Methods of Identification of Leaders and Followers in Online Social Networks - A Survey

[1] Rajeshri Puranik, [2] Sharad Pokley
[1] Assistant Professor Priyadarshini Institute of Engineering and Technology, Nagpur
[2] Associate professor Kavikulguru Institute of Technology and Science.,Ramtek

Abstract - Now a day's Social Network has a great influence on our lives. Directed links in social media could represent anything from intimate friendships to common interests, or even a passion for breaking news or celebrity gossip. Such directed links determine the flow of information and hence indicate a user's influence on others—a concept that is crucial in sociology and viral marketing. Identifying leaders and followers in online social networks is important for various applications in many domains such as advertisement, community health campaigns, administrative science, and even politics.

Here we study the role of actors and a set of relations between actors i.e. leaders and followers, and there relations. For our study purpose we have selected two well-known popular SNS(Social Networking Sites) face book and twitter. Also discussed different methods, (LUCI Model, Leader Rank Approach, Binary Approach , Integrated Text Mining Model , Content based Metrics and Social Network Metrics Based Approach) for the identification of leaders and followers

Keywords:--- Directed Links, Influence, Network Coefficient, Social Actors, Social Network.

I. INTRODUCTION

The 'social network' consists of a set of actors and a set of relations between actors.A social network is a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics.Social networks and the analysis of them is an inherently interdisciplinary academic field which emerged from social psychology, sociology, statistics, and graph theory.

*A **social networking site*** is a platform to build social networks or social relations among people who share similar interests, activities, backgrounds or real-life connections. The mathematical analysis of network is successfully used since many decades to an engineered system. Due to the advancement of the internet and social networks ,the underlying assumptions in the analysis of networks have changed dramatically. The plenty of such social network data and the multiplicity of social networks is changing the face of research of social network. This change represents both opportunities and challenges for mathematical and statistical modeling.

The contribution of mathematical sciences to the new opportunity is the significant amount of recent work focused on the development of random graph models that capture some of the qualitative properties observed in large scale network data. Many attributes of social network are easily understood by using the newly developed mathematical models. Degree of connectivity is one of the attributes of network which makes world small, in which very distant parts of a population are connected via short paths. These short paths are easy to find by using decentralized search algorithms.

Now a days social networking sites have very much importance in our daily life. .There are various entities which affects these social networking sites like Leaders and followers, Whistle blowers, and so on . In this paper we have discussed the importance of Opinion Leaders and identification methods of these leaders and followers in social networking sites viz Face book & Twitter.

The outline of the paper is as follows. In section II we have discussed the need of opinion leaders in social networking sites. Importance of identification of leaders is discussed in section III. Section IV includes the various methods used for identification of leaders and followers. Section VI concludes the paper.

II. NEED OF LEADERS IN SOCIAL NETWORKING SITES:

Leaders- Leaders in social networks are users whose opinions are highly influential on those of others.

Kotler defined Leaders as "people who can influence members in the social community because of special techniques, knowledge, personalities and other uniqueness"

Rogers describes opinion leaders as people with
- high social participation
- high social status
- and a high social responsibility

According to Robertson leaders are:
- more directive
- more innovative
- and more professional

In 1944 Lazarsfeld & Katz and his team has found that communication does not directly flow to the mass but is actually interpreted first by opinion leaders and then forwarded to the rest of the people. They have described this process as the "Two-Step-Flow of Communication".

The Two-Step-Flow of Communication tells the following points:
- Information is transferred not only by the (mass) medium but also through interpersonal communication
- There are people between the medium and the interpersonal communication network which are called opinion leaders
- The influence of such opinion leader is significantly larger than that of the medium

Importance of Identifying leaders in online social networks

Leaders are identified in online social networks as they pass on new ideas to the rest of the population in a very effective way, even to the introvert people who shy away from learning new things. Identifying leaders in online social networks is important for various applications in varied fields such as advertisement, community health campaigns, administrative science, and even politics [3].

In advertisement field leaders are given free samples of the products for mouth publicity. They recommend the product to the masses. Due to their effective publicity and recommendation, the product is adopted by the population.

In community health campaigns the leaders make the population aware of smoking hazards. Identify leaders to spread awareness of HIV prevention and anti-smoking helps to reach large number of people than to apply them to random individuals. New ideas are introduced to leaders who in turn pass on the ideas to the rest of the population.

