ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 45, No. 11, pp. 1203-1209, 2018. 11
https://doi.org/10.5626/JOK.2018.45.11.1203

소셜 네트워크 상에서 역방향 경로 활성화 기법 기반 역방향 영향력 극대화 연구

(Reverse Path Activation-based Reverse Influence Maximization in Social Networks)

아시스 탈룩더 † 아누팜 쿠마르 바이라기 † 김 도 현 † 홍 충 선 ††
(Ashis Talukder) (Anupam Kumar Bairagi) (Do Hyeon Kim) (Choong Seon Hong)

요 약 영향력 극대화(Influence Maximizaion)기법은 소셜 네트워크에서 바이럴 마케팅에 대한 영향력 있는 사용자를 찾는 것을 다루지만, 역방향 영향력 극대화(Reverse Influence Maximizaion) 기법은 영향 비용 극대화 영역의 새로운 연구 방향으로 기회비용을 처리한다. 영향력 극대화 기법은 이러한 시드 노드를 대상으로 네트워크에서 영향이 극대화되는 방식으로 작은 시드 집합을 추정한다. 일반적으로 시드 노드는 영향력 극대화 문제에서 처음에 활성화되는 노드로 가정한다. 그러나, 우리는 활성화 된 노드가 추후 활성화 될 외부 노드에 영향을 주는 것과 유사한 방식으로 다른 노드의 영향을 받아야한다고 주장한다. 역방향 영향력 극대화 문제는 모든 시드 노드를 활성화하기 위해 활성화해야하는 최소 노드 수로 정의되는 시드 비용을 찾는 문제이다. 이에 본 논문에서는 최소 기회 비용을 찾기 위해 Active Reverse Path 기반 역방향 영향력 극대화 모델을 제안한다. 본 모델은 Voting 모델과 전통적인 Independent Cascade 모델을 기반으로 한다. 아울러 잘 알려진 세 가지 소셜 네트워크의 실제 데이터 셋을 활용하여 모델을 시뮬레이션 하였으며, 그 결과 제안하는 모델이 기존 역방향 영향력 극대화 모델보다 우수한 성능을 보였다.
키워드: 영향력 극대화, 역방향 영향력 극대화, 기회 비용, 재배 비용

Abstract Influence Maximization (IM) deals with finding influential users for viral marketing in social networks, whereas Reverse Influence Maximization (RIM), a new research direction in the influence-maximization domain, deals with seeding cost, also known as opportunity cost. The IM estimates a small seed set in such a way that by targeting those seed nodes, the influence is maximized in the network. Generally, the seed nodes are assumed to be activated initially in the IM problem. However, we argue that seed nodes need to be influenced by some of their in-neighbor nodes in a similar way how an activated node influences its out-neighbors to be activated. The RIM problem finds the seeding cost, which is defined by the minimum number of nodes that must be activated in order to activate all the seed nodes. In this paper, we propose an Active Reverse Path-based Reverse Influence Maximization (ARP-RIM) model to find the minimum seeding cost. Our model is based on the voting model and the classic Independent Cascade model. We simulate our model with three real datasets of three popular social networks. The experimental result shows that the ARP-RIM model outperforms existing RIM models.
Keywords: influence maximization, reverse influence maximization, opportunity cost, seeding cost

· This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2017R1A2A2A05000995).
· 이 논문은 제44회 한국소프트웨어종합학술대회에서 'Active Reverse Path Based Reverse Influence Maximization in Social Networks'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 경희대학교 컴퓨터공학과
ashis@khu.ac.kr
anupam@khu.ac.kr
doma@khu.ac.kr
†† 종신회원 : 경희대학교 컴퓨터공학과 교수(Kyung Hee Univ.)
cshong@khu.ac.kr
(Corresponding author임)

논문접수 : 2018년 3월 23일
(Received 23 March 2018)
논문수정 : 2018년 7월 18일
(Revised 18 July 2018)
심사완료 : 2018년 7월 19일
(Accepted 19 July 2018)

정보과학회논문지 제45권 제11호(2018. 11)

1. Introduction

Social Networks not only gain a potential research interest in recent years but also become an attractive medium in business applications (*e.g.* viral marketing) due to their increasing popularity and proliferation. In viral marketing, identifying influential users in the network is essential and the influence is defused in the social network from user to user in *word-of-mouth* [1] effect. The *Influence Maximization (IM)* problem calculates such a small set of influential seed users that maximizes the spread of influence in the network [1].

