A new viral marketing strategy with the competition in the large-scale Online Social Networks

Canh V. Pham[1], Dung. Ha[1], Dung Q. Ngo[1], Quang C. Vu [1] and Huan X. Hoang[2]

***Abstract*— The problem of Influence Maximization (IM) on social networks proposed firstly by Kempe et al. (2003) has been researched and developed with many cases. However, the IM in limited time while unwanted users are restricted is still a new potential subject. In this paper, we conducted research the problem on the model of information diffusion name Locally Bounded Diffusion and tested some useful heuristic algorithms. The results of the experiment on some real datasets of social networks show that the algorithm meta-heuristic generated better output than the others.**

I. Introduction

With the fast development and steady of the Online Social Networks (OSNs), such as Facebook, Twitter, Google+, etc. OSNs have become the most common utilized way for information propagation. OSNs provide a nice platform for information diffusion and fast information exchange among their users.

The field of Influence Maximization (IM) has received a lot of research interests in recent years. This problem asked to find k users on OSNs to initiate to spread of information such the number of users be affected is the maximum. The problem was firstly proposed by Kempe et. al [1] in two diffusion models which are *Independent Cascade* (IC) and *Linear Threshold* (LT) model and they also proved that it is NP-hard, and designed a greedy algorithm can obtained $1 - 1/e$. Although extensive related works have been conducted on the IM problem [2], [3], [4], [9], [10], [14], [15], most of them are based on such an assumption that without the existence of unwanted target users whom we do not want information come to. In reality, on OSNs exits the group of users who have opposite viewpoints and benefits with us and they create a negative impact to oppose for information received.

Considering the following example that highlights a basic need for every organization that uses OSNs. There are two mutual competitive companies A and B. The A has been deploying a large advertisement, even via the Internet. They drew a marketing blueprint on several social networks but the A tried to hide everything against every one of the B as long as its possible. Constantly, the advertising information of A can reach to the B after a time. Thus, the A needs a solution help them fast imply the marketing strategy to much many users except unwanted users (from B) to gain the best consumption more quickly than B within t hop.

Motivated by the above phenomenon, in this paper, we formulate a new optimization, called Influence Maximization while Limited unwanted target users (IML), to find seeding set S to Maximize Influence and the Influence to unwanted is under some certain threshold after at most d time (hop). The total influence is total active user and the unwanted users are referred as those whom we do not want the information come to.

Our contributions in this paper are summarized as follows:

- First attempt to study the Influence Maximization while Limited unwanted (IML) target users under LBD model.
- Prove d-MIL is NP-Complete and show it can not be approximated in polynomial time with a ratio $e/(e-1)$ unless $NP \subseteq DTIME(n^{O(\log n \log n)})$.
- Conduct our experiments on real-world datasets, and design some heuristic algorithms to find the solution, results showed that meta-heuristic algorithm better than the other.

Related work. The target is to spread the desired information for as many people as possible on OSNs. Kempe et al. [1] first formulated the Influence Maximization (IM) problem which asks to find a set of users who could maximize the influence. The influence is propagated based on a stochastic process called Independent Cascade Model (IC) in which a user will influence his friends with probability proportional to the strength of their friendship. The author proved that the problem was NP-hard and proposed a greedy algorithm with approximation ratio of $1 - 1/e$. After that, a considerable number of works studied and designed new algorithms for the problem variants on the same or extended models [2], [3], [4], [9], [10], [14], [15]. Huiyuan et al. [2] proposed a problem to maximize the positive news in propaganda rather than maximizing the users affected. They said that to maximize positive things in many cases had more beneficial than maximizing the number of people affected. They used the Cascade Opinion (OC) model to solve the problem. On the other hands, Zhang et al. [3] recommended to maximize the influence of information to a specific user by finding out the k most influential users and proved that it was NP-hard problem and the function is *submodular*. They also launched an effective approximation algorithm. Zhuang et al. [4] have studied the *IM* problem in the dynamic social network model over time. In addition, there were several other studies: Chen et al. [14] investigated *IM* problem on a limited

