STM JOURNALS
Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720 (Online)
Volume 5, Issue 3
www.stmjournals.com

Finding Influential Users in Online Social Networks: A Tree-based Approach

Hilal Ahmad Khanday*, Rana Hashmy
Department of Computer Sciences, University of Kashmir, Srinagar, Jammu and Kashmir, India

Abstract
In online social networking, people are getting involved in communications, but the real problem that comes to the mind is whether these communications can bring revolutions, and if so, who are the best people in given social networks that are really holding the remote control i.e. they have the power to bring these revolutions. This problem and the likes can be modelled as detecting influential users in online social networks. There can be many different ways to find influential users in social networks. Here, we propose a novel method based upon tree data structure to determine the influential users in online social network. We identify influential nodes based on their positions in the network and assume that the nodes are homogeneous. This approach identifies the specific users who most influence others' activity and does so considerably better than other methods.

Keywords: *Social networks, social network metrics, influence, influential users*

Author for Correspondence *E-mail: hilalhyder@gmail.com*

INTRODUCTION

With the advent of online social networks, today's era has become an e-hub; from business to education, home to office, and from payment of bills to the payment of salary, everything has been wrapped up by these online social networks. With the advent of such applications and their usage, every aspect of life has dealings with these networks in one or the other way. Human beings are psychologically an entity with extraordinary potential to make and take changes to/from environment; and modeling this behavior of human beings is among the foremost tasks that need to be dealt with. The adventure of online social networks changed traditional social influencing to modern mode of social influencing. In traditional social networks, these users with some sort of influence are ordinarily known to the masses, so these influential can naturally be pointed out and consulted to favor or to contradict any opinion related to any walk of life but in online social networks, influential users are blind to every user, so the evocation of these unknown influential users becomes a vital task. Once detected, these influential users can be put to the task of boosting the domain we are dealing with. In this paper we find influential users in social networks by using tree-based approach. The proposed method provides solution by constructing state space tree wherein each node of the given online social network is considered along with the communication it does.

RELATED WORK

Social networking sites such as Facebook, Google^{+}, LinkedIn, Myspace and Twitter [1, 2, 3, 4] are providing valuable social information about their entities (e.g., individuals, corporations, collective social units, or organizations), contacts and their relationships (e.g., kinship, friendship, common interest, beliefs, or financial exchange). A social network user p can create a personal profile in Facebook, add other Facebook users as friends, and share messages among his friends. Similarly, p can connect with friends and family members (e.g., find and connect with people from p's high school or hometown) in Google+, categorize his friends into different circles, and share ideas and thoughts with other Google+ users. Moreover, p can also create a professional profile in LinkedIn, establish connections to other LinkedIn users, and exchange messages among his connections. There are many users in these social networks, and each user may have different number of users connected/linked to his network. Among all users in these social networks, some users can (i) have a strong relationship or (ii) be very significant or important to others depending on different parameters such as the number of interactions made between them.

In the last few years, there have been many attempts towards the application of different data mining techniques on social network data for extracting useful information [5, 6, 7], and [8]. These efforts have used number of messages transferred between the nodes as the primary characteristic for finding strong friends [9], popular friends [10], or significant friends [11]. Conversely, one may like to think in terms of connectivity of nodes rather than focusing on the sheer amount of messages transferred between them. For example, a Facebook user may want to identify those prominent friends who have high impact (e.g., in terms of knowledge or expertise about a subject matter) in the social network. As another example, a LinkedIn user may want to get introduced to those second-degree connections who have rich experience in some profession. Similarly, a Twitter user may also be interested in following (and subscribing twitter feed from) those who are highly influential in the whole network.

Domingos et al. [12, 13] took influence maximization problem as an algorithmic problem and proposed an influence propagation model which specifies a joint distribution over all nodes' behavior globally. Kempe et al. [14] proposed two popular stochastic influence propagation models: The Independent Cascade Model and the Linear Threshold Model; that can explicitly model the step-by-step dynamics of influence propagation progress. Many other studies have put forward several initiatives to identify key individuals and group leaders [15, 16, 17]. Some used SNA metrics and algorithms, such as centrality measures and Page Rank algorithm [18], while others used text mining techniques [19]. A computational model that integrates data mining with social computing was proposed [20] to help users to discover influential friends from the social networks. Nancy et al. [21] explore the association rule between a course and gender in the Facebook 100 university dataset. Yu et al. [22] introduced the scheme for association rule learning of personal hobbies in social networks. Initial research used association rule learning to identify influential users and predict user participation in online social networks [23].

