A Cost Optimized Reverse Influence Maximization in Social Networks

Ashis Talukder, Md. Golam Rabiul Alam, Nguyen H. Tran, and Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, South Korea
Email: {ashis, robi, nguyenth, cshong}@khu.ac.kr

***Abstract*—In recent years, Influence Maximization (IM) has gained great research interest in the field of social network research. The IM is a viral marketing based approach to find the influential users on the social networks. It determines a small seed set that can activate a maximum number of nodes in the network under some diffusion models such as Linear Threshold model or Independent Cascade model. However, previous works have not focused on the opportunity cost defined by the minimum number of nodes that must be motivated in order to activate the initial seed nodes. In this work, we have introduced a Reverse Influence Maximization (RIM) problem to estimate the opportunity cost. The RIM, working in opposite manner to IM, calculates the opportunity cost for viral marketing in the social networks. We have proposed the Extended Randomized Linear Threshold RIM (ERLT-RIM) model to solve the RIM problem. The ERLT-RIM is a Linear Threshold (LT)-based model which is an extension to the existing RLT-RIM model. We also have evaluated the performance of the algorithm using three real-world datasets. The result shows that the proposed model determines the optimal opportunity cost with time efficiency as compared to existing models.**

***Index Terms*—reverse influence maximization, opportunity cost, RIM, viral marketing, influence maximization, linear threshold model, social network.**

I. Introduction

In this day and age, social networks have become the ideal platform for disseminating and exchanging information, ideas, new reports, trends etc. due to the rapid growth of the number of social sites and their usage. Information origination and diffusion in the social networks are growing sharply every day and hence, social networks are becoming the most attractive medium for marketing and research field as well [1]. Social network research has been conducted in many directions and viral marketing based Influence Maximization (IM) is a prominent direction among them.

In the last decade, the IM has become attractive social network research field. The IM problem estimates influential users in the social network such that the spread of influence is maximized. In other words, it selects a seed set such that the total number of activated nodes is maximized assuming that the seed nodes are initially activated [2]. In the activation process, all the activated nodes try to activate their out-neighbors, that is, motivated users try to influence their friends or followers to make some specific decision. The process is also called viral marketing where influence spreads in the *word-of-mouth* effect [3]. It mirrors the human behavior in real-life scenarios that people always consult with the family members, friends, colleagues, or other experts before taking any decision ($viz.$ any purchase decision) [4]. The enormous applications of the IM problem include finding community leaders [5], profit maximization or maximizing product adaptation [6], [7], searching experts in some fields [8], rumor spread and detection [9], [10], e-commerce and media industry [11], contamination and outbreak detection [12], online recommendation [13] etc.

Most of the researchers in this field have conducted research to identify the influential seed set subject to maximize the influence [3], [14], [15] or maximize the profit [6], [16] or maximize product adaptation [7]. But estimating the minimum cost of influence maximization has not been addressed deeply. Many authors have described the cost of motivating the seed users of influence maximization problem in a trivial way. They have just offered free sample products or free tickets of a concert to the seed users [3], [6], [7]. The approach is not rational since those influential users are human and they also might be motivated by some other icon people they follow. Some authors have tried to find influential users with a cost budget and they considered the cost of all activated users including seeds. Zhu *et al.* [15] have balanced the influence and the profit but did not consider the cost in their profit maximization research. But none of the studies has addressed the cost of activating those influential users. Unlike the previous studies, we have tried to compute the minimum opportunity cost to activate the influential target users.

In this research, we have formulated a Reverse Influence Maximization (RIM) problem to calculate the opportunity cost of viral marketing in the social network. The opportunity cost [17] is identified by the minimum number of nodes that we need to activate in order to motivate the given set of target nodes in a social network. In IM problem, a small seed set of users that maximizes the spread of influence is determined [2]. On the other hand, RIM finds the minimum number of in-neighbors that are necessary to activate a given set of target nodes. Thus, the RIM works in the reverse fashion as compared to IM problem as illustrated in the Fig. 1. This is the logic of the naming of RIM. In this work, we propose the Extended Randomized Linear Threshold RIM (ERLT-RIM) model which is a variation of the classical Linear Threshold (LT) [2] model and an extension to the existing RLT-RIM model [18]. The ERLT-RIM considers already activated nodes

978-1-5386-3416-5/18/$31.00 © 2018 IEEE

and commonality discount which are not previously addressed. The performance evaluation of the proposed model has been accomplished employing the datasets of three popular real-life social networks: Facebook and Twitter and Epinions. The result shows that the proposed model has faster running time with better opportunity cost as compared to that of previous models.

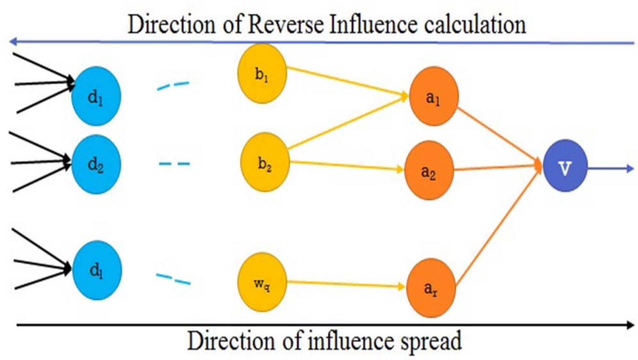


Fig. 1. Basic working principle of RIM compared to IM

The rest of the paper is organized in the following order. In section II, we have reviewed the state-of-the-art of influence maximization research. The RIM problem formulation has been provided in section III. In section IV, the solution models of the RIM problem along with the proposed model have been stated elaborately. The performance evaluation of the proposed ERLT-RIM model has been presented in the section V. Finally, the concluding remarks and future scope have been provided in section VI.

