Reverse Influence Maximization for the Competitive Market in Dynamic Social Networks

Ashis Talukder, Sarder Fakhrul Abedin, Md. Shirajum Munir, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, Yongin 17104, South Korea.
Email: {ashis, saab0015, munir, cshong}@khu.ac.kr

Abstract

Social networks are growing faster, and there is a massive number of active users every day. Therefore, people are using social networks as a potential platform for marketing. Influence Maximization (IM) is such an approach to identify influential users for viral marketing in social networks. Most of the IM algorithms deal with viral marketing profit, which is the maximum number of nodes that can be activated by initial seed nodes. On the other hand, the minimum number of nodes that are required to activated initial seed nodes is called viral marketing cost or seeding cost, which is not focused in most of the existing studies. Therefore, in this paper, we propose a Reverse Influence Maximization (RIM)-based model for seeding cost optimization under the Competitive market in Dynamic (CD-RIM) social networks. The experimental results show that the proposed model outperforms existing RIM models.

1. Introduction

As social networks are expanding dramatically, and a massive number of people are interacting through social networks every day, people are using social networks as a potential medium of marketing. Influence Maximization (IM) is such a tool to identify influential nodes in the social networks for viral marketing [1], [2], [3], [4], [5], [6]. Kempe *et al.* [2] proposes widely accepted Linear Threshold (LT) and Independent Cascade (IC) models. Guney *et al.* [7] introduce the Sample Average Approximation (SAA) model for influence spread maximization. Further, Liu *et al.* [8] propose a time-bounded profit maximization. Yang *et al.* [9] propose a novel maximizing activity profit (MAP) model. However, most of the influence and profit maximization models assume either the seed nodes are initially activated or offer some free products to activate seed nodes.

Thereafter, some Reverse Influence Maximization (RIM) models are proposed to optimize the seeding cost, which is the minimum number of nodes that are required to activated seed nodes [4], [5], [10], [11]. These studies also identify some challenges such as convergence, basic network components (BNC), NP-Hardness, and insufficient influence. However, existing RIM models fail to produce optimized seeding cost and resolve RIM-challenges simultaneously.

Therefore, in this paper, we propose a Susceptible-Infection-Recovered (SIR)-based seeding cost minimization model under the Competitive market for Dynamic (CD-RIM) social networks. The CD-RIM model estimates the marginal seeding cost by using the SIR model and optimizes the cost by a Minimum Set Cover (MSC) approach. The key contributions of this paper are as follows:

1) We propose a SIR model-based RIM approach for seeding cost minimization under the Competitive market in Dynamic (CD-RIM) social networks.
2) The SIR and the greedy Set Cover models resolve all the RIM issues efficiently. Moreover, the greedy optimization contributes to a lower seeding cost.
3) Finally, the empirical results on two real social networks show that the prosed model outperforms the existing models.

2. System Model

For a particular product of multiple brands, let us assume that instances of social networks are represented by directed graphs $G_t(V_t, E_t)$ at any time slot t, where $1 \leq t \leq T$. The sets V_t, and E_t indicate the set of social network users and social ties at time t, respectively. We indicate in-neighbor and out neighbor sets of a node v by $n_t^{-1}(v)$ and $n_t(v)$ at time t of the network instance G_t, respectively.

The marginal seeding cost $\delta(v)$, of a seed node v, is the optimal number of influencer nodes that are required to ensure the activation of v in whole contagion observation time T. With a given seed set S of size k, the total seeding cost $\delta(S)$, is produced by combining the marginal costs of all $v \in S$ in total time T.

Definition 1 (The CD-RIM Problem). Given an initial social network instance $G_1(V_1, E_1)$ and a seed set S of size k, the CD-RIM problem estimates the seeding cost $\delta(S)$ in a competitive market, in T time slots. The cost is defined by the minimum number of nodes that must be activated in order to activate all the seednodes in S. □

3. The Proposed CD-RIM Model

The proposed RIM-based CD-RIM is illustrated in Figure 1, and Figure 2.

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2019-2015-0-00742) supervised by the IITP (Institute for Information & communications Technology Promotion) and by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2015-0-00557, Resilient/Fault-Tolerant Autonomic Networking Based on Physicality, Relationship and Service Semantic of IoT Devices).
*Dr. CS Hong is the corresponding author.

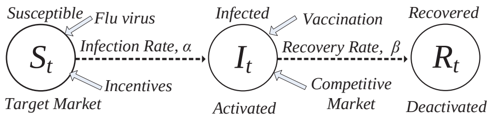


Figure 1: The SIR mechanism is used in the CD-RIM model.