Those individuals who keep themselves away from the authorities due to their shy nature benefit due to this system. In administration it is necessary to identify the leaders and followers to make an effective team for better work performance. This identification will further become beneficial in developing the product. During US election community leaders play an important role of the preparing and organizing voters to increase the turnout in election.

In the pharmaceutical industry leaders, physicians with specialized knowledge are identified for giving knowledge of new treatments. For this purpose a significant portion of marketing budget is spent. After gaining knowledge these leaders and physicians provide expert advice to large community of medical practioners who learn new treatments from them.

III. METHODS FOR IDENTIFICATION OF LEADERS:

The researchers have proposed various methods for Identification of Leaders in on line social networks.

LUCI Model:

This method [1] is used to identify the leaders and followers in online social networks using the Longitudinal User Centered Influence (LUCI) model, which takes input as user inter-action information and classifies users into four categories: extrovert leaders, introvert leaders, followers, and neutrals. The interaction information used in this study contains only the timestamp of interaction, the identifier of sender and receiver [1].

The interactions are rarely initiated by the introvert leaders while extrovert leaders frequently initiate the interactions. The followers are the users for whom; the number of interactions initiated by them with their friends depends on how many interactions they received. Neutrals initiate interactions rarely and use the social network very inconsistently [1].

The FJ influence model [1] is considered as the basis for this LUCI model. According to this model, the interaction behavior of a particular user at a time t is a linear function of the inter-action behavior and the combined interactions of the neighbors in the previous time period $t-1$. The model also assigns a weight to the degree to which a user's neighbors' interaction behaviors in the previous time period $t-1$ influences its inter-action behavior in the current time period t. This assigned weight is called as the network coefficient. The network coefficient of a user is denoted as ρ. The model [1] also assigns a weight to the degree to which a user's interaction behavior in the previous time period $t-1$ influences its interaction behavior in the current time period t. This assigned weight is called the ego coefficient. The ego

coefficient of a user is denoted as γ. Low ρ, Positive γ: extrovert leaders, Low ρ, Negative γ: introvert leaders, Low ρ, Zero γ: neutrals, high ρ and negligible γ : followers.

IV. LEADER RANK APPROACH:

In many online applications, users are able to select other users to be their sources of information [4]. These user-user relations can be represented by a network with directed links pointing from fans to their leaders. The direction of the link corresponds to votes from fans for their leaders. The popular leaders would have a large number of incoming links which are known as in-links. This is considered as the convention as it matches the direction of random walk in this algorithm, though the flow of the information is in the opposite direction i.e. from leaders to fans. The aim of the approach is to rank all the users based on the network topology.

A network of N nodes and M directed links has been considered by the authors [4]. Nodes correspond to users and links are established according to the relations among leaders and fans. A ground node which connects to every user through bidirectional links has been introduced in order to rank the users. The network thus becomes strongly connected and consists of N + 1 nodes and M + 2N links. One unit of resource is assigned to each node except for the ground node which is then evenly distributed to the node's neighbors through the directed links to initiate the ranking process. The process gets repeated until attaining a steady state. This process is equivalent to random walk on the directed network mathematically.

The stochastic matrix P with elements $p_{ij} = a_{ij}/k_i^{out}$ represents that, in the next step, the probability that a random walker at i goes to j. If node i points to j, then, $a_{ij} = 1$ and 0 otherwise, while k_i^{out} denotes the out-degree, i.e. the number of leaders, of i. Thus, this probability flow corresponds to the vote from fan i to leader j. By denoting using $s_i(t)$ [6] the score of node i at time t,

$$s_i(t + 1) = \sum_{j=1}^{N+1} \left(\frac{a_{ji}}{k_j^{out}}\right) s_j(t)$$

The initial scores are $s_i(0) = 1$ for all node i and $s_g(0) = 0$ for the ground node. Since the network is strongly connected, the presence of the ground node makes P irreducible. The ground node also ensures the co-existence of loops of size 2 and 3 from any node, which denotes P6 is positive, which means all elements of P6 are greater than zero. Since P_n value is positive for some natural number n, thenon-negative P is primitive.

