Knapsack-based Reverse Influence Maximization for Target Marketing in Social Networks*

Ashis Talukder
Dept. of Computer Science and Engineering
Kyung Hee University, South Korea
ashis@khu.ac.kr

Choong Seon Hong†
Dept. of Computer Science and Engineering
Kyung Hee University, South Korea
cshong@khu.ac.kr

ABSTRACT

With the proliferation of social networks, the Influence Maximization (IM), which identifies the influential users for target marketing, gains enormous research interest in recent years. However, most of the existing IM studies ignore the seeding cost estimation, which is the minimum number of nodes that are required to activate the seed nodes. Moreover, the existing Reverse Influence Maximization (RIM) models are incapable of resolving the challenging issues appropriately. Therefore, in this paper, we propose an efficient Knapsack-based RIM solution (KRIM) under the Linear Threshold model, which returns the optimized seeding cost. The experimental results show that our model outperforms the existing RIM models.

CCS CONCEPTS

• **Information systems → Social networking sites**; *Collaborative and social computing systems and tools*; • **Human-centered computing → Social media**; • **Networks →** *Network simulations*;

KEYWORDS

influence maximization, reverse influence maximization, seeding cost, social network, target marketing.

ACM Reference Format:
Ashis Talukder and Choong Seon Hong. 2019. Knapsack-based Reverse Influence Maximization for Target Marketing in Social Networks. In *The 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19), April 8–12, 2019, Limassol, Cyprus.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3297280.3297621

1 INTRODUCTION

Social Network has been turned out to be the most convenient means of communication at present and hence, has become a very powerful medium for viral and target marketing. The Influence Maximization (IM) is such a technique that maximizes the profit, which is the maximum number of individuals that can be activated by the seed users when they are activated initially. The Linear Threshold (LT) and Independent Cascade (IC) models are considered to be the pioneering works [4], and after that, many models are proposed. Leskovec *et al.* [5] propose a heuristic cost-effective lazy forward (CELF) method for outbreak detection. This work is extended by Goyal *et al.* [2] as CELF++, which outperforms most of the greedy models. Further, Bhagat *et al.* [1] propose a profit maximization scheme by maximizing the product adaptation. However, most of the IM models either assume the seed nodes are initially activated or give a free sample product and hence, ignore the seeding cost, which is the minimum number of nodes that are required to activate all the seed nodes. Thus, Reverse Influence Maximization (RIM) models [6, 8] are introduced to find the seeding cost. However, the existing RIM models fail to exhibit the expected efficiency.

Therefore, in this paper, we propose a Knapsack-based Reverse Influence Maximization (KRIM) model, which estimates the seeding cost of target marketing. The KRIM model uses the Linear Threshold (LT) model in reverse order for the node activation process and the greedy Knapsack technique to optimize the cost. The key contributions of this paper include:

(1) The proposed KRIM model gives optimized seeding cost while addressing the RIM-challenges efficiently.
(2) We update the LT model to apply in reverse order for cost calculation and provide the approximation ratio of the greedy optimization model as well.
(3) Finally, the simulation of the KRIM model on real dataset shows that the proposed model outperforms the existing RIM models.

2 PROBLEM FORMULATION

Consider a social network represented by a directed graph $G(V, E)$, where V denotes the set of social network users, and E represents the social relations among them. The set $n(v)$ and $n^{-1}(v)$ denote the out-neighbors and in-neighbors sets of a node v, respectively, and w_{uv} is the influence weight the node u on the node v.

In the LT model, a node v is activated if the combined influence coming from all the active in-neighbors of the node v is no less than the node's threshold value, θ_v [4], that is,

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

where, x_u indicates whether an in-neighbor u is active ($x_u = 1$) or not ($x_u = 0$).

*This work was supported by the Institute for Information & the 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), and also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B01015320).
†Professor Dr. Hong is the corresponding author.

