Susceptible-Infection-based Cost-effective Seed Mining in Social Networks

Ashis Talukder
Computer Science and Engineering
Kyung Hee University, Yongin 17104, South Korea
email: ashis@khu.ac.kr

Choong Seon Hong*
Computer Science and Engineering
Kyung Hee University, Yongin 17104, South Korea
email: cshong@khu.ac.kr

***Abstract*—The aim of Influence maximization (IM) techniques is to mine social networks to find a small set of influential seed users that maximize the viral marketing profit. On the other hand, the Reverse Influence Maximization (RIM) maximizes the profit by minimizing the viral marketing cost. Here, the cost is estimated by the lowest number of nodes which are needed to activate seed nodes. On the other hand, the profit is computed by the highest number of nodes that can be influenced by seed users. However, most of the existing works assume that the seed nodes are either initially activated or offered free products for motivation. Thus, most of the studies do not address the seed activation cost. Therefore, in this research, we propose a Susceptible-Infection-based Greedy Reverse Influence Maximization (SIG-RIM) model to maximize the profit by minimizing the seeding cost. The proposed SIG-RIM model employs the Susceptible-Infected (SI) mechanism in reverse order to compute the seeding cost and a greedy technique to optimize the cost. Moreover, the SIG-RIM model tackles RIM challenges more efficiently. Finally, we conduct the performance evaluation of our model with real datasets of two popular social networks, and the result shows that the proposed model outperforms state-of-the-art models.**

***Index Terms*—influence maximization; reverse influence maximization; viral marketing; seeding cost; social network mining.**

I. Introduction

A. Background

We are the witness of the booming proliferation of social networks, which are currently used by almost all Internet users. Besides sharing news, trends, ideas, etc., social networks are used as a powerful medium for marketing. Nowadays, almost all the small and large business organizations are highly leveraging social networks in the perspective of *viral marketing* [1], [2], [3]. For instance, a *Facebook* or *Twitter* post of a *YouTube* video can make the video viral by any celebrity's fan-followers and in return, by the followers of followers. Then, the originator of the video would earn a considerable amount of incentives from *Google* for million-time views of the video in the *YouTube*.

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Grand Information Technology Research Center support program (IITP-2018-2015-0-00742) supervised by the Institute for Information & communications Technology Promotion (IITP).

***Dr. CS Hong** is the corresponding author.

B. Motivation

The *Influence Maximization (IM)* is a social network mining tool to find the influential seed users for viral marketing, and by targeting them, the profit is maximized in social networks [4]. We define the profit by the highest number of nodes that can be motivated by seed users when they are initially activated. However, most of the IM models assume that the seed users are initially active [4], [5], [6] or offer some free sample products [7], [8] and thus, do not address the fact that some other influential users could also activate seed nodes. Although some authors [9], [10] consider the activation cost of all the nodes activated by the seed users, still the seed activation cost is ignored.

On the other hand, the *Reverse Influence Maximization (RIM)* techniques maximize the viral marketing profit by minimizing the cost, i.e., minimizing the seeding cost, which is given by the least amount of nodes that are needed to activate the seed users [11], [12], [13], [14]. However, their models are not fully capable of resolving RIM-challenges such as the insufficient influence, setting the stopping criteria of the node activation process, considering three Basic Network Components (BNC), and the NP-Hardness of the problem.

C. Contributions

In this paper, we propose a *Susceptible-Infection-based Greedy RIM (SIG-RIM)* solution to maximize the viral marketing profit by minimizing the seeding cost in social networks. The SIG-RIM model estimates the marginal seeding cost by using the *Susceptible-Infection (SI)* model applying it in reverse order. Then, a greedy approach is employed to minimize the marginal seeding cost. The key contributions of the proposed SIG-RIM model are stated below.