[1] Canh V. Pham, Dung .Ha, Dung Q. Ngo and Quang C. Vu are with the Faculty of Technology and Information Security, People's Security Academy, Hanoi, Vietnam. E-mail: maicanhki@gmail.com, dungha.hvan@gmail.com, quocdung.ngo@gmail.com,quangvc@gmail.com

[2] Huan X. Hoang is with University of Technology and Engineering, Vietnam National University, Hanoi, Vietnam. E-mail: huanhx@vnu.edu.vn

time; Gomez-Rodriguez et al. [15] studied *IM* problem for continuous time. Researches on IM with various contexts and various models received many attentions, but the diffusion of information problem, in addition to spreading the positive information still faced with the misinformation. How to spread the positive information while the misinformation limited? To solve it, Ceren et al. [5] launched the problem selecting k users to convince them aware of good information so that after the campaign, amount of use influenced by the misinformation was the least. By using Model-Oblivious Independent Campaign Cascade, they proved the problem be NP-hard and the objective function was *submodular*. Nguyen et al. [6] gave the decontamination problem of misinformation by selecting a set of users with sources of misinformation I assumed to have existed on the social network at the rate of $\beta \in [0, 1]$ after T time. They launched the different circumstances of the I and the T, but they only solved the case I was unknown. On preventing infiltration to steal information on OSNs, Pham et al. [13] have built a Safe Community for the purpose of protecting all users in an Organization. In problems of detecting misinformation source on social networks, Nguyen et al. [8] assumed that the exist a set of misinformation sources I, they purposed of finding the largest number of users in I who started to propagate that information. Nevertheless, the predictions were likely confused because they did not know the order of real time start to spread misinformation. Zhang et al. [9] studied the problem of limited resources that often was incorrect information while maximized the positive source of misinformation on OSNs under Competive Activation model. In this study, they were considered a model of misinformation and good information and presence on the social network, they also proved to be NP-complete problem and could not be approximated with rate $e/(e-1)$ unless $NP \subseteq DTIME(n^{O(\log \log n)})$.

In these researches, no one focused on the spread of information with the limiting of information to the set of ones whom we did not want the information reach to (called *unwanted users*). While positive information is desired to propagate to more and more users, we also face with the existence of unlike users on OSNs. Because every time they receive the positive information, they can be able to conduct the activities, propagation strategies that opposes to our benefits.

II. Model and problem definitions

A. Network and information diffusion model

We are given a social network modeled as an undirected graph $G = (V, E)$ where the vertices in V represent users in the network and the edges in E represent social links between users. We use n and m to denote the number of vertices and edges. The set of neighbors of a vertex $v \in V$ is denote by $N(v)$ and $d(v) = |N(v)|$ is degree of node v. Existing diffusion models can be categorized into two main groups [1]: Threshold model and Independent Cascade model.

Threshold Model. In this model, each node v has a threshold $\theta_v \in [0, 1]$, typically drawn from some probability distribution. Each connection (u, v) between nodes u and v is assigned a weight $w(u, v)$. Initially, nodes in network is not influenced (the state of each node is inactive). For a node v, let $N^a(v)$ be the set of neighbors of v that are already influenced (active). Then v is influenced if $\sum_{u \in N^a(v)} w(u, v) \geq \theta_v$.

Independent Cascade model. Whenever a node u is influenced, it is given a single chance to activate each of its neighbor v with a given probability $p(u, v)$.