PROPOSED METHOD

A given online social network is considered in its graph-based representation G(V,E), wherein V, the set of nodes represents users in the given online social network and E, the set of edges, represents the communication that occurs among the users within the given social network. In order to provide a solution to the problem of finding influential users, we construct an m-ary tree. For constructing an m-ary tree, root has to be the starting point. To obtain the root, we calculate the degree of every node of graph. The node having maximum degree is selected to act as the root of the tree to be constructed. Since root is among the important nodes of tree, we hold an election for finding a node, which will lead the tree as its root. This can be depicted as:

$\forall a_i \mid a_i \,\varepsilon\, G\ (V, E) \wedge \exists\, a_i \mid \deg_{(a_i)} >= \deg\ (V\text{-}a_i)$

Any node a_i, having maximum degree in the given network, will be used as the root node in the tree formation. In case there are multiple nodes with the same maximum degree, arbitrarily one of them is selected as root node.

The next step towards the solution is to generate the children of this root node. It is achieved by adding all the adjacent members (neighbors) of the given root node as children to it (root node). Technically, it can be shown as:

$\forall a_i\ (a_i \,\varepsilon\, \text{neighbor (root)} \rightarrow \text{root-child}: = a_i)$

Then, at the level 1 of the constructed tree, all nodes or children of root are elaborated by their instant neighbors. From level 2 onwards, only those nodes are elaborated which have not been elaborated till now anywhere from the root to present node. This process of elaboration/ expanding of the nodes continue until we are left with no node from the graph that has not been elaborated. Just we need to be sure that a node at any level is elaborated only once.

Once we end up with a tree consisting of all the nodes from the social network, we compute the leaf frequency of every node in the tree. Eventually, we perform the following test on each node for checking whether it is influential or not.

If the leaf frequency of any node is greater than maximum height of the tree +1, then the node is categorized as influential, i.e.,
$(\forall a_i \mid a_i \varepsilon G(V, E)) \wedge (\exists a_i \mid \text{leaf-frequency}_{(a_i)} > (\text{TreeDepth}+1))$

All those nodes which pass the above test form the final list of influential users. An algorithmic description for the above said approach is given below:

Function Influential_User_Tree_Approach (Social Graph G (V, E))
Compute_degree (G);
root <-Select_root(G);
expand(root, G);
while (expand_flag!=0)
{
expand (T, G);
}
Calculate_leaf_frequency (T, G);
D<-depth (T);
Select_influentials();

End function
A very short description of the functions used in the above algorithm is given below:

I. Compute_degree*(G):* This function finds the degree of every node of given graph, G.
II. Select_root(G): It finds out root of the tree as: *return (max(deg(node, G)).*
III. *expand(root,G):* This function produces all children of root by using G and then sets the expand_flag to 1.
IV. *expand(T,G):* Here we produce all children of tree by using G. If there is no such node which has remained out from expansion we set the flag accordingly.
V. *Calculate_leaf_frequency(T,G):* Calculates the total number of occurrences for each node as the leaf of tree.
VI. *depth(T):* It returns the maximum height of tree T.
VII. Select_influentials(): It categorizes the node as influential if its leaf frequency is greater than depth of tree plus one. i.e. > depth() +1.

EXPERIMENTATION AND RESULTS

Two real datasets viz. Karate Club network dataset and Krackhardt network dataset have been mainly used whilst implementing the above algorithm. On executing our proposed method of tree-based influential user detection, we maintained the following heuristics in the construction of the tree: (1) The node with highest degree will act as the root node of tree to be formed. (2) Each node will be elaborated just once. (3) If a node occurs multiple times at any given level, arbitrarily, one of its occurrences is chosen as the seed for the elaboration of the node.

Karate Club Network Dataset

Following state space tree (Figure 1) was obtained when we executed our method on Karate Club Network.

The real threshold for the above dataset = height of above tree + 1=5+1=6

Nodes springing their appearance as leaves more frequently than the above threshold will be treated as influential users in the given social network. We conclude that nodes 1, 2, 33, 34 are influential as they cross the above threshold. If we compare it with the ground truth, nodes 1, 2 belong to first community and nodes 33, 34 belong to second community and are actually powerful enough to enforce influence on other members of the network.

Krackhardt Network Dataset

Krackhardt network has 21 nodes and 20 edges. On applying our method on Krackhardt network, following state space tree was obtained (Figure 2).

Since the height of the tree is 4, so the threshold for the network is 5 (4+1). Nodes springing their appearance as leaves more frequently than the above threshold will be treated as influential users in the given social network. From the above tree it is clear that only node 14 appears more than 5 times as leaf node and hence is the only influential node in the network which coincides with the ground truth of the network data. We tested the implementation of our algorithm on a variety of other synthetic datasets and the results were encouraging.