1) We propose the SI-based Greedy RIM approach (SIG-RIM) to maximize the viral marketing profit by minimizing the seeding cost. The SIG-RIM model jointly employs the SI model in reverse order and greedy optimization technique.
2) We introduce a variant of the traditional SI model to employ it in reverse order, and use it in the node diffusion process.
3) The use of SI and greedy techniques resolve all the RIM issues efficiently.

4) Finally, we perform simulation of our SIG-RIM model by using real datasets of two popular social networks, and the results show that the SIG-RIM model outperforms existing techniques.

The remaining part of the paper is organized as follows: literature review and the problem formulation are provided in section II and section III, respectively. The SIG-RIM model is proposed in section IV, and the performance analysis of the proposed model is stated in section V. Finally, section VI covers concluding remarks.

II. Related Works

Influence Maximization (IM) techniques gain huge research interest after the emergence of the classical *Independent Cascade (IC)* and *Linear Threshold (LT)* models, proposed by Kempe *et al.* [4]. The IM technique under either of the models exhibits 63% performance ratio with sub-modularity optimization.

Many authors apply influence maximization methods to maximize the profit in social networks. Bhagat et al. [7] present such a study in which product adoption is optimized to enhance the profit. Lu et al. [8] maximize the profit by product adoption as well. However, the authors show that a user adopts any product not only influenced by viral marketing, but also by evaluating the product price. Unlike previous studies, multiple products are taken into account in [15] for profit maximization in social networks. Zhu et al. [10] show that the high price of the product disrupts the influence diffusion in the social network. Therefore, we cannot maximize profit and influence simultaneously in the network. However, none of the above studies considers the viral marketing cost, i.e., the seed activation cost in their studies. Zhou et al. [16] consider the cost of all the activated nodes; however, do not find the seed activation cost.

After that, Talukder et al. [12] propose a new paradigm of Reverse Influence Maximization (RIM) to maximize the profit by minimizing the cost especially, the seeding cost. In their *Random RIM (R-RIM)* and *Randomized LT-based RIM (RLT-RIM)* methods, influence is diffused in a reverse manner. RIM techniques estimate nodes that activate seed nodes whereas, the IM model determines nodes that are activated by active seed users. An extended work is proposed in [14] and a reverse path activation-based RIM model is proposed in [17]. Further, RIM studies also identify some challenging issues, e.g., insufficient influence, the stopping criteria of the diffusion process, considering three basic network components (BNC), the NP-Hardness of the problem. However, existing RIM models cannot provide the optimal seeding cost while handling all the challenges simultaneously.

Therefore, in this paper, we propose the SI-based Greedy RIM (SIG-RIM) model to maximize the viral marketing profit by minimizing the seeding cost. Moreover, the SIG-RIM model deals with RIM challenges efficiently and outperforms the existing RIM models at the same time.

TABLE I
LIST OF PARAMETERS

Symbols	Meaning
$\mathcal{G}(\mathcal{V}, \mathcal{E})$	Social network
$\mathcal{V}$	Nodes or social network users
$\mathcal{E}$	Edges or Social relationships
$n(v)$	Set of out-neighbors of the node v
$n^{-1}(v)$	Set of in-neighbors the node v
$\mathcal{S}$	Seed set
k	Seed set size, $k = \|\mathcal{S}\|$
$\Delta(v)$	Marginal seeding cost set of the node v
$\Delta(\mathcal{S})$	Seeding cost set of all the seeds in $\mathcal{S}$
$\delta(v)$	Marginal seeding cost of the node v, $\delta(v) = \|\Delta(v)\|$
$\delta(\mathcal{S})$	Seeding cost of all the seeds in $\mathcal{S}$, $\delta(\mathcal{S}) = \|\Delta(\mathcal{S})\|$
$\mathcal{S}_t$	Susceptible population
$\mathcal{I}_t$	Infected population
α	Infection rate (infection probability)
t	Iteration variable for hops
T	The total number of hops (iterations)
p_t	The cascade influence of up to t hops (iteration)
d	Maximum degree in $\mathcal{G}$
C	The complexity of the SIG-RIM algorithm

III. The Problem Formulation

In this section, we formulate the Reverse Influence Maximization (RIM) problem to minimize the seeding cost or opportunity cost. We take a social network given by a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, in which a node $v \in \mathcal{V}$, is a social network user, and a link $(u, v) \in \mathcal{E}$ specifies that users u and v are connected in the social network. We also denote the in-neighbor set as $n^{-1}(v)$ and the out-neighbor set as $n(v)$ of a node v.