Most influence maximum papers assume that the probabilities $p(u, v)$ or weight $w(u, v)$ and thresholds θ_v are given as a part of the input. However, they are generally not available and inferring those probabilities and thresholds has remained a non trivial problem [11]. Hence, in this work, we use a simple diffusion model is *Locally Bounded Diffusion* Model [10] defined as follow:

Locally Bounded Diffusion (LBD) [10] Let $S_0 \in V$ be the subset of vertices selected to initiate the influence propagation, which we call the seeding. We also call a vertex $v \in S_0$ a seed. The propagation process happens in round, with all vertices in S_0 are influenced at round $t = 0$. At a particular round $t \geq 0$, each vertex can be active or inactive and each vertex $'$s tendency to become active increases when more of its neighbors become active. If an inactive vertex u has more than $\lceil \rho d(u) \rceil$ active neighbors at time t, then it becomes active at round t, where ρ is the *influence factor*.

B. Problem Definition

The paper focus the value of the objective function after d hop. Denote function $\delta_d(.)$ is total active users after t hop and $L_i(.) = |N^a(t_i)|$ is the information leakage i.e the number of neighbor of t_i has been actived. Considering that influence can be propagated at most d hops, We study the Maximizing Influence while Limited unwanted target users (d-IML) problem defined as follow:

Definition 1 (d-IML problem): Given an social network represented by a directed graph $G = (V, E)$ and an under LBD model. Let $T = \{t_1, t_2, .., t_p\}$ be the set of $|T| = p$ unwanted users. Our goal is to chose the set seeding of users $S \subset V$ at most k-size that maximizes influence such that the total influence come to t_i after d round (hop) less than threshold for preventing information leakage τ_i i.e: $|N^a(t_i)| < d(t_i)\tau_i$.

III. NP-Complete and Inapproximation

In this section, we first show the NP-Completeness of IML problem on LBD model by reducing it from Maximum Coverage problem. By this result, we further prove the inapproximability of d-IML which is NP-hard to be approximated within a ratio of $e/(e-1)$ unless $NP \subseteq DTIME(n^{O(\log \log n)})$.

Theorem 1: d-IML is NP-Comlete in **LBD** model.

Proof: We consider of the decision version of d-IML problem that asks whether the graph $G = (V, E)$ contains a set $k-size$ of seed user $S \subset V$ that number active node at

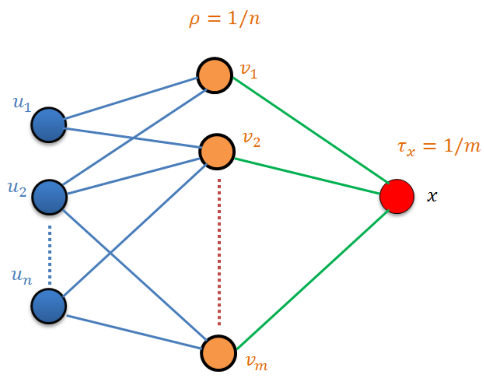


Fig. 1. Reduction from MC to 1-IML

least K, such that $\sum_{u \in N^a(t_i)} w(u, t_i) < \tau_i$ within at most d (hop) rounds. To prove d-IML is NP-Complete, we prove the following two tasks:

1) d-IML $\in$ NP.
2) d-IML is NP-Hard.

Given $S \subset V$, we can calculate the influence spread of S in polynomial time under LT model after d hop. This implies d-IML is NP.

Now we prove a restricted class of d-IML instance is NP-hard, $d = 1$. To prove that 1-IML is NP-hard, we reduce it from the decision version of Maximum Coverage problem defined as follows.

Maximum Coverage. Given a positive integer k, a set of m element $\mathcal{U} = \{e_1, e_2, \ldots, e_m\}$ and a collection of set $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. The set may some element in common. The *Maximum Coverage* problem asks to find a subset $S' \subset S$, such that $| \cup_{S_i \in S'} S_i|$ is maximized with $|S'| \leq k$. The decision of this problem asks whatever the input instance contains a subset S' of size k which can cover at least t elements where t is a positive integer.

Reduction. Given an instance $I = \{\mathcal{U}, \mathcal{S}, k, t\}$ of the maximum coverage, we construct an instance $G = (V, E, \mathcal{S}, \mathcal{U}, \theta)$ of the 1-IML problem as follows.

