Online Influence Maximization Using Rapid Continuous Time Independent Cascade Model

Annu kumari
Research Scholar
Department of Computer
Science and Engineering ,ASET
Amity University Uttar Pradesh ,India
kumariannu.amity98@gmail.com;
anumanish98@gmail.com

Dr.Shailendra Narayan Singh
Department of Computer Science
and Engineering, ASET
Amity University Uttar Pradesh, India
snsingh36@amity.edu;
sns2033@gmail.com

Abstract*-Do one really know the meaning of Online Influence(OI) Maximization? Do you know why do we need to calculate the influence of social networking sites? How to measure the influence of online maximization? How does it really works? If you have been pondering for the answers of the questions then this paper will assist you to identify and connect with influences of online maximization. Influence Maximization is the problem in which subset of seed nodes are found within the social networks which maximizes influence on other nodes in their ties and relationships. Influence maximization has been developed to find out how influence gets propagated through its network. The concept of Influence Maximization lies in the selection of minimal set of seed nodes which propagates maximum of its influenciality within a network. This paper firstly comprises of previous models used in appropriate selection of seed nodes i.e. Linearly Threshold Model(LT), Classic Cascade Independent s Model(IC) , Extended Classic Independent Model(EIC). Then, I proposed a new Model Rapid Continuous Time (RCT) Independent Cascade Model that can be used in the Classic Independent Model(IC) .

Keywords: Online Influence(OI), Linear Threshold(LT), Independent Cascade Model, Rapid Continuous Time (RCT) Independent Cascade Model

I. INTRODUCTION

Today's era is the era of online marketing and the Online Influence(OI) Maximization plays a very crucial in applications of viral marketing, analyzing infection diffusion in community, calculating social capital ranking scores and many more. Communities involved can be described using Graph Model and various algorithms which signifies the propagation of influence from one node to its neighbor nodes. Various models have been proposed in the past literature and most popular and the basic models are Linearly Threshold (LT) Model, Classic Cascade Independents(IC) Model and many algorithms have been proposed too such as Simulated Annealing, Greedy Algorithm, Genetic Algorithm[7].Influence Propagation in terms of time and space plays a crucial role in social networking marketing. Online Social networking sites like Facebook, Twitter, Pinterest all having different functions and users, connect its users into a virtual society. Each user within the network is termed as a vertex and the relationships are the edges relating these vertices. The communication taking place between them is either one way or two way totally depending on its relationship. Facebook and Friendster are nowadays have become successful as they are effectively connecting people in huge numbers in very short time span for long lasting. But, these are also becoming a large dissemination and marketing platforms, as huge amount of information and ideas is influenced in huge population in small period of time span. Marketing and dissemination of social networks has to cope up with many problems and challenges.

Strategies in Viral Marketing comprises of inviting some initial users by a company. Now these initial users are considered as "seed nodes." These initial users are given samples new products and technologies to use. Positive and good feedback is expected from these seed nodes by the company on

978-1-5090-3519-9/17/$31.00 ©2017 IEEE

social networking sites. Using power of word-of-mouth, these seed nodes(customers) will influence their neighbors to buy or use those new products and technologies given as sample to them. Now these "seed nodes" have to influence their neighbors social networking sites. Subsequently, influence is propagated to their social network neighbors, which were themselves affected by the influence and this process carries forward. But, the main challenge is the proper selection of seed nodes is to maximize the profit of the company. This paper firstly comprises of previous models for the selection of seed nodes i.e. Linearly Threshold Model(LT), Classic Cascade Independent Model(IC) , Extended Classic Independent Model(EIC). Then, I proposed a new Model Rapid Continuous Time (RCT) Independent Cascade Model that can be used in the Classic Independent Model(IC) .

