A Susceptible-Infection-based Greedy Model for Mining Cost-effective Seeds in Social Networks

Ashis Talukder, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, Yongin 17104, South Korea.
Email: {ashis, cshong}@khu.ac.kr

Abstract

Nowadays, viral marketing has gained huge research interest in social network research with the growth of remarkable social networks. Identifying or mining cost-effective influential users in social networks is very important in viral marketing. The Influence Maximization (IM) provides profit effective influential seed nodes, and on the other hand, the Reverse Influence Maximization (RIM) gives cost-effective influential seed nodes in the social network. However, most of the IM methods have a drawback that they do not consider the seeding cost estimation, whereas, most of the RIM techniques fails to handle the challenging issues properly. Thus, we propose a Susceptible-Infection (SI)-based Greedy RIM (SIG-RIM) model for finding cost-effective influential nodes in the social networks. The proposed SIG-RIM model handles the RIM issues efficiently and provides cost-optimization for viral marketing. The evaluation on real dataset reveals that the proposed SIG-RIM model performs better than the existing models in terms of the estimated cost and running time.

1. Introduction

Nowadays, social networks are considered to be the most attractive medium for marketing epically, for viral and target marketing [1]. It is very important to mine the influential users in viral or target marketing [2]. Influence Maximization (IM) is such a mining tool which identifies a small set of seed users who can maximize the spread of influence or profit in the network [3], [4], [5], [6]. The profit is the maximum number of nodes that can be activated by the active seed users [3]. The classical Linear Threshold (LT) and Independent Cascade (IC) models were proposed by Kempe *et al.* [3] as the seminal work in this domain. After that, many researchers have proposed different variations of the classical LT and IC models as well as new models for influence and profit maximization. Chen *et al.* [7] propose a heuristic model, named as LT-based directed acyclic graphs (LDAG) model, which exhibits better performance than most of the greedy models. Leskovec *et al.* [8] propose a cost-effective contagion-based model for outbreak detection. Bhagat *et al.* [9] propose a product adoption-based profit maximization model. However, none of the above models has considered seed activation cost or seeding cost which is given by the minimum number of nodes that must be activated in order to activate initial seed nodes [2], [5].

Thereafter, Reverse Influence Maximization (RIM) models are proposed to estimate the seeding cost [2], [4], [10]. However, the existing Random RIM (R-RIM) and Randomized Linear Threshold RIM (RLT-RIM) models are incapable of handling the RIM challenges such as setting stopping condition, handling network components, and insufficient influence [2].

Therefore, in this paper, we propose a Susceptible-Infection (SI)-based Greedy RIM solution (SIG-RIM) to mine the minimum cost seed users in social networks. The SIG-RIM model estimates the marginal seeding cost by using Susceptible-Infection (SI) model and a greedy approach to minimize the marginal seeding cost. The key contributions of this paper are as follows:

1) We propose an SI-based Greedy RIM approach (SIG-RIM) for minimizing the seeding cost.
2) The use of the SI and the greedy techniques resolves all the RIM issues efficiently. The greedy optimization contributes to a lower seeding cost.
3) Finally, the simulation results of the proposed model on the dataset of a real social network show that the prosed model outperforms the existing models.

2. System Model

Let us consider a social network given by a graph $G(V, E)$, where, a node $v \in V$, indicates a social network user and an edge $(u, v) \in E$ means that the users u and v are friend in the social network. Let the $n^{-1}(v)$ and $n(v)$ are the in-neighbor and out-neighbor sets of a node v. For a given seed set S of size k, we will have to find the optimized seeding cost set $\Lambda(S)$, and associated seeding cost $\lambda(S) = \left|\Lambda(S)\right|$.

Definition 1 (**The RIM Problem**). Given a social network $G(V, E)$ and a seed set S of size k, the RIM problem is to minimize the viral marketing cost $\lambda(S)$, defined by the minimum number of nodes that must be activated in order to activate all the seed nodes in S. □

3. The Proposed SIG-RIM Model

The working principle of the proposed Susceptible-Infection-based Greedy RIM (SIG-RIM) model is illustrated in Figure 1, and Figure 2.

A. The SI Model

As shown in Figure 1, the SI model includes a set S_t that indicates a susceptible set. The members of this set are the possible candidates for infection if they come in contact with any infected node [11]. The model also includes an infected set I_t to represent members of the susceptible set that are exposed to some disease and got infected by the disease with an infection rate α [11]. At each hop, we consider the inactive in-neighbors are susceptible candidates and activated in-neighbors as infected population.

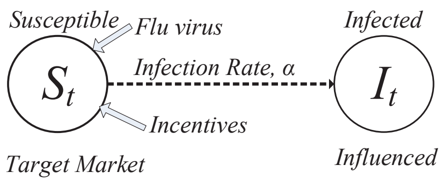

Figure 1: The SIG-RIM model using the SI method.

B. Resolving Challenges

In the SIG-RIM model, the stopping criterion is efficiently set by using the influence decay function stated in [5]. Considering different cases and insufficient influence are unnecessary due to the use of the SI model instead of the LT model. Finally, the greedy optimization resolves the NP-Hardness issue of the RIM problem efficiently.

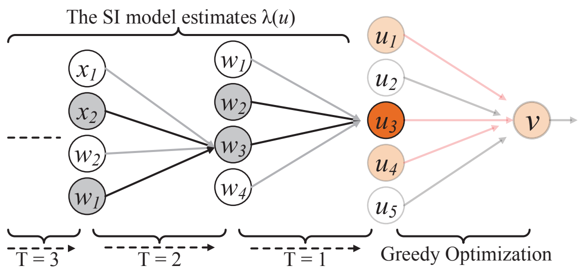

Figure 2: The basic working principle of the SIG-RIM model.