- *The set of vertices:* add one vertex u_i for each subset $S_i \in \mathcal{S}$, once vertex v_j for each $e_j \in \mathcal{U}$, and a vertex x is a unwanted users.
- *The set of edges:* add an edge (u_i, v_j) for each $e_j \in S_i$ and connect x to each vertex v_j.
- *Threshold for prevent leakage information and Factor influence:* We assign threshold for prevent leakage information for vertex x is $\tau_x = 1/m$. The factor influence $\rho = 1/n$.
- Finally, set $d = 1, K = t$.

The reduction is illustrated in Fig. 1.

Suppose that $\mathcal{S}^*$ is a solution to the maximum coverage instance, thus $|\mathcal{S}^*| \leq k$ and it can cover at least t elements in $\mathcal{U}$. By our construction, we select all nodes u_i corresponding to subset $S_i \in \mathcal{S}^*$ as seeding set S. Thus, $|S| \leq k$. Since $\mathcal{S}^*$ cover at least t elements e_j in $\mathcal{U}$ so S influence at least t vertices v_j corresponding to those e_j and total incoming active neighbour $N^a(v_j)$ at least 1. Due to $d(v_j) \leq n$, we have:

$$N^a(v_j) \geq 1 = \frac{1}{n}.n \geq \lceil \rho d(v_j) \rceil$$

implies v_j is active. Hence, there are at least $t = K$ nodes in the 1-IML has been active.

Conversely, suppose there is seeding $S, |S| \leq k$ such that the number of active node at least K. We see that $v_j \notin S, j = 1, 2, .., m$ because in this case the number of neighbor of x are active at least 1. Thus,

$$|N^a(x)| \geq 1 = m.\frac{1}{m} = d(x)\tau_x$$

Hence, $S \subseteq \{u_1, u_2, \ldots, u_n\}$. Then $\mathcal{S}^*$ can be collection of subset S_i corresponding to those $u_i \in S$. Hence the number of elements which it can cover is at least $K = t$. ■

Based on above reduction, we further show that inapproximation of IML in the following theorem.

Theorem 2: The IML problem can not be approximated in polynomial time within a ratio of $e/(e-1)$ unless $NP \subseteq DTIME(n^{O(\log \log n)})$.

Proof: Supposed that there exits a $e/(e-1)$-approximation algorithm $\mathcal{A}$ for d-IML problem. We use the above reduction in proof of Theorem 1 then $\mathcal{A}$ can return the number of active nodes in G with seeding size equal to k. By our constructed instance in Theorem 1, we obtain the Maximum Coverage with size t if the number of active nodes in optimal solution given by $\mathcal{A}$ is K. Thus algorithm $\mathcal{A}$ can be applied to solve the Maximum Coverage problem in polynomial time. This contradict to the NP-hardness of Maximum Coverage problem in [12]. ■

Although the objective function is submodular, propagation of influence is constraint by the leakage information. Hence, we can not give an algorithm for approximately with the ratio is $1 - 1/e$ as Kemp [1]. In this section, we introduce an Greedy Algorithm for IML problem.

IV. Algorithms

A. Linear Programming approach

One advantage of our discrete diffusion model over probabilities is that the exact solution can be using mathematical programming. Thus, we formulate the IML problem as an $0-1$ Integer Linear Programming (ILP) problem below.

$$\texttt{maximize} \sum_{v \in V \setminus T} x_v^d \tag{1}$$

$$\texttt{st:} \sum_{v \in V \setminus T} x_v^0 \leq k \tag{2}$$

$$\sum_{u \in N(v)} x_w^{i-1} + \lceil \rho.d(v) \rceil x_v^{i-1} \geq \lceil \rho.d(v) \rceil x_v^i,$$

$$\forall v \in V, i = 1..d \tag{3}$$

$$\sum_{v \in N(t_i)} x_v^d < d(t_i).\tau_i \tag{4}$$

$$x_v^i \geq x_w^{i-1}, \forall v \in V, i = 1..d \tag{5}$$

$$\text{where} \quad x_v^i = \begin{cases} 1 & \text{if } x \text{ is active at round (hop) } i \\ 0 & \text{otherwise} \end{cases}$$

The objective function (1) of the ILP is to find the number of node is active. The constraint (2) is number of set seed is bounded by k; the constraints (3) capture the propagation model; the constraint (4) limit leakage information come to unwanted user; and the constraint is simply to keep vertices active once they are activated.