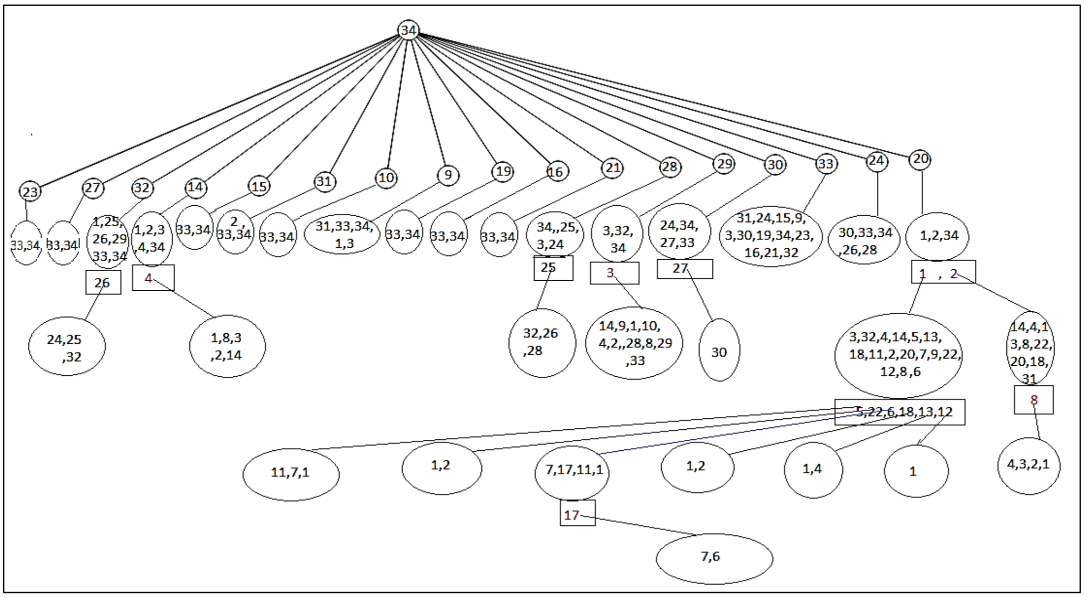

Fig. 1: *State Space Tree as Obtained on Karate Club Network Dataset.*

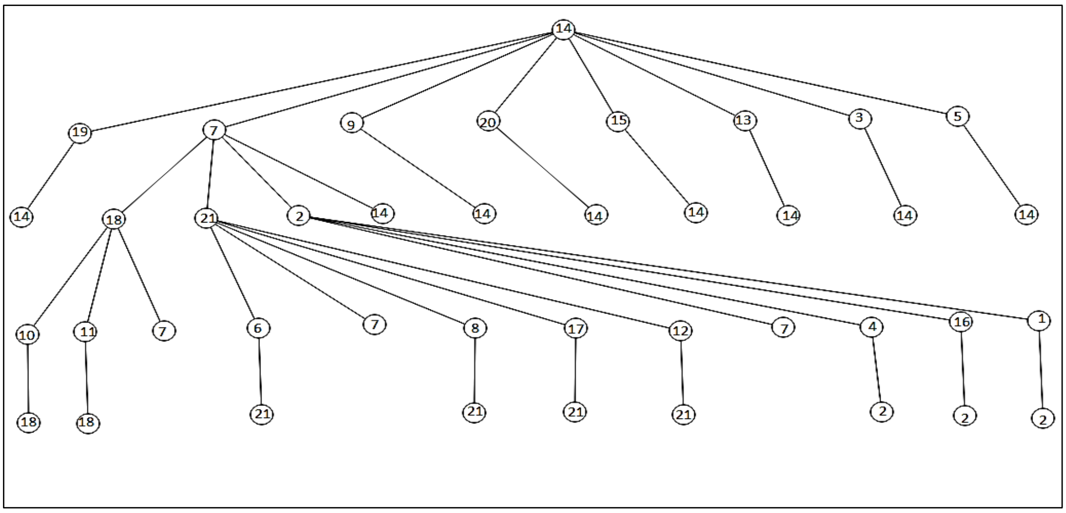

Fig. 2: *State Space Tree as Obtained on Krackhardt Network Dataset.*

CONCLUSION

Finding of influential users is one of the key problems in online social network analysis and there have been many models that have been proposed so far. Our approach investigates each node and elaborates it exactly once. This approach is quite deterministic, and no hardcoded parameter is used. Besides the threshold used in our method is dynamic in nature as it adapts itself in accordance with the given online social network. The method produces very good results by displaying only such individuals as influential users that can be verified from the ground truth of the dataset. The most complex aspect of our approach is social tree generation especially when the network size is too big. Finally, we conclude that there is lot more scope for further research in finding influential users in social networks as the problem is hard and in infancy stages.