SAC '19, April 8–12, 2019, Limassol, Cyprus

ACM ISBN 978-1-4503-5933-7/19/04.
https://doi.org/10.1145/3297280.3297621

For a given seed set S ($k = |S|$), we aim at finding the minimum number of nodes $\gamma(S)$, that are required to activate all seed nodes in S. At first, the optimized marginal seeding cost set, $\Gamma(v)$ is determined for all seed nodes, $v \in S$, up to T hops, repeatedly for each activated node as,

$$minimize \sum_{u \in n^{-1}(v)} x_u \tag{2}$$

$$s.t. \sum_{u \in n^{-1}(v)} w_{uv} x_u \geq \theta_v, \tag{3}$$

$$x_u \in \{0, 1\}, \tag{4}$$

$$w_{uv} \in (0, 1]. \tag{5}$$

Finally, the optimal seeding cost set, $\Gamma(S)$ and seeding cost, $\gamma(S)$ of the whole RIM problem are given by:

$$\Gamma(S) = \bigcup_{v \in S} \Gamma(v) \tag{6}$$

$$\gamma(S) = |\Gamma(S)| \tag{7}$$

*Definition 2.1 (**RIM Problem**).* Given a social network $G(V, E)$ and a seed set S, the RIM problem is defined by finding the minimum number of nodes, $\gamma(S)$ that must be activated in order to activate all the seed nodes in S.

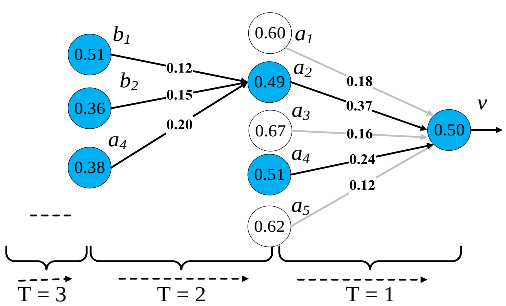


Figure 1: The working principles of the KRIM model.

3 THE PROPOSED KRIM MODEL

In this section, we state the general working principles of the proposed Knapsack-based Reverse Influence Maximization (KRIM) model as depicted in Figure 1.

3.1 Meeting the Challenges

The first challenge is to set up an effective stopping criterion, which is up to two hops ($T = 2$) of seed nodes in the existing models [6, 8]. However, the proposed model sets the terminating condition effectively by influence decay concept. At each hop, the minimum w_{uv} (e.g., in Figure 1, 0.24 is selected at $T = 1$, and 0.12 at $T = 2$) is selected and are multiplied. When the product reaches some negligible value (say, $< 10^{-6}$), the KRIM algorithm stops. The next challenge is to handle three Basic Network Components (BNC). The existing RIM models consider three BNCs such as Case A (seed node with no in-neighbors), Case B (seed node with one hop in-neighbors), and Case C (seed node with multiple hop in-neighbors). However, Case C is the combination of Case A and Case B, and thus, the KRIM considers only first two cases.

The insufficient influence is the next challenge to address. In Figure 1, all the in-neighbors of a_2 have not enough aggregated influence to activate it, and this situation is called insufficient influence. The issue is mostly intractable, and therefore, the node is assigned a lower threshold to mitigate the insufficient influence effect. Even if the situation arises, the in-degree of the node v is considered to be the cost, $\gamma(v) = \left|n^{-1}(v)\right|$. The last challenge is to handle the NP-Hardness of the problem.

Theorem 3.1. *The RIM problem under KRIM model is NP-Hard.*

Proof. We consider the Knapsack problem stated in [8], and compare it with the KRIM problem stated in equations (2) - (7). Let us consider the Knapsack size to be the threshold value θ_v in the RIM problem, the item weights in the Knapsack problem to be influence weights w_{uv}, the profit of an item is 1 if it is selected, otherwise 0, and finally, replace the objective function of the RIM problem stated in equation (2) as:

$$maximize \quad - \sum_{u \in n^{-1}(v)} x_u \tag{8}$$

Thus, the Knapsack problem, which is a well-known NP-Hard problem [3], is reduced to the RIM problem and hence, the RIM problem under KRIM model is also an NP-Hard problem. □