In IM, the seed nodes are assumed to be activated initially and then the activated nodes try to activate their inactive out-neighbors to take some decision (*e.g.* purchase any product or service). However, most of the studies do not analyze the cost of activating the seed users. Thereafter, Talukder *et al.* [2], [3] introduce and extend the novel *Reverse Influence Maximization (RIM)* problem which estimates the optimized opportunity cost [4] or seeding cost. However, their method fails to handle major RIM challenges which include handling basic network structures (BNC), stopping criterion, and insufficient influence.

Thus, in this paper, we propose an Active Reverse Path-based RIM (ARP-RIM) model which is based on Independent Cascade (IC) model [1], Voting model [5] and Greedy approach. The proposed model handles the challenges of the RIM problem in a better way as compared to the existing models. We evaluate the performance of the proposed model using the datasets of three real networks and the results show that the model outperforms existing algorithms.

The organization of the rest of the paper is as follows: Section II describes literature overview. Section III and IV state problem formulation and the ARP-RIM model, respectively. Performance analysis is stated in Section V. Finally, Section VI presents conclusion and future scope.

2. Literature Review

The seminal work in influence maximization is conducted by Kempe et al. [1] who formulate two classical models named Linear Threshold (LT) and Independent Cascade (IC) models to maximize influence in social networks. They show both the LT and IC-based IM problems to be NP-Hard and they employ the sub-modularity technique to solve the optimization problem.

Leskovec et al. [6] propose the Cost Effective Lazy Forward (CELF) model, an approximation model that uses the heuristic technique for outbreak detection. A simple path-based model is proposed by Goyal et al. [7] and the model shows better performance than many existing models in terms of seed quality, memory utilization, and running time. Chen et al. [8] introduce a heuristic model in which nodes with the higher degrees are chosen first as seed nodes. The model enhances the accuracy of the classical models [1] and the running time of the CELF model [6] simultaneously.

Barbieri et al. [9] propose Topic-aware IC model (TIC) and Topic-aware LT model (TLT) which have high accuracy in influence calculation. Kim et al. [10] formulate a time-critical influence model and their IC-based model gives good quality seeds and outperforms baseline greedy models.

However, none of the previous studies addresses the problem of optimizing the opportunity cost or seeding cost which is the cost of activating seed nodes given by the minimum number of nodes that must be motivated in order to activate a given set of target nodes. Generally, all of the IM models assume that the seed nodes are initially activated whereas some authors propose free sample product for seed activation [11], [12]. Thereafter, a Reverse Influence Maximization (RIM) model is proposed to compute the opportunity cost or seeding cost [2]. The authors argue that the seed users must be activated by some other influential users they follow, likewise the IM process. Further, they mention several RIM challenges such as stopping criteria, handling three Basic Network Structures (BNC), and insufficient influence. However, their LT-based RIM models are incapable of resolving the challenges properly.

Thus, in this paper, we propose an Active Reverse Path-based RIM (ARP-RIM) model which meets the RIM challenges more efficiently and addresses the problem of finding optimized seeding cost as well.

3. Problem Formulation

Let us assume that we have a social network $G(V,E)$ with $n=|V|$ users and $m=|E|$ social relationships among them. Two sets D_v^{in} and D_v^{out} for each node are defined as in-neighbor and out-neighbors set of v, respectively. The associated in-degree and out-degree are expressed as d_v^{in} and d_v^{out}, respectively.