II. Existing Study

Influence maximization (IM) technique estimates the influential users in the social networks using viral marketing approach. Many research works have been conducted in this direction in past years and yet it is a potential research trend. The problem has been well-formulated by Kempe *et al.* [2] in 2003. They have proved that the IM problem is NP-Hard and proposed two classical models named *Linear Threshold (LT)* and *Independent Cascade (IC)* models wihich ensure approximation ratio of $(1 - \frac{1}{e})$.

Leskovec *et al.* [12] have proposed a heuristic approximation model for outbreak detection using IM. Their Cost-Effective lazy Forward (CELF) model shows better result than the standard greedy algorithms. The authors in [19] have performed influence maximization in the trust network with considering both positive and negative influence of trusted and dis-trusted neighbors respectively. Their proposed LT based algorithm outperforms the popular CELF model by about 35%. Chen *et al.* [20] have also worked with negative influence and their Maximum Influence in Arborescences with Negative opinion(MIA-N) model has faster running time with the same level of approximation as a greedy model. Goyal *et al.* [21] have formulated a simple path-based algorithm that has shown the better result than many existing models in the scale of running time, memory utilization, and seed quality. A heuristic approach, degree centrality, has been introduced in [14] and the model has enhanced the accuracy of classical models [2] and the running time of CELF model simultaneously.

Chen *et al.* [14] have proposed a linear time model named Local Directed Acyclic Graph (LDAG) model which is scalable to extremely large networks. It has exhibited better approximation ratio as compared to greedy models. A Greedy Algorithm based on Users' Preferences (GAUP) has been proposed in [8]. It finds user preferences by employing Singular Value Decomposition (SVD) first, and then, mines the top-k influential users by the greedy approach. An optimized and linear time algorithm, called Non-backtracking Random Walk (NBRW) has been introduced by Pan *et al.* [22]. The NBRW algorithm calculates non-backtracking walk in network communities first and then, estimates node's influence by its traversing number. The algorithm has better performance than many existing models, especially the LT model.

Tong *et al.* [23] have studied influence maximization in dynamic social network taking uncertainty into account. They have formulated the Dynamic Independent Cascade (DIC) model with adaptive and Heuristic greedy algorithms (A-Greedy and H-Greedy) with $(1 - \frac{1}{e})$ performance bound.

Zhang *et al.* [24] have formulated a novel heuristic model to find influential users (termed as a Key Opinion Leader, KOL) in messenger based online social networks. The scenario is challenging since the number of friends in messenger based apps ($e.g.$ WeChat) is very small. Their proposed model has shown lower running time.

Some researchers have considered influence maximization as profit maximization in their research. The authors in [6] have presented such works. They have considered multiple products rather than a single product in the profit maximization. Generally, previous works have involved only single product but the fact is that any company does not produce only one product but multiple products. Zhu *et al.* [15] have traced a difference between influence and profit in profit maximization research. They have proposed the Balanced Influence and Profit (BIP) method to tackle the scenario that profit and influence cannot be maximized simultaneously. Nguyen *et al.* [16] have integrated multiple social networks in their work with a fact that multiple networks can support each other to propagate influence among the networks. The authors in [7] have maximized product adaptation and their Linear Threshold model with Color (LT-C) achieves standard approximation margin of $(1 - \frac{1}{e})$.

However, none of the above researches have addressed the problem of determining the opportunity cost. To find the opportunity cost, the RIM problem was first introduced by Talukder *et al.* [18] and they proposed Random-RIM (R-RIM) and Randomized Linear Threshold RIM (RLT-RIM) models to solve the RIM problem as well. However, none of the existing R-RIM and RLT-RIM models considers the already activated nodes and commonality discount that can decrease opportunity cost.

Thus, we propose ERLT-RIM model which includes the above features to have better performance in cost calculation.

TABLE I
LIST OF PARAMETERS

Symbol	Meaning
$G(V,E)$	Social Network.
V	Set of social Network users.
E	Social relationship among users.
$n(v)$	Out-neighbor set of v.
$n^{-1}(v)$	In-neighbor set of v.
S	The target set.
k	Size of the target set S.
$\Gamma(v)$	Opportunity cost set of node v defined by a set of minimum number of nodes need to be activated in order to activate v.
$\Gamma(S)$	Opportunity cost set of all the nodes of the set S.
$\sigma(v)$	Opportunity cost of node v defined by minimum number of nodes need to be activated in order to activate v, $\sigma(v) = \|\Gamma(v)\|$.
$\sigma(S)$	Combined opportunity cost of all the nodes of set S, $\sigma(S) = \|\Gamma(S)\|$.
x_u	Whether a second hop node is activated or not.
y_u	Whether a first hop node is activated or not.
t_u	Whether a target node is activated or not.
θ_v	Activation threshold of node v,
w_{uv}	Influence probability/weight of node u to v.
C_b, C_w, C_a	Best, worst and average case complexity.

III. PROBLEM FORMULATION

In this section, we define the RIM problem along with mathematical formulation and detailed system model. We consider a social network $G(V,E)$, where each vertex represents a social network user and each edge a social relationship between two such users. The influence weight w_{uv} is the influence weight of user u to v. We define $n(v)$ and $n^{-1}(v)$ as the out-neighbor and in-neighbor set of v respectively. Each node v in the network is assigned an activation threshold θ_v. As stated in Linear Threshold (LT) model, the node v is activated if the influence coming from all the active in-neighbors is no less than a given threshold θ_v, *i.e.* if, $\sum_{u \in n^{-1}(v)} w_{uv} x_u \geq \theta_v$ [2]. Here, x_u indicates whether a node u is activated ($x_u = 1$) as second hop in-neighbor node of any target node or not ($x_u = 0$) and the same definition holds for y_u and t_u for first hop in-neighbors and target node respectively. For a given target set S of k influential users, the RIM aims at finding the opportunity cost set and the opportunity cost, denoted by $\Gamma(S)$ and $\sigma(S)$ respectively. The summary of all the parameters is stated the Table I.