C. The SIG-RIM Model

In order to estimate the marginal seeding cost $\lambda(v)$, we first, estimate the cost $\lambda(u)$ for all $u \in n^{-1}(v)$ up to T hops as shown in Figure 2. At the first hop, we have,

$$S_1 = n^{-1}(u), \tag{1}$$

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

Now, the members of the S_1 are exposed to infection (i.e. some incentives are offered), and they are infected with a rate α. The infected population at any hop t $(1 \leq t \leq T)$ is given by:

$$I_t = \alpha S_{t-1} \tag{3}$$

Here, α is the infection rate. The new susceptible set S_t is the inactive in-neighbors of all the nodes of the set I_t. The number of hops (T) is determined by the influence decay function [5], [12], [13]. Then, we select the majority number of in-neighbors $u \in n^{-1}(v)$ by greedy optimization such that the aggregated (marginal) cost, $\lambda(v) = \left|\Lambda(S)\right|$ is minimum [14] [15]. Finally, the total seeding cost is given by:

$$\lambda(S) = \left|\Lambda(S)\right| = \bigcup_{v \in S} \Lambda(v), \tag{4}$$

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

Input: $G(V, E), S$
Result: $\Lambda(S), \lambda(S)$

1 $\Lambda(S) := \emptyset$;
2 **for** *each seed* v *in* S **do**
3     $\Lambda(v) := \emptyset$;
4     $S_t = n^{-1}(v)$;   /* Initialization */
5   **while** $u \in S_t$ **do**
6   **while** $t := 1$ *to* T **do**
7   Calculate I_t with rate α ;   /* Infect */
8   Update new S_t for next hop;
              /* Susceptible */
9   **end**
10   **end**
11     $\Lambda(u) :=$ list all I_t;
12     $\Lambda(v) :=$ Aggregate cost by selecting majority number of $u \in n^{-1}(v)$ s.t. $\cup\Lambda(u)$ is minimum;
       /* Marginal Cost set by Greedy method */
13     $\Lambda(S) := \Lambda(S) \cup \Lambda(v)$;   /* Seeding cost set */
14 **end**
15 $\lambda(S) = \left|\Lambda(S)\right|$;   /* Final Seeding cost */
16 *return* $\lambda(\mathcal{S}), \Lambda(\mathcal{S})$;

D. The SIG-RIM Algorithm

The proposed SIG-RIM model is expressed in the Algorithm 1. The susceptible and infected population are computed hop by hop in lines $6 - 9$. All the cost, $\lambda(u)$ are computed in line 11. The majority number of in-neighbors are selected in line 12 using the greedy method. The line 13 and 15 estimate the final seeding cost set and seeding cost, respectively.

E. The Complexity of the SIG-RIM Model

The complexity of the proposed SIG-RIM algorithm is

$$C = k(d^2 + d)T \leq O(kTd^2). \tag{5}$$

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

Proof. The Knapsack problem, which is a famous NP-Hard problem, can be converted into the RIM problem under SIG-RIM model [5]. Thus, the RIM model under SIG-RIM model is NP-Hard. □

4. Performance Evaluation

Here, we evaluate the performance of the proposed SIG-RIM model with a comparative analysis with two existing models using Facebook[1] dataset.

Table I: Dataset Summary

Social Network	Nodes	Edges
Facebook	4,039	88,234

[1] https://snap.stanford.edu/data/egonets-Facebook.html

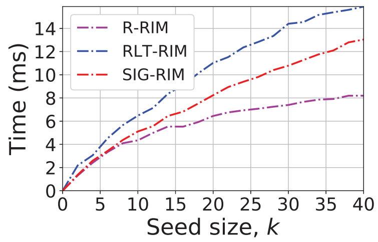

Figure 4: The running time for different values of k.

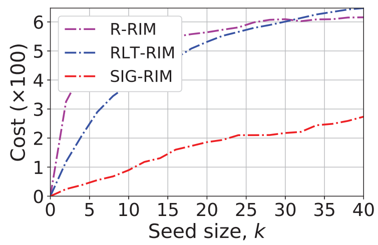

Figure 3: The Seeding cost for different values of k.

A. Data Collection and Simulation Setup

For simulation, the real dataset of Facebook is collected from SNAP dataset collections [16] as shown in Table I. We employ Monte Carlo (MC) technique for simulation. The infection rate α is taken by Trivalency model [10].

B. Performance Analysis

Here, the performance analysis of the proposed model is stated with a comparative study with the existing R-RIM and RLT-RIM models.

1) Seeding Cost: The estimated seeding costs of the proposed and existing models are depicted in Figure 3. The proposed model uses the greedy technique for cost optimization whereas; the existing models apply stochastic process. As a result, the proposed SIG-RIM model returns significantly optimized seeding cost as compared to both the existing models as shown in the figure.

2) Running Time: On the other hand, the proposed SIG-RIM algorithm exhibits better running time than that of the RLT-RIM model as illustrated in Figure 4. However, the running of the SIG-RIM 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.

3) Handling RIM Challenges: The SIG-RIM model expertly handles the stopping criteria as compared to the existing models. The issues like the three basic network components and insufficient influence do not arise with the SI model. The greedy approximation technique properly addresses the NP-Hardness issue.

5. Conclusion

In this paper, we introduce a Susceptible-Infection-based Greedy Reverse Influence Maximization (SIG-RIM) model to mine the cost-effective users for viral marketing in social networks. The proposed algorithm not only exhibits better seeding cost but also handles the RIM challenges more efficiently than the existing RIM models.

Acknowledgment

This research was partially 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) and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B01015320).

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

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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

[14] 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.

[15] 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.

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