The IM problem finds a small set S of k seed nodes that maximizes the influence, $\sigma(S)$, in the network and the influence is given by the number of nodes that are activated by the seed users when the seed nodes are set initially activated [1]. The seed set of the IM problem is considered to be the target set in the RIM problem [2]. The RIM problem estimates the seeding cost, $\gamma(S)$, which is defined by the minimum number of nodes that must be motivated in order to activate the target nodes.

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

4. Active Reverse Path (ARP-RIM) Model

In this section, we formulate the proposed Reverse Path Activation model to solve the RIM problem. We start with calculating the marginal seeding cost for each target node $v \in S$.

A. Active Reverse Path

In order to estimate the marginal cost, $\gamma(S)$, of a node v, we consider all the reverse paths from the each of the in-neighbors, $u \in D_v^{in}$.

Definition 2. Reverse Path (RP): If a node sequence, $P_{u_1}=\{\mathrm{u_1 \to u_2 \to \dots \to u_p}\}$ is a path in G, then $P_{u_p}=\{\mathrm{u_p \to u_{p-1} \to \dots \to u_1}\}$ is a reverse path in G.

The RPs starting at u_3 are $\{\mathrm{R_{u_3}}\}=\{\{\mathrm{u_3 \to w_3 \to z_5 \to y_3 \to x_1}\},\{\mathrm{u_3 \to w_3 \to z_7 \to y_4 \to x_4}\},\{\mathrm{u_3 \to w_3 \to z_7 \to y_7}\}\}$ (see figure 1). Let us assume that the nodes u_3, w_3, z_5, and z_7 are activated by the IC model but y_3, and y_4 are not activated. Since y_3, and y_4 are not activated, rest parts of the RPs (x_1, x_4, and x_7) are ignored. Hence, the ARPs starting at u_3 are $\{\mathrm{AR_{u_3}}\}=\{\{\mathrm{u_3 \to w_3 \to z_5}\},\{\mathrm{u_3 \to w_3 \to z_7}\}\}$. Therefore, the seeding cost set, $\Gamma(u_3)=\{\mathrm{u_3, w_3, z_5, z_7}\}$, and the seeding cost, $\gamma(u_3)=4$.

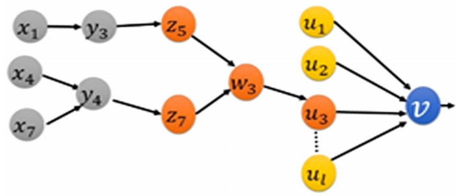

Fig. 1 Reverse Path (RP) and Active Reverse Path (ARP).

Definition 3. Active Reverse Path (ARP):

If a node sequence, $\{\mathrm{R_{u_1}}\}=\{\mathrm{u_1 \to u_2 \to \dots \to u_{i-1} \to u_i \to u_{i+1} \to \dots \to u_l}\}$ is a reverse path in G, and the node activation process (by IC model) activates all the consecutive nodes until the node u_i and the node u_i is not activated, then $\mathrm{AR_{u_1}}=\{\mathrm{u_1 \to u_2 \to \dots \to u_{i-1}}\}$ is the associated Active Reverse Path.

We find Active Reverse Path (ARP) for all the reverse paths by using IC model [1] applied in reverse order. Every inactive node on the reverse path is given a single chance to be activated by its activated neighbors, with a biased coin toss with a probability. If the toss results in head, the node is activated, and node remains inactive otherwise. The probability, p, may be a constant value [1], [13], chosen randomly from a range $(0.01, 0.1)$ or can be calculated by Weighted Cascade Model [1], [13], [14], e.g. $p=\frac{1}{d_v}$, where d_v is the degree of v. However, we employ the Tri-valency Model [13], [14], [15], which suggests p to have a value randomly chosen from a set, e.g. $\{0.001, 0.01, 0.1\}$.