Wee see that the number of constraint is up to $\mathcal{O}(d.n^2)$ and the number of variables is bounded by $\mathcal{O}(d.n)$. Although solve ILP can provide the optimal solution, it can not be applied for larger network.

B. Greedy algorithm

In this section, we introduce a straightforward greedy algorithm in Algorithm 1. We denote $L_{t_i}(S)$ is the total influence to t_i respect to seeding sets S after d hop *i.e.,* $L_{t_i}(S) = |N^a(t_i)|$. The greedy algorithm sequentially selects a node u into the seed set S that the number of actived neibour of t_i does not exceed $\rho.d(t_i)$ and maximizes the flowing influence marginal gain:

$$\Delta_d(S, u) = \delta_d(S + \{u\}) - \delta_d(S). \tag{6}$$

Because of conditions of limited to leak information to unwanted users, greedy algorithm does not guarantee approximately ratio $1 - 1/e$ as in the greedy algorithm in [1].

Algorithm 1: Greedy Algorithm (GA)

Data: $G = (V, E), \rho, \tau$ set unwanted users
$U = \{t_1, t_2, \ldots, t_p\}$, p, k;
Result: Seeding S;
begin
  $S \leftarrow \emptyset$;
  $L \leftarrow \emptyset$;
  while $|S| \leq k$ **do**
    $\delta_{max} \leftarrow 0$;
    foreach $u \subset V - S$ **do**
      if $\forall t_i \in T, L_i(S + \{u\}) < d(t_i).\tau_i$ **then**
        if $\delta_{max} < \Delta_d(S, u)$ **then**
          $\delta_{max} \leftarrow \Delta_d(S, u)$;
          $v \leftarrow u$;
        end
      end
      if $\delta_{max} = 0$ **then**
        Return S;
      end
      $S \leftarrow S + \{v\}$;
    end
    Return S;
  end
end

C. Meta-heuristic algorithm

In the next section, we introduced a meta-heuristic algorithm to improve, we designed the algorithm combine the influence marginal gain and evaluation information leakage. Accordingly, we used a heuristic function $f(v)$ to evaluate the fitness of user v which defined as follows:

$$f(v) = \frac{\Delta_d(S, v)}{1 + \frac{1}{p} \sum_{t_i \in T} l_{t_i}(v)} \tag{7}$$

where $l_{t_i}(v) = \frac{L_{t_i}(S + \{v\})}{\tau_i}$ is the normalized leakage level at t_i after adding v to seed set S.

Algorithm 2: Meta-Heuristic (MH) algorithm

Data: $G = (V, E), \rho, \tau$ set unwanted users
$U = \{t_1, t_2, \ldots, t_p\}$, p, k;
Result: Seeding S;
begin
  $S \leftarrow \emptyset$;
  $L \leftarrow \emptyset$;
  while $|S| \leq k$ **do**
    $\delta_{max} \leftarrow 0$;
    $f_{max} \leftarrow 0$;
    foreach $u \subset V - S$ **do**
      if $\forall t_i \in T, L_i(S + \{u\}) < d(t_i).\tau_i$ **then**
        if $f_{max} < f(u)$ **then**
          $fmax \leftarrow f(u)$;
          $v \leftarrow u$;
        end
      end
      if $f_{max} = 0$ **then**
        Return S;
      end
      $S \leftarrow S + \{v\}$;
    end
    Return S;
  end
end

V. Experiment

In this section we perform experiments on OSNs to show the efficiency of propagation and compare performance of greedy algorithms with optimal solution given by ILP.