Let us now consider a profit maximization scenario under a single product. For a given set $\mathcal{S}$ of k seed nodes, the IM method maximizes the profit by maximizing the number of activated nodes by the seed nodes when they are primarily enabled [7], [8], [15]. On the other hand, the RIM problem estimates the seeding cost, which is given by the least amount of nodes required to influence the seed nodes [12], [14]. That is, the RIM technique maximizes the profit by minimizing the seeding cost $\delta(\mathcal{S})$, which is the main objective of this paper. Further, we assume that if a node is activated to influence the seed node, it does not change the decision. That is an activated (infected) node never be considered to be in the target market (susceptible) again. From now on, we use cost minimization and profit maximization interchangeably.

Definition 1 (RIM Problem)**.** Given a social network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and a seed set $\mathcal{S}$, ($|\mathcal{S}| = k$), the RIM problem aims at maximizing the profit by minimizing the seeding cost $\delta(\mathcal{S})$, which is the least amount of nodes that can influence all the seed nodes, $v \in \mathcal{S}$. □

IV. The Proposed Solution Framework

Here, we propose a Susceptible-Infected-based Greedy solution to the RIM problem (SIG-RIM). The SIG-RIM model employs the Susceptible-Infected (SI) mechanism in reverse order in the node activation process, and the greedy method for cost optimization. The traditional SI model is modified to be applied in reverse order to determine which nodes infect a given seed node.

A. Justification of the SI Model

In the desired scenario of cost minimization (profit maximization), we assume that all the nodes are initially inactive (susceptible). Then, an inactive (susceptible) node can be activated (infected) by the influence (exposure to contagion disease) of an activated (infected) node in the RIM process. However, an activated (infected) node never becomes inactive (susceptible) again. Thus, the SI model can adequately represent the cost minimization scenario, as illustrated in Fig. 1. Moreover, the use of the SI model gives the proposed model leverage in solving RIM challenges intelligently.

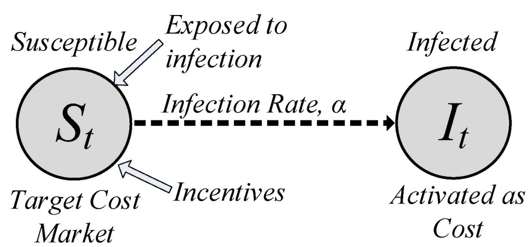


Fig. 1. The SI model is applied in viral marketing.

B. The SIG-RIM Model

The partial seeding cost set $\Delta(u)$ for all $u \in n^{-1}(v)$ is estimated first by using a SI diffusion model applied in reverse order and then, the partial costs are optimized by using a greedy model to find the least marginal seeding cost set $\Delta(v)$, of all $v \in \mathcal{S}$, as stated in Fig. 2[1].

1) The Partial Cost Set, $\Delta(u)$*:* Initially, we assume that the node u is the only infected node, and all the in-neighbors of the node u are the potential markets (susceptible) and hence we have,

$$\mathcal{S}_1 = n^{-1}(u), \tag{1}$$

$$\mathcal{I}_1 = \{u\}, \tag{2}$$

Now, we determine the infected nodes by exposing the susceptible population to an infection. Nodes are infected (activated) with probability α in the SI mechanism. We consider that these infected nodes are responsible to activate the node u. The contagion process is examined up to the T-th hop of the node u. The infected nodes participate in seed marketing (activating) and are included in seeding cost set. The list of activated nodes at any hop t with activation probability α is given by,