To solve the RIM problem, we decompose the network into k *Basic Network Components (BNC)* [18]. Every BNC contains a target node v as the only one node in the target layer and v has either zero, one or two hops of in-neighbors and are named as BNC-A, BNC-B and BNC-C respectively. Here, $a_i \in n^{-1}(v)$, are the first hop in-neighbors and b_i are the second hop in-neighbors of v as illustrated in the Fig. 2.

We start the RIM problem formulation with calculating the marginal opportunity cost set of each target node $v \in S$ (*i.e.* for each BNC). It is denoted by $\Gamma(v)$ and given by [18]:

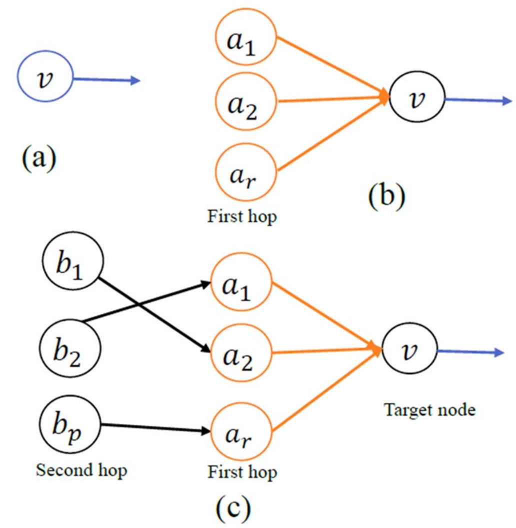


Fig. 2. Basic network components of RIM

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

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

$$x_u \in \{0,1\}, \; w_{uv} \in (0,1] \tag{3}$$

Then the opportunity cost set $\Gamma(S)$ is constituted by combining all marginal cost sets of all nodes $v \in S$ and is given by:

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

Finally, the opportunity cost $\sigma(S)$ is given by:

$$\sigma(S) = |\Gamma(S)| \tag{5}$$

The equations from (1) to (5) give the detailed mathematical formulation of the RIM problem and the formal definition of the RIM problem is stated below.

Definition 1. RIM Problem: Given a social network $G(V,E)$ and a target set S of size k, the RIM problem estimates the opportunity cost $\sigma(S)$, which is defined by the minimum number of nodes that must be activated in order to activate all the target nodes in S. □

IV. SOLUTION FRAMEWORKS OF RIM

In this section, we first discuss the challenges of RIM problem and techniques to resolve these challenging issues. Then, we propose the ERLT-RIM model along with detailed discussion of the existing R-RIM and RLT-RIM models. We investigate the complexity analysis of these algorithms as well.

Algorithm 1: R-RIM Model

Input: $G(V, E), S$
Result: $\Gamma(v)$
$TOC = \emptyset$ $MOC = \emptyset$;
for $each\ v\ \in\ S$ **do**
  $MOC = n^{-1}(v)$; /* first hop neighbors */
  for $each\ u\ \in\ S$ **do**
    $MOC = MOC \cup n^{-1}(u)$; /* 2nd hop ... */
  end
  $TOC = TOC \cup MOC$; /* Total opportunity cost set */
end
$m = Select\ a\ number\ between(1, |TOC|)\ randomly$;
$\Gamma(S) = Select\ m\ nodes\ from\ TOC\ randomly$;
$return\ \Gamma(S)$;

Algorithm 2: MarginalCost

Input: $G(V, E), S,\ v,\ p,\ q$
Result: $\Gamma(v)$
$Calculate\ the\ set\ n^{-1}(v)$;
$active = 0$;
if $n^{-1}(v) = \emptyset$ **then**
  $t_v = 1$;
  $return\ v$; /* checking of BNC-A */
end
$influence = 0,\ inn\_set = n^{-1}(v),\ \Gamma(v) = \emptyset$;
while $inn\_set \neq \emptyset$ **do**
  if $influence \geq \theta_v$ **then**
    $active = 1,\ q_u = 1$;
    $break$;
  end
  $u = Select\ a\ node\ from\ 'inn\_set'\ randomly$;
  $p_u = 1$;
  $influence = influence + w_{uv}$;
  $\Gamma(v) = \Gamma(v) \cup u;\ inn\_set = inn\_set - u$;
  /* Selects in-neighbors according to LT Model */
end
$return\ \Gamma(v)$;

A. Challenges

The first challenge is related to the stopping criteria of the cost computation. We must settle on how many numbers of predecessor hops up to which the cost calculation will be continued. Generally, single hop is considered in influence maximization process but it may have inaccuracy and lower chance to influence the target nodes. On the other hand, multiple hops would incur tremendous estimation complexity. Thus, we propose a generalized model of a maximum 2-hop model for the RIM problem. For instance, the first hop in-neighbors a_i and the second hop in-neighbors b_i are considered to activate the target node v as stated in the Fig. 2.

The second challenge is to handle three BNCs. The decomposition process involved in the RIM solution model results in three *Basic Network Components (BNC)*: the target node v, having zero hops (A), one hop (B) and two hops (C) of in-neighbors. The case BNC-A is trivial and we are just compelled to offer free sample product to the target node v like [7]. Since v has no in-neighbors, we just set $\sigma(v) = |\Gamma(v)| = |\{v\}| = 1$. The case BNC-B is the basic unit of calculation and BNC-C is a combination of multiple instances of BNC-A and BNC-B. Thus, it is enough to design only BNC-A and BNC-B to solve the RIM problem.

The third challenge is to handle the insufficient influence, as stated in the Fig. 3, which happens when all the in-neighbors have not enough aggregated influence to activate the target node v, that is, if

$$\sum_{u \in n^{-1}(v)} w_{uv} * x_u < \theta_v. \tag{6}$$

It totally depends on the inherent network structure and threshold values of the nodes. The inherent network structure includes how many numbers of friends a user might have, how many in-neighbors and out-neighbors a node might have and hence it is intractable. Thus, we just set thresholds to some relatively smaller values to avoid the major effect of insufficient influence. The final challenge is the NP-Hard nature of the RIM problem.

