ISSN 2229-5518

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, Bhopal, M.P., India

Abstract— In today's world online social network (OSNs) are exclusive social and web portent affecting perceptions and behaviours of their users and helping them to maintain/create social presence. It is remarkable to analyze the growth and evolution of online social networks where either we consider the marketing strategy or new services being offered, also from a scientific viewpoint, since their erection and evolution may share similarities with tangible social network. Several techniques for analyzing social networks have been developed, to gauge quantitative properties (e.g., defining metrics and measures of fundamental characteristics of the networks) or qualitative aspects (e.g., studying the attachment model for the network evolution and the link prediction problem). However, online social network analysis poses novel challenges both to computer and Social scientists.

Social Influential user (SIU) is a unique web and social phenomenon affecting tastes and behaviours of their users and helping them to maintain/create friendships. There can be many different ways to calculate the power of user. In this section, we are describing novel method to calculate the influence power of users in online social network. In our Research we focussing on user activity based influence power. The influence power of a document is greater when activities are more of other users on the document and with more activities of other users on the documents that are *reproduced* from the original document and also the *degree of friends of friend*. Therefore, to the computational point of view we are going to calculate document power as well as user's power in existing network

Index Terms- Online social networks; Influential user; Active User; Viral marketing.

1 INTRODUCTION

SOCIAL networks are being analysed in methodically here which gives views of social engagement in terms of theory of networks, consisting of action nodes (representing individual users within the network) and link (which represent associations between the individuals, such as Social presence, friendship, organizational position etc.) These networks are usually represented in a social network diagram, where nodes as points and links are represented as lines.

More importantly, the Internet has changed the way people communicate. As in previously information was shared on face-to- face communication or paper, fax. The digitalized new world makes it easier to communicate and float information. Intially Internet brought us e-mail, but as the Internet is maturing the growth of the new Social network day by day. Social Network [1] connects millions of people whom build relationships online, and is growing at an amazingingly. Marketing is also embracing these new communication channels, using social networks to strengthen their business. Since its commercialization in the 1990s, is steadily penetrating almost all dimensions of modern human life.

The scope of this work is at all the levels (i.e. at micro, meso and macro level) of online social networks. This Research work proposes a new mechanism for to identify key persons or customers to recommending products online. According to the current trend in moving marketers toward social recommender systems and the difficulties in finding influential people in this environment, the influence maximization problem in social networks is considered as a controversial problem. This problem is formally defined in this work. The problem deals with first selection of small subsets and then infecting of nodes with a new idea.

This Research work is organized in rest 4 sections. Section two literature works in field of identification of influential users in online social network is analysed, in preparation for model and algorithm which are proposed in section three. Section four provides solutions and describes proposed algorithms for measuring influence in the online social network. In order to validate the proposed algorithm in this Research, a general framework of online social network is simulated. Finally, conclusions and future work will be discussed in Section 5.

ISSN 2229-5518

2 LITERATURE SURVEY

2.1 Technique based on Structural Measures of Online Social Networks

Social network research community describes a variety of structural measures for the identification of existing Influence in a network. This paper will briefly summarize the well-known centrality method and number of link topological ranking measures.

2.2.1 Centrality Measures.

Structural location of the node is advantageous to find the relative significance in the graph. There are various types of centrality [2] measures of a node used to find the importance in the social network structure.

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- EigenVector Centrality
- Edgevector Centrality

2.2.2 Link topological ranking measures

- HITS (Hyperlink-Induced Topic Search)
- PageRank algorithms

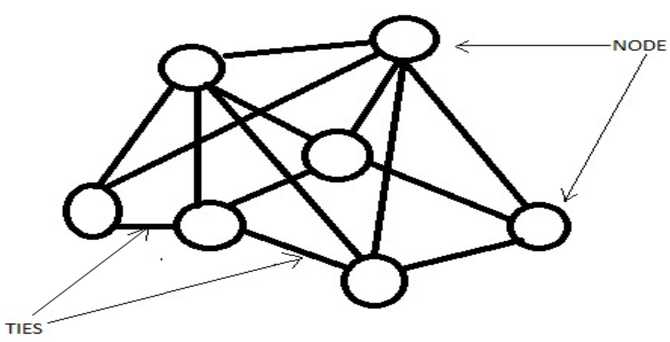


Figure 2.1 Example of Social Network as a Graph

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.

