

# Finding Active Influential User in Multiple Online Social Network

Praneetha G N<sup>1</sup>, Bhavatarini N<sup>2</sup>, Tejaswini B S<sup>3</sup>, Chaithra B M<sup>4</sup>, Kavyashree k<sup>5</sup>

Assistant Professor, Department of Information Science & Engineering, Sapthagiri College of Engineering,  
Bengaluru, India<sup>1,2,3,4,5</sup>

**ABSTRACT:** People are very much dependant on Online Social Network for various purposes such as movie review, finding experts, finding opinions about any movies, places or products. Influence Maximization is the problem of finding a small set of influential users in the online social network so that their influence in the social network is maximized. Now-a-days, people are found in more than one social network such as Facebook, Twitter, LinkedIn etc. So, finding influential user in multiple online social network is very much important. Rather than just finding influential user in multiple online social network, finding active influential user would be much more beneficial. The proposed method outperforms the previous methods of finding influential user.

**KEYWORDS:** Online Social Network, Influence Maximization, Fuzzy K-means algorithm, Active User

## I. INTRODUCTION

Social network can be analysed with the help of graphs using a technique called as Social Network Analysis. It is represented as a graph where nodes represents individuals and edges represents interactions or relations between them. The various applications of Social Network Analysis are Data Mining, Behavioural analysis, Link Prediction, modelling of network and also in Recommender systems. It is also used in business applications such as analysis of customer behaviour and marketing and analysis of customer interaction.

One of the applications of Social Network Analysis is in Viral Marketing. There are three types of marketing i.e., Direct Marketing, Mass Marketing and Viral Marketing. Direct Marketing refers to marketing to each and every individual. Mass Marketing refers to marketing to more number of people. Viral Marketing refers to “word-of-mouth” marketing i.e., it uses its customer to market the product. Customer will influence his friends for buying a product and if he gets convinced, he buys a product and starts influencing his friends and so on.

In many markets, customers are strongly influenced by the opinions of their friends. Viral marketing takes advantage of this to promote a product by marketing it primarily to those with the strongest influence in the market. Viral marketing uses the customers in the market to promote a product. This way of marketing is very much beneficial than direct marketing and mass marketing. Further people trust and act on recommendations from friends and they further influence their friends. This is referred to as influence propagation [2]. For example, influence propagation can help decide which sports to watch, which item to purchase and so on. Hence, influence propagation has become an important mechanism for viral marketing. This further motivates the researchers to carry out studies on different aspects of the influence propagation problem. Influence Maximization problem is a problem of finding a small set of nodes that maximizes the spread of influence.

The processing time of older model increases exponentially as network size increases. Now-a-days users are found in multiple online social networks rather than one. User found in more than on network is referred as crossing users. These crossing users are very much important in propagating information in multiple online social networks. In the previous methods proposed, only influential users were taken into account, but in the proposed method the “active” influential user are considered. Active users can be decided based on various factors such as number of posts, likes, views and shares. Fuzzy k-means algorithm is being used to solve Influence Maximization Problem i.e., to find

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 5, May 2018

influential users in multiple online social networks so the influence reach is maximum. The processing time of this algorithm increases linearly as network size increases, which is better than IC and LT model.

## II. RELATED WORK

Domingo's and Richardson [1] gave the first algorithm to deal with influence propagation problem. They built probabilistic model of influence for mining the data on knowledge-sharing websites. Knowledge-sharing sites are the sites where customer review products and advise each other about the products. Customer's have two types of values: intrinsic value and network value. Intrinsic value of a customer is the value of a customer based on the products he is likely to purchase and the network value of a customer is high when he is expected to have a very positive influence on other's probabilities of purchasing the product. A customer with high network values is the one who is worth of marketing. It concluded that by building the probabilistic models and applying those models to the knowledge sharing websites, solved the influence propagation problem and their method is scalable to large networks. But, the method mined a network from single source and not from multiple sources.

Then, Kempe et al. [2] studied influence propagation using two fundamental propagation models, named Independent Cascade (IC) Model and Linear Threshold (LT) Model. Here, a social network is modelled as a graph with nodes representing individuals and edges representing connections or relationship between two individuals. Influences are propagated in the network according to the model, such as the independent cascade (IC) model. Kempe et al. proved that the optimization problem is NP-hard, and present a greedy approximation algorithm guaranteeing that the influence spread is within the optimal influence spread. The main disadvantage of the algorithm is that it was less efficient.

Barbieri et al. [3] also studied the problem on influence propagation from a topic modelling perspective. The standard propagation models suffer various limitations when it comes to model real-world cascades. Observations made were: (i) users have different interests, (ii) items have different characteristics and (iii) similar items are likely to interest the same users. Thus, topic-modelling perspective is proposed to jointly learn item characteristics, user's interests and social influence, resulting in new propagation models that experimentally are proven to be more accurate in describing real-world cascades. They extended the classic IC and LT models to be topic aware. The propagation models obtained are dubbed Topic-aware Independent Cascade (TIC) model and Topic-aware Linear Threshold (TLT) model. A new influence propagation model was developed, named AIR (Authoritativeness-Interest- Relevance) model. Instead of considering user-to-user influence, this model focuses on user authoritativeness and interests in a topic, leading to a drastic reduction of the number of parameters of the model.

## METHODOLOGY

The main objective of this work is to find influential seed users from multiple online social networks (OSN). In order to achieve this objective, the design and implementation of the project is done in phases.