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

Proof. The Knapsack problem defined in the following equations (7) to (9) can be reduced to the RIM problem.

$$\text{maximize} \sum_{u=1}^{N} x_u p_u \tag{7}$$

$$s.t. \quad \sum_{u=1}^{N} x_u w_u \leq M, \tag{8}$$

$$x_u \in \{0, 1\} \tag{9}$$

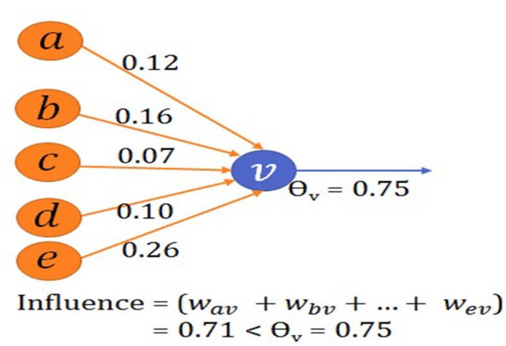


Fig. 3. Insufficient influence: node v is not activated since the aggregated influence, 0.71, of all its in-neighbors, $n^{-1}(v) = \{a, b, c, d, e\}$, is less than the node threshold, $\theta_v = 0.75$.

Algorithm 3: RLT-RIM Model

Input: $G(V, E), S$
Result: $\sigma(S), \Gamma(S)$
1 $\Gamma(S) = \emptyset$; /* First hop neighbors */
2 **for** *each* $v \in S$ **do**
3 $\quad \Gamma(S) = \Gamma(S) \cup MarginalCost(G, S, v, y, t)$ by equation (1) to (3); /* Equation (4) */
4 **end**
5 $S_1 = \Gamma(S)$; /* Second hop neighbors */
6 **for** *each* $v \in S_1$ **do**
7 $\quad \Gamma(S) = \Gamma(S) \cup MarginalCost(G, S_1, v, x, y)$ by equation (1) to (3); /* Equation (4) */
8 **end**
9 *return* $\sigma(S) = |\Gamma(S)|$; /* Equation (5) */

Here w_u and p_u are the weights and associated profits of N items respectively. Now let us compare the parameters of two models:

- let us consider the influence weights w_{uv} in RIM as the item weights w_u in the knapsack problem,
- let us substitute the objective function of the RIM problem by $maximize \ -\sum_{u \in n^{-1}(v)} x_u$ in the equation (1).
- let us consider node's threshold θ_v in RIM as the knapsack size M. As long as the aggregated influence does not reach to the node threshold, a new in-neighbor is chosen in RIM. Similarly, as long as the bag has enough free space, a new item is picked up in Knapsack problem.
- the constraint in the equation (2) of RIM problem now can be considered as the counterpart equation (7) of the Knapsack problem.

The cost minimization in RIM problem is now equivalent to the profit maximization in Knapsack problem and hence, the Knapsack problem is reduced to the RIM problem which is NP-Hard [25]. Thus, the RIM is NP-Hard as well. $\square$

B. R-RIM and RLT-RIM Models

Here, we discuss the R-RIM and RLT-RIM [18] with detailed working philosophy and complexity analysis for the comparison purpose.

1) The R-RIM algorithm: The Random RIM (R-RIM) estimates the Marginal Opportunity Cost (MOC) set by taking first hop in-neighbors of all k target nodes as stated in the line 3 of the Alg. 1. Then, the Total Opportunity Cost (TOC) is calculated by considering all the members of MOC along with the second hop in-neighbors of all $v \in S$ as stated in lines 4 to 7. Finally, it picks up a random number of nodes from *TOC* as the opportunity cost set, $\Gamma(S)$ and the final opportunity cost as $\sigma(S) = |\Gamma(S)|$ by lines 9 and 10.

The complexity of the algorithm is defined in terms of the number of in-neighbor nodes required to be processed. The complexity of the R-RIM model is given by $O(kd^2)$, where d is the maximum number of in-degrees in the network.

Algorithm 4: ExtendedMarginalCost

Input: $G(V, E), S, v, p, q$
Result: $\Gamma(v)$
1 *Calculate the set* $n^{-1}(v)$;
2 **if** $n^{-1}(v)$ = $\emptyset$ **then**
3 $\quad t_v = 1$;
4 $\quad$ *return* v ; /* Case A: No incoming node */
5 **end**
6 $\Gamma(v) = \emptyset$;
7 *Calculate* $n(u)$ *for all* $u \in \ n^{-1}(v)$;
8 **for** *each* $u \ \in \ n^{-1}(v)$ **do**
9 $\quad$ *Calculate* $|n(u)|$;
10 $\quad w_{uv} = \frac{1}{|n(u)|}$; /* Influence weight */
11 **end**
12 $alreadyset = \Gamma(S) \cap n^{-1}(v)$;
13 $newinn = n^{-1}(v)$;
14 $inf\_sum = 0.0$;
15 **while** $alredayset \ \neq \emptyset$ **do**
16 $\quad$ **if** $inf\_sum \geq \theta_v$ **then**
17 $\quad\quad active = 1$;
18 $\quad\quad break$;
19 $\quad$ **end**
20 $\quad$ *Select* $z \in alreadyset$ *randomly*;
21 $\quad p_z = 1$;
22 $\quad inf\_sum = inf\_sum + w_{zv}$;
23 $\quad newinn = newinn - z$;
24 $\quad already = alreadyset - z$; /* include already-activated node */
25 **end**
26 **while** $newinn \neq \emptyset$ **do**
27 $\quad$ **if** $inf\_sum \geq \theta_v$ **then**
28 $\quad\quad active = 1$;
29 $\quad\quad break$;
30 $\quad$ **end**
31 $\quad$ *Select* $z \in alreadyset$ *randomly*;
32 $\quad p_u = 1$;
33 $\quad newinn = newinn - u$;
34 $\quad inf\_sum = inf\_sum + w_{zv}$;
35 $\quad \Gamma(v) = \Gamma(v) + u$; // Include new in-neighbors
36 **end**
37 *return* $\Gamma(v)$;