$$\mathcal{I}_t = \alpha \mathcal{S}_{t-1} \mathcal{I}_{t-1}, \tag{3}$$

Thereafter, the target market (susceptible) for the next hop is updated with all the in-neighbors of activated nodes. Again, an activated node can never be in the target market again and thus, is excluded from the target market. In other words, any infected node is neither recovered nor considered as a susceptible candidate again. Therefore, the target market for the next hop is given by,

$$\mathcal{S}_t = \left[\cup_{u \in \mathcal{I}_t} \left\{ n^{-1}(u) \right\} \right] - \Delta(u), \tag{4}$$

where, $\Delta(u)$ is the set of all activated nodes in all previous $t-1$ hops. Again, we include only the inactive in-neighbors of activated individuals in the target market for the next hop, we can simplify (3) as,

$$\mathcal{I}_t = \alpha \mathcal{S}_{t-1} \tag{5}$$

We keep track of the activated nodes at each hop t and aggregate them to estimate the partial seeding cost set, $\Delta(u)$ as,

$$\Delta(u) = \underset{t}{\cup} \mathcal{I}_t \tag{6}$$

2) The Greedy Optimization: Next, a majority number of in-neighbors of v are selected such that their combined seeding cost is minimum and this cost is the optimal marginal seeding cost $\delta(v)$. We employ the contagion threshold, 0.5 [18], [19], [20], [21], which indicates that a node v is infected if at least half of its in-neighbors are infected. For instance, in Fig. 2, nodes u_1, u_3, and u_4 are selected greedily for their aggregated cost being minimum (say).

Finally, the optimized seeding cost for all the seed nodes is estimated as follows:

$$\Delta(\mathcal{S}) = \underset{v \in S}{\cup} \Delta(v), \tag{7}$$

$$\delta(\mathcal{S}) = \left| \Delta(\mathcal{S}) \right|, \tag{8}$$

C. The Infection Rate (α)

We consider the *infection rate* as the infection (activation) probability which is convenient for implementation [22]. In our model, the values of α is computed by *Tri-valency model* [5], [17], [23], [24], [25] where $\alpha \in \{0.1, 0.01, 0.001\}$.

D. Stopping Criteria for the SI model

The node activation process in $\Delta(u)$ estimation terminates if either of the following two conditions (lines 7 - 9 in Algorithm 1) arises first:

1) If the contagion dies out i.e., if no node is infected at any hop and there is no susceptible node for the next hop.
2) If the *cascade influence*, p_t reaches to some negligible value (say, 10^{-6}) [26] at any hop t.

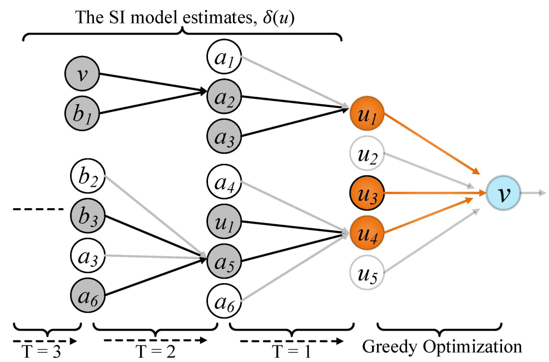


Fig. 2. Working strategy of the SIG-RIM model.

[1]The graph in Fig. 2 is drawn hop-wise for better realization and thus, has some repeated nodes.

E. The SIG-RIM Algorithm

The proposed SIG-RIM model is presented in Algorithm 1. Here, $\Delta(u)$ is estimated in lines $4 - 25$, in which the terminating conditions are stated in lines 7 - 9. The contagion is examined in line 11 - 18 and the susceptible set is updated in lines $21 - 23$ for the next hop. For each hop, the influence is decayed α times as shown in line 19.

The greedy optimization is performed in line 26 and the marginal seeding cost set, $\Delta(v)$ is calculated by line 27. Finally, the seeding cost set, $\Delta(S)$, and the seeding cost, $\delta(S)$, are computed in lines 28 and 29, respectively.