Now, we compute the cost set of each ARP, AR_{u_i}, given by the size of the path, in turn, which is measured by the number of nodes that it connects as shown in the Eq. (1).

$$\Gamma(AR_{u_i})=\{\mathrm{w}|\mathrm{w} \in \mathrm{AR_{u_i}}\} \tag{1}$$

In the next step, we compute the seeding cost set of each node $u_i \in D_v^{in}$ (see Fig. 2) by combining the seeding cost sets of all the active reverse paths starting at u_i and is given by:

$$\Gamma(u_i) = \cup_{AR_{u}\in\{\mathrm{AR}_{u}\}} \Gamma(AR_{u_i}) \quad (2)$$

B. Marginal Seeding Cost

Now, we have all target nodes v with all their in-neighbors $D_v^{\mathrm{in}} = \{\mathrm{u}_1, \mathrm{u}_2, \cdots, \mathrm{u}_l\}$ and their associated seeding cost sets, $\Gamma(u_i)$, as illustrated in Fig. 2.

According to the voting method, we select $\lfloor \frac{1}{2} d_v^{\in} \rfloor + 1$ number of nodes in S^v, from the set D_v^{in}, such that, the aggregated seeding cost of the selected nodes is minimized as stated in the Eq. (3).

$$S^v = \mathrm{argmin}_{|S^v| = \lfloor \frac{1}{2} d_v^{\mathrm{in}} \rfloor + 1} \left| \cup_{u_i \in D_v^{\mathrm{in}}} \Gamma(u_i) \right| \quad (3)$$

We compute marginal seeding cost of a target node, $\Gamma(v)$, by combining all the marginal seeding costs, $\Gamma(u)$, $u \in S^v$, and is given by:

$$\Gamma(v) = \left[\cup_{u \in S^v} \Gamma(u) \right] \cup \{\mathrm{v}\} \quad (4)$$

We apply the greedy approach to choose the nodes instead of subset problem and this improves the running time by reducing the exponential time problem to a linear time problem.

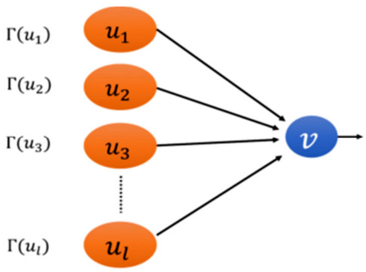

Fig. 2 Target node activation by voting.

C. Seeding Cost of the Target Set

The seeding cost set, $\Gamma(S)$, is computed by combining the marginal seeding cost sets of all $v \in S$ and is given by:

$$\Gamma(S) = \cup_{v \in S} \Gamma(v) \quad (5)$$

The final seeding cost, $\gamma(S)$, of the target set S is given by:

$$\gamma(S) = |\Gamma(S)| \quad (6)$$

Theorem 1. The RIM Problem is NP-Hard.

Proof: The ARP-based RIM problem includes IC model to find the ARPs and at the last step it selects $\lfloor \frac{1}{2} d_v^{\mathrm{in}} \rfloor + 1$ number of in-neighbors out of D_v^{in}, to activate v by the greedy approach which is simply the well-known subset problem. Kempe et al. [1] has proved that the IC model is NP-Hard and the subset problem is also a well known NP-Hard problem [16]. Thus, the RIM problem is NP-Hard as well. □

D. The ARP-RIM Algorithm

The ARP-RIM algorithm, as stated in the Alg. 1, finds the seeding cost set of all in-neighbors of all target nodes (lines 1 to 12) firstly. Then, it calculates the optimized marginal seeding cost set, $\Gamma(v)$, for all $v \in S$ by voting technique (lines 13 to 18) and finally, it combines all marginal cost sets to estimate the desired seeding cost, $\gamma(S)$, by lines from 19 to 24.

All the ARPs can be computed in $O(n+m)$ time by a breadth or depth first search and all the marginal seeding costs, $\Gamma(v)$, can be computed in $O(d)$ time by the greedy method, where d = maximum degree in G. Therefore, the proposed algorithm has complexity $O(kd(n+m))$.