2.2 Technique based on Community Mining

Kempel et al.[13] state that optimization problem of influence maximization is NP-Hard. Accordingly to solve this type of problem, Greedy algorithms with provable approximation may give better outcome. But Greedy methods are expensive in computation, so as a result it is not practicable to social network. *Yu Wang et al.[14]* proposed novel method called "Community based Greedy algorithm for mining top-*K* influential nodes" which divides the network into a number of communities, and then selects individual community to find top-*K* influential nodes. The community structure is a main property of social network features: Individuals within a community have frequent contact; in contrast, individuals across communities has much less contact with each other and thus is less likely to influence each other. This property suggests that it might be a good approximation to identify influential nodes within communities instead of the whole network. This work gives several directions to expand research of location based social network to find influential over time.

2.3 Techniques based on Content Mining in Online Social Networks

All Seung-Hwan Lim et al. [11], proposed new method to identify content power user in Blog network. This work provide concept of Document Content Power (DCP) and User Content Power (UCP).

G. Alan Wang et al. [10], present a novel algorithm ExpertRank: A topic-aware expert finding algorithm for online knowledge communities to find expert or influential person on particular topic.

Topic Level Influence, Lu Liu et al. [9] brings out the idea of mining the strength of direct and indirect influence. They proposed a generative graphical model to identify topic-level direct influence.

4 Technique based on Link Polarity in online social network

Keke Cai et al. [12] proposed "Opinion Oriented Link Analysis Model (OOLAM)".

They study kind of influence inclined among Users in Network.

In particular, three kinds of influence personae which take place widely in social network include:

- Positive Persona,
- Negative Persona,
- Controversy Persona.

2.5 Technique used in Micro Blog Marketing

Fei Hao et al.[15] analyse the influence of nodes in a microblog network and proposed the "Community Scale-Sensitive Max Degree (CSSM)", an algorithm for maximizing the influ-

ISSN 2229-5518

ence when placing advertisements.

Influence of a node depends on following three matrices:

- Centrality based on node degree.
- Sum of neighbour's degree
- Attributes of nodes.

The attributes of a node in the micro-blog network, includes Activity degree, Interaction degree and Social Prestige.

2.6 Techniques based on the Diffusion Model of the Online Social Network

The diffusion model of the social networks also helps in the identifying the influential user in networks. The diffusion models are classified into three categories: linear threshold model, independent cascade model and a model that combines both the features of the linear threshold model [7] and independent cascade model [8].

- Linear Threshold Model.
- Independent Cascade Model.

Table 2.1 Key features of various Influential identification techniques.

Name of Technique	Year launched	Key Features
Centrality Measures[3][4][5]	1966	Identify the relative importance of Nodes in Network.
HITS Algorithms	1998	Identify Authority Pages and Hub Pages in Network.
PageRank Algorithms[6]	1998	Maintain a single metric for information of all Web Pages.
Linear Threshold Model [7]	1978	Focuses on Threshold (whole) behaviour of Nodes.
Independent Cascade Model[8]		Focuses on Individual's Interaction in Network.
Community Modelling [14]	2010	Efficient over Greedy method and orthogonal to existing algorithms of Influential detection.
Topic Level Algorithm[9]	2010	Consider the presence of Indirect Influence with Direct Influence in Online Social Network.
ExpertRank Algorithm [10]	2013	Document based relevance and Authority of Individuals.
Content Power User[11]	2011	Illustrate the dynamic nature of Online Social Networks.
CSSM Algorithm[15]	2012	Includes Activity degree, Interaction degree and Social prestige of the user.
OOLAM Algorithm[12]	2011	Opinion consistency and Opinion creditability are used to capture the persona of user.

3 PRAPOSED WORK

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, Betweenness 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 behavioural aspects of online social network. In our approach documents power represents content quality and degree of friends/neighbours of user describe the quantity of spreading information due to influence power.

3.1 Definition Used in identification of Influential Users

Table 3.1 summarizes the terminologies and symbols used in this Research work. Ui represents user i. Di represents the set of documents owned by Ui, and Di,j represents document j of user i. Document Power (DocP) is defined as the content power of a document, and DocP (Di,j) represents the document power of Ui .

User power (UserP) is defined as the sum of indegree of all the friends who republished the document published by user Ui. The UserP(Ui) represents the user power of user i. Action Type (AT) represents the types of actions (i.e., comment, like, and share) a user can perform in the online social network. An action of type k is denoted as Ak. When computing the content power of a document, different weights may be assigned depending on the types of actions. The weight for Ak is denoted as WAk.

Table 3.1- Summary of terminology used.