2) The RLT-RIM algorithm: The RLT-RIM algorithm is stated in the Alg. 3 which iteratively calls Alg. 2 to calculate the marginal cost of each target node. The RLT-RIM is a variation of the Linear Threshold (LT) model. It randomly selects an in-neighbor u of a target node v and aggregates its influence wight w_{uv} to compare whether the combined influence of selected in-neighbors have reached the node v's threshold θ_v or not. If it reaches the threshold then the node v is activated and all the selected in neighbors are included in marginal cost set $\Gamma(v)$ as mentioned in the lines 2 to 4 of the Alg. 3. Finally, all the marginal cost sets are then merged together to find the opportunity cost set $\Gamma(S)$ and the final

Algorithm 5: ERLT-RIM Model

Input: $G(V, E), S$
Result: $\sigma(S), \Gamma(S)$

1 $x_u = 0$;
2 $h_u = 0$;
3 $t_u = 0$;
4 $\Gamma(S) = \emptyset$;
5 **for** *each* $u \in S$ **do**
6 $\quad t_u = 1$;
7 **end**
8 **for** *each* $v \in \mathcal{S}$ **do**
9 $\quad \Gamma(S) = \Gamma(S) \cup ExtendedMarginalCost(G, S, v, h, t)$; /* First hop in-neighbor cost */
10 **end**
11 $S_1 = \Gamma(S)$;
12 $\Gamma(S) = \emptyset$;
13 **for** *each* $v \in S_1$ **do**
14 $\quad \Gamma(S_1) = \Gamma(S_1) \cup ExtendedMarginalCost(G, S, v, x, h)$; /* Second hop in-neighbor cost */
15 **end**
16 $\Gamma(S) = \Gamma(S_1)$;
17 **for** *each* $v \in \Gamma(S)$ **do**
18 $\quad \Gamma(S) = \Gamma(S) - \{v | y_v = 1 \ OR \ t_v = 1\}$; // Commonality discount
19 **end**
20 $\sigma(S) = |\Gamma(S)|$;
21 $return \ \sigma(S)$;

opportunity cost is given by $\sigma(S) = |\Gamma(S)|$ as described in the lines 6 to 9 of Alg. 3.

In the best case (C_b), the algorithm selects only one in-neighbor node from both the first hop and second hop neighbors to activate the associated target node. The best case complexity is given by:

$$C_b = k + k.1 + k.1.1 = 3k = O(k) \tag{10}$$

The worst case (C_w) happens when the algorithm selects all the in-neighbors in both first hop and second hop to activated the target node and is represented by:

$$C_w = k + kd + k.d.d = k + kd + kd^2 = O(kd^2) \tag{11}$$

In the average case (C_a), the algorithm picks up an expected number of in-neighbors with equal probability ($\frac{1}{d}$) from both the first and second hop neighbors in order to activate the target nodes and hence the complexity is given by:

$$C_a = k + k.\frac{d-1}{2} + k.\frac{d-1}{2}.\frac{d-1}{2} \approx O(kd^2) \tag{12}$$

C. The Proposed Extended RLT-RIM algorithm

This ERLT-RIM algorithm is a linear threshold (LT)-based model and an extension of the existing RLT-RIM model [18] for finding opportunity cost as stated in the Alg. 4 and Alg. 5. Two levels of extensions have been incorporated into the ERLT-RIM model and are described below.

1) Considering already activated nodes: If there are some previously activated (activated as other node's in-neighbors) in-neighbors of a node, these nodes are considered first for node activation by picking them one by one randomly until the target node is activated with their aggregated influence as stated in the lines 12 to 25 of the Alg. 4. If the target node is not activated with all the already-activated in-neighbors, then new inactive nodes are selected randomly as stated in the lines 26 to 36 of the Alg. 4. This feature reduces the opportunity cost while ensuring the node's activation.

2) Commonality discount: There may arise some cases where a node u may be activated as second hop node of some target node v_1 and also activated as first hop and/or a target layer node of some other target node v_2. Then, the node u will not be added to the opportunity cost set. We describe this extended level of optimization as *commonality discount* as illustrated in the Fig. 4. This has been accomplished by the lines 16 to 19 of the Alg. 5.