Algorithm 1: ARP-RIM Model

Input: $G(V, E)$, S
Output: $\gamma(S)$
1. $Q = \varnothing, \Gamma(u) = \varnothing, \mathrm{Pr} = \{0.001, 0.01, 0.01\}$;
2. **for** each $u \in D_v^{\mathrm{in}}$ **do**
3. InsertQ(u);
4. $\Gamma(u) = \Gamma(u) \cup \{\mathrm{u}\}$;
5. **while** $Q \neq \varnothing$ **do**
6. DeleteQ(u);
7. **if** u is activated by at least one $w \in D_u^{\mathrm{in}}$ with probability $p \in \mathrm{Pr}$ **then**
8. InsertQ(w);
9. **end if**
10. $\Gamma(u) = \Gamma(u) \cup \{\mathrm{w}\}$;
11. **end while**
12. **end for**
13. $\Gamma(v) = \varnothing$;
14. **for** 1 to $(\lfloor \frac{1}{2} d_v^{\mathrm{in}} \rfloor + 1)$ **do**
15. $u = \mathrm{argmin}_{u \in D_v^{\mathrm{in}}} |[\Gamma(v) \cup \{\mathrm{u}\}] - \Gamma(\mathrm{v})|$;
16. $\Gamma(v) = \Gamma(v) \cup \{\mathrm{u}\}$;
17. $D_v^{\mathrm{in}} = D_v^{\mathrm{in}} - \{\mathrm{u}\}$;
18. **end for**
19. $\Gamma(S) = \varnothing$;

```
    for each v∈S do
        Γ(S) = Γ(S) ∪ Γ(v);
    end for
    γ(S) = |Γ(S)|;
    return γ(S);
```

5. Performance Evaluation

We evaluate the performance of the proposed ARP-RIM model using real datasets of three popular networks: ego-Facebook, ego-Twitter, and web-Google as shown in Table 1. We apply *Monte Carlo (MC)* simulation [1] on an *Intel Core i3* machine with 8 GB RAM. A set of Python programs are executed times on all three the datasets and we take the average values of all the parameters. Finally, we compare the results with that of existing Random-RIM (R-RIM) and Randomized LT-based RIM (RLT-RIM) models [2].

Seeding Cost: Fig. 3, Fig. 4, and Fig. 5 depict that the proposed ARP-RIM has lower seeding cost than any of the existing R-RIM and RLT-RIM algorithms. Moreover, the existing models have drastic fluctuation in the result due to the random nature but our algorithm does not suffer from this flaw. Again,

Table 1 Dataset Description

Networks	Nodes	Edges
ego-Facebook[1)]	4,039	88,234
ego-Twitter[2)]	81,306	1,768,149
web-Google[3)]	875,713	5,105,039

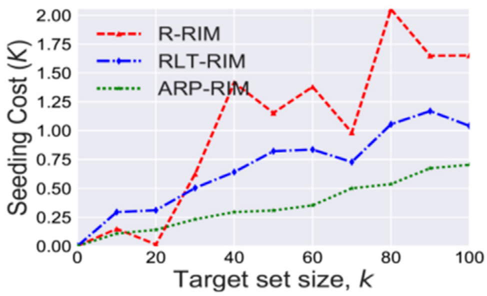

Fig. 3 Seeding cost for different target set size, k (Facebook dataset)

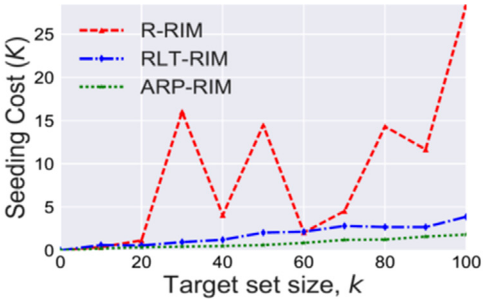

Fig. 4 Seeding cost for different target set size, k (Twitter dataset)

