COMPARATIVE STUDY OF DETERMINING INFLUENTIAL USER(S) IN ONLINE SOCIAL NETWORKS

Shruti Nigam[1], Chetan Agrawal[2], Anurag Jain[3]
[1]PG Scholar, [2]Assistant Professor, [3]HOD
Deptt. of Comp. Sc. Engg., Radharaman Institute of Technology & Science
Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, M.P., India

ABSTRACT
Internet has become now an engagement between known and unknown users using Social networks. However these user(s) has impacted the business also in terms of Marketing and Sharing/Spreading the information across the world. Online Social network has become communication medium for Infinite Users and Social web-sites are providing the platform to them for spreading the content base & user base information all across, which is accessible globally. Sales and marketing is also one of the key elements in online social network from business prospective as well. This paper presents a small/large-scale measurement study and analysis of the structure of multiple online social networks. We scrutinize data gathered from four well-liked online social networks: Facebook, Twitter or KDD and those techniques based on content published on the social network. Here we are trying to identify the maximum influential user resultant from direct and indirect influenced users along with the content or document scattering method as a quick marketing strategy across the world on all type of socially affected users. Also when it comes to Viral marketing technique social network plays very important role to determine the influenced user(s). The main objective of this paper is to lessen the complexity of determining the influenced user from previous work and maximize the marketing using degree of users who can determine the highest influential users.

INDEX TERMS—*Online social networks; Influential user; Active User; Viral marketing*

I. INTRODUCTION

Now a day's Social networking plays vital roles in each individual level and organization level. Currently in 20th Century today everyone is connected to a world called "internet". Social media such as Facebook, Twitter etc provides platform to all type of users to connect with world. However everyone has their own set of areas to work with these given platform, similar one can use these platform as a base of business model in terms of sales i.e. for Promotion/advertisements. Particularly in social sciences, the collective desire to associate in a community has been a well-researched phenomenon for a long time. Already, about 400 years earlier Christ, Aristotle [1] described human beings as zoon politicon– a character with the primary need of searching and creating communities. Therefore, the general idea of social networks is not really new. Information Technology has changed everything and is increasingly making businesses more effective. It has also changed the rules of game in terms of globalization, increased competition, how people work and how information is shared.

More importantly, the Internet has changed the way people communicate. Where information previously was shared and spread on paper, fax, telephone or face-to-face communication. The digitalized new world makes it easier to communicate and share information. The early days of the Internet bring us e-mail, but as the Internet is maturing the growth of the new Social network (Media) has exploded. Social Network connects millions of people whom build associations online, and is

increasing at an astonishing pace. Marketing and PR people are embracing these new communication channels, using them to improve the market performance of their business. Since its commercialization in the 1990s, is steadily acute almost all extent of modern human life.

Online Social Network is an arrangement of individuals or organizations coupled to any of dependencies such as friendship, fellowship, Co authorship, Collaboration etc [2]. A relationship or link is defined as a pattern of social dealings between two or more persons that involve in significant communication and awareness of the probable behavior of the other person. Online social networking provides a stage to build the communities of special interests virtually. It gives a new aspect to business models, inspires viral marketing, provides trend analysis and sales forecast in market, assists counterterrorism efforts and acts as a foundation for information sources. In the figure 1.1 general benefits of Online Social Network described:

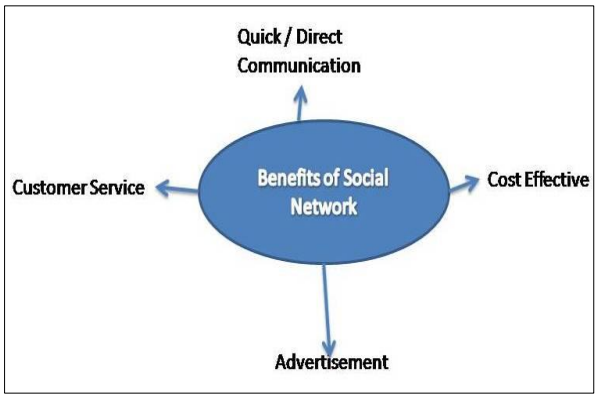

Figure1.1 General benefits of Online Social Network.

Online social networks can be thought as a number of nodes (or persons) connected by a series of links (or relationships). In section 2 the Related Area of the work done in this field is explained. Section 3 describes the problem statement and in the section 4, proposed algorithm is explained. Section 5 of the paper concludes the work and last section of the paper includes the future vision in this research area.