II. LITERATURE REVIEWS ON RESEARCHES

A. The phenomena such as mob, riots or strikes[1], as a true physical existing turned into psychological happenings, led to the analysis of influence propagation through social media. Thus, it can be stated that, individual decision are taken as the result of social interactions. The same concept applies to the propagation of advertisements, innovations, rumors[2], thus leading this topic to become a part of research in Marketing fields. The main challenge in Research area was to define an appropriate balance between word-of-mouth propagation and marketing efforts. These both are the strategies to be done through social networking sites defined by strong and weak ties and relationships.

B. Domingos and Richardson [3] were the first who studied and proposed an algorithm for viral marketing. They did not considered the users as independent individuals .They selected customers network value based on their intrinsic values of each users, they defined an expected total profit achieved from all the customers network value, whom they influenced directly or indirectly.

C. Kleinberg, Tardos, Kempe[4] were the one who proposed and formulated the optimized discrete concept. A graph is figured as a social network site. This graph consists of nodes or vertices representing users, customers or individuals. They also contain edges representing the relationship or strong ties and weak ties between the individuals. Stochastic Cascade Model was used to check out the influence propagation in the network. [4] used mainly three Models i.e., The Linearly Threshold Model, The Cascade Independent Model and the Weight Cascade Concept. Consider a graph of network connected socially, a Cascade Independent Model and a few seed nodes as 's', maximizing influence theory says to find a small number seed nodes 's' which influences maximum number of seed nodes.

They proved that the discrete optimization problem is NP-Hard. It resulted as an optimized solution which is 63% of (1-1/e) of Cascade Independent Model (IC) and Linearly Threshold Model (LT). To get an approximation result to optimal within the bound (1-1/e) using Hill Climbing Approach and Greedy Approach. But, the major drawback lies in the efficiency. Calculating influence of given size of seed nodes like 15000 nodes proved to be challenging. They ran Monte- Carlo simulations many times to get accuracy but failed. In a large network setup it is difficult to find out small number of seed nodes.

D. Many recent studies aimed to resolve the efficiency issues. Kimura and Saito[5] also proposed an algorithm to find Influence Cascade(IC) Model based on shortest path but they couldn't made it as they did not resolved the efficiency issues discussed by[4].

III. CONCEPTS OF VIRAL MARKETING

This paper firstly comprises of previous models for the selecting seed nodes i.e. Linearly Threshold Model(LT), Classic Cascade Independent Model(IC) , Extended Classic Independent Model(EIC). Then, I proposed a new Model Rapid Continuous Time (RCT) Independent Cascade Model that can be used in the Classic Independent Model(IC) .

A. Domingoset.al[3] was the first to propose the algorithm in the study of Viral Marketing. Nowadays, ideas are spreading like viruses. This was the original concept of Viral Marketing which enforced researchers to research on this interesting topic. Viral Marketing and Direct Marketing are totally two different concepts of Marketing having opposite Marketing Strategies. Direct Marketing evaluates each customers independently which is opposite to that of viral marketing.

Let us consider an example for a Company or an Organization, how do they select their seed nodes(customers or users). Consider the term A as investment by a Company to a user, and return expected be B, and the profit P can be determined by the Company when user buys purchases the product of a Company. We calculate profit as P=B-A. When the value of profit will be positive value then only the company will consider that particular user as potential user. The value of B i.e. the expected return is totally different in the case of direct marketing and viral marketing. Direct Marketing calculates the expected return value considering each customer as an individual and independent of rest of the users. The Company considers only those users whose decision of purchase are independent and are directly purchased or not being affected by other users appreciation and persuasion. Thus, direct marketing considers only those customers as valuable and potential users who purchases products direct from the company itself as an independent decision.

But, Domingos et.al[3] enquired that customers existing in a society are not independent , but are surely connected to each other via social network. Thus, the marketing decision must not be an individuals decision but may be surely affected by a network value. Thus, he calculated the expected return of a user as the sum of an individual added to their network value. Thus, the network value can be defined as the purchase action of customer which are influenced directly or indirectly. He argued on the concept that customer with low individual value having lower investment A is equally valuable customer for the company, than to the user having high network value in his social network. In this way, the purchases and opinions of individual affect his/her connected users in viral marketing.