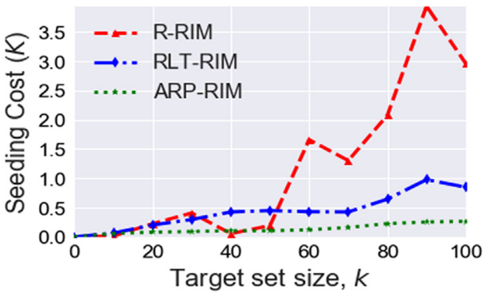

Fig. 5 Seeding cost for different target set size, k (Google dataset)

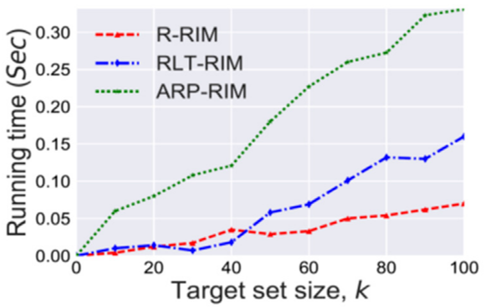

Fig. 6 Running time for different target set size, (Facebook dataset)

by applying the IC model, the ARP-RIM model resolves the issue of stopping criteria which is a major drawback of the existing algorithms.

Running time: On the other hand, the running time of our algorithm is $O(kd(n+m))$ which is slight higher than that of the existing models having running time $O(kd^2)$. The experimental results also reveal the same fact as shown in the Fig. 6, Fig. 7, and Fig. 8. This is due to the higher running time to calculate the ARPs. It contributes the major part,

1) https://snap.stanford.edu/data/egonets-Facebook.html
2) https://snap.stanford.edu/data/egonets-Twitter.html
3) https://snap.stanford.edu/data/web-Google.html

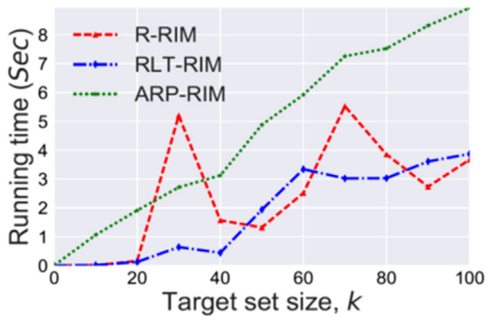

Fig. 7 Running time for different target set size, (Twitter dataset)

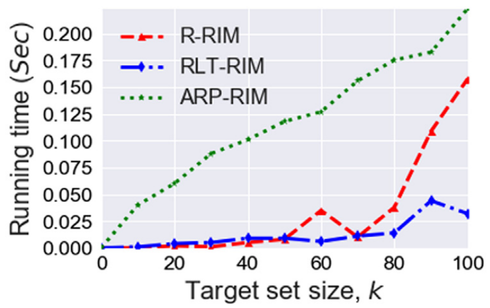

Fig. 8 Running time for different target set size, (Google dataset)

$O(n+m)$, to the running time of the proposed ARP-RIM algorithm. However, in spite of having little higher running time, our algorithm not only gives better and optimized seeding cost but also meets the RIM challenges more efficiently.

6. Conclusion

In this paper, we introduce an Active Reverse Path-based solution (ARP-RIM) to the RIM problem to find the seeding cost of viral marketing in the social networks. The ARP-RIM model jointly employs Independent Cascade (IC) model and Voting model along with Greedy optimization. The proposed model resolves challenging issues of the RIM problem more efficiently and yet provides optimized opportunity cost or seeding cost. The experimental results show that it outperforms the existing algorithms though it consumes little more time. The future work may focus on the reducing the running time.

References

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the Spread of Influence Through a Social Network," *Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2003*, pp. 137-146, 2003.

[2] 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," *Proc. of the 19th IEEE Asia-Pacific Network Operations and Management Symposium (APNOMS 2017)*, pp. 354-357, 2017.

[3] A. Talukder, M. G. R. Alam, H. T. Nguyen, and C. S. Hong, "A Cost Optimized Reverse Influence Maximization in Social Networks," *Proc. of the IEEE/IFIP Network Operations and Management Symposium (NOMS 2018)*, 2018.