Ui	**=**	**User i of the Online Social Network**
Di,j	**=**	**Document j of user i**
DocP(Di,j)	**=**	**Document Power of (Di,j)**
UserP(Ui)	**=**	**User power of Ui**
AT	**=**	**{A1, A2, A3…} Action**

ISSN 2229-5518

		Type
Ak	**=**	**Action of type k (document influence measures)**
Ak'	**=**	**Action of type k (user influence measures)**
WAk	**=**	**Weight of Ak**

Step 1. Assigning of weights:

In our approach weights for calculation of influential power of content or documents are assigned as in following manner,

wA1 = 1 for user activity mention (No_at),

wA2 = 2 for user activity comment (No_cmt),

wA3 = 3 for user activity retweet (No_rt).

Here we choose wA1< wA2< wA3 because we assumed that user can retweet only when he/she having highest degree of association/influence with content and subsequently less for comments and mention (at action by user).

Step 2. Computation of Docp, UserP

Computation of Document Power (DocP)

$$DocP(D_{i,j}, A_k) = \sum\nolimits_{A_k \in AT} W_{A_k} * Count\ (D_{i,j}, A_k)$$

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

Computation of User Power (UserP)

$$UserP\ (U_i) = \sum\nolimits_{A_{k'} \in AT} Friends\ Degree$$

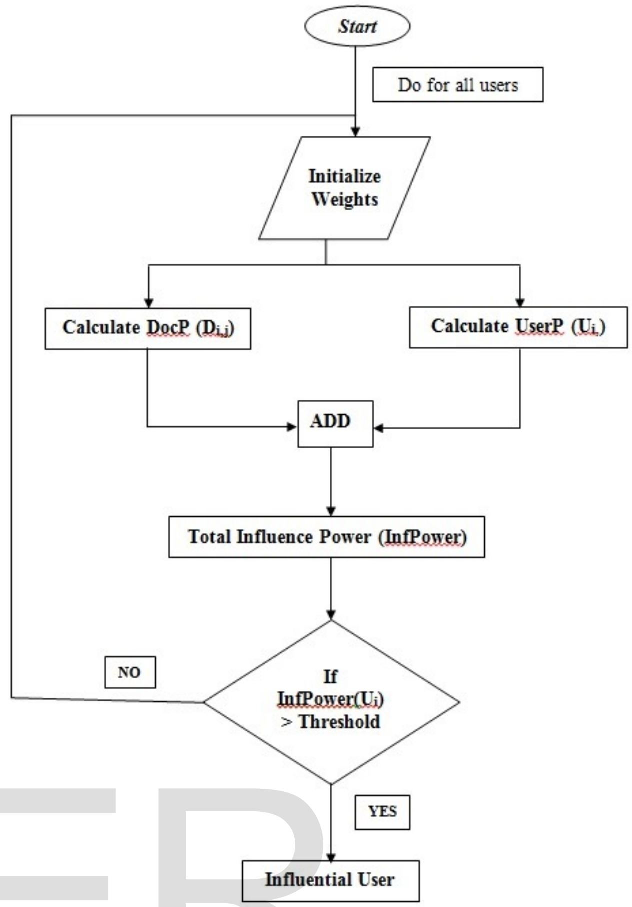
Computation of Total Influence Power of User (InfPower)

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

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

Figure 3.2: flowchart of proposed Influence power calculation approach.

Below flowchart explains the Social Infuluential user (SIU) identification using user's data. In first step we insert all the users data, based on each user data and using defined weight values for them we will calucate the Content or Document powere and User Power.Then summations of Document an User powers will resultant into Infulential user power.In the end we do verify the threshold value of the indentify infulential user i.e. not a mandatory process to do end if we work with less number and easily identifiable users.



4 Experimental Result Analysis

In this Research work dataset of KDD CUP 2012, is used. These experiments were performed on 5000 records of 3341 distinct users and social connections between different users. For evaluation dataset is stored in MySQL 6.3

4.1 Top 10 Influential Users

In first experiment we identify top 10 influential users by using the proposed approach in previous chapter. Here selection of weights as per wA1< wA2< wA3. Here number of active friend represents those who create activities regular basis on that particular user.

Table 4.1 Top 10 Influential Users

Influential User ID	No of like		No of Share		No of Comments		No of Active friends of user
	Max	Ave	Max	Ave	Max	Ave	
1031058	0	0	1	1	1	.33	146
1029816	0	0	3	2.5	1	.5	59
1029209	0	0	1	1	3	2	52
1025350	0	0	1	.4	1	.6	162
1002406	0	0	1	1	0	0	122
101212	2	.5	9	2.2	172	36.8	43
1024224	0	0	5	3	6	4.5	32
1024799	0	0	1	1	0	0	81
1010226	0	0	8	4.8	6	3	11
1029876	0	0	6	4	3	2	26
Average		0.05		2.09		4.9	73.4