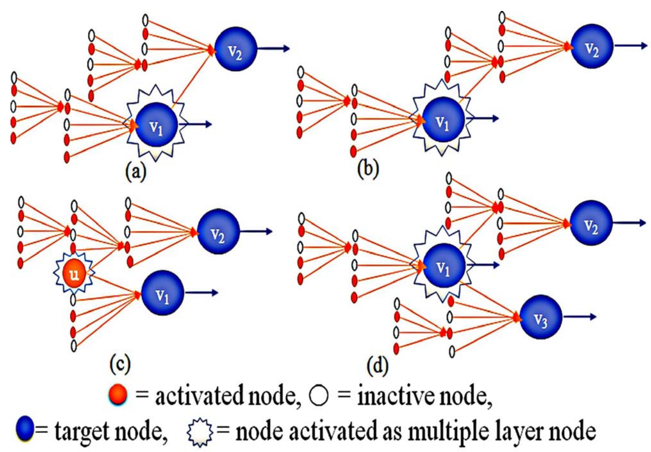


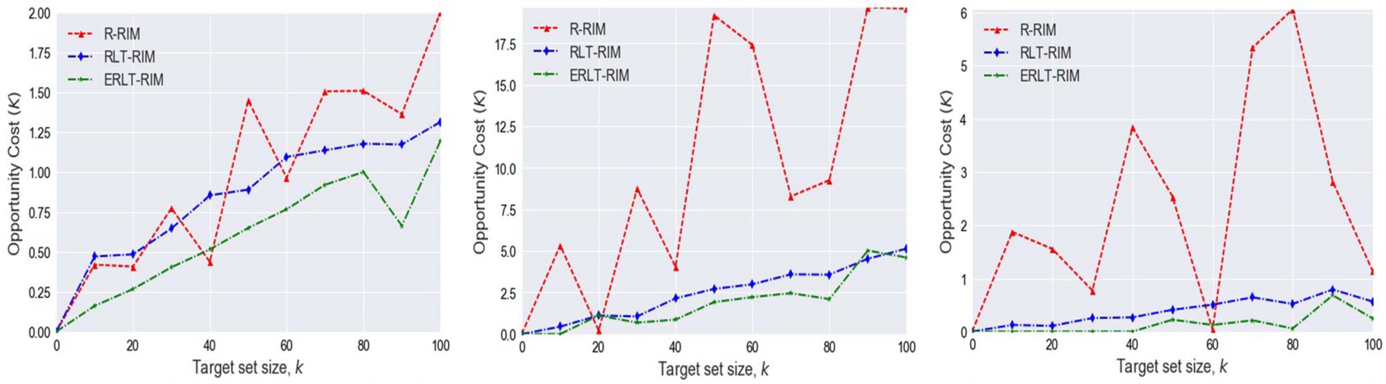
Fig. 4. Commonality discount (a) node v_1 is activated as both first hop and target layer node, (b) node v_1 is activated as both second hop and target layer node, (c) node u is activated as both second hop and first hop node, (d) node v_1 is activated as second hop, first hop, and target layer node.

Definition 2. Commonality discount: Once a node u is activated as a first hop and/or target layer node of any target node, it must not be included in the opportunity cost set $\Gamma(S)$ even if it is activated as a second hop node of one or more target nodes earlier or afterward. □

The commonality discount can be jointly computed by the following equation:

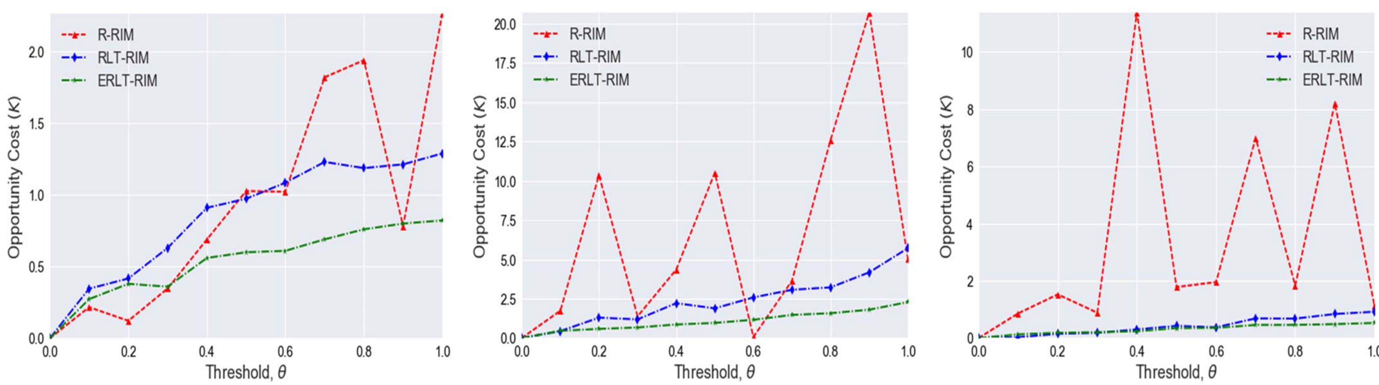
$$\Gamma(S) = \Gamma(S) - \{u | y_u = 1 \ OR \ t_u = 1\} \tag{13}$$

The complexity of ERLT-RIM has the same asymptotic order $O(kd^2)$ as that of RLT-RIM.



(a) Facebook dataset. **(b)** Twitter dataset. **(c)** Epinion dataset.

Fig. 5. The opportunity cost for different values of k for a) Facebook dataset, b) Twitter dataset, and c) Epinions dataset.



(a) Facebook dataset. **(b)** Twitter dataset. **(c)** Epinion dataset.

Fig. 6. The opportunity cost for different values of θ ($k = 100$) for a) Facebook dataset, b) Twitter dataset, and c) Epinions dataset.

V. Performance Evaluation

In this section, we have evaluated the performance of the proposed ERLT-RIM model using real datasets of three popular social networks: Facebook, Twitter, and Epinions. We have compared the result with the existing R-RIM and RLT-RIM models described in the previous section.

A. Data collection

We have collected real datasets of three prominent social networks. First one is Facebook[1] dataset. The entries of the dataset are given by an edge list and the ID of each node is anonymized by replacing the original ID with a new value for ethical issue. It has $4,039$ nodes and $88,234$ edges. The second dataset is taken from another widely used social network named Twitter[2] and is given by and edge list as well. There are $81,306$ nodes and $1,768,149$ edges in the Twitter dataset. The last one is Epinions[3] dataset where there are $75,879$ nodes and $508,837$ edges as summarized in the Table II.

