ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

Product Promotion in Social Networks by Query Approach

Nagalakshmi R
Pg Scholar
Department of computer science and engineering
Anna university regional campus
Madurai, Tamilnadu, India
Lakshminags3093@gmail.com

Shanmuga priya E
Assistant professor
Department of computer science and engineering
Anna university regional campus
Madurai, Tamilnadu, India
vijaylaya2000@gmail.com

Abstract:- **Now a days Social networking plays a major role in business marketing. To promote business through viral marketing Influence maximization is introduced. The influence is valuable only if it reaches the specific user. But finding the specific user from others in the wide spread social network is not accomplished with influence maximization. In order to focus on maximizing the influence to specific user, a query based approach is discussed in this paper. A fast greedy-based approximation method based on the proposed expectation model is discussed in this paper. We investigate the Relationship between paths of specific user for the expectation model. Experiments are conducted with real life datasets and the results are compared with the existing methods. The results shows that with the proposed method, high accuracy is achieved and faster than the existing method.**

I. INTRODUCTION

In recent years, online social networks have great impact on individuals and helps in marketing such as Twitter and Face book. The amount of information to be propagated into this social market is increasing day by day. There are lots of researches to use the online social network as a marketing platform and to handle the influence of information. Influence maximization (IMAX) is one of the NP-hard research problems that focus on finding the seed node that would spread the influence to many users in the social network.

Viral marketing is initiated from the 'word of mouth' communication to promote the product. Influence maximization provides a way to obtain maximum profit from social network users. But influence maximization has been always not effective in viral marketing, as some items are useful to some specific users only. For example, if a men's perfume is to be promoted through the social network, then the specific users are male and female users who wish to purchase it for male users. In such cases, the other users are not considered by the marketer. A better strategy in this case is to maximize the number of specific users, but to identify the specific users from others is the problem in influence maximization. To overcome this, a homogeneous graph of targets is considered.

In order to overcome the problems in influence maximization, a query based approach of influence maximization without predefined labels, namely IMAX query processing is discussed in this paper. In this paper, IMAX query processing represents the social network in the form of a graph where node represents the individual involved in the social network and an edge will represent the relation between any two individuals. The existing methods for influence maximization are not based on query processing. The proposed expectation model is to spread the influence from the seed node to other uninfluenced nodes based on independent maximum influence path. With this expectation model efficient processing of the IMAX query processing is done. This paper has made the following contributions: i) the limitations of existing researches are identified. This paper proves that the influence maximization is an NP-hard and the IMAX query problem is sub modular, ii) a new expectation model to efficiently spread influence of seed is proposed, iii) to process the IMAX query, a greedy based method of approximation is introduced.

II. SYSTEM ARCHITECTURE

The user details are maintained in the user account information. All the details of the online users can be accessed by the administrator also based on the product to be promoted.

A. Influence Maximization problem:

In this paper, a directed graph G represents the social network. G= (V,E) where V is a set of nodes that represents users and E is the set of directed edges which represents the relationship between users. For every edge (x,y) ∈ E, has a weight denoted as p(x,y).