IV. VARIOUS PROPAGATION MODELS

Domingos et.al[3] considered only those customers who purchase decisions were not individual i.e. users whose decisions were not independent and their purchase actions were directly or indirectly were dependent on the opinions of connected users.

We can model these people as networks. There are basically two kinds of models in socialized networks namely Linearly Threshold model and Classic Independent Model. Many more models have been designed , considering these two models as the basic model. These models aim to reflect the relationships and ties between the nodes in social networks. Considering various parameters in a social network G as directed graph comprising the entire socialized network , N is the set of vertices , consists of node ‘n’ representing a single node in N in the network. Each node ’n ’ can have only two states i.e. ‘active’ or ‘inactive’. We can say that an ‘active‘ status is representing a node ‘n’ which are affected by other nodes. W is the set of edges that contains all the relationship between nodes having different weights reflecting their relationship strength between two nodes. Maximizing Influence problem can be stated as: Selecting seed nodes size ‘k’ from N, then what can be the minimal and optimized seed node used to maximize the seed nodes that gets affected in the social network. The propagation of influence from seed nodes between the other nodes exist until no more nodes are affected. For this model, we have to maintain two constraints. The first one is, the change in the status of the node is considered to be irreversible. In each and every step , we are considering the nodes to be in two states only i.e. active or inactive. When a node/vertex changes its status from inactive status to active status , it further cannot return to the active state. The other constraint is that the propagation of inactive status node to convert into an active is increasing in a monotonous form as its neighbor active nodes. Thus, the probability of inactive to be active is also increasing. Now considering these two models we are going to discuss various propagation models.

A. LINEARLY THRESHOLD MODEL(LT)

[1,6] were the first to propose the Linearly Threshold (LT) Model which simulated the propagation of influence taking place in social networks. This model basically comprises of a threshold value which is made specific. This threshold value signifies the difficulty in switching from inactive status to active status. If a threshold value is large it is more likely to switch its status. In this model, node ‘p’ is connected to a set of neighbor nodes M. Each node ‘p’ by an edge having having weight d(p,q). Consider, also that the total weight in set M must not be greater than 1 i.e.[$\sum d(p,q) \leq 1$]. Each node ‘p’ has a threshold of θp such that its values lies between 0 and 1i.e.($0 \leq \theta p \leq 1$), which is the minimum requirement for its active neighbor nodes set. When this value of total weights

of all its active neighbor nodes is more than θp then only inactive nodes switches its status to active. Because of the first constraint, the changes in the status are irreversible thus, if the node is in active state at step ‘s’, then it will remain active at step ‘s+1’ also. We can consider any constant value of θp ranging between 0 to 1. This threshold value can be achieved from a real time social network structure or can be randomly set to a constant value say as 0.3. This θp is reflecting the capacity of a node to adopt a new idea when it comes under the influence of various neighbor nodes.

