

# Least Cost Influence by Mapping Online Social Networks

<sup>1</sup>Bessmitha S, <sup>2</sup>Shajini N

<sup>1</sup>PG Student, Dept of CSE, <sup>2</sup>Associate Professor, Dept of CSE

---

**Abstract:** The online social network has become popular for sharing the information. Online social networks exhibit many platforms to create awareness of new products. In recent time Least cost Influence problem to find minimum number of seed user is most important topic in online social networks. The eventual target is to find the least advertising cost set of users which produce enormous influence. In existing many diffusion models are used. In this paper the stochastic threshold model is used to find the seed user in multiple online social networks to maximize the influence. This model decreases the processing time comparing to the other models.

**Keywords:** Stochastic threshold model, influence, diffusion, multiple networks, online social networks.

---

## I. INTRODUCTION

A social network is a social structure made up of a set of social actors such as individuals or organizations, 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. A Social network can be defined as a set of socially relevant individuals with some patterns of interactions or contacts among them, which are connected by one or more online relations. Online social networks (OSN) provide platforms for users to share information and to communicate with their families and friends. With the prosperous development of online social networks, many marketers have exploited the opportunities and attempt to select influential users within online social media to influence other users through online.

The growth of online social networks like facebook, twitter, gmail and so on has created major communication medium. The people of 83% believe platforms like twitter, facebook help them make new friends. Each day facebook users spend 10.5 billion minutes on the online social network. 50% of facebook users have more than 100 friends, 340 million tweets are sent out every day on twitter. An online social network can be defined as a set of socially relevant individuals with some patterns of interactions or contacts among them, which are connected by one or more online relations. Online social networks provide platforms for users to communicate and share information with the relatives and friends effectively, connecting users across the world. The various online social networks are fundamentally classified by their node degree distributions. The most frequently mentioned categories of online social networks are randomly distributed, scale free and complex networks.

Online social networks (OSN) are self-organizing, emergent, dynamic, and complex. They are most commonly characterized as scale-free and small-world effect. The scale-free network was defined as a social network whose degree distribution is at least infinitely close to a power law form indicating that the network is independent of the social network scale. The main feature of a scale-free social network is inhomogeneity. Most of the nodes in a social network have few physical link connections, while the connections of a few nodes far exceed the average number of physical link connections within the social network. An online social network is the small-world effect. It describes that almost any two nodes in the online social network even non-neighboring nodes can be connected by a very short path. In addition, small-world online social networks are governed by two important parameters clustering, coefficient and the diameter.

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.

Further people trust and act on recommendations from friends and their further influence their friends. This is referred to as influence propagation. Influence propagation has become an important mechanism for viral marketing. This further motivates the researchers to carry out extensive studies on various 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.

**Table 1: list of Abbreviation**

LCI	Least Cost Influence
OSN	Online Social Networks
LT	Linear Threshold
IC	Independent Cascade

## II. RELATED WORK

In existing the single network was used. In this it is very difficult to find the overlapping users. While considering the single network the information are spread inaccurately. The influence in only one network fails to identify the most influential users [1]. There are also works on Linear Threshold (LT) model in which the active user can be found only up to a threshold value. Then the greedy algorithm is used in which the improvements are slow and not scalable. Here, a social network is modeled 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 [2]. The researches have started to explore multiplex networks [3], [4] which studied the connection between offline and online networks. In existing it focus on information sharing but it does not focus on solving least cost influence problem.

## III. PROPOSED SYSTEM

The Least Cost Influence (LCI) problem is solved by coupling multiple networks into single network. The seed user is found from multiple online social networks. The seed user refers to the user who is active for more time in online social network. The seed user will propagate massive influence. To find the seed user the stochastic threshold model and improved greedy algorithm is used. In particular, the improvement factor scales up with the size of the network which allows the algorithm to run on very large networks with millions of nodes. It improves the scalability in large network and has better marketing.

### A. Network initialization:

A social networking service is a platform to build social networks or social relations among people who share similar personal and career interests, activities, background or real life connection. Create one or more social network for the information sharing in the online social network so it helps us to gather information from one user to another user.

The accounts are initialized by registering all the individual elements into database and common password is given. By giving network id and password login to the network and the information are shared.