II. RELATED AREA

Social networks have been studied for valuable information not only to systematize these networks for maximum efficiency is accomplished but also to understand the role of individual nodes or groups in the network. However materialization of social networks in recent years has sought the attention of researchers in this area. In this section of paper we discuss about the previous findings and facts to select influential nodes in a connected network based on their formation as well as the activities.

A. Based on Structure of Network:

Trying to identify powerful node (influential node) in online social network is challenging task to research communities. Most early studies in this area focused on structural properties of networks (topology, connections etc.) and they quantify influencing power of nodes by centrality measures and by ranking the links of different nodes or groups.

Structural location of the node is advantageous to find the relative significance in the graph. There are various types of centrality measures of a node used to find the importance in the social network structure

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Centrality
- Edge vector Centrality

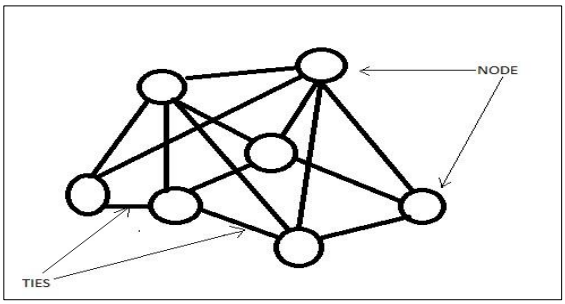

Figure 2.1 Example of Social Network as a Graph.

Figure 2.1 illustrate the social network as a Graph G. In this nodes are acts as vertices and relationship links among nodes represents edges of graph.

- Degree Centrality: The degree centrality of any node in a graph or network is defined by the count of edges that are incident with it, or the count of nodes adjacent to it. Degree centrality of node in case of directional networks given in two ways as in-degree and out-degree. The out-degree centrality is defined as [4]:

$$C_{DO}(i) = \sum_{j=1}^{n} a_{ij} \quad 2.1$$

Where a_{ij} is 1 in the binary adjacency matrix A if a link from node *i* to *j* exists, or it is 0. Similarly, the in-degree centrality is defined as:

$$C_{DI}(i) = \sum_{j=1}^{n} a_{ji} \quad 2.2$$

Where *i* describe the node *i* and a_{ji} is 1 if a link from node *j* to *i* exists, or it is 0.

- Closeness Centrality: Closeness centrality is a measure that identifies how fast it will take to flow information from node to all other nodes sequentially**.** Beauchamp explains, nodes occupying a central position with respect to closeness are very productive in distributing information to the other nodes [3]. Hakimi and Sabidussi developed a measure of closeness centrality as:

$$C_C(i) = \frac{1}{\sum_{j=1}^{n} d(i,j)} \quad 2.3$$

Where $d(i,j)$ represent the distance between node *i* and *j*, which is the measured minimum length of any path connecting *i* and *j* [5].

- Betweenness Centrality: Betweenness centrality counts the number of times a node founds (acts) as a bridge along the shortest route between two other nodes. It was introduced by Linton Freeman [6] as:

$$C_B(i) = \frac{\sum_{i \neq j \neq l} g_{jl}(i)}{g_{jl}} \quad 2.4$$

Where $g_{jl}(i)$ is the number of shortest route linking the two nodes j and l containing node i.

- Eigenvector Centrality: Eigenvector centrality measures the influence of a node within a network. It is based on the concept that links to high-scoring nodes attract more in comparison to the low-scoring nodes. A variant of Eigenvector centrality measure is Google's Page Rank [7].
Let A again be the binary adjacency matrix of the network and $\vec{x}$ be the principal eigenvector corresponding to the maximum Eigen value θ. The eigenvector centrality for a node i can be defined as a single element of the eigenvector, calculated as:

$$C_E(i) = x_i = \frac{1}{\theta} \sum_{j=1}^{n} a_{ji} x_j \quad 2.5$$

- Edge vector Degree Centrality. Weighted digraph describes how often a user (node) called another one, or how many text messages sent by him/her. In this case, the element of an adjacency matrix a_ij describe the numeric weights of a connection from node i to j. Each weighted graph can be converted into a multigraph, where the same pair of nodes can be connected by multiple edges. Edge-weighted degree centrality is defined as [4]:

$$C_{ED}(i) = \sum_{j=1}^{n} (a_{ij} + a_{ji}) \quad 2.6$$

Comparative analysis of the Influential User determination techniques is given below:

Table 1: Different Influential User Identification Techniques

Name of Technique	Year	Key Features
Centrality Measures	1966	Identify the relative importance of Nodes in Network.
HITS Algorithms[8]	1998	Identify Authority Pages and Hub Pages in Network.
PageRank Algorithms[7]	1998	Maintain a single metric for information of all Web Pages.
Linear Threshold Model[9]	1978	Focuses on Threshold (whole) behavior of Nodes.
Independent Cascade Model		Focuses on Individual's Interaction in Network.
Community Modeling[10]	2010	Efficient over Greedy method and orthogonal to existing algorithms of Influential detection.
Topic Level Algorithm[11]	2010	Consider the presence of Indirect Influence with Direct Influence in Online Social Network.
ExpertRank Algorithm[12]	2013	Document based relevance and Authority of Individuals.
Content Power User[13]	2011	Illustrate the dynamic nature of Online Social Networks.
OOLAM Algorithm[14]	2011	Opinion consistency and Opinion creditability are used to capture the persona of user.

III. PROBLEM STATEMENT

There are various techniques for influential user identification as discussed in literature survey but all those suffered from different deficiencies. Structural techniques (topology based for example closeness centrality, between's centrality, degree centrality etc. are not capable in recent scenario of online social network due to high computational complexities and the huge size of online social networks lead to incorrect results, than the activity based influence calculation techniques introduced. These techniques are resolve the network size issue but completely ignored the structural power of user in network. We argue that the "influential users in the online social network are the ones who induce many (direct and indirect) activities on their contents and also having high node degree". So to calculate influential power we propose hybrid method using both structural and behavioral aspects of online social network. In our approach documents power represents content quality and degree of friends/neighbors of user describe the quantity of spreading information due to influence power.
This study focuses on how to identify the best nodes/users available in online social networks, that if those nodes are infected with an idea, they can infect more nodes than their rivals. This problem is called influence maximization because influence is modeled as the ability of a node to infect more nodes in the network. Here, question is posed that, how to rank nodes based on their importance of infecting more nodes if that node individually gets infected.

IV. PROPOSED ALGORITHM

We argue that the influential users in the online social network are the ones who induce many (direct and indirect) activities on their contents and also having high node degree. So to calculate influential power we propose hybrid method using both structural and behavioral aspects of online social network. In this approach documents power represents content quality and the degree of friends/neighbors of user describe the quantity of spreading information due to influence power.

A. Computation of Document Power (DocP)

The DocP is termed as a degree of influence of a document on other users in the online social networks. When an action is performed by the user to a particular document (such as writing a

comment, liking, tagging, mentioning etc), it depicts that content of the document has some influence on the user(s). This study leads to a new method of quantifying the DocP. The main idea is that the DocP can be computed by adding up the weighted frequencies of other user's activities induced on particular document.

A document can have either direct influence on other users who access the document directly, or indirect influence on others who access the document in someone else's post on online social network (for e.g. wall in Facebook, blog etc.) that is reproduced through share, retweet, trackback or scrap of the original document. One may adjust the weights, w_D and w_I, which will change the degree of relative importance of direct and indirect influence. DocP is calculating as:

$$DocP(D_{i,j}, A_k) = \sum_{A_k \in AT} W_{A_k} * Count\ (D_{i,j}, A_k) \quad 4.1$$

$$Count\ (D_{i,j,}, A_k) = Frequency\ of\ A_k \quad 4.2$$

B. User Power (UserP)

UserP is defined as a degree of influence of a user on other users in the online social networks. The UserP is computed by adding up the in-degree of all the friends'/neighbours, who republish the document of user. We found that if the set of initial users (represented as nodes in the network) is chosen according to degree centrality measures, the number of reached users is substantially higher than selecting the initial set of customers according to other centrality measures across different treatments. Thus, it is these measures that the designer of a marketing campaign should use in order to achieve a wide dissemination of marketing messages. The InfP of a document tends to increase as it is exposed to the huge audience, which may result in high degree friends produce more influence power in compared to relatively lower.

$$UserP\ (U_i) = \sum_{A_{k'} \in AT} Friends\ Degree \quad 4.3$$

C. Computation of Total Influence Power of User (InfPower)

Influence power of user is sum of document power and user power. InfPower of user is calculated as in following manner:

$$InfPower(U_i) = DocP(D_{i,j}, A_k) + UserP\ (U_i) \quad 4.4$$

Table 2: Summary of terminology used.