REFERENCES

1. Fan W, Yeung KH. Virus propagation modelling in Facebook. *Proceedings of the 2010 Int'l Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE 2010;* pp. 331–335.
2. Lin W, Kong X, Yu PS, et al. Community detection in incomplete information networks. *Proc WWW 2012, April 16–20, 2012, Lyon, France;* pp. 341–350.
3. Nasirifard P, Hayes C. Tadvise: a twitter assistant based on twitter lists. *Proc. Soc Info*. 2011; pp. 153–160.

4. Yang X, Ghoting A, Ruan Y, Parthasarathy S. A framework for summarizing and analysing twitter feeds. *Proceedings of the 18th ACM international conference on Knowledge discovery and data mining; 2012*, pp. 370-378.
5. Alsaleh S, Nayak R, Xu Y. Finding and matching communities in social networks using data mining. *Proc. ASONAM*. 2011; pp. 389–393.
6. Baatarjav E-A, Dantu R. Unveiling hidden patterns to find social relevance. *Proc. IEEE Social Com*, 2011; pp. 242–249.
7. Ferreira LN, Pinto AR, Zhao L. QK-means: a clustering technique based on community detection and K-means for deployment of cluster head nodes. *Proc. IJCNN*. 2012.
8. Leung CK-S, Carmichael CL, The EW. Visual analytics of social networks: mining and visualizing co-authorship networks. *Proc. HCII-FAC*. 2011. pp. 335–345.
9. Cameron JJ, Leung CK-S, Tanbeer SK. Finding strong groups of friends among friends in social networks. *Proc. SCA*. 2011. pp. 824–831.
10. Jiang F, Leung C, Tanbeer S K. Finding popular friends in social networks. *Proceedings of Second International Conference on Cloud and Green Computing*, 2012.
11. Leung S, Tanbeer S K. Mining social networks for significant friend groups. *Proceedings of International Conference on Database Systems for Advanced Applications*, 2012, pp. 180–192.
12. Domingos P, Richardson M. Mining the network value of customers. *Proceedings of the Seventh ACM International Conference on Knowledge Discovery and Data Mining*, 2001.
13. Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. *Proceedings of the eighth ACM International Conference on Knowledge Discovery and Data Mining*, 2002.
14. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. *Proceedings of the ninth International Conference on Knowledge Discovery and Data Mining*, 2003.
15. Schwartz D, Rouselle T. Using social network analysis to target criminal networks. *Trends in Organized Crime*. 2009; 12(2): 188–207p.
16. Hoppe B, Reinelt C. Social network analysis and the evaluation of leadership networks. *The Leadership Quarterly*. 2010; 21 (4): 600–619p.
17. Duijn P, Klerks P. Social network analysis applied to criminal networks: Recent developments in Dutch law enforcement. In: Masys AJ, (editor). *Networks and Network Analysis for Defence and Security, ser. Lecture Notes in Social Networks*. Springer; 2014. pp. 121–159.
18. Heidemann J, Klier M, Probst F. Identifying key users in online social networks: A PageRank based approach. In Association for Information Systems, 2010.
19. Rıos SA, Munoz R. Dark web portal overlapping community detection based on topic models. *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics*. 2012, pp. 2:1–2:7.
20. Tanbeer S K, Leung C, Cameron J. *DIFSoN: Discovering Influential Friends from Social Networks. IEEE Fourth International Conference on Computational Aspects of Social Networks*, 2012.
21. Nancy P, Ramani R, Jacob S. Mining of Association Patterns in Social Network Data (Face Book 100 Universities) through Data Mining Techniques and Methods. In: Meghanathan N, Nagamalai D, Chaki N, editors, *Advances in Computing and Information Technology*. Berlin/Heidelberg, Germany: Springer, 2013. Volume 178, pp. 107–117.
22. Yu X, Liu H, Shi J, Hwang JN, Wan W, Lu J. Association Rule Mining of Personal Hobbies in Social Networks. In: *Proceedings of the 2014 IEEE International Congress on Big Data (Big Data Congress)*, Anchorage, AK, USA, pp. 310–314, 27 June –2 July 2014.
23. Erlandsson F, Borg A, Johnson H, Bródka P. Predicting User Participation in Social Media. In: *Advances in Network Science*. Springer International Publishing: Cham, Switzerland, 2016; pp. 126–135.

Cite this Article
Hilal Ahmad Khanday, Rana Hashmy, Finding Influential Users in Online Social Networks: A Tree-based Approach. *Journal of Artificial Intelligence Research & Advances*. 2018; 5(3): 66–70p.