A. Dataset

arXiv-Collaboration. The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its GR-QC section. If an author i co-authored a paper with author j, the graph contains a undirected edge from i to j. If the paper is co-authored by k authors this generates a completely connected (sub)graph on k nodes [16].

Gnutella. A sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. There are total of 9 snapshots of Gnutella network collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts [16].

TABLE I
BASIC INFORMATION OF NETWORK DATASETS

Network	nodes	edges	Type	Avg. Degree
ArXiv-Collaboration	5,242	28,980	Direct	5.53
Gnutella	6,301	20,777	Direct	3.30

In each graph, we used the method in [1] to assign the diffusion weight to each edge and then normalize the weights of all incoming edges of a node v to let it satisfy that $\sum_{u \in N^{in}(v)} w(u,v) \leq 1$. For each network, we randomly selected 100 unwanted users.

B. Experiment results

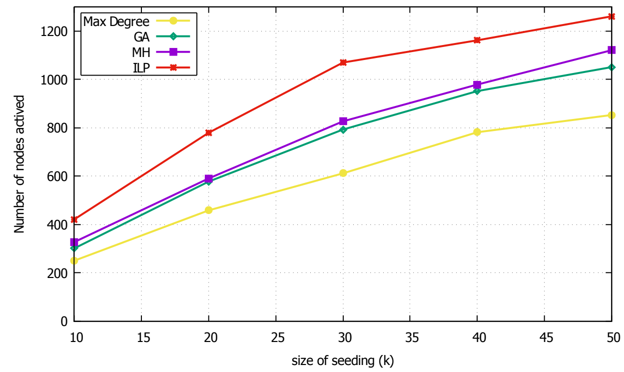
In this part, we describe comparison algorithms: Max degree algorithm, Greedy Algorithm (GA), Meta-heuristic (MH) algorithm and ILP. Max Degree method is the greedy algorithm that chose the vertex v that had maximum degree when the information leaked to unwanted users less than the threshold leakage.

In the experiment, we tested the performance of algorithms with $d = 4, k = \{10, 20, 30, 40, 50\}$ and $\rho = \{0.2; 0.4; 0.6\}$. We espcially compared between these algorithmsresults with optimal solution given by ILP.

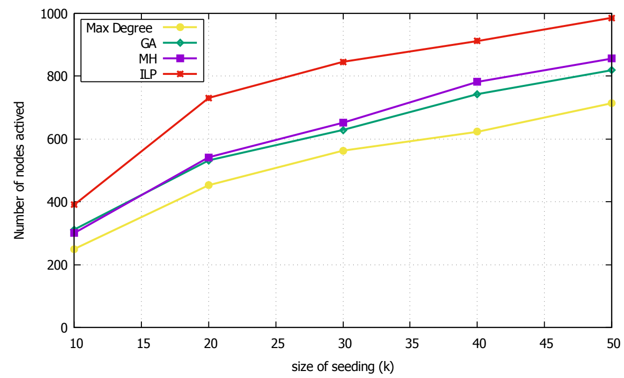
We solved the ILP problem on Gnutella network [16], with $d = 3$, The ILP was solve with CPLEX version 12.6 on Intel Xeon 3.6 Ghz, 16G memories and setting time limit for the solver to be 48h. For $k = 10, 20$ the solver return the optimal solution. However, for $k = 30, 40$ and 50, the solver can not find the optimal solution within time limit and return sub-optimal solutions.

1) Solution Quality: The algorithms gave different resutls when k changes. With all values of $\rho = \{0.2; 0.4; 0.6\}$, MH resulted better than GA and Max degree. For more details, MH typically showed to be better than Max Degree, when k became larger, the distance between MH and Max Degree also became larger. As can be seen from these graphs, the MH seems to be better than Max Degree and GA. Especially, In the comparison with Max Degree, MH gave the moderately better result that can be seen in the distance between MH and Max Degree when the values of k increased. On the other hand, MH and GA generated the same results with small ones of k. The graphs illustrated that MH line and GA line had the same trends and the gap between them was unclear when k changed from 10 to 30. It could be really different when the values of k grew up. In these cases of k from 30 to 50, the number of nodes activated of MH was bigger than the one of GA. It meaned that MH went better than GA when the size of seeding set was larger. Moreover, MH reached closer ILP than the others. The trending of red line and purple line seemed to be convergened with large value of k. The changing rate got from the percentage of 64% to 80%.