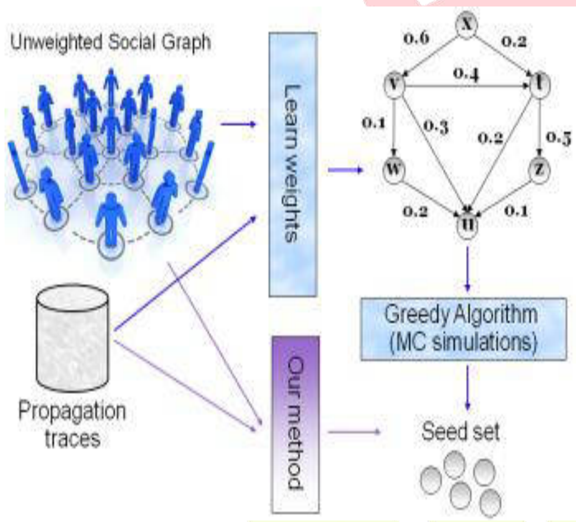


Fig.1 System Architecture

A. Independent cascade (IC) model:

Nodes in the graph may be active or inactive. Influenced nodes as said to be active and nodes not influenced are said to be inactive. At discrete time steps an inactive nodes can change into active node. At time t, every node x that is already active makes neighbor (inactive) node active at time t-1 with probability P_{xy}. If it fails to make nodes active then IC model do not try again. Influence diffusion model: It is assumed that influence is propagated from seed node. Let S ⊆ V be a set of seeds such that every seed(s) in set S is influenced at time 0 initially. Let S_t ⊆ V be a set of nodes influenced at time t by a node in S_{t-1}.

B. Target-aware viral marketing:

In promoting a product in social network, the user's (target) involved exist in three kinds. They are i)the target users interested in the product, ii) non-target users who can influence their friends by introducing the product to them and iii) non-target (immune) users who are not influenced and also not introduce the product to their friends. To effectively process the IMAX query with high accuracy, a novel preprocessed structure is needed to have a space based on a concrete and effective expectation model for influence spread.

C. Independent maximum influence path set (IMIPS):

In a social network each user is linked with many other users. As users are considered as nodes, each node has many paths to reach other nodes. The process of calculating the influence spread in these groups of interrelated nodes with various connecting paths is complicated. To simplify the influence spread among paths, independence between paths is made, to reduce the complication of calculating of influence spread. Paths may share destination and source too. Two paths are said to be independent, if they do not share any other node except the source and the destination. If P1 and P2 are independent paths but they do not share the source. If the source of P1 is a seed, the other nodes along the path P1 can be influenced by the seed but the nodes in the path P2 cannot be influenced by the seed.

D. Influence probability:

The influence probability of a node y ∈ V means that y is influenced by the node in the seed set S and is denoted as p(S, y). For every node $y \in V$, if all the paths are initiated from seed node and have y as destination are said to be independent to each other. By the IC model, the influence probability of y is calculated as

$$p\,(S, y) = 1 - \prod_{P \in \pi(S,y)} (1 - p(P)),$$

Where S is the set of the seed nodes and π(S, y) is the set of all the paths from a seed in S to y.

III. PROPOSED ALGORITHM

A. Independent Maximum Influenced paths based on Expectation model

The difficulty of estimate the influence spread libel in one node that can influence the node throughout the various paths. For each two paths that may share the source and destination and also that the paths are independent. For example, suppose that the two paths R,S are independent and they do not share any of its source.

Let the influence probability p(S,v) is to represent the node v is influenced by a node in the given seed set S.Based on the IC model, the probability of node is computed as

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

$p(S,v) = 1 - \prod_{p \in \pi(S,v)} 1 - p(P)$

Next, our focal point is to identifying the path which is representing the influences between two independent nodes. Based on the Dijkstra's algorithm,$p^1(i,j)$is computed.

A New Expectation Model:

One node can influence the node among one of its paths in IMIP's.consider a circumstance that a node become a fresh seed node in the greedy algorithm. As we explained, the new seed node is achieving maximum influences when starting the PMIA heuristics. However it is expensive to compute the results on processing time.

A.1 Error Analysis:

This IMIP method covers only a stable number of independent paths from a seed node to another node there should be an fault on the IMIP model.Error of IMIP model steadily small in our observations.

A.2 Influence tree:

Let us produce a new efficient way that is handling the concerns of multiple seed nodes.Let us consider a non seed set v≤ V and seed set S≤ v. Fig.1 shows an example of IMIPs and an influence tree. The original graph is shown Fig. 1a and the IMIPs from the seeds (s1, s2, s3) to node v are shown in Fig. 1b. Consider that we look at each IMIP from left to right in Fig. 1b to build the influence tree of v. Then, the last IMIP is < s3; u1; u2; u4; v >. In addition, Fig. 1c shows the situation that we are looking at the last IMIP to build the influence tree of v. To process the last IMIP, we find the common part < u2; u4; v > in Tv. Then, we copy u1; s3; ðs3; u1Þ, and ðu1; u2Þ to the position before u2 of the common part. After processing the last IMIP, the influence tree of v is built as described in Fig. 1d.