A. The SIR Model

Figure 1 presents the function of the SIR method, which we apply in reverse order to estimate the seeding cost. We consider susceptible candidates (S_t) in the SIR model as the target market in the CD-RIM model. Similarly, we assume that the infected population (I_t) presents activated nodes, and the recovered population (R_t) indicates deactivated nodes due to competitive market. We also let the infection rate (α) and the recovery rate (β) are the activation and deactivation probabilities in the CD-RIM model, respectively.

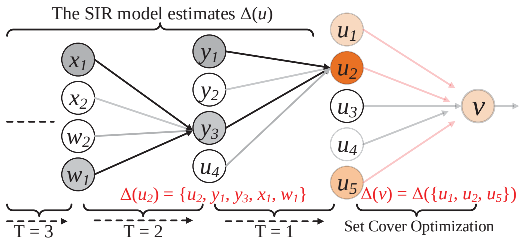


Figure 2: The basic working principle of the CD-RIM model.

B. The CD-RIM Model

The proposed CD-RIM model is shown in the Algorithm 1. In order to estimate the marginal seeding cost set $\Delta(v)$, we first, estimate the cost set $\Delta(u)$ for all $u \in n_t^{-1}(v)$ up to T hops as shown in Figure 2. At the first hop, we have,

$$S_1 = R_1 = \emptyset, \tag{1}$$

$$I_1 = \{u\} \tag{2}$$

In the next time slots or hops t $(1 < t \leq T)$, we generate new susceptible as,

$$S_t = \left\{ \bigcup_{u \in I_{t-1}} n_t^{-1}(u) \right\} \cup A_t - D_t - \Delta(u) - \Delta_R(u) \tag{3}$$

where, A_t and D_t are added and deleted nodes by link prediction algorithm [12], respectively. Here, $\Delta(u)$ contains already infected nodes and $\Delta_R(u)$ holds total recovered nodes. Then, at any hop t, the infected nodes with probability α, and recovered nodes with probability β are given by,

$$I_t = \alpha S_t I_{t-1} \tag{4}$$

$$R_t = \beta I_t \tag{5}$$

$$I_t = \alpha S_t I_{t-1} - \beta I_t \tag{6}$$

This process continues up to T hops and the value of T is determined by the influence decay function [11], [13], [14]. We get $\Delta(u)$ by collecting I_t form all time slots by (6). For example, $\Delta(u_2) = \{u_2, y_1, y_3, x_1, w_1\}$, as shown in Figure 2. Similarly, we estimate $\Delta(u)$, for all $u \in n_1^{-1}(v)$.

Now, we select the majority (as contagion threshold is 0.5 [15]) number of in-neighbors $u \in n^{-1}(v)$ by Set Cover mechanism such that the aggregated marginal cost, $\delta(v) = |\Delta(S)|$ is minimum [16] [17]. Finally, the total seeding cost is given by:

$$\delta(S) = |\Delta(S)| = \bigcup_{v \in S} \Delta(v). \tag{7}$$

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

Input: $G_1(V_1, E_1), S$
Result: $\Delta(S), \delta(S)$

$\Delta(S) := \emptyset$;
for *each seed* v *in* S **do**
    $\Delta(v) := v, S_1 = A_1 = D_1 = \emptyset$;
    while $t := 1$ *to* T **do**
        $S_t = \left\{ \bigcup_{u \in I_{t-1}} n_t^{-1}(u) \right\} \cup A_t - D_t - \Delta(u) - \Delta_R(u)$;
        Calculate I_t with rate α ; `/* Infect */`
        Calculate R_t with rate β ; `/* Recover */`
    end
    $\Delta(u) :=$ list all I_t, $\Delta_R(u) :=$ list all R_t;
    $\Delta(v) :=$ Aggregate cost by selecting majority number of $u \in n_t^{-1}(v)$ s.t. $\cup\Delta(u)$ is minimum;
    `/* Marginal Cost set by Greedy method */`
    $\Delta(S) := \Delta(S) \cup \Delta(v)$; `/* Seeding cost set */`
end
$\delta(S) = |\Delta(S)|$; `/* Final Seeding cost */`
return $\delta(S), \Delta(S)$;

C. The Complexity

The complexity of the proposed CD-RIM algorithm is,

$$C \leq \underbrace{k}_{\text{seeds}} (\underbrace{ndT}_{\text{SIR}} + \underbrace{d^2}_{\text{Set Cover}}) \approx O(kndT). \tag{8}$$

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

Proof. The Minimum Set Cover problem, which is a famous NP-Hard problem, can be reduced to the CD-RIM problem [18]. Thus, the RIM model under the CD-RIM model is also NP-Hard. □

4. Performance Evaluation

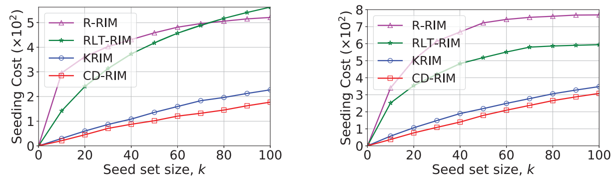
In this section, we compare the experimental results of the CD-RIM model with Random RIM (R-RIM), Randomized LT RIM (RLT-RIM), Knapsack-based RIM (KRIM) methods.