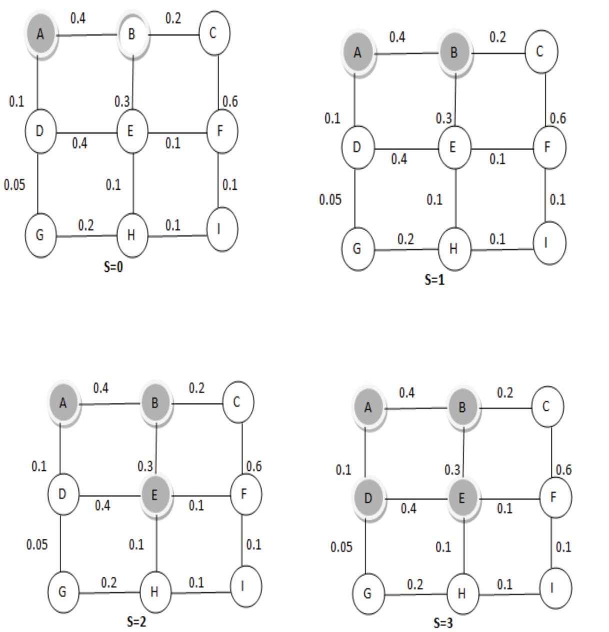

FIGURE 1: PROPAGATING INFLUENCE USING LINEARLY THRESHOLD MODEL

In the above figure, we are assuming that all the nodes are having the same threshold value of 0.3. The weights between the two nodes are represented using the edges between them. Considering node A, as the seed node with status active at s=0. Now, node A is trying to propagate its influence on its neighbor vertex B and vertex D. The AB edge is having a weight of 0.4 and edge AD is having a weight of 0.1. In the above scenario, only node B is going to switch toactive status from inactive status. This is because the total summation of weight of node B is larger than the threshold value of 0.3. Thus, resulting node B to be active status from inactive status at step s=1.

Now, we see still the total weight of active neighbor of node D is 0.1. and node D remains inactive at step s=1. Continuosly following same steps, node E switches to actives tatus by the influence of node B at step 2. Node D is still in inactive state at step 1. It will too switch to active state as the influence propagation will proceed up. At s=2,vertex E switches to active state, thus the total summation of weight of active neighbor vertices of vertex D is larger than threshold, thus vertex D also becomes active at s=3. Thus, at step 3 there are no more nodes satisfying the above condition and therefore propagation stops. We can observe many things from this propagation model. Firstly, though thereare few people who are not willing to accept new technologies at the beginning and thus are more likely to change their minds as more of their neighbors are accepting the new terminologies and technologies. We can represent these people by node D from the above example. We can also note one more thing that node C and node F are also having total weights higher than thethreshold value, but still they are not active . This is because their neighbor nodes are not enough powerful to influence them. People who are eager to accept new technologies can be reprsented by node C and node F.

Thus, we can see that selection of inappropriate seed nodes leads to failure in propagating the influence and discover these customers. Thus, prior concern must be given to consider the appropriate seed nodes.

B. CASCADE INDEPENDENT MODEL(IC)

This model mainly comprises the concept of probability theory. Goldernerg Libai and Muller[2] proposed the Simple Independent Cascade Model. This model is explained as: Select a seed node initially at step s=0. If node A becomes active at step ‘s’, then itwill continually try to influence its neighbor at step ‘s+1’ with some probability ‘P’. Now, node A will behaving only one chance to influence its neighbor nodes which would be inactive and further it will not get another chance to influence its other inactive neighbors nomatter whether it succeds or not. And if the inactive node is being connected with more than one inactive nodes , these inactive nodes will be continuosly affected by the influence of newly active nodes in a random sequence. Thus, the process stops if no more nodes can be affected. The influence propagation in IC Model has been decsribed in the following two figure i.e. figure2 and figure3, each representing the output in a single run. In both figures node A is being selected as the seed node. It will popagate its influence to each active neighbor nodes having some propagation probabilities between them. We will see each run will give a different result. We can see in figure 2 , that influence propagation reaches to four other nodes(node D, node E, node F, node H) and in figure 3 the influence propagates to (node B, node C, node E, node F, node H). Thus, we see the results are different in both the runs. Leading to a conclusion that we need to run this model multiple times to obtain a result of accurate estimation with appropriate selection of seed nodes.