B. Computing the influence probability:

Based on the IMIP model, the influence probability of the node v in the seed set is computed. It is denoted by $p_v(s)$ .The copied nodes are consist by a influence tree.we can compute influence probability as recursively by calling influ(v,root(v)) based on the IC model.

ALGORITHM 1. INFLUENTIAL GREEDY ALGORITHM (G =(V,E),K,T)

Input G :An input graph, k size of a seed set,T:a set of targets

Output s :Output seed set

1: **begin**
2: S = ϕ;
3: for i = 1 to k do
4: $s = argmax_{v \in V}\left(\sigma_T(S \cup \{v\}) - \sigma_T(S)\right)$;
5: S = S ∪ $\{s\}$;
6: return S;

C. Query processing using local regions

To process an IMAX query efficiently, first we identify which nodes strongly influence targets, and consider such nodes as candidates for optimal seeds. Then, we approximate the optimal seeds with the candidates.

C.1 Identifying local Influencing regions:

Suppose given IMAX query may large, it is important to considered how to professionally identify candidates, optimal seeds, which is strongly influences the targets in enquiry processing time. Algorithm 2 requires only the process of local influencer influence the node of v in the given seed set.

ALGORITHM 2 INFLU(V,I)

Input *v*; a node in V, *I* : a copied node in T_v

Output the influence probability of *i* when *S* is aseed set under the IMIP model

1: begin
2: if *i is a leaf* then
3: return 1;
4: else
5: p=1;
6: for $n \in IN(i)do$
7: $p = p(1 - p(n - i)influ(v,n))$
8: $p = p - 1$;
9: return p;

C.2 Approximating Optimal Seeds:

Sum of influences of targets is observed from section 4.2.1.Recall the submodular function σ_T^*.The objective function has (1-1/e) approximation ratio based on the IMIP model.

C.3 Preprocessed Structure:

Based on the IMIP model,we can create a novel data structure for calculating the spread of

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

influence. Let us consider p nodes which are the local influencers of node v. We define the local influence tree of v as the influence tree of v when the p local influencers are seeds. In preprocessing time, for each node v, we build the local influence tree of v with information computed by Algorithm 4. As we mentioned in Section 4.1, v's local influence tree can be easily computed by traversing all IMIPs from v's local influencers to v once. The cost of building the local influence trees of all nodes is much smaller than that of Algorithm 4

ALGORITHM 3 STORING LR(G,H,δ)

Input: G = (V,E); an input graph,h:the maximum number of an IMIPs,δ:a parameter in (0,1-$\sqrt[h]{1-a}$)

```
1.begin
  for v ∈ V do
    compute p^1(u,v)s,t.
     u ∈ V ∩ p(p^1(u,v)) > 1-ʰ√(1 − a);
    ℵ(v) = {u|u ∈ V, p(p^1(u,v))>1-ʰ√(1 − a);
                 for u ∈ ℵ(v)do
          insert p^1(u,v) into π^h(u,v);
          for t = 2 to h do
          h' = t;
          p(p^1(u,v) > bound;
         insert u into x(v);
```

IV. EXPERIMENTAL RESULTS

Several experiments were conducted using various comparison methods and real time dataset. IMIP model and Incremental updating mainly focused to test the efficiency of the method. In the current system, the IMIP is proposed. CELF++ is a greedy algorithm which helps in identifying the local influencing provinces. CD model is one of the greedy method which is probabilistic model that identifies the historical action done by user.