### B. Coupling network:

The multiple networks are coupled into single network by using star lossless coupling. The lossless coupling scheme preserves all properties of original networks to achieve high-quality solutions. The star lossless coupling is the one in which multiple inputs are given and it gives one output. In star loss less coupling the coupling is done by synchronizing the network edges. The networks are coupled by connecting intermediate node. The intermediate node act as the gateway and all the information are transferred through the intermediate node. The lossless coupling scheme returns solution with higher quality. To understand the benefit of taking consideration of overlapping users and coupled network, influencing a fraction of the nodes in all networks by selecting seeds from each network and taking the union to compare with seeds achieved from lossless coupling scheme.

### C. Stochastic threshold model:

The stochastic threshold is the threshold at which the analyst can be confident that if one peak for a heterozygote is above this threshold. This increases confidence in homozygous calls made for single source samples and alleles attributed to

each contributor in mixture analysis. The stochastic threshold is the point at which variation begins a rapid increase. In this model the heuristic of choosing nodes with probability distribution is done. The first node if chosen it will activate a large fraction of network. This model identifies a subset of individuals whom adopt a service or a product and in turn influence others, producing an effect which spreads in the social network as much as possible.

#### D. Finding seed user:

The seed users are the users that actively participate in network. In multiple networks the seed user is found based on all online social networks they participate. Seed user can trigger massive influence diffusion. The seed user is found by implementing greedy algorithm. The greedy algorithm is designed to achieve optimal solution. A scalable greedy algorithm is used to solve the LCI problem. The greedy algorithm runs much faster in the Lossless coupled networks. In particular, the improvement factor scales up with the size of the network which allows the algorithm to run on very large networks with millions of nodes. Considering multiplex networks instead of a single network can effectively choose the most influential users. In particular, the improvement factor scales up with the size of the network which allows the algorithm to run on very large networks with millions of nodes. Considering multiplex networks instead of a single network can effectively choose the most influential users.

In greedy algorithm the threshold value is used. Given a system of coupled networks  $G$ . Initialize a heap. Add the active users having higher marginal gain with similar interest in to the heap. Counter value is updated for every addition of new users in to the heap. Repeat the process until there is no active user. Returns the node of maximum marginal value from a max heap after removing least marginal value user from the heap. The counter displays the updated value. The users who are using the network above threshold value are inserted in to the heap. The heap sorts the active user by extracting the maximum user. The heap is a special case of balanced binary tree data structure where root node is compared with its children and is sorted as  $key(\alpha) \geq key(\beta)$ . The seed users are found and the information is shared through the seed user to the different network. The lossless coupling scheme achieves the best result in both networks. When the networks is considered as a standalone network and choose seeds individually, the seeds size is relatively larger than choosing from the coupled network. The sizes decrease this improvement is also due to the information diffusion across several networks by the overlapping users.