U_i	=	User *i* of the Online Social Network
$D_{i,j}$	=	Document j of user i
DocP($D_{i,j}$)	=	Document Power of ($D_{i,j}$)
UserP(U_i)	=	User power of U_i
AT	=	{A_1, A_2, A_3...} Action Type
A_k	=	Action of type *k (document influence measures)*
$A_{k'}$	=	Action of type *k (user influence measures)*
W_{Ak}	=	Weight of A_k

V. CONCLUSION

Our methodology is quiet simple in comparison to the previous methods because it is based on activities of users performed to the documents (contents). This information is easily available in diffusion history or activity log of online social network and can be maintained with low cost. In addition, we conclude that our approach maintained "quality with quantity". A person who having less number of friends (degree) but publishes quality documents also have chances to select as influential user.

VI. FUTURE WORK

The future direction to this work is to development of approaches by considering temporal aspects. Another major area for extending works by basis of temporal and location based methodologies for identification of influencers in online society, additionally not only calculates influence power but to type of influence (negative or positive) using link polarity concepts.

REFERENCES

[1] H.U. Buhl: Online communities. Wirtschafts informatik 50 (2) (2008), pp. 81–84.
[2] Charu C. Aggarwal: Social Network Data Analytics. Springer Science Business Media, 1-14, LLC 2011.
[3] M. Beauchamp: An improved index of centrality. Behavioural Science 10, 161–163, 1965.
[4] S. Hakimi: Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research 12 (1965).
[5] G. Sabidussi: The centrality index of a graph. Psychometrika 31 (1966) 581–603.
[6] L.C. Freemann: A set of measures of centrality based on betweenness. Sociometry 40 (1977) 35–41.
[7] S. Brin, L. Page: The anatomy of a large scale hyper textual web search engine. WWW Conference, Australia, 1998.
[8] J. Kleinberg. Auth. sources in hyperlinked environment. CM-SIAM Symposium on Discrete Algorithms, 1998.
[9] M. Granovetter: Threshold models of collective behaviour. American Journal of Sociology, vol. 83, no. 6, pp. 1420-1443, November 1978.
[10] J.Goldenberg, B. Libai, and E. Muller: Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, pp. 211-223, August 2001.
[11] Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang: Mining Topic-level Influence in Heterogeneous Networks. CIKM'10, October, 2010, Toronto, Canada.
[12] G. Alan Wang, Jian Jiao, Alan S. Abrahams, Weiguo Fan, Zhongju Zhang: ExpertRank: A topic-aware expert finding algorithm for online knowledge communities. Decision Support Systems 54 (2013), pp. 1442–1451.
[13] Seung-Hwan Lim, Sang-Wook Kim, Sunju park and Joon Ho Lee: Determining Content Power Users in a Blog Network: An Approach and Its Applications. IEEE Transactions on Systems, Man, and Cybernetics – part-A: System and Human, Vol41, NO.5, September 2011, pp. 853-862.
[14] Keke Cai, Shengha Bao, Zi Yang, Jie Tang, Rui Ma, Li Zhang, Zhong Su: OOLAM: an Opinion Oriented Link Analysis Model for Influence Persona Discovery. WSDM'11, February 9–12, 2011, Hong Kong, China.

Authors

Shruti Nigam, Studying Master of Engineering in CSE at Radharaman Institute of Technology & Science, Bhopal, INDIA. She has studied her Bachelor of Engineering in CSE at Technocrats Institute of Technology, Bhopal, INDIA. Her research area of interest is Cyber Wireless Networks, Social Network and Data Mining.

Chetan Agrawal, Studied Master of Engineering in CSE at TRUBA Institute of Engineering & Information Technology, Bhopal, INDIA. He has studied his Bachelor of Engineering in CSE at BANSAL Institute of Science & Technology, Bhopal, INDIA. Currently He is working as Assistant professor in CSE department at RADHARAMAN Institute of Technology & Science Bhopal M.P. INDIA. His research area of interest is Cyber Security, Network Security, Wireless Networks, and Data Mining and Cloud Computing.