Table 1 Statistics of Our Data Sets

Data Set	Wiki - Vote	Epinions	Slashdot	Amazon	Pokec
Node	7K	76K	77K	262K	1.6M
Edge	104K	509K	906K	1.2M	30.6K
Degree	14.6	6.7	11.7	4.7	18.8
Data Set	**Gowalla**	**Diggs**	**Flixster**		
Node	197K	279K	0.8K		
Edge	1.9K	1.7K	11.8K		
Degree	9.7K	6.2	15		
Action	6.4M	3M	8.2K		

For consideration ,Eight real datasets such as Flixster,Wikivote,Digg,Amazon,Slashdot,Pokec,Epinions,Gowalla Elections are included based on Wiki-votes . It promotes the adminship and readily available edge as of u to v while user u vote on user v. Slashdot is machinery related news site where any relationships between users. Epinions is who- faith – whom social network. Amazon is one of the co-purchasing network, where readily available edge as of u to v while both edges are purchased frequently.Gowalla is locality-based site where the users can carve up their locations to their friends.Movie reviews Flixster is one of the social network where users share their movie comments with friends.Digg is online information network and Pokec is the Slovakia network service.The dataset of Pokec contains real profile data.These profile data is to specifying the factual targets. Table 2 defines the dataset statistics.In this table the average degree of nodes is denoted by Degree and the number of deed logs denoted by Action. Based on the model, the Flixster result is shown along with three datasets due to space limitation. The Digg and Gowalla results are as similar tendency to Flixster results.

A. Generating queries:

For experiments, we create a syntactic query through three parameters.First,to select the nodes randomly as a element of whole targets to be chosen. Randomly selected nodes are denoted by p1

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

Table 2 The Sensitivity Tests of the Parameters

α(×0.001)	5	16.25	27.5	38.75	50
IS	1211.77	1202.65	1197.94	1195.98	1193.02
R(s)	0.1278	0.003385	0.0213	0.016	0.01375
β	**0**	**1**	**1.5**	**2.0**	**2.5**
IS	1211.77	1211.77	1211.77	1211.6	1211.68
R(s)	0.1646	0.1278	0.1139	0.105	0.09915
h	**1**	**2**	**3**	**4**	**5**
IS	1210.47	1211.91	1211.7	1211.68	1211.77
R(s)	0.107	0.12245	0.12635	0.1278	0.1278
δ(×0.0001)	**5**	**30**	**55**	**80**	**105**
IS	1211.77	1211.12	1210.6	1208.32	1206.44
R(s)	0.1278	0.1107	0.099	0.06465	0.046

Then, for residual part. We can select the nodes consistently through breadth-first search method.

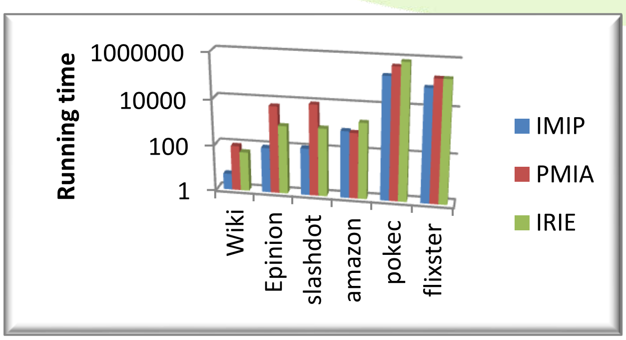
Table 3 Influence Spread and Running Time with Real Targets in Pokec, WC, k ¼ 50

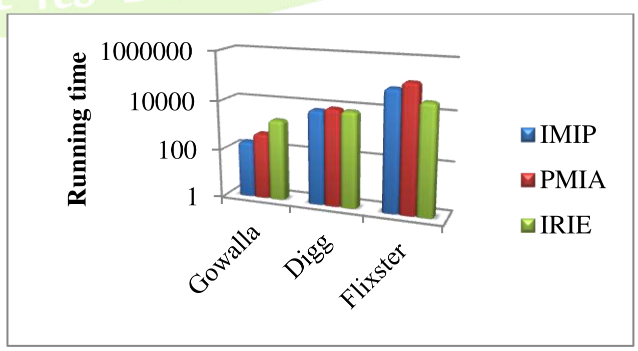
IS	Men	Women	Adults	Non-Adults
IMIP	1968.1	22539.6	25411.9	17620.5
PMIA	20267.1	22931.7	25511.3	17782.9
IRIE	19580.9	22168.7	25418.3	17321.7
IS	**Men**	**Women**	**Adults**	**Non-Adults**
IMIP	8.652	9.55	10.152	7.805
PMIA	474.69	539.636	613.081	466.488
IRIE	279.692	283.592	285.372	272.611