Table 1: Dataset description

Social Networks	Nodes	Edges
Facebook	$4,039$	$88,234$
Wiki-Vote	$7,115$	$103,689$

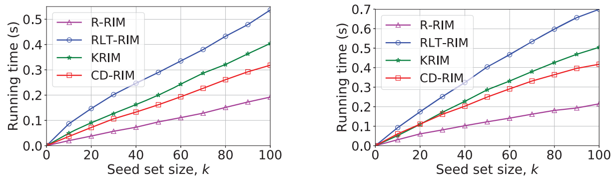
A. Dataset Collection and Experiment Setup

For the experiment, we use two real datasets such as Facebook and Wiki-Vote datasets [19], as shown in Table 1. We employ Monte Carlo (MC) technique for the experiment. The infection rate α is taken by Trivalency model [4], [20] and $\beta \in [0, \alpha)$.



(a) Facebook dataset (b) Wiki-vote dataset

Figure 3: Seeding cost for k = 1 to 100, for various datasets.



(a) Facebook dataset (b) Wiki-vote dataset

Figure 4: Running time for k = 1 to 100 for various datasets.

B. Seeding Cost

The proposed model uses the greedy technique for cost optimization, whereas; the existing R-RIM and RLT-RIM models apply stochastic process. As a result, the CD-RIM model returns significantly optimized seeding cost as compared to both the existing models, as shown in the figure. Again, the KRIM model has a slightly worse seeding cost than the proposed model as depicted in Figure 3.

C. Running Time

Again, the proposed CD-RIM algorithm exhibits better running time than that of the RLT-RIM and KRIM models, as depicted in Figure 4. Further, the running time of the proposed model is slightly higher than that of the R-RIM model, which is a simple random method. However, the R-RIM model produces the worst seeding cost.

D. Handling RIM Challenges

The CD-RIM model handles the stopping criteria the most efficiently as compared to the existing models by using influence decay function. The handling BNC and insufficient influence issues do not arise with the SIR model. The greedy Set Cover approximation technique resolves the NP-Hardness issue appropriately.

5. Conclusion

In this paper, we introduce a Susceptible-Infection-Recovered (SIR)-based Reverse Influence Maximization (RIM) model for seeding cost minimization under the Competitive market in Dynamic (CD-RIM) social networks. The proposed CD-RIM model not only produces optimized seeding cost but also addresses the RIM challenges more efficiently than the existing RIM models.

References

[1] 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.

[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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] E. Güney, "On the optimal solution of budgeted influence maximization problem in social networks," *Operational Research*, vol. 19, no. 3, pp. 817–831, 2019.

[8] Y. Liu, Z. Liu, S. Xie, and X. Li, "Time optimal profit maximization in a social network," in *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. Springer, 2019, pp. 264–272.

[9] W. Yang, J. Yuan, W. Wu, J. Ma, and D.-Z. Du, "Maximizing activity profit in social networks," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 117–126, 2019.

[10] 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.

[11] 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, no. 1, pp. 44 182–44 198, 2019.

[12] W. Liu, X. Chen, B. Jeon, L. Chen, and B. Chen, "Influence maximization on signed networks under independent cascade model," *Applied Intelligence*, vol. 49, no. 3, pp. 912–928, 2019.

[13] S. Feng, X. Chen, G. Cong, Y. Zeng, Y. M. Chee, and Y. Xiang, "Influence maximization with novelty decay in social networks." in *AAAI*, 2014, pp. 37–43.

[14] 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.

[15] A. Talukder, M. G. R. Alam, N. H. Tran, D. Niyato, G. H. Park, and C. S. Hong, "Threshold estimation models for linear threshold-based influential user mining in social networks," *IEEE Access*, vol. 7, no. 1, pp. 105 441–105 461, 2019.

[16] 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.

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

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

[20] A. Talukder and C. S. Hong, "Susceptible-infection-based cost-effective seed mining in social networks," in *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2019, pp. 1–6.