## III. PROPOSED SYSTEM

The work is divided into three phases.

### 3.1.1 Phase I: OSN Data Processing

In this phase, the user information such as his identity, relationship with other users in the network, activities like time at which recent message sent, recent message received etc is read from different OSN. Following are the tasks performed in this phase to process the OSN data.

1. The user activity data from different online social network is collected and the data set is generated.
2. The data from different OSN is combined to make a single OSN data. The combined data is written to a file for further processing.
3. The users who are common in all the OSNs are extracted and listed.
4. The user activity data from all OSN is parsed and information is extracted.

### Phase 2: Data Aggregation and Generation of Data Points for clustering

In this phase, the user activity data from combined OSN data file is read and data points of each user are generated for clustering. The list of data points used for clustering the users is given in table 1.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 5, May 2018

The various data points considered to form clusters are number of posts, views, likes and shares of each user. Activeness factor for each user is calculated using the formula

$$\text{Activeness factor} = \sum a_i * c_i$$

Where  $a_i$  is the values of attribute of each user and  $c_i$  is the cumulative value for each attribute. The value of  $i$  can range from 0 to  $n$  where  $n$  is number of attributes. Activeness factor can be normalized using the formula.

$$E_i = \frac{e_i - e_{\min}}{e_{\max} - e_{\min}}$$

USER ID	NO. OF POSTS	NO. OF VIEWS	NO. OF LIKES	NO. OF SHARES	ACTIVENSS FACTOR
1	5	500	250	20	2610
2	5	350	100	10	1510
3	8	400	150	60	2116

Table 3.1 Data Points for Clustering

### Phase 3: Fuzzy k-means Clustering and Finding Seed Users

In this phase, the users are clustered with their data points using Fuzzy K-Means Clustering Algorithms and their membership association value with each clusters are found. The procedure to cluster the data points and find seed users are given below.

1. The user data points generated in phase 2 are fed to the Fuzzy K-Means Clustering algorithm.
2. After the clustering is performed, the membership association value for all the users in the combined OSN is generated.
3. From the User Membership Association Value, the common user who has got highest associativity with the given cluster is chosen as the seed user for that cluster. Similarly, the seed users are found for all the clusters. For each cluster we can also find more than one highest associative common user if more number of seed users is to be found.

## IV. RESULTS

The proposed method is implemented using Fuzzy k-means clustering algorithm. Instead of considering just influential user across multiple online social network, we consider activeness of that particular user with the parameter called as activeness factor mentioned in table1. The activeness factor is normalized and then the normalized value is taken for clustering. The processing time for the proposed method is less when compared to older methods as shown in figure 4.1

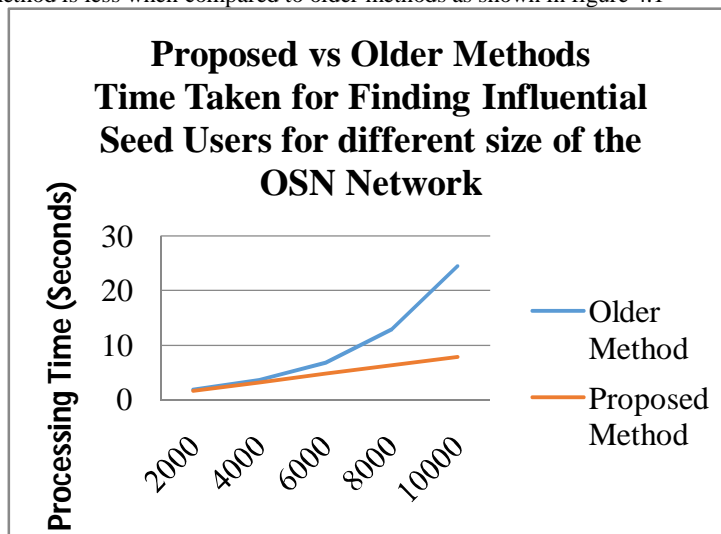


Fig 4.1 Graph showing Processing time

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 5, May 2018

## IV. CONCLUSION

The study of Influence Maximization is being studied from years and solutions like Independent Cascade (IC) model and Linear Threshold (LT) Model was proposed. But this model neglected the effect of active users in the network and processing time of these model increased exponentially as network size increases. Hence, we proposed a new approach towards Influence Maximization in multiple online social networks i.e, using clustering technique. Fuzzy k means clustering is being used to find the influential seed user from multiple online social networks. And also we are considering active influential user rather than just influential user. Processing time increases linearly as network size increases. Hence, processing time is reduced when compared to the diffusion models like Independent Cascade (IC) model and Linear Threshold (LT) model.

## REFERENCES

- [1]. Kempe, J. M. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. Of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'03)*, 2003, pp.137-146.
- [2]. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance, "Cost-effective outbreak detection in networks," in *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'07)*, 2007, pp. 420–429
- [3]. W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks", in *Proc. of the 15<sup>th</sup> ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'09)*, 2009.
- [4]. W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model", in *Proc. of the 10th IEEE Int. Conf. on Data Mining (ICDM'10)*, 2010..
- [5]. Y. Shen, T.N. Dinh, H. Zhang, and M. T. Thai. "Interest-matching information propagation in multiple online social networks in *CIKM*, 2012.
- [6]. Dung T. Nguyen, Huiyuan Zhang, Soham Das, My T. Thai, "Least cost influence in multiplex social networks: Model Representation and Analysis, 2012 IEEE 13<sup>th</sup> International Conference on Data Mining.