Epidemiological Reverse Influence Maximization in Social Networks with Negative Influencing

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

Abstract

Influence Maximization (IM)-based profit maximization estimates a seed set of users that maximizes the profit in social networks. On the other hand, Reverse Influence Maximization (RIM) finds the seeding cost of the same process. The profit is measured by the maximum number of users that can be activated by the seed users and the seeding cost is given by the minimum number of users which should be activated in order to activate the seed or target users. The major drawback of most of the existing IM models is that the seed users are assumed to be activated initially. However, seed users may also be influenced by some other icon users they follow. Moreover, none of the state-of-the-art models has considered negative influencing. Thus, in this paper, we propose an epidemic model-based RIM model with negative influencing (EN-RIM) for seeding cost optimization. The simulation shows that the EN-RIM model outperforms many existing models.

1. Introduction

Influence Maximization (IM) has been a potential viral marketing model in social networks research for many years. The IM aims at finding a small set of seed users that can maximize the influence or profit in the network. However, seeding cost of viral marketing is estimated by the Reverse Influence maximization (RIM) in which influence is propagated in a reverse manner. The seeding cost is the minimum number of individuals that might be activated so that all the target users are activated. The IM and RIM together can provide the *cost-benefit-analysis* and thus, both have research and business importance.

Kempe *et al.* [1] propose Linear Threshold (LT) model and Independent Cascade (IC) models as a pioneering work in this area of research. Leskovec *et al.* [2] propose the popular Cost-Effective Lazy Forward (CELF) model for outbreak detection and the model outperforms many existing greedy models. Gardner *et al.* [2] employ *Susceptible-Exposed-Infected-Recovered (SEIR)* model for viral marketing. However, none of the above addresses the finding seeding cost for viral marketing.

Talukder *et al.* [3] introduce RIM model for the first time and an extension is proposed in [4]. However, none of the existing RIM models can handle the RIM challenges ($e.g.$ stopping criteria, Basic Network Components, insufficient influence) properly. Moreover, none of the above IM and RIM models considers negative influencing. Thus, We propose an epidemic-based RIM model with negative influencing (EN-RIM) for profit maximization in social networks.

2. The Epidemiological Model

The Epidemiological model is based on how infectious diseases spread among the individuals and has been applied in viral marketing in social networks for many recent years [2], [5], [6], [7], [8], [9]. The *Susceptible-Exposed-Infected-Recovered (SEIR)* model is the most widely used epidemic model in which the population is partitioned into different groups. The groups are illustrated in Fig.1(a) and are described as follows [8]:

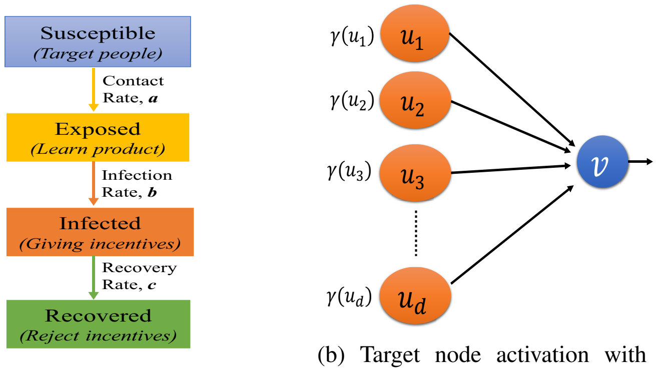
1) *Susceptible (S):* This group of individuals are not yet infected but are vulnerable to the disease. They are the population for profit maximization.
2) *Exposed (E):* This set includes the subset of S which are contacted to the disease but not yet infected. They can be considered to be the individuals that are exposed to some advertisement or personal influencing.
3) *Infected (I):* This group of individuals has got the disease and any susceptible candidates may be infected if they are exposed to this class. In profit maximization application, an individual is infected means a product is adopted and in case of cost optimization, a incentive is given.
4) *Recovered (R):* This class of individuals had the disease previously but they have recovered the disease. In profit maximization application, these users decide to adopt the product but finally, they reject to adopt it for negative influencing.

3. System Model

We design an epidemic model to minimize seeding cost using RIM with Negative Influencing (EN-RIM). The EN-RIM model is applied for viral marketing in social networks which is represented by a directed graph $G(V, E)$ with an individual set V and their social interaction set E. We introduce the out-neighbor set $n(v)$ and in-neighbor set $n^{-1}(v)$ for each node v.