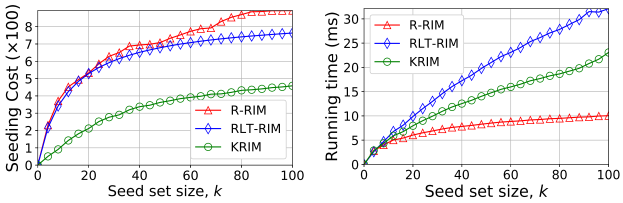
Algorithm 1: The KRIM Algorithm

Input: $G(V, E), S$
Result: $\gamma(S), \Gamma(S)$
$\Gamma(S) := \emptyset$; /* Initialization */
for *each* $v \in \mathcal{S}$ **do**
  $\Gamma(v) := \emptyset$; /* Initialization */
  for $t := 1$ *to* T **do**
    Calculate the cost by LT model applied in reverse order, considering trivial and general cases by (9);
    Calculate the cost by LT model applied in reverse order, considering trivial and general cases by (10);
    Combine both in $\Gamma(v)$;
  end
  $\Gamma(S) := \Gamma(S) \cup \Gamma(v)$; /* Final seeding cost set */
end
$\gamma(S) := |\Gamma(\mathcal{S})|$; /* Final seeding cost */
return $\gamma(S)$;

3.2 The KRIM Model

Here, we propose the Knapsack-based RIM (KRIM) model, which employs the Linear Threshold model in reverse order and selects the in-neighbor nodes with higher influence weight greedily. The LT model which is generally applied forwardly in IM problem to determine which nodes are activated by a seed node v. However, we modify the LT model to employ in a retrograde manner to determine which nodes are required to activate the seed node v.

The algorithm iterates up to T hops of in-neighbors of v to estimate the optimized marginal seeding cost set, $\Gamma(v)$. The in-neighbor nodes which are previously activated are considered first to activate a node v at any hop $t \in T$, and if $\Gamma(S)$ is the set of activated nodes up to now, then, the set of already activated nodes is given by $n^{-1}(v) \cap \Gamma(S)$. Thereafter, the inactive in-neighbors $(n^{-1}(v) \setminus \Gamma(S))$ are considered to activate a node v, if necessary. Moreover, the



(a) Seeding cost. **(b) Running time.**

Figure 2: Performance evaluation of the KRIM model.

Knapsack method is used to select a node u with maximum w_{uv} in order to reduce the cost. In case of already activated and newly inactive in-neighbors, a node u is selected by equation (9) and (10), respectively.

$$u = \underset{u \in \{n^{-1}(v) \cap \Gamma(S)\}}{arg\ max} \quad w_{uv} \tag{9}$$

$$u = \underset{u \in \{n^{-1}(v) \setminus \Gamma(S)\}}{arg\ max} \quad w_{uv} \tag{10}$$

Every time a node u is selected, the influence weight of u is aggregated, and the aggregated influence is compared with the node's threshold value, θ_v, whether it is activated or not. The process continues for each of the activated nodes in the next hop $t + 1$, and so on. The whole process is illustrated in Figure 1 and Algorithm 1.

3.3 The Performance Bound

Here, we describe the approximation ratio of the greedy Knapsack technique along with the complexity of the proposed KRIM model.

THEOREM 3.2. *The KRIM algorithm is a 2-* ***approximation*** *algorithm, that is,*

$$\gamma \leq 2\gamma^* \tag{11}$$

PROOF. In Theorem 3.1, we have proved that the Knapsack problem could be reduced to the RIM problem under KRIM algorithm. Therefore, the estimated cost γ is bounded by the optimal cost γ^*, *i.e.*, $\gamma^* \leq \gamma$, and can be maximum $2\gamma^*$ [3], and hence, $\gamma \leq 2\gamma^*$. □

Complexity: If d is the average in-degree in the graph G, the complexity of the proposed KRIM model is $O(kTd^2)$.