Theorem 1. *The RIM problem is NP-Hard under the SIG-RIM model.*

Proof. The greedy selection used in the SIG-RIM algorithm can be considered as Knapsack optimization. At each step, the node u with the lowest $\Delta(u)$ is selected then, the cost is updated with respect to all the selected nodes and this process continues for $\left\lfloor |n^{-1}(v)| \right\rfloor + 1$ in-neighbors. That is, the RIM problem under the SIG-RIM algorithm is a variant of the traditional Knapsack problem. The Knapsack problem is a well-known NP-Hard problem [27], [28], and therefore, the RIM problem under the SIG-RIM model is also NP-Hard. □

F. The Approximation Ratio

The approximation ratio (i.e., performance bound) of the proposed model is stated in the theorem below.

Theorem 2. *The greedy model used in the SIG-RIM model is a 2-approximation algorithm, i.e.,*

$$\delta \leq 2\delta^* \tag{9}$$

where, δ and δ are the estimated and optimal cost, respectively.

Proof. According to the Theorem 1, the SIG-RIM model is a variation of Knapsack technique. Thus, the estimated cost of selected in-neighbors are greater than the optimal cost ($\delta \geq \delta^*$) and can be at best the double of the optimal cost ($\delta \leq 2\delta^*$) [29], and hence,

$$\delta^* \leq \delta \leq 2\delta^*$$
$$i.e.,\ \delta \leq 2\delta^*.$$
□

G. Complexity

The SI model takes $O(Td^2)$ time for the execution up to T hops, and the greedy optimization takes $O(d^2)$ time. Therefore, the SIG-RIM model's total complexity is expressed as,

$$C \leq k(T.d^2 + d^2) \approx O(kTd^2), \tag{10}$$

where, d is the average degree in the network, $\mathcal{G}$.

V. PERFORMANCE EVALUATION

The performance of the proposed SIG-RIM model is evaluated by comparing with that of existing models using two real datasets such as Epinions and Twitter.

Algorithm 1: The **SIG-RIM** Model

Input: $\mathcal{G}(\mathcal{V},\mathcal{E})$, $\mathcal{S}$
Result: $\Delta(\mathcal{S})$, $\delta(\mathcal{S})$

1 $\Delta(\mathcal{S}) := \emptyset$;
2 **for** $v \in \mathcal{S}$ **do**
3 $\quad \Delta(v) := \emptyset$;
4 $\quad$ **while** $u \in n^{-1}(v)$ **do**
5 $\quad\quad p_t := 1,\ I_t := \emptyset,\ \mathcal{S}_t := n^{-1}(u)$;
$\quad\quad$ /* Initialization by (1), (2) */
6 $\quad\quad$ **while** *true* **do**
7 $\quad\quad\quad$ **if** $S_t == \emptyset \parallel p_t \leq 10^{-6}$ **then**
8 $\quad\quad\quad\quad$ *break*; /* Terminating condition */
9 $\quad\quad\quad$ **end**
10 $\quad\quad\quad \Delta(u) := \mathcal{I}_t := \{u\}$; /* Initialization */
11 $\quad\quad\quad$ **while** $w \in \mathcal{S}_t$ **do**
12 $\quad\quad\quad\quad$ **if** $w \notin \Delta(u)$ **then**
13 $\quad\quad\quad\quad\quad$ **if** *w is activated with probability α* **then**
14 $\quad\quad\quad\quad\quad\quad \Delta(u) := \Delta(u) \cup \{w\}$;
$\quad\quad\quad\quad\quad\quad$ /* Infected */
15 $\quad\quad\quad\quad\quad\quad \mathcal{I}_t := \mathcal{I}_t \cup \{w\}$;
16 $\quad\quad\quad\quad\quad$ **end**
17 $\quad\quad\quad\quad$ **end**
18 $\quad\quad\quad$ **end**
19 $\quad\quad\quad p_t = p_t * \alpha$; /* Influence decay */
20 $\quad\quad\quad \mathcal{S}_t := \emptyset$;
21 $\quad\quad\quad$ **while** $y \in \mathcal{I}_t$ **do**
22 $\quad\quad\quad\quad \mathcal{S}_t := \mathcal{S}_t \cup \{y\}$; /* Update by (4) */
23 $\quad\quad\quad$ **end**
24 $\quad\quad$ **end**
25 $\quad$ **end**
26 $\quad \mathcal{A}$:= Select the majority number of $u \in n^{-1}(v)$ s.t. min $\Delta(\mathcal{A})$;
27 $\quad \Delta(v) := \Delta(\mathcal{A})$; /* Marginal cost set */
28 $\quad \Delta(\mathcal{S}) := \Delta(\mathcal{S}) \cup \Delta(v)$; /* Cost set by (7) */
29 **end**
30 $\delta(\mathcal{S}) = |\Delta(\mathcal{S})|$; /* Final seeding cost by (8) */
31 *return* $\delta(\mathcal{S}), \Delta(\mathcal{S})$;