A. The Proposed EN-RIM Model

We apply epidemiological SEIR model at each hop of neighbors of a target node v. Our main goal is to minimize the cost so that the profit is maximized and if an individual is infected, it is considered to provide some cost incentives in exchange of influencing the seed users and a unit of cost is incurred. An individual is recovered means she changes her decision due to negative influencing $e.g.$ for a higher



(a) State transmission.

(b) Target node activation with optimized cost [11].

Figure 1: The SEIR model for cost optimization.

price than her evaluation. The recovered individual rejects incentives and does not take part in viral marketing. The EN-RIM model first, finds the seeding cost, $\gamma(u)$, of all the in-neighbors of v and then computes the optimized marginal seeding cost $\gamma(v)$. For a node $u \in n^{-1}(v)$, we assume $N_i = |n^{-1}(u)|$, no individual has been recovered yet *i.e.* $R = 0$ initially. Then, we start the process more assuming that the node u is activated or infected *i.e* $I = 1$. In every step, all the in-neighbors of u are considered to be susceptible and hence, the contact rate, $a = 1$. Thus, the susceptible and exposed individuals are given by:

$$S(t) = -\left|n^{-1}(u)\right|, \tag{1}$$

$$E(t) = aS(t) = S(t), \tag{2}$$

where, $t = 1$ indicates the infection calculation time quantum for the first hop of node u and so on. Next, we calculate the number of infected and recovered individuals at the end of time slot t by the following equations:

$$I(t) = bSI - cI, \tag{3}$$

$$R(t) = cI, \tag{4}$$

where, b and c are infection rate and recovery rate respectively. Let us now consider I_t is the set of infected individuals after the time t and the seeding cost $\gamma(u)$ of u is calculated by:

$$\gamma(u) = \sum_t I_t \tag{5}$$

Next, the target individual v is infected with epidemic threshold $\theta = 0.5$ [10] and the marginal cost, $\gamma(v)$ is computed by selecting majority number of in-neighbors of v such that $\sum_t \gamma(u)$ is minimum as depicted in Fig.1(b). Finally, the optimized seeding cost, $\gamma(S')$ is given by:

$$\gamma(S') = \left|\Gamma(S')\right| = \left|\cup_{u \in S'} \Gamma(u)\right|. \tag{6}$$

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

Algorithm 1: EN-RIM Model

```
Input: G(V, E), S'
Result: Γ(S), γ(S)
Γ(S') = ∅;
for v ∈ S' do
    I = 1;
    b, c = Learning Rates(G, v);  /* Find rates */
    while u ∈ n^{-1}(v) do
        Γ(u) = u;
        S_0 = -|n^{-1}(v)|;          /* Equation (1) */
        I_0 = bS.1;                  /* v is infected */
        R_0 = cI_0;                  /* Equation (4) */
        I_0 = I_0 - R_0;             /* Equation (3) */
        t = 1
        InsertQ(I_0 nodes from n^{-1}(v));
        while Q ≠ ∅ do
            w = DeleteQ();
            Γ(u) = Γ(u) ∪ {w};
            S_t = -|n^{-1}(u)|;      /* Equation (1) */
            I_t = bS_t I_{t-1};
            R_t = cI_t;              /* Equation (4) */
            I_t = I_t - R_t;         /* Equation (3) */
            InsertQ(I_t nodes from n^{-1}(v));
            t = t + 1;
        end
        Γ(v) = Select majority number of
          in-neighbors of v with minimum cost;
    end
    Γ(S') = Γ(S') ∪ Γ(v);
end
γ(S') = |Γ(S')|;                     /* Final seeding cost */
return γ(S');
```

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

Proof. The IC model can be reduced to the RIM problem defined in the equations (1) to (6). The IC model is NP-Hard [1] and therefore, RIM problem is NP-Hard as well. □

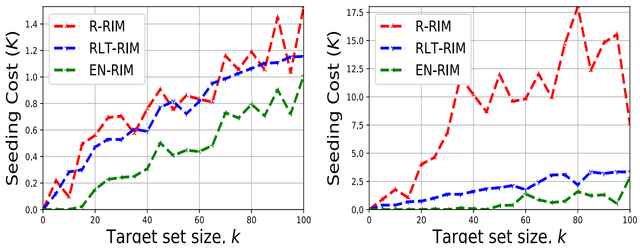
B. Terminating Condition

The calculation continues to hops until an infected node does not have any in-neighbors *i.e.* there is no susceptible candidate. However, we can limit the value of t by employing $b < c$, which indicates that the disease dies out after some finite time. This implements the fact that influence reduces with distance.