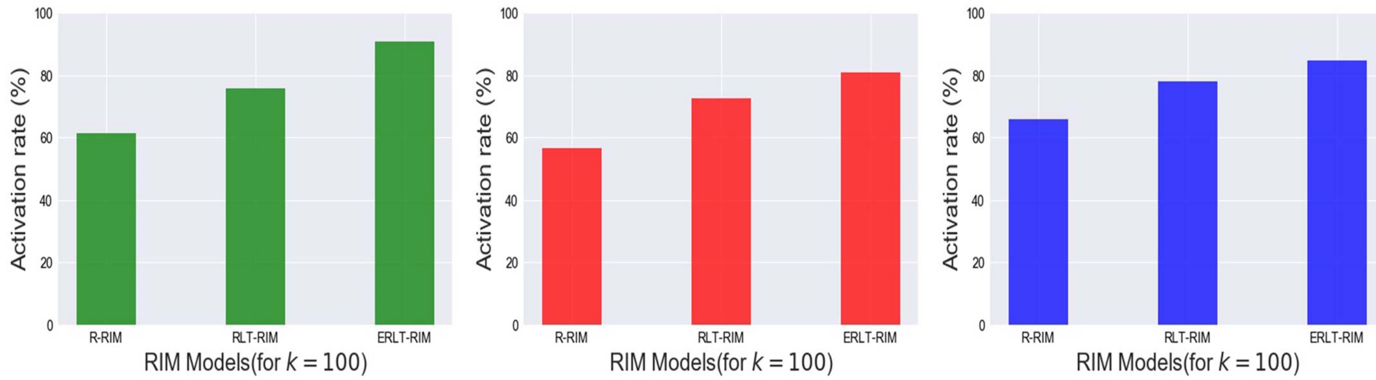
[1]https://snap.stanford.edu/data/egonets-Facebook.html
[2]https://snap.stanford.edu/data/egonets-Twitter.html
[3]https://snap.stanford.edu/data/soc-Epinions1.html

TABLE II
Dataset description

Networks	**Facebook** [26]	**Twitter** [26]	**Epinions** [27]
Nodes	$4,039$	$81,306$	$75,879$
Edges	$88,234$	$1,768,149$	$508,837$
Average cluster coefficient	0.6055	0.5653	0.1378

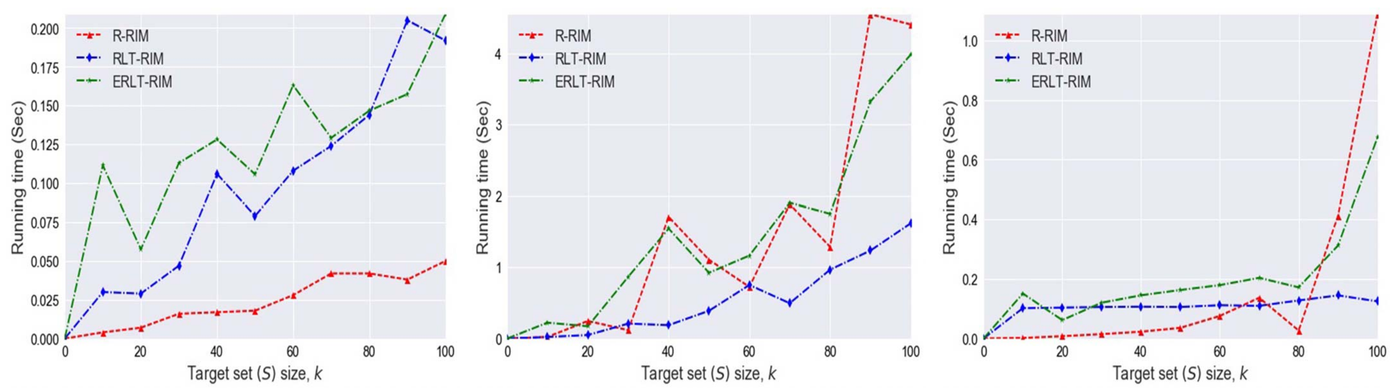
B. Simulation setup

We have evaluated the performance of the proposed ERLT-RIM model using Python programs executed on an Intel(R) Core (TM) i3-4150 CPU @ 3.50GHz, 3.50GHz machine with 8 GB RAM. We have employed *Monte Carlo (MC)* simulation [2] for all the datasets. Each algorithm is executed $1,000$ times and the average of all the calculated values is taken for the analysis. The target set S would be generated by any of IM algorithms $e.g.$ LT or IC model. However, for simplicity, we have taken k target nodes randomly. We have applied *degree centrality* [2] technique to estimate the influence weight w_{uv} of necessary edges. In the experiment, we have generated threshold values, θ_v to each node v by the *Heuristic Individual (HI)* threshold model [28]. The beauty of the HI model is that it gives smaller threshold values when



(**a**) Facebook dataset. (**b**) Twitter dataset. (**c**) Epinion dataset.

Fig. 7. Activation rate of different RIM Models of a) Facebook dataset, b) Twitter dataset, and c) Epinions dataset for $k = 100$.



(**a**) Facebook dataset. (**b**) Twitter dataset. (**c**) Epinion dataset.

Fig. 8. The running time for different values of k for a) Facebook dataset, b) Twitter dataset, and c) Epinions dataset.

degree centrality is employed to estimate the influence weight w_{uv}. Thus, it is favorable to decrease the effect of insufficient influence.

C. Performance analysis

In this section, we have discussed the performance analysis of the proposed ERLT-RIM algorithm by comparing with the performance of the existing R-RIM and RLT-RIM algorithms. We have compared the opportunity cost, running time and the effect of insufficient influence as activation rate of each of the algorithms.

1) Opportunity cost: The ERLT-RIM model exhibits the lower cost margin than R-RIM and RLT-RIM models as revealed in the Fig. 5. This is due to the working principles of the algorithms. The R-RIM model just selects a random number of in neighbors as opportunity cost whereas the RLT-RIM aggregates the influence weight each time it picks up a new node and compares the aggregated influence weight with the threshold of the target node. If the aggregated influence weight reaches the threshold, the target node is activated. When the target node is activated it stops probing more in-neighbors. On the other hand, ERLT-RIM works in the same manner as RLT-RIM with extra two levels of optimization. At every step, it first includes already activated in-neighbors and finally, it excludes commonality discount. This makes the ERLT-RIM more economical than both the R-RIM and RLT-RIM models.

The Fig. 6 illustrates the effect of different threshold values on the opportunity cost for the mentioned datasets while k is fixed to 100. The opportunity cost increases with the increase of threshold values. The proposed ERLT-RIM again outperforms both the existing models, that is, the proposed model is the most economical as compared to the R-RIM and RLT-RIM models.