In that method, we can choose the target node with probability p2.we can select another nodes among probability p3.we repeat this process until the residual part of whole targets is done. The connected targets are controlled by p1 and p3.In that connected graph, the exits targets are represented by p2.For example if p3=1 and p1=1, all targets choosing randomly.

we perform the sensitivity test for IMIP based on following parameters;$\alpha, \beta, \gamma, \delta$.Given $\beta = 1.0$ $\alpha = 0.005$,$\gamma = 5, \delta = 0.005$.

The Result shows if α gets larger then the running time and influence spread gets shorter, respectively. The sensitivity test result for β is bigger then the influence spread is more or less not changed and running time gets smaller and shorter. we compare the proposed method with unlike maximum count of IMIP's.The influence spread is somewhat low when γ=1.The influence spread is slightly high when $\gamma >$ 1.The parameter δ which is used to let alone computing IMIP's whose authority is to boot small.





ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 15, March 2016

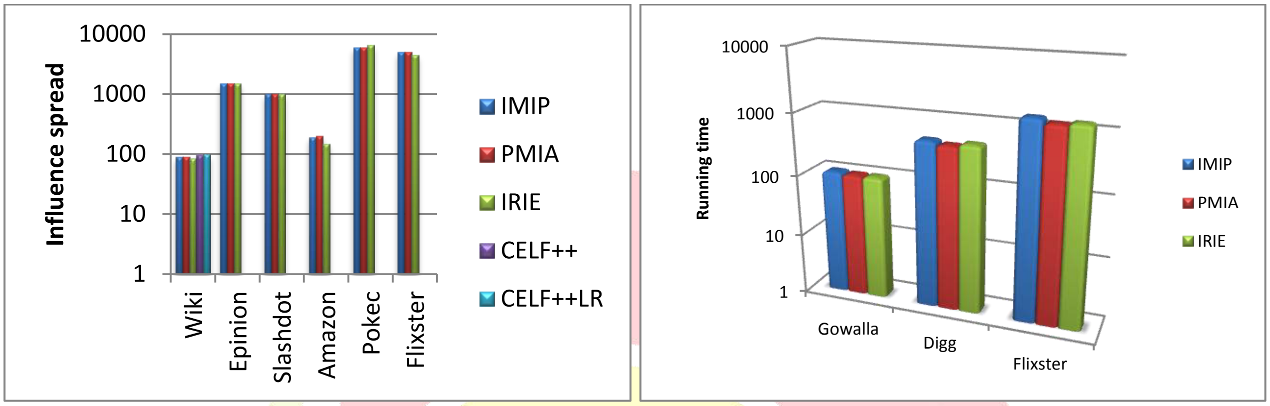


Fig.2 Influence spread analysis (k ¼ 50)

V. CONCLUSION

In the proposed work, we create IMAX query processing to improve the authority on specific users in online social networks. We intend the IMIP model for estimating the value of objective function. Our identifying neighborhood regions method is so effective and also our proposed method is enormity faster than the existing methods. In future, a mixture of targets such as users in same university or same community based on inert profiles of users for IMAX query processing.

VI. REFERENCES

1. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003, pp. 137–146.
2. F.-H. Li, C.-T. Li, and M.-K. Shan, "Labeled influence maximization in social networks for target marketing," in Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust, Int. Conf. Social Comput., 2011, pp. 560–563.
3. W. Lu and L. Lakshmanan, "Profit maximization over social networks," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 479–488.
4. P. Domingos and M. Richardson, "Mining the network value of customers," in Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2001, pp. 57–66.
5. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2007, pp. 420–429.
6. A.Goyal, W. Lu, and L. V. Lakshmanan, "CELF++: Optimizing the greedy algorithm for influence maximization in social networks," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 47–48.
7. Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1039–1048.
8. W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2009, pp. 199–208.
9. W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038.