C. Tuning Parameters: a, b, and c

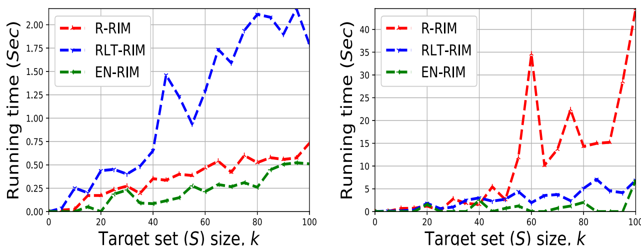
In our application, all the susceptible individuals are exposed to the infected individuals and hence $a = 1$.

We employ a heuristic technique to learn the infection rate, b and recovery rate, c. For each node in the susceptible set, a coin is tossed with a probability taken from *Tri-valency model* [12] and heads are counted. This process continues



(a) Facebook data. (b) Twitter data.

Figure 2: The seeding cost for different k values.



(a) Facebook data. (b) Twitter data.

Figure 3: The running time (in sec) for different k values.

for a random number of times for in-neighbors at different hop of v and the expected value is taken as infection rate. Similarly, we calculate the recovery rate. The line 4 in the Alg. 1 estimates such b and c values.

D. The EN-RIM Algorithm

The EN-RIM algorithm, as stated in the Alg. 1, tunes the value of a and b in line 4. Then, it finds the optimal marginal seeding cost set $\Gamma(v)$ in lines 5-23 and finally, the seeding cost $\gamma(S')$ is computed in line 25.

The training phase has complexity $O(ktd)$ and the marginal cost calculation requires $O(k)$ time and hence the complexity of the whole algorithm is given by $O(k(ktd + k)) \approx O(k^2td))$, where d = maximum degree in G. The running time can be improved by reducing the value of k in the training phase, however, it results in *running time-training quality* trade-off.

4. Performance Evaluation

We evaluate the performance of the proposed EN-RIM model by using Python programs running on two datasets: Facebook[1] and Twitter[2], as stated in Table I.

Table I: Dataset description

Networks	Nodes	Edges
ego-Facebook	4,039	88,234
ego-Twitter	81,306	1,768,149

The seeding cost of the proposed EN-RIM model is better than that of both R-RIM and RLT-RIM models as shown in the Fig. 2. There are some fluctuations available in the result due to the presence of randomness in the algorithms.

Moreover, Fig. 3 depicts that the running time is also better than that of existing models. As we employed the feature of influence decay with distance, the value of k becomes smaller and this improves the running time. However, running time can be reduced by compromising the training quality of parameters. In our simulation, we balanced the trade-off between running time and learning quality of parameters.

5. Conclusion

In this paper, we propose an epidemic model for viral marketing where negative influence is considered. The proposed EN-RIM model employs SEIR model in reverse order to find the optimized seeding cost where recovery feature is used as negative influencing. The simulation shows that the proposed EN-RIM exhibits better seeding cost and improved running time as well, as compared to that of existing R-RIM and RLT-RIM models.

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

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2017R1A2A2A05000995).
***Dr. CS Hong** is the corresponding author.

References

[1] 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.

[2] 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.

[3] 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, pp. 354–357.

[4] A. Talukder, M. G. R. Alam, H. T. Nguyen, 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.

[5] H. S. Rodrigues, "Application of sir epidemiological model: new trends," *arXiv preprint arXiv:1611.02565*, 2016.

[6] J. Woo, J. Son, and H. Chen, "An sir model for violent topic diffusion in social media," in *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. IEEE, 2011, pp. 15–19.

[7] H. S. Rodrigues and M. J. Fonseca, "Viral marketing as epidemiological model," *arXiv preprint arXiv:1507.06986*, 2015.

[8] J. T. Gardner, K. Sohn, J. Y. Seo, and J. L. Weaver, "A sensitivity analysis of an epidemiological model of viral marketing: when viral marketing efforts fall flat," *Journal of Marketing Development and Competitiveness*, vol. 7, no. 4, p. 25, 2013.

[9] D. J. Watts, J. Peretti, and M. Frumin, *Viral marketing for the real world*. Harvard Business School Pub., 2007.

[10] 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," *Korean Institute of Information Scientists and Engineers (KIISE)*, pp. 888–890, 2016.

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

[12] A. Arora, S. Galhotra, and S. Ranu, "Debunking the myths of influence maximization: An in-depth benchmarking study," in *Proceedings of the 2017 ACM International Conference on Management of Data*. ACM, 2017, pp. 651–666.