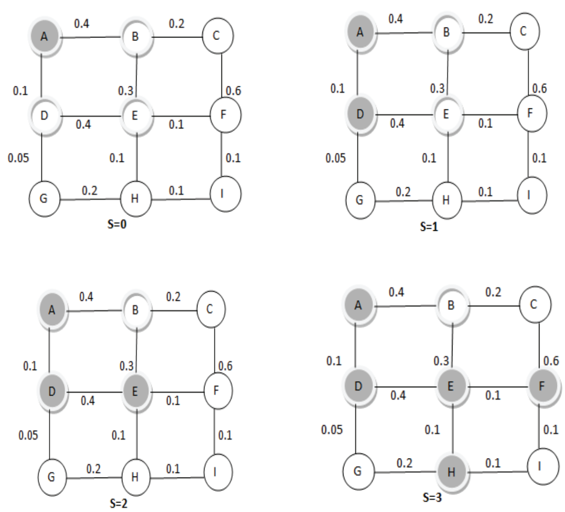

Figure 2: PROPAGATING INFLUENCE IN CASCADE INDEPENDENT IN FIRST RUN

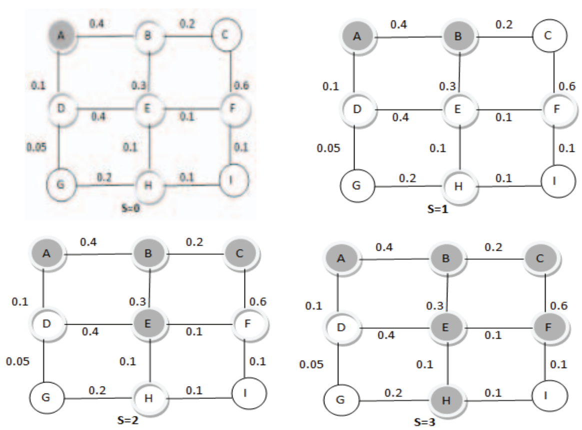

Figure 3: PROPAGATING INFLUENCE IN CASCADE INDEPENDENT MODEL IN SECOND TRY

C. EXTENDED CASCADE INDEPENDENT MODEL(EIC)

Wang,Qian and Lu[8] were the one who extended the Cascade Independent Model and proposed the Extended Cascade Independent Model(EIC). In the EIC Model, onlu one probability does not determine the influence propagation rather the propagation process is basically divided into two process, one the spreading probability from an active node and the other is the adopting probability from the inactive node being influenced. We have two more terms p(s) and p(a), where p(s) signifies the the probability of the active node to decide whether to spread the influence to its neighbor nodes and p(a) signifies the adopting probability of the neighbor nodes. And the new propagation probability is calculated by the multipier product of p(s) and p(a) i.e. $p=p(a) \times p(s)$.The Extended Independent Model was a good extension of Classic Independent Model. But, this model had few drawbacks i.e. the propagation influence process is static and all the nodes are given only one chance to propagate influence.

D. CONTINUOUS RAPID TIME INDEPENDENT CASCADE MODEL(CRT)

To overcome the drawbacks of EIC Model I Proposed a new cocept to be called as Continious Rapid Time Independent Cascade Model. What if we consider the propagation process on continious basis and rapid dynamic time? This can be explained as : The propagation influence must keep on propagating by the active nodes untill and an unless there are no more further inactive neighbor node. The nodes when not succeded in one run may also be given an another chance to influence its neighbor inactive nodes.But , a time quantum must be fixed for every run. Once the first run for the active nodes completes it will be given another chance of fixed time quantum to propagate influence when not suceeded. This slice of time quantum must be repeated when have got a chance of their single run.Thus, this concept will surely improve the probability of time propagation of EIC Model. This is explained below: Consider all edges have a probability of 0.3. at s=0, consider node A as The seed node. At s=1,node A influences node B and node B changes its status from inactive to active. At s=2, node C and node E are activated by the influence propagation of node B. At s=3, though Node A fails to influence node n at s=0, but node A gets another chance and opputunity suceeds. We can see a major differnce in Classic Independent Model and the Continuous Rapid Model that if a node fails in first run, it is still given a second and third time slice which will guarantee its influence.

