The GRIM: Target Marketing in Social Network

Ashis Talukder, Md. Golam Rabiul Alam, Anupam Kumar Bairagi, Sarder Fakrul Abedin,
Hoang T. Nguyen and Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, South Korea
Email: {ashis, robi, anupam, saab0015, nguyenth, cshong}@khu.ac.kr

***Abstract*—Influence maximization (IM) is a popular research area for viral marketing in social network. In this research we propose a novel problem, Reverse Influence Maximization (RIM) problem for target marketing along with a greedy solution GRIM model under Linear Threshold (LT) model. The GRIM model determines the opportunity cost returned by minimum number of nodes that must be activated in order to motivate a target seed set. We also perform simulation to evaluate the performance of the algorithm using two real world datasets.**

I. Introduction

The Influence Maximization (IM) for viral marketing has gained remarkable interest in social network research. The IM calculates a fixed size seed set that can activate the maximum number of nodes in the network [1]. Here we propose a novel problem for target marketing, named *Reverse Influence Maximization (RIM)* problem. It estimates the opportunity cost [2] which is the minimum number of nodes that must be activated in order to motivate a given seed set. These targeted seed members are suppose to be prominent entity. For example, Usain Bolt's Facebook photo using a Sumsung cellphone can influence his millions of fans to buy the same phone. Thus the GRIM model also have great business value like the IM.

In this research we prove that the RIM problem is NP-Hard and propose a Knapsack based greedy solution under classical LT model [1]. We also evaluate the performance of the GRIM model with two real social network datasets.

The rest of the paper provides literature review, problem formulation, GRIM model, performance evaluation and conclusion in the consecutive sections.

II. Literature Review

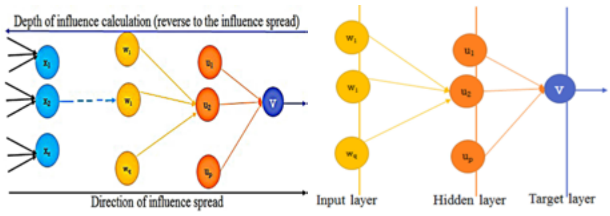
To the best of our knowledge the IM problem was first introduced by [3] in 2001. Then Kempe et al. in [1] give a great shape with two classical models such as Linear threshold (LT) and Independent cascade (IC) models providing $(1-\frac{1}{e})$ performance approximation.

In [4], the authors have proposed a heuristic approximation for outbreak detection using IM. Goyal et al.[5] extend it which has $35-55\%$ faster running time. Chen et a. introduce a degree discount (DC) heuristic in [6]. It improves the accuracy of [1] and the running time of [4] simultaneously. Most recently, the authors in [7] has given a running time improvement by an innovative idea of stop and stare sampling.

But none of the above has addressed the problem of finding the opportunity cost for target marketing that we propose.

III. Problem Formulation

Assume a social network is given by a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where each vertex is a users and ech edge is a social relationship. A user u influences v with weight w_{uv}. The node v is activated if the influence coming from all the active in-neighbors is no less than a given threshold θ_v that is $\sum_{u\in n^{-1}(v)} w_{uv} * x_u \geq \theta_v$ [1]. Here x_u indicates whether a node u is active ($x_u = 1$) as input layer node or not ($x_u = 0$) and same definition holds for h_u and x_u for hidden and target layer nodes respectively. For a given seed set $\mathcal{S}$ of $\mathcal{K}$ influential customers, the RIM aims at finding opportunity cost set denoted by $\Gamma(\mathcal{S})$ and the opportunity cost $\sigma(\mathcal{S}) = |\Gamma(\mathcal{S})|$.



(a) Influence in reverse order (b) Reduced 3-layer network

Figure 1: The RIM reduce the calculation level by a neural network analogy.

The RIM decomposes the network into $\mathcal{K}$ neural networks containing v as the only one output layer node as depicted in the Fig. 1. The marginal opportunity cost $\sigma(v)$ for each target node v is computed by:

$$\sigma(v): \quad \text{minimize} \quad \sum_{u\in n^{-1}(v)} x_u \tag{1}$$

subject to-,

$$\sum_{u\in n^{-1}(v)} w_{uv} * x_u \geq \theta_v, \tag{2}$$

$$x_u \in \{0,1\}, w_u \in (0,1] \tag{3}$$

Then the final opportunity cost and is given by:

$$\sigma(S) = \bigcup_{v\in S} \sigma(v) \tag{4}$$

IV. Solution Framework of RIM

We propose the GRIM model which is Knapsack based greedy solution of the RIM problem.