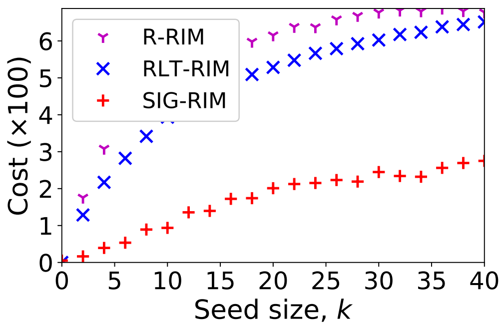
A. Data Collection

We collect two datasets of well recognized social networks such as *Epinions*[2] and *Twitter*[3] from the Stanford large network dataset collection [30]. The summary of the datasets is represented in Table II.

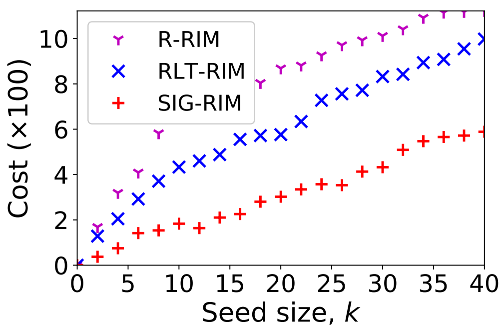
B. Simulation Setup

In our simulation, we execute Python codes on an Intel Core i5 machine with 8 GB RAM. We use the *Monte Carlo (MC)* technique for the simulation and the mean value of each parameter is used for the comparative study [4]. The seed set, $\mathcal{S}$ is generated randomly and we take $\alpha \in \{0.1, 0.01, 0.001\}$

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

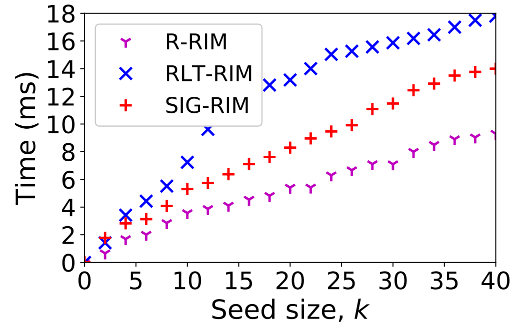


(a) Epinion dataset.

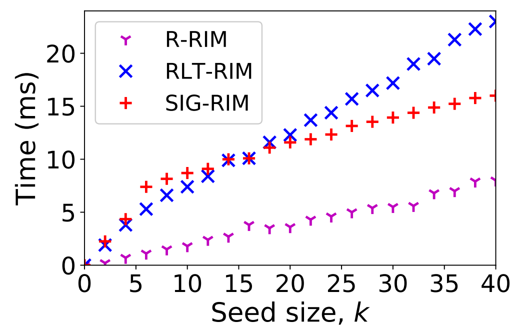


(b) Twitter dataset.