Table 4.2: Average Share count in Dataset

INF USERS	Average Share in Base(Content Power)	Average Share in Proposed Approach(SIU)
INF-1	16	1
INF-2	0	2.5
INF-3	7.2	1
INF-4	0	.4
INF-5	.25	1
INF-6	20.5	2.2
INF-7	2.2	3
INF-8	1.45	1
INF-9	0	4.8
INF-10	1.33	4

Findings:

By analysing table 4.1, it is noticed that the selection of influential user based on proposed approach is better choice because it includes both kind of power (structure and behaviour). By applying only one approach we can not cover whole networks power users. Structure power of user shows how many persons get influenced and behaviour power represents how much influence of particular one in network. The distribution of influence power is another measure advantage of proposed approach. Concept of friends of friend widely spread the message in to network if the selection done appropriately.

4.3 Comparison of republishing action parameter:

To model the behaviour of this top 10 influential users share counts are analysed. Main objective behind this experiment is to check is number of republishing action are responsible for maximum influence power?

To comparison we apply two approaches as first proposed by Seung-Hwan Lim et al.[11] and our proposed approach.

Table 4.2 consist average share count in base approach and the Average Share count in proposed approach.

Figure 4.3: Average Retweet counts in base and proposed approach.

Findings: Figure 4.3 shows that in base approach average value of share action is high for every influential user but our approach not fully satisfies this and the influence power also very high for low average share actions.

This is proved that by taking only documents power (behaviour) we not able to select most influential. In our approach it's guaranteed that infection spread widely and it is the foremost demand of viral marketing.

ISSN 2229-5518

4.4 Comparison with pre-existing approach

In this experiment we apply the same dataset on Seung-Hwan Lim et al. approach and our proposed approach.

The objective of this experiment to calculate similarity between influential users set.

We select four combinations (25, 50, 75 and 100) of top influential users and by using jaccard index similarity is computed. Jaccard index measures similarities between two sets, and is defined by the size of the intersection divided by the size of the union of the sets.

Figure 4.4: Comparison with existing approach

In this study, however, we are not saying that our approach is the finest among the various techniques for identifying the Influential users in the online social network. However, we emphasis mainly on the following facts.

- To maximize the spread of infection there is combination of both structural and behavioural aspects are foremost choice.
- The influential power should calculate based on the average active user counts.
- Viral marketing based organizations should focus on the identification of the users based on the user's actions as well as structural demographics.

5 Conclusion & Future Work

This research work work, presents novel approach for identification of the influential users in the online social networks.This novel approach is hybrid form of structural properties and behaviour of the users in online social network.

Structural properties mainly include topological features along with links between users inside social boundaries.Behavioural aspects, fully dependent on user's activities inside network, published contents and actions of users depicting the behaviour of individual in online social network.

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] Charu C. Aggarwal: Social Network Data Analytics. Springer Science Business Media, 1-14, LLC 2011.

[2] M. Beauchamp: An improved index of centrality. Behavioural Science 10, 161–163, 1965.

[3] S. Hakimi: Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research 12 (1965).

[4] G. Sabidussi: The centrality index of a graph. Psychometrika 31 (1966) 581–603.

[5] L.C. Freemann: A set of measures of centrality based on betweenness. Sociometry 40 (1977) 35–41.

[6] S. Brin, L. Page: The anatomy of a large scale hyper textual web search engine. WWW Conference, Australia, 1998.

[7] M. Granovetter: Threshold models of collective behaviour. American Journal of Sociology, vol. 83, no. 6, pp. 1420-1443, November 1978.

[8] 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.

[9] Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang: Mining Topic-level Influence in Heterogeneous Networks. CIKM'10, October, 2010, Toronto, Canada.

[10] G. Alan Wang, Jian Jiao, Alan S. Abrahams, Weiguo Fan, Zhongju Zhang: ExpertRank: A topic-aware expert finding algorithm for online knowledge communities. Decision Sup-port Systems 54 (2013), pp. 1442–1451.

[11] 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.

[12] 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

[13] D. Kempel, J. Kleinberg, and E. Tardos: Maximizing the spread of influence through a social network". In ACM SIGKDD, pp 137-146, 2003.

[14] Yu Wang, Gao Cong, Guojie Song, Kunqing Xie: Community-based Greedy Algorithm for Mining Top-K Influential Nodes in Mobile Social Networks. KDD'10, July 25–28, 2010, Washington.

[15] Fei Hao, Min Chen, Chunsheng Zhu, Mohsen Guizani: Discovering Influential Users in Micro-blog Marketing with Influence Maximization Mechanism. Globecom 2012 - Ad Hoc and Sensor Networking Symposium.

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 Software testing.

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