According to the Perron-Frobenius theorem [4], P has the eigen value 1 which is maximum limit, with an unique Eigen-vector. The proof of permittivity and convergence are outlined in Supporting Information (SI). The score $s_i(t)$ for all i thus converges to a unique steady state which is indicated as $s_i(tc)$. Here, the convergence time is tc. At the steady state, distribute the score of the ground node to all other nodes to conserve scores on the nodes of interest are evenly distributed. Thus the final score of a user to be the leadership score S is defined as

$$S_i = s_i(tc) + (s_g(tc)/N)$$

where $s_g(tc)$ is the score of the ground node at steady state. There are several advantages of applying Leader Rank in ranking based on the above properties. The advantages include: (i) parameter-freeness, (ii) wide applicability to all kinds of graph, (iii) independence of the initial conditions and (iv)Convergence to a unique ranking.

V. BINARY APPROACH

In this approach [3], each user of a social networking site is considered as a node. This approach determines whether a node is leader or not. Social network analysis has been used in order to understand the behavior of the nodes. The centrality measure is used in this approach and it determines whether a node is a follower or a leader. The value of the centrality measure of the leaf nodes is lower than the inner nodes. The leaf nodes are very important especially in terrorist cells as the operations are executed by them. This binary concept [3] is used to highlight the distinctiveness of these nodes.

The three centrality measures considered by D. M. Akbar Hussain [3] are degree, betweenness and closeness. The first measure degree is the number of direct connections to other members of the network.

Other than degree, the well-known measures are betweenness and closeness. Betweenness measures the extent to which a node can be an intermediate node in the interaction between the other nodes. Thus, the nodes which are located on many shortest paths between other nodes will have higher between-ness comparing to other nodes. Closeness is the measure of the time taken by the information to spread from a given node to other nodes in the network.

It is clear from the Fig 1 that the nodes 1,5 and 6 can either be a leader or a follower where as other nodes have various positions to be of some importance. From the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 3, March 2017

fig.1 it is difficult to decide the leader. So, binary approach is used. In binary approach, the node with least spill over is considered to be at higher level in the hierarchy. Hence, the node with least spill over is considered to be the leader.
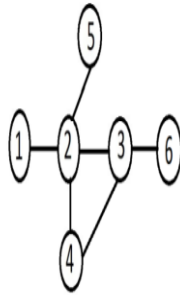

Fig 1. Network of 6 nodes

The main idea behind the binary approach [3] is that the leader is the node that has the lowest or null spill over. The subsequent nodes may have other roles in the network de-pending on the degree of spill over. The node having the least amount of spill over indicate the node sitting at a higher level, and also having less amount of communication with the rest of the network which is a typical case of a leader in the terrorist network.

The binary approach [3] is discussed mainly based on exploring the terrorist network. Apart from the standard measures like degree, betweenness and closeness several other measures also exists. The process of determining the role of a leader or a follower is hard to determine with standard centrality measures.

Integrated Text Mining Model: The opinion leaders affect other members via word of mouth (WOM) on social networks[6]. Expertise, novelty, and richness of information are three important features of opinion leaders, which are obtained from WOM documents (Kim and Han, 2009). Thus, this research proposes a text mining--based approach in order to identify opinion leaders by evaluating features of expertise, novelty and richness of information from contents of posts.

Expertise
This can be evaluated by comparing their posts with the controlled vocabulary base [9].
For member i, words are collected from his or her posted documents and member vector i is represented as fi=(w1, w2, …wj, …, wN), where wj denotes the frequency of word j used in the posted documents of user i. N denotes the number of words in the controlled vocabulary. Then

normalize the member vector by his or her maximum frequency of any significant word. The degree of expertise can be calculated by the Euclidean norm.[6]

$$\mathbf{exp_i} = \left\| \frac{\mathbf{f_i}}{\mathbf{m_i}} \right\|$$

Novelty
The authors utilized Google trends service (http://www.google.com/trends) to obtain the first--search time tag for significant words in documents. Thus, each significant word has its specific time
tag taken from the Google search repository. For example, the first--search time tag for the search
term, Nokia N81, is 2007 and for Nokia WindowsPhone 8 is 2011 [6]. The equation of novelty for a member is shown:

$$\mathbf{novi} = \frac{e_h + 0.66 \times e_m + 0.33 \times e_l}{e_h + e_m + e_l}$$

where e_h, e_m and e_l is the number of words that belong to the groups of high, normal and low
novelty, respectively.