Fig. 3. Seeding cost of considered models for $k = 1$ to 40, for a) Epinions dataset, and b) Twitter dataset.



(a) Epinion dataset.



(b) Twitter dataset.

Fig. 4. Running time of considered models for $k = 1$ to 40, for a) Epinions dataset, and b) Twitter dataset.

[23]. The comparative study is done with R-RIM, and RLT-RIM models [12].

TABLE II
DATASET DESCRIPTION

Social networks	Nodes	Edges
Epinions	75,879	508,837
Twitter	81,306	1,768,149

C. The Result Analysis

Here, we evaluate the proposed SIG-RIM method in terms of the estimated seeding cost, running time, and the efficiency to resolve RIM-issues.

1) Seeding Cost: Fig. 3 shows the estimated seeding cost of all the considered algorithms for the Epinions, and Twitter datasets. Here, the cost is computed for the seed sets of different size, $k = 1$ to 40.

The R-RIM method randomly estimates the cost, and the RLT-RIM applies the LT model with randomly selected nodes. On the other hand, the proposed SIG-RIM method determines the marginal cost by using the SI model in reverse order and then, uses a greedy model to optimize the cost. Therefore, the proposed model returns the most economical seeding cost, which is about $2-3$ time lower than that of the existing models, for both the datasets. The enhanced performance of the proposed model is due to the use of greedy optimization, which is not employed in the existing models. Thus, the simulation results exhibit that the SIG-RIM technique beats the current models in terms of the estimated seeding cost. Again, the Twitter dataset presents a higher cost due to its higher number of social links (higher average degree) as compared to another network.

2) Running Time: Fig. 4 depicts the running time of different methods for different seed sets with size, $k = 1$ to 40, for both the datasets.

The figure unveils that, on average, the R-RIM model requires the lowest running time, the RLT-RIM needs the highest running time. However, the running time of the proposed technique lies in between that of these models. This running time pattern is due to the use of a fully stochastic procedure in the R-RIM model, which selects the nodes randomly. The RLT-RIM model also selects a node randomly. However, every time the RLT-RIM method selects a node, it aggregates influence weights and compares with the node's threshold value for activation. In contrary, in the proposed model, the node activation is performed for T hops, and then, the greedy model also takes a handsome amount of time. Again, for the Twitter dataset, all the models explore a higher number of nodes for the higher average degree and thus, require more time.

3) Handling RIM Challenges: The application of the SI model invalidates not only the concept of BNCs involved in the existing R-RIM and RLT-RIM models but also removes the insufficient influence effect because of not using threshold values in the node activation process in the proposed model. The SIG-RIM model discussed in subsection IV-D properly sets the terminating condition of the node activation process by either influence decay function [26] or with the death of contagion [23].

Finally, the greedy approximation technique addresses the NP-Hardness issue due properly. We also discuss the performance bound of the greedy approximation in Theorem 2. Thus, the SIG-RIM model tackles the RIM challenges efficiently as compared to the existing models.

VI. Conclusion

In this paper, we propose a social network mining tool, a Susceptible-Infection-based Greedy Reverse Influence Maximization (SIG-RIM) technique to maximize the viral marketing profit by minimizing the seeding cost. We modify the traditional Susceptible-Infection (SI) technique to employ it in reverse order in the node activations process. Then, a greedy approximation approach is employed to optimize the seeding cost. The use of the greedy optimization contributes to achieving the most economical seeding as compared to the existing models. We simulate the proposed model with datasets of two real networks, e.g., Epinions, and Twitter. The empirical result shows that the proposed method provides a superior seeding cost and the faster running time as compared to existing models. Moreover, our model addresses challenging issues more efficiently than existing models.

References

[1] J. Zhu, Y. Liu, and X. Yin, "A new structure-hole-based algorithm for influence maximization in large online social networks," *IEEE Access*, vol. 5, pp. 23 405–23 412, 2017.

