Robles, “ToRank: Identifying the most influential suspicious domains in the Tor network”, *Expert Syst. Appl.*, Vol. 123, pp. 212–226, 2019, doi: 10.1016/j.eswa.2019.01.029.
- [12] Z. Z. Alp and Ş. G. Öğüdücü, “Influence Factorization for identifying authorities in Twitter”, *Knowledge-Based Syst.*, Vol. 163, pp. 944–954, 2019, doi: 10.1016/j.knosys.2018.10.020.
- [13] Z. Jianqiang, G. Xiaolin, and T. Feng, “A new method of identifying influential users in the micro-blog networks”, *IEEE Access*, Vol. 5, No. c, pp. 3008–3015, 2017, doi: 10.1109/ACCESS.2017.2672680.
- [14] K. Almgren and J. Lee, “Who Influences Whom: Content-Based Approach for Predicting Influential Users In Social Networks”, In: *Proc. of Int. Conf. Adv. Big Data Anal.*, pp. 89–95, 2015.
- [15] P. I. B. and A. A. Y. E. E. L. V. S. Zamyatina, “Estimation of social network user’s influence in a given area of expertise”, In: *Proc. of J. Phys. Conf. Ser.*, Vol. 803, No. 1, 2017, doi: 10.1088/1742-6596/755/1/011001.
- [16] F. Erlandsson, P. Bródka, A. Borg, and H. Johnson, “Finding influential users in social media using association rule learning”, *Entropy*, Vol. 18, No. 5, 2016, doi: 10.3390/e18050164.
- [17] N. Ayman, T. F. Gharib, M. Hamdy, and Y. Afify, “Influence Ranking Model for Social”, In: *Proc. of In International Conf. on Advanced Machine Learning Technologies and Applications*, 2020, pp. 928–937, doi: 10.1007/978-3-030-14118-9.
- [18] S. Peng, G. Wang, and S. Yu, “Mining mechanism of top-k influential nodes based on voting algorithm in mobile social networks”, In: *Proc. of - 2013 IEEE Int. Conf. High Perform. Comput. Commun. HPCC 2013 2013 IEEE Int. Conf. Embed. Ubiquitous Comput. EUC 2013*, pp. 2194–2199, 2014, doi: 10.1109/HPCC.and.EUC.2013.314.
- [19] A. E. Cano, S. Mazumdar, and F. Ciravegna, “Social influence analysis in microblogging platforms-A topic-sensitive based approach”, *Semant. Web*, Vol. 5, No. 5, pp. 357–372, 2014, doi: 10.3233/SW-130108.
- [20] S. K. Imminni, “Pyquery : A Search Engine for Python Packages and Modules”, 2015.
- [21] T. L. Saaty, “How to handle dependence with the analytic hierarchy process”, *Math. Model.*, Vol. 9, No. 3–5, pp. 369–376, 1987, doi: 10.1016/0270-0255(87)90494-5.
- [22] T. Saaty and L. Vargas, “Models, methods, concepts & applications of the analytic hierarchy process”, in *International Series in Operations Research & Management Science*, Vol. 175, pp. 1–20, 2012.
- [23] S. Brin, R. Motwani, L. Page, and T. Winograd, “What can you do with a Web in your Pocket?”, *IEEE Data Eng. Bull.*, Vol. 21, pp. 37–47, 1998.
- [24] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” *World Wide Web Internet Web Inf. Syst.*, Vol. 54, No. 1999–66, pp. 1–17, 1998, doi: 10.1.1.31.1768.
- [25] A. N. Langville and C. D. Meyer, “Deeper Inside PageRank”, *Internet Math.*, Vol. 1, No. 3, pp. 335–380.
- [26] S. Ye, J. Lang, and F. Wu, “Crawling online social graphs”, In: *Proc. of Adv. Web Technol. Appl. - Proc. 12th Asia-Pacific Web Conf. APWeb 2010*, 2010, pp. 236–242, 2010, doi: 10.1109/APWeb.2010.10.
- [27] R. a Hanneman and M. Riddle, “Introduction to Social Network Methods”, *Riverside, CA Univ. California, Riverside. On-line Textb.*, Vol. 46, No. 7, pp. 5128–30, 2005, doi: 10.1016/j.socnet.2006.08.002.
- [28] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, “Is the sample good enough? Comparing data from twitter’s streaming API with Twitter’s

- firehose”, In: *Proc. of 7th Int. Conf. Weblogs Soc. Media, ICWSM 2013*, pp. 400–408, 2013.
- [29] J. V. Cossu, V. Labatut, and N. Dugué, “A review of features for the discrimination of twitter users: application to the prediction of offline influence”, *Soc. Netw. Anal. Min.*, Vol. 6, No. 1, 2016, doi: 10.1007/s13278-016-0329-x.
- [30] E. Mu and M. Pereyra-Rojas, “Understanding the Analytic Hierarchy Process”, in *Practical Decision Making*, No. 2012, p. 110, 2017
- [31] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?”, In: *Proc. of 19th Int. Conf. World Wide Web, WWW '10*, pp. 591–600, 2010, doi: 10.1145/1772690.1772751.
- [32] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, “Measuring user influence in twitter: The million follower fallacy”, In: *Proc. of ICWSM 2010 - Proc. 4th Int. AAAI Conf. Weblogs Soc. Media*, pp. 10–17, 2010.
- [33] M. A. Al-Garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, G. Mujtaba, M. U. Shahid Khan, and S. U. Khan, “Analysis of online social network connections for identification of influential users: Survey and open research issues”, *ACM Comput. Surv.*, Vol. 51, No. 1, 2018, doi: 10.1145/3155897.