A. Meeting the Challenges

We set the *limit of calculation* with a simple 3-layer neural network analogy as shown in the Fig. 1. The decomposition results in *three cases*. The target node v has zero (A), one (B) or multiple layer (C) of in-neighbors. The Case A is trivial and we set $\sigma(v) = |\Gamma(v)| = |\{v\}| = 1$. The Case B is the basic unit of calculation and the Case C is the combination of multiple A and B cases. We set θ_v to some smaller value to avoid *insufficient influence*. A node u may be activated as input layer node of a target node v_1 and also activated as hidden layer and/or a target layer node of another target node v_2. Then the node u will not be added in the opportunity cost set and this optimization is called *commonality discount*.

Theorem 1. *The RIM problem is NP-Hard.*

Proof. The RIM can be reduced to Knapsack problem by considering node's threshold as Knapsack size, influence weights as item weights and substituting the objective function of the RIM problem by a maximize format of the negation of the equation (1). □

Algorithm 1: GRIM Model

Input: $\mathcal{G}(\mathcal{V}, \mathcal{E}), \mathcal{S}$
Result: $\sigma(\mathcal{S}), \Gamma(\mathcal{S})$
$\Gamma(\mathcal{S}) = \emptyset$; /* Hidden to output layer */
for *each* $v \in \mathcal{S}$ **do**
  $\Gamma(S) = \Gamma(S) \cup \Gamma(v)$ by equation (1) to (3);
end
$\mathcal{S}_1 = \Gamma(\mathcal{S})$, $\Gamma(\mathcal{S}) = \emptyset$; /* Input to hidden layer */
for *each* $v \in \mathcal{S}_1$ **do**
  $\Gamma(\mathcal{S}_1) = \Gamma(\mathcal{S}_1) \cup \Gamma(v)$ by equation (1) to (3);
end
$\Gamma(\mathcal{S}) = \Gamma(\mathcal{S}_1)$; /* Commonality discount */
for *each* $v \in \Gamma(S)$ **do**
  $\Gamma(\mathcal{S}) = \Gamma(\mathcal{S}) - \{v | h_v = 1 \ OR \ t_v = 1\}$;
end
$return \ \sigma(\mathcal{S}) = |\Gamma(\mathcal{S})|$;

B. The GRIM algorithm

The GRIM algorithm is stated in the Alg. 1. The complexity of the algorithm is given in equation (5) where $\mathcal{D}$ = maximum number of neighbors (degree) in the network.

$$\mathcal{T} = O(\mathcal{K}.\mathcal{D}^2) \tag{5}$$

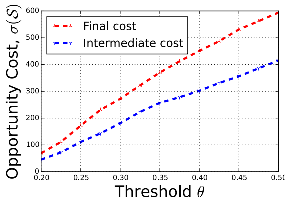
V. Performance Evaluation

We evaluate the performance of GRIM model for two real datasets.They are scaled and stated in the Table I.

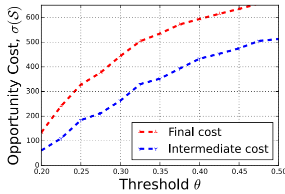
Table I: Scaled dataset description

Networks	ego-Facebook [8]	Epinions [9]
Nodes	1000	1000
Edges	13,969	40,240

The GRIM model exhibits the expected trend of cost increasing with the increase of threshold when the seed size is constant as portrayed in the Fig. 2 for both the datasets. It also shows good running time in the Fig 3 since it involves fixed number of layers and hence lesser nodes and edges.

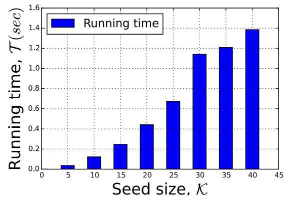


(a) Facebook data.

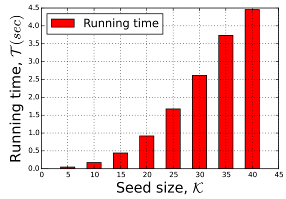


(b) Epinions data.

Figure 2: The opportunity cost for different threshold values.



(a) Facebook data.



(b) Epinions data.

Figure 3: The running time $\mathcal{T}$, (in sec) for different seed sizes.

VI. Conclusion

In this research we propose the Reverse Influence Maximization (RIM) problem along with greedy solution GRIM model to estimate the opportunity cost for target marketing in social network. The GRIM model show good running time (order of $\mathcal{K}.\mathcal{D}^2$) with reasonable opportunity cost.

The main limitation of this research is that the GRIM model can not always provide optimal solution for its greedy nature but yet provides a feasible solution.

References

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[2] B. Ryan, *Strategic accounting for management*. Cengage Learning EMEA, 1995.

[3] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001, pp. 57–66.

[4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 420–429.

[5] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 47–48.

[6] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

[7] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks," *arXiv preprint arXiv:1605.07990*, 2016.

[8] J. J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks." in *NIPS*, vol. 2012, 2012, pp. 548–56.

[9] M. Richardson, R. Agrawal, and P. Domingos, "Trust management for the semantic web," in *International semantic Web conference*. Springer, 2003, pp. 351–368.