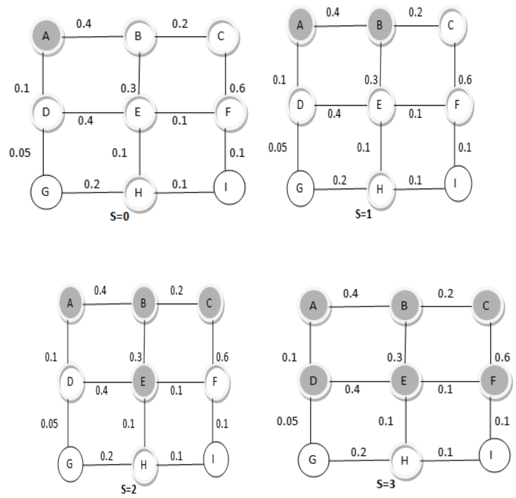

Figure 4: CONTINUOUS RAPID TIME INDEPENDENT CASCADE MODEL

V. CONCLUSIONS

Our paper comprises of various models have been discused like Linearly Threshold Model, Classic Cascade Independent Model,Extended Cascde Independent Model. To overcomefew drawbacks of Extended Independent Cascade Model I develpoed new Model called Continuous Rapid time Model.This Model says about introducing time quantum to give every active node another chance to propagate influence to inactive neighbor nodes which overcomes the static and one time behaviour of EIC Model.

REFERNCES

[1] M.Granovetter,"Threshold Models Of Collective Behaviour ", 'The American Juornal Of Sociology',83,pp.1420-1443,1978.

[2] J.Golderberg,"Talk of the Network:A Complex Systems Looking At The Underlying Process Of Word-Of- Mouth", 'Marketting Letters',pp.211-223,2001.

[3] M.Richardson and P.Domingos,"Mining Knowledge Sharing Sites For Viral Marketing",'In Proceedings Of The 8th ACM SIGKDD Conference On Knowledge Discovery And Data Mining ',pp.61-70,2002.

[4] D.Kempe, J.M. Kleinberg, E.Tardos,"Maximizing The Spread Of Influence Through A Social Network",'In Proceedings Of the 9th ACM SIGKDD Conference On Knowledge Discovery And Data Mining',pp.137-146,2003.

[5] M.Kimura and K.Saito,"Tractable models For Information Diffusion In Social Network",'In Proceedings Of 10th European Conference On Principles And Practice Of Knowledge Discovery In Database',pp.259-271,2006.

[6] Schelling TC,"Micromotives And Macrobehaviour", 1978.

[7] J.David, Borja Ayerdi, Maneul Grana, Michal Wozniak," A New Heuristic For Influence Maximization In Social Networks ",'In Proceedings Of Logic Journal Of IGPL Advanced Access ', Published August 4,2016.

[8] Qian Z, Wang Z, Lu S,"A probability Based Algorithm For Influence Maximization In social Networks",'In Proceedings of The 5th Asia-Pacific Symposium On Internetware, Changsha, China',ACM,pp.1-7,2013.

[9] L.Leskovec, A.Krause, C.Guestrin, C.Faloutsos, J.Vanbriensen and N.Glance, "Cost Effective Outbreak Detection In Networks", 'In Proceedings of KDD',2007.

[10] P.Kumaran, S.Chitrakala,"Topic Sensitivity Based Community Formation In Online Social Network", 'In Proceedings Of International Conference On Circuit Power And Computing Technologies ',pp.978-1-5090-1277,2016.

[11] M.Sachan, T.A.Faruquie, D.Contractor and L.V. Subramaniam,"Using Content And Interactions For Discovering Communities in Social Networks",'In Proceedings Of 21st International Conference WWW',pp.331-340,2012.

[12] H.Zhang, M.T,Thai, T.N.Dinh,"Maximizing The Spread Of Positive Influence In Online Social Networks",'Distibuted Computing Systems ICDCS, IEEE 33rd International Conference ,IEEE', pp.317-326,2013.

[14] A.Goyal, L.V.Lakshamana and F.Bonchi," Learning Influence Probabilities In Social Networks",'In Proceedings Of Third ACM International Conference On Web Search And Data Mining', pp.241-250,2010.

[15] H.Li,S.S.Bhowmick,A.Sun, and J.Cui,"Conformity Aware Influence Maximization In Online Social Networks", VLDB JOURNAL ,Vol. 24,pp.117-141,2015.