VI. RICHNESS OF INFORMATION

In general, a long document suggests some useful information to the users [8].
Thus, richness of information of posts can be used for the identification of opinion leaders.
$$\mathbf{ric = d + g,}$$
where d is the total number of significant words that the user uses in his or her posts and g is the total number of multimedia objects that the user posts.
Finally integrated expertise, novelty and richness of information from the content of posted.
Documents for the identification of leaders [8]. As each feature has its own distribution and range, normalized each feature to a value between 0 and 1. Thus, the weights of opinion leaders based on the quality of posts become the average of these three features [8].

$$\mathbf{ITM} = \frac{\mathbf{Norm\ (nov) + Norm(exp) + Norm(ric)}}{\mathbf{3}}$$

Finally, this proposed text mining model compared with four quantitative approaches, i.e., involvement, *Degree Centrality, Closeness Centrality* and *Betweenness Centrality*, evaluated by human judgment. In these experiments, the involvement approach is proved to be the best one among the quantitative approaches. The text mining approach outperforms its quantitative counterparts as the richness of document information

provides a similar function to the qualitative features of opinion leaders. The proposed text mining approach further measures opinion leaders based on features of novelty and expertise.

Content based Metrics and SocialNetwork Metrics Based Approach:

Study In [5] has identified highly influential users i.e, leaders by using both content-based metrics and social network metrics in a large social network. In addition to collection of the textual content of the tweets the authors have analyzed the network relationship among users. A network link is defined as a connection between two users who are associated in one of the following ways:

(1) Targeted tweet (TT), (2) Re-tweet (RT), (3) Modified tweet (MT) [5].

The 36-day time frame was used to study evolution of network activities. Using the network relationship, computed the *Betweenness Centrality* to represent the influence score of each user. The *Betweenness Centrality* of a node i measures the proportion of the number of shortest paths that pass through node i to the number of all shortest paths between a pair of users [7]. To enable visualization and summarization of the collected data developed an interface that presents to users daily counts of the collected tweets as well as top 10 leaders in different categories.

VII. CONCLUSION:

In this paper we have discussed three approaches for identification of Leaders and Followers[2] and two for identification of only Leaders.

Binary Approach: The centrality measure is used in this approach and it determines whether a node is a follower or a leader. The three centrality measures considered are degree, betweenness and closeness.

Integrated Text Mining Model :This research proposes a text mining--based approach to evaluate features of expertise, novelty and richness of information from contents of posts for identification of opinion leaders.

Content based Metrics and Social Network Metrics Based Approach: Highly Influential Users
can be treated as opinion leaders in Social Network. This approach [5] has identified Highly Influential Users by using various metrics and techniques.

LUCI Model:
This study has used user interaction base to categories the users into four types of introvert leaders, extrovert leaders, followers and neutrals by calculating Ego coefficient and Network Co-efficient.

Leader Ranking Method used ranking approach based on the network topology

REFERENCES

[1] M. ZubairShafiq, Student Member, IEEE, Muhammad U. Ilyas, Member, IEEE, Alex X. Liu, Member, IEEE, and HayderRadha, Fellow, IEEE,"Identifying Leaders and Followers in Online Social Networks",2013

[2] Jeba Emilyn J, Keerthana R, " Community detection and identifying leaders and followers in online social networks" International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 245 ISSN 2229-5518

[3] D. M. Akbar Hussain, Member IEEE, IDA, IAENG," Identifying Leader or Follower using a Binary Approach"2010.

[4] LinyuanL¨u, Yi-Cheng Zhang, Chi Ho Yeung, Tao Zhou," Leaders in Social Networks, the Delicious Case,"2011

[5] Wingyan Chung, Daniel Zeng, Nathan O'Hanlon, "Identifying Influential Users in Social Media:A Study of U.S. Immigration Reform" Twentieth Americas Conference on Information Systems, Savannah, 2014

[6] Chihli Hung, Pei--Wen Yeh, "Identification of Opinion Leaders Using Text Mining Technique in Virtual Community**"**

[7] Jackson, M. O. 2008. Social and Economic Networks, Princeton University Press: Princeton, NJ.

[8] Agarwal, N., Liu, H., Tang, L. and Yu, P. S. 2008.Identifying the Influential Bloggers in a Community. Proceedings of WSDM, 207--217

[9] Li, F. and Du, T. C. 2011. Who Is Talking? AnOntology--Based Opinion Leader Identification Framework for Word--of--Mouth Marketing in OnlineSocial Blogs. Decision Support Systems, 51, 2011:190--197