4 PERFORMANCE EVALUATION

The Monte Carlo simulation is used to evaluate the performance of the KRIM model using the real dataset of the Epinions[1] social network (see Table 1). The *degree centrality* [4] technique is employed to compute the w_{uv}, and the Heuristic Individual (HI) model [7] to generate θ_v. The estimated result is compared with that of the existing R-RIM and RLT-RIM models [6, 8].

Table 1: Dataset Description

Social Networks	# of Nodes	# of Edges
Epinions	75, 879	508, 837

Figure 2(a) depicts the seeding cost estimated by different models for the Epinions dataset. The KRIM returns the optimized seeding cost which is, on an average, 2 − 2.5 times lower than that of R-RIM and RLT-RIM models. This remarkable performance is due to the use of greedy optimization, and therefore, our proposed model outperforms the baseline and existing models.

The proposed algorithm exhibits a quite adequate running time as shown in Figure 2(b). The running time of our model is sandwiched between that of existing two models. The KRIM model is about 20 − 50% faster than the RLT-RIM model; however, it is slightly slower than the R-RIM model.

[1] https://snap.stanford.edu/data/soc-Epinions1.html

5 CONCLUSIONS

In this research, we propose a Knapsack-based Reverse Influence Maximization (KRIM) model to estimate the optimized seeding cost for target marketing in social networks. The proposed model employs the Linear Threshold model in retrograde order to estimate the seeding cost and the greedy Knapsack technique to optimize the cost. The performance evaluation using real dataset popular social network shows that the KRIM model outperforms the existing RIM models as well as resolves the RIM challenges more efficiently.

ACKNOWLEDGMENTS

The authors would like to thank Md. Golam R. Alam, Sarder F. Abedin, Anupam K. Bairagi, Md. Shirajum Munir, Dr. Thant Z. Oo, and Dr. Nguyen H. Tran for insightful comments on the idea.

REFERENCES

[1] Smriti Bhagat, Amit Goyal, and Laks VS Lakshmanan. 2012. Maximizing product adoption in social networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 603–612.

[2] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*. ACM, 47–48.

[3] Dorit S Hochbaum. 1996. *Approximation algorithms for NP-hard problems*. PWS Publishing Co.

[4] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. 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, 137–146.

[5] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 420–429.

[6] Ashis Talukder, Md Golam Rabiul Alam, Anupam Kumar Bairagi, Sarder Fakhrul Abedin, Md Abu Layek, Hoang T Nguyen, and Choong Seon Hong. 2017. An approach of cost optimized influence maximization in social networks. In *Network Operations and Management Symposium (APNOMS), 2017 19th Asia-Pacific*. IEEE, 354–357.

[7] Ashis Talukder, Md Golam Rabiul Alam, Anupam Kumar Bairagi, Sarder Fakhrul Abedin, Hoang T Nguyen, and Choong Seon Hong. 2016. Threshold Estimation Models for Influence Maximization in Social Network. In *Korean Institute of Information Scientists and Engineers (KIISE)*. KIISE, 888–890.

[8] Ashis Talukder, Md Golam Rabiul Alam, Nguyen H Tran, and Choong Seon Hong. 2018. A cost optimized reverse influence maximization in social networks. In *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 1–9.

Ashis Talukder is currently persuading his Ph.D. degree in the Department of CSE, Kyung Hee University, under the supervision of Professor Dr. Choong Seon Hong. He is an Assistant Professor of the Department of MIS, University of Dhaka, Bangladesh. His research interests include social networks, influence maximization, network optimization, and data mining. He is a member of IEEE, IEEE ComSoc, KIISE, AIS, and ISOC.

Choong Seon Hong received his Ph.D. degree from Keio University in 1997. He has been working as a Professor of the Department of CSE, Kyung Hee University since September 1999. His research interests include the future Internet, ambient intelligent technology, wireless networks, network security, and network management. He is a member of ACM, IEEE, IEICE, IPSJ, KIISE, KICS and KIPS.