2) Influence Factor ρ: From fig. 1 to fig.3, the values of ρ grew up with $0.2, 0.4, 0.6$, respectively. From these graphs, it was clear to realize the number of active usders went out from being approximately equal to 1000 down to bellow 700 and typically down to bellow 400. It mean that when ρ increased, the information diffusion decreased.



(a) ArXiv-Collaboration



(b) Gnutella

Fig.1. The actived nodes when the size of seeding set varies while $d = 4, \rho = 0.2$.

VI. CONCLUSIONS

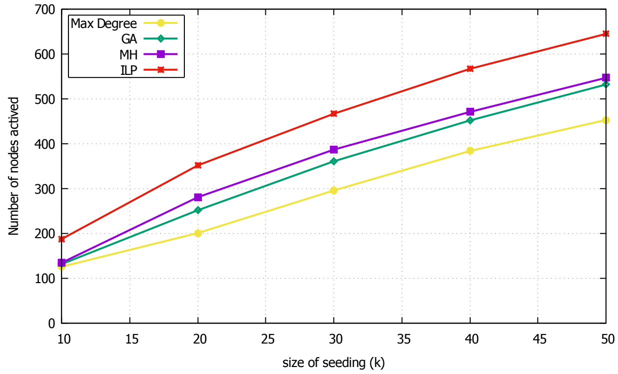
In order to propose a viral marketing solution while there exists the competition between organizations that have benefit collisions, we built the problem of maximization influence to users whereas limits the information reach to unwanted ones in constrained time. We proved it be an NP-complete and not be approximated with $1 - 1/e$ rating number. We also recommended an efficient solution MH to solve the problem. The experiment via social networks data showed that our algorithm got the better result and reached closer to optimized solution than several algorithms.

REFERENCES

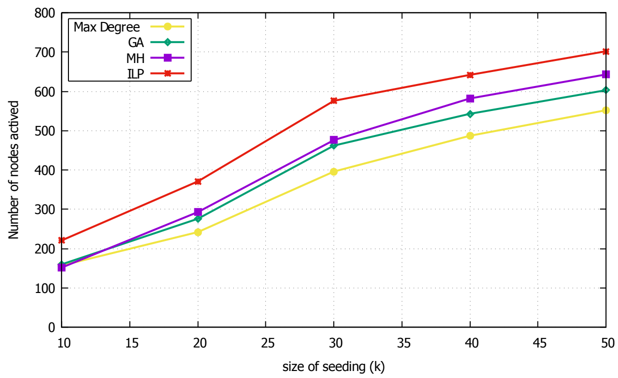
[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of inuence through a social network. In Proc. 9th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD 03, pages 137-146, New York, NY, USA, 2003.

[2] Huiyuan Zhang, Thang N. Dinh, and My T. Thai .: Maximizing the Spread of Positive Influence in Online Social Networks, in Proceedings of the IEEE Int Conference on Distributed Computing Systems (ICDCS), pages 317 – 326, 2013.

[3] J Zhang, P Zhou, C Cao, Y Guo L .: Personalized Inuence Maximization on Social Networks, In Proceedings of the 22nd ACM international conference on Conference on Information and Knowledge Management, pages 199–208, 2011.