[2] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks," in *Proceedings of the 2016 International Conference on Management of Data*. ACM, 2016, pp. 695–710.

[3] 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 *Korean Computer Congress (KCC 2015)*. KIISE, 2015, pp. 1214–1216.

[4] 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.

[5] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 88–97.

[6] 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.

[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] 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.

[9] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "A billion-scale approximation algorithm for maximizing benefit in viral marketing," *IEEE/ACM Transactions on Networking (TON)*, vol. 25, no. 4, pp. 2419–2429, 2017.

[10] 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.

[11] A. Talukder and C. S. Hong, "A susceptible-infection-based greedy model for mining cost-effective seeds in social networks," in *Korean Computer Congress (KCC 2019)*. KIISE, 2019, pp. 1533–1535.

[12] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, M. A. Layek, H. T. Nguyen, and C. S. Hong, "An approach of cost optimized influence maximization in social networks," in *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2017, pp. 354–357.

[13] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, H. T. Nguyen, and C. S. Hong, "The grim: Target marketing in social network," in *The International Symposium on Perception, Action, and Cognitive Systems (PACS 2016)*. KIISE, 2016, pp. 24–25.

[14] A. Talukder, M. G. R. Alam, N. H. Tran, and C. S. Hong, "A cost optimized reverse influence maximization in social networks," in *2018 IEEE/IFIP Network Operations and Management Symposium (NOMS 2018)*. IEEE, 2018, pp. 1–9.

[15] 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.

[16] J. Zhou, J. Fan, J. Wang, X. Wang, and B. Cheng, "Cost-efficient influence maximization in online social networks," in *Advanced Cloud and Big Data (CBD), 2017 Fifth International Conference on*. IEEE, 2017, pp. 232–237.

[17] A. Talukder and C. S. Hong, "Active reverse path based reverse influence maximization in social networks," in *Korea Software Congress (KSC 2017)*. KIISE, 2017, pp. 1203 – 1205.

[18] 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.

[19] A. Talukder and C. S. Hong, "A heuristic approach for viral marketing cost optimization in social networks," in *Korea Software Congress (KSC 2018)*. KIISE, 2018, pp. 1082–1084.

[20] 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.

[21] A. Talukder, D. H. Kim, and C. S. Hong, "A heuristic mixed model for viral marketing cost minimization in social networks," in *IEEE International Conference on Information Networking (ICOIN 2019)*. IEEE, 2019, pp. 141–146.

[22] P. Krishna, A. Mohan, and K. Srinivasa, *Practical Social Network Analysis with Python*, 1st ed. Springer International Publishing, 2018.

[23] A. Talukder and C. S. Hong, "Epidemiological reverse influence maximization in social networks with negative influencing," in *Korean Computer Congress (KCC 2018)*. KIISE, 2018, pp. 1277–1279.

[24] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 918–923.

[25] A. Talukder, A. K. Bairagi, D. H. Kim, and C. S. Hong, "Reverse path activation-based reverse influence maximization in social networks," *Journal of The Korean Institute of Information Scientists and Engineers (JOK)*, vol. 45, no. 11, pp. 1203–1209, 2018.

[26] A. Talukder, M. G. R. Alam, N. H. Tran, D. Niyato, and C. S. Hong, "Knapsack-based reverse influence maximization for target marketing in social networks," *IEEE Access*, vol. 7, pp. 44 182–44 198, 2019.

[27] A. Talukder and C. S. Hong, "Knapsack-based reverse influence maximization for target marketing in social networks," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, ser. SAC '19. New York, NY, USA: ACM, 2019, pp. 2128–2130. [Online]. Available: http://doi.acm.org/10.1145/3297280.3297621

[28] E. Horowitz and S. Sahni, *Fundamentals of computer algorithms*, 2nd ed. Computer Science Press, New York, 2012.

[29] D. S. Hochbaum, *Approximation algorithms for NP-hard problems*. Boston, MA, USA: PWS Publishing Co., 1996.

[30] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, jun 2014.