2) Activation rate: Activation rate measures the effect of insufficient influence in the social network. The Fig. 7 illustrates the percentage of activated nodes for different social networks for the target set size, $k = 100$. The figure reveals that the RLT-RIM is better than R-RIM but ERLT-RIM outperforms both the R-RIM and RLT-RIM due to extra computation and optimization adapted.

3) Running time: On the other hand, The R-RIM model shows better running time than that of RLT-RIM and ERLT-RIM models as depicted in the Fig. 8. This is due to extra calculations and comparisons involved in the ERLT-RIM and RLT-RIM models. Each time a node is chosen, the algorithm checks whether the aggregated influence reaches to the target node's threshold or not. If the target node is not activated it iteratively probes other nodes consuming more time than R-RIM model which just selects a number of nodes randomly. ERLT-RIM model involves more computation due to two levels of optimization.

Some fluctuations are present in the results due to the random nature of the all three algorithm, especially shown in the Fig. 5, Fig. 6, and Fig. 8.

VI. Conclusion

In this research, we have solved the Reverse Influence Maximization (RIM) problem in which the viral marketing approach is applied in reverse order as compared to that of general influence maximization problem. The RIM estimates the opportunity cost which is the minimum number of in-neighbors that must be activated so that all the target nodes are activated. To solve the RIM problem, we have proposed the ERLT-RIM model which is the extension of the existing RLT-RIM model. The ERLT-RIM model estimates the opportunity cost by employing two more levels of optimazation such as, considering already activated nodes and commonality discount. The simulation of the ERLT-RIM model on three datasets of real networks ($e.g.$ Facebook, Twitter, and Epinions) shows that the proposed model out performs the existing R-RIM and RLT-RIM models by providing an economical opportunity cost with a faster running time and better activation rate as well.

The future scope includes stating the approximation ratio of the proposed algorithms to enrich the performance bound.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2015-0-00567, Development of Access Technology Agnostic Next-Generation Networking Technology for Wired-Wireless Converged Networks).

*Dr. CS Hong is the corresponding author.

References

[1] J.-R. Lee and C.-W. Chung, "A query approach for influence maximization on specific users in social networks," *IEEE Transactions on knowledge and data engineering*, vol. 27, no. 2, pp. 340–353, 2015.

[2] 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.

[3] W. Lu and L. V. Lakshmanan, "Profit maximization over social networks," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 479–488.

[4] K. Burke, "As consumer attitudes shift, so must marketing strategies," in *Twelfth ACM Conference on Electronic Commerce (EC'11)*, 2003, pp. 157–166.

[5] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Discovering leaders from community actions," in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 499–508.

[6] H. Zhang, H. Zhang, A. Kuhnle, and M. T. Thai, "Profit maximization for multiple products in online social networks," in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 2016, pp. 1–9.

[7] S. Bhagat, A. Goyal, and L. V. Lakshmanan, "Maximizing product adoption in social networks," in *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012, pp. 603–612.

[8] J. Zhou, Y. Zhang, and J. Cheng, "Preference-based mining of top-k influential nodes in social networks," *Future Generation Computer Systems*, vol. 31, pp. 40–47, 2014.

[9] J. Kostka, Y. A. Oswald, and R. Wattenhofer, "Word of mouth: Rumor dissemination in social networks," in *International Colloquium on Structural Information and Communication Complexity*. Springer, 2008, pp. 185–196.

[10] A. Talukder, R. Kamal, A. K. Bairagi, M. G. R. Alam, S. F. Abedin, M. A. Layek, H. T. Nguyen, and C. S. Hong, "Rumors in the social network: Finding the offenders using influence maximization," in *KoreanComputer Congress (KCC)*. KCC, 2015, pp. 1214–1216.

[11] Y. Wang, W. Huang, L. Zong, T. Wang, and D. Yang, "Influence maximization with limit cost in social network," *Science China Information Sciences*, vol. 56, no. 7, pp. 1–14, 2013.

[12] 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.

[13] A. Goyal and L. V. Lakshmanan, "Recmax: Exploiting recommender systems for fun and profit," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1294–1302.

[14] 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.

[15] Y. Zhu, Z. Lu, Y. Bi, W. Wu, Y. Jiang, and D. Li, "Influence and profit: Two sides of the coin," in *2013 IEEE 13th International Conference on Data Mining*. IEEE, 2013, pp. 1301–1306.

[16] D. T. Nguyen, S. Das, and M. T. Thai, "Influence maximization in multiple online social networks," in *Global Communications Conference (GLOBECOM), 2013 IEEE*. IEEE, 2013, pp. 3060–3065.

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

[18] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, M. A. Layek, H. T. Nguyen, and C. S. Hong, "A cost optimized influence maximization in social networks," in *2017 IEEE The 19th Asia-Pacific Network Operations and Management Symposium (APNOMS 2017)*. IEEE, 2017.

[19] S. Ahmed and C. Ezeife, "Discovering influential nodes from trust network," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM, 2013, pp. 121–128.

[20] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate." in *SDM*, vol. 11. SIAM, 2011, pp. 379–390.

[21] A. Goyal, W. Lu, and L. V. Lakshmanan, "Simpath: An efficient algorithm for influence maximization under the linear threshold model," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 211–220.

[22] J. Pan, F. Jiang, and J. Xu, "Influence maximization in social networks based on non-backtracking random walk," in *IEEE International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2016, pp. 260–267.

[23] G. Tong, W. Wu, S. Tang, and D.-Z. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Transactions on Networking*, 2016.

[24] Y. Zhang, Y. Bai, L. Chen, K. Bian, and X. Li, "Influence maximization in messenger-based social networks," in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[25] E. Horowitz and S. Sahni, *Fundamentals of computer algorithms*. Computer Science Press, 1978.

[26] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in neural information processing systems*, 2012, pp. 539–547.

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

[28] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, H. T. Nguyen, and C. S. Hong, "Threshold estimation models for influence maximization in social network," in *Korean Institute of Information Scientists and Engineers (KIISE)*. KIISE, 2016, pp. 888–890.