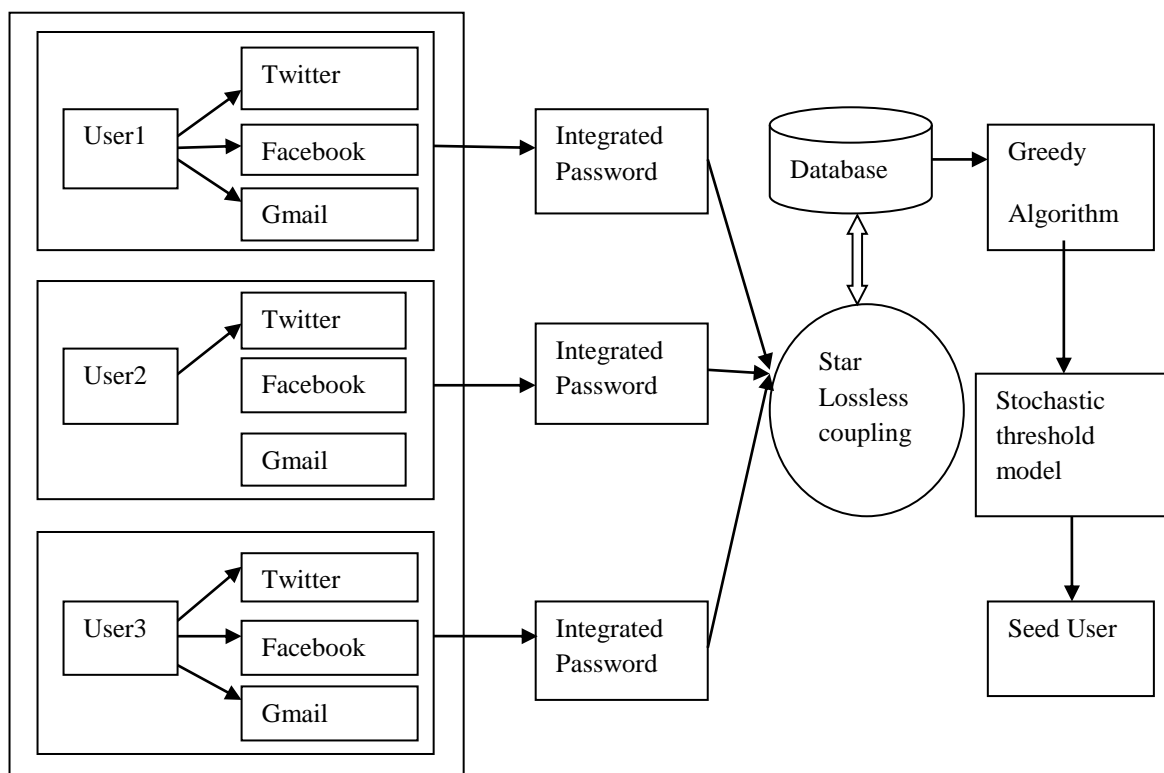


Fig 1: Architecture diagram

The Fig.1 explains the system architecture of the proposed system to find seed user. The seed users are the one who active in online social network for long time. It is mainly used in viral marketing. When a company produces a new product they advertise the new products in online social networks. While advertising these products in single network its cost is very high and the seed user is found inaccurately because the seed user found in one network will be highly active in other network. To avoid this multiple network is used. The same user uses multiple networks. So to find the seed user in multiple networks the networks are coupled into single network. In this project facebook, gmail and twitter are considered. The user registers to multiple network and give the integrated password. The multiple networks of the different user are coupled using star lossless coupling. After coupling all the multiple networks into single network all these information are stored in the database. From the database by using greedy algorithm the performance of each user is analyzed and the seed user is found. In greedy algorithm the users above threshold value are inserted into the heap and heap sorts all the user and maximum user will be the seed user.

#### IV. CONCLUSION

The study of Influence Maximization is being studied from years. But this model neglected the effect of active users in the network and processing time of these model increased as network size increases. Hence, we proposed a stochastic threshold model for Influence Maximization in multiple online social networks. This model has better marketing and reduces the processing time.

##### Advantages:

- Improves the scalability in large networks.
- Less processing time.
- It triggers a massive influence.
- It has better marketing.

##### Disadvantages:

- It relies on low degree of nodes.
- Less realistic marketing.
- Processing time is high.

#### REFERENCES

- [1] Huiyuan Zhang, Dung T. Nguyen, and My T. Thai, "Least Cost Influence Across Multiple Social Networks,"IEEE/ACM Trans. Netw., vol.24, no.2, April 2016.
- [2] Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proc. KDD, 2003, pp. 137–146.
- [3] O. Yagan, D. Qian, J. Zhang, and D. Cochran, "Information diffusion in overlaying social-physical networks," in Proc. CISS, 2012, pp. 1–6.
- [4] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, and J. Han, "Event based social networks: Linking the online and offline social worlds," in Proc. KDD, 2012, pp. 1032–1040.
- [5] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in Proc. SIGKDD, 2010, pp. 1029–1038.
- [6] H. Zhang, T. N. Dinh, and M. T. Thai, "Maximizing the spread of positive influence in online social networks," in Proc. IEEE Int. Conf. Distrib.Computing Syst., 2013, pp. 317–326.
- [7] T. N. Dinh, H. Zhang, D. T. Nguyen, and M. T. Thai, "Cost-effective viral marketing for time-critical campaigns in large-scale social networks," IEEE/ACM Trans. Netw., vol. 22, no. 6, pp. 2001–2011, Dec.2013.