[4] Windheim, M., Greve, E. and Krause, D., "Decisive Economies and Opportunity Cost of Modular Product Structure Alternatives: An Empirical Case Study," *Proc. of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1636-1640, 2017.

[5] H. Zhang, S. Mishra, M. T. Thai, J. Wu, and Y. Wang, "Recent Advances in Information Diffusion and Influence Maximization in Complex Social Networks," *Journal of Opportunistic Mobile Social Networks*, Vol. 37, No. 1.1, 2014.

[6] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective Outbreak Detection in Networks," *Proc. of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2007*, pp. 420-429, 2007.

[7] A. Goyal, W. Lu, and L. V. Lakshmanan, "SIMPATH: An Efficient Algorithm for Influence Maximization under the Linear Threshold Model," *Proc. of the 11th IEEE Conference on Data Mining (ICDM) 2011*, pp. 211-220, 2011.

[8] W. Chen, Y. Wang, and S. Yang, "Efficient Influence Maximization in Social Networks," *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009*, pp. 199-208, 2009.

[9] N. Barbieri, F. Bonchi, and G. Manco, "Topic-Aware Social Influence Propagation Models," *Proc. of the 12th IEEE International Conference on Data Mining (ICDM) 2012*, pp. 81-90, 2012.

[10] J. Kim, W. Lee, and H. Yu, "CT-IC: Continuously Activated and Time-restricted Independent Cascade Model for Viral Marketing," *Journal of Knowledge-Based Systems*, Vol. 62, pp. 57-68, 2014.

[11] S. Bhagat, A. Goyal, and L. V. Lakshmanan, "Maximizing Product Adoption in Social Networks," *Proc. of the 5th ACM International Conference on*

Web Search and Data Mining 2012, pp. 603-612, 2012.
[12] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "Stop-and-Stare: Optimal Sampling Algorithms for Viral Marketing in Billion-scale Networks," *Proc. Of the ACM International Conference on Management of Data 2016*, pp. 695-710, 2016.
[13] A. Arora, S. Galhotra, and S. Ranu, "Debunking the Myths of Influence Maximization: An In-Depth Benchmarking Study," *Proc. of the ACM International Conference on Management of Data 2017*, pp. 651-666, 2017.
[14] W. Chen, C. Wang, and Y. Wang, "Scalable Influence Maximization for Prevalent Viral Marketing in Large-scale Social Networks," *Proc. of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2010*, pp. 1029-1038, 2010.
[15] K. Jung, W. Heo, and W. Chen, "IRIE: Scalable and Robust Influence Maximization in Social Networks," *Proc. of the 12th IEEE International Conference on Data Mining (ICDM) 2012*, pp. 918-923, 2012.
[16] E. Horowitz, S. Sahni, and S. Rajasekaran, Computer Algorithms, 2nd Ed., Pp. 499, Computer Science Press, New Yeork, 1998.

Ashis Talukder received the BS and MS degrees in Computer Science and Engineering from University of Dhaka (DU), Bangladesh. He is an Assistant Professor, Department of Management Information Systems (MIS), University of Dhaka, Bangladesh. Currently, he is pursuing Ph.D. in the Department of Computer Science and Engineering at Kyung Hee University, South Korea. His research interests include social networks, influence maximization, network optimization, and data mining. He is a member of KIISE, and IEEE communication society.

Anupam Kumar Bairagi received his B.Sc. and M.Sc. degree in Computer Science and Engineering from Khulna University (KU), Bangladesh. Currently, he is pursuing his PhD degree from Kyung Hee University (KHU), South Korea, for which he was awarded a scholarship in 2014. He is a faculty member in the discipline of Computer Science and Engineering, Khulna University, Bangladesh. His research interests includes wireless resource management, cooperative communication and game theory.

김 도 현
정보과학회논문지
제 45 권 제 4 호 참조

홍 충 선
정보과학회논문지
제 45 권 제 3 호 참조