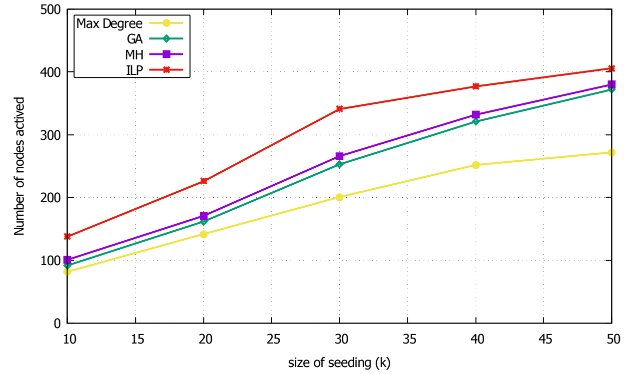


(a) ArXiv-Collaboration

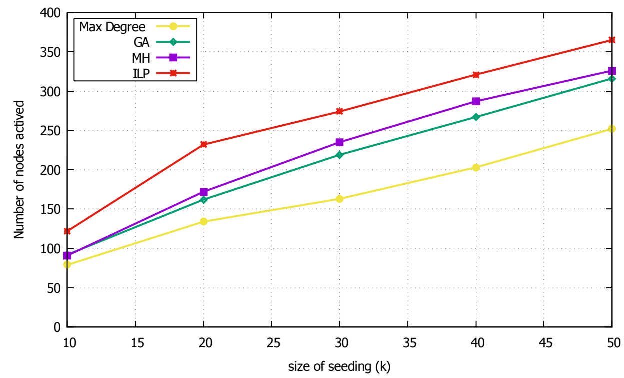


(b) Gnutella

Fig.2. The actived nodes when the size of seeding set varies while $d = 4, \rho = 0.4$.



(a) ArXiv-Collaboration



(b) Gnutella

Fig.3. The actived nodes when the size of seeding set varies while $d = 4, \rho = 0.6$.

[4] Honglei Zhuang, Yihan Sun, Jie Tang, Jialin Zhang, and Xiaoming Sun. Influence Maximization in Dynamic Social Networks. In Proceedings of IEEE International Conference on Data Mining (ICDM), pages 1313 – 1318, 2013.

[5] Ceren Budak, Divyakant Agrawal, Amr El Abbadi. Limiting the Spread of Misinformation in Social Networks. Proceedings of the 20 th international conference on World Wide Web (WWW'11) pages 665–674, New York, NY, USA, 2011.

[6] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, Containment of Misinformation Spread in Online Social Networks, in Proceedings of ACM Web Science (WebSci), pages 213–222, 2012.

[7] T. N. Dinh, Y. Shen, and M. T. Thai, The walls have ears: optimize sharing for visibility and privacy in online social networks, in Proceedings of ACM Int Conference on Information and Knowledge Management (CIKM), Pages 1452-1461, 2012.

[8] D. T. Nguyen, N. P. Nguyen, and M. T. Thai, Sources of Misinformation in Online Social Networks: Who to Suspect?, in Proceedings of the IEEE Military Communications Conference (MILCOM), pages 1 – 6, 2012.

[9] H. Zhang, X. Li, and M. Thai, Limiting the Spread of Misinformation while Eectively Raising Awareness in Social Networks, in Proceedings of the 4th International Conference on Computational Social Networks (CSoNet), pages 35–47, 2015.

[10] T. N. Dinh, H. Zhang, D. T. Nguyen, and M. T. Thai.: Cost-effective Viral Marketing for Time-critical Campaigns in Large-scale Social Networks, IEEE/ACM Transactions on Networking (ToN), pages 2001 – 2011, 2013.

[11] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, Learning influence probabilities in social networks, in Proceedings of the third ACM international conference on Web search and data mining, WSDM 10, pages 241250, 2010.

[12] Feige, U.: A threshold of ln n for approximating set cover. Journal of the ACM (JACM) 45(4), pages 634652, 1998.

[13] Canh V. Pham, Huan X. Hoang, Manh M. Vu.: Preventing and Detecting Infiltration on Online Social Networks, in Proceedings of the 4th International Conference on Computational Social Networks (CSoNet), pages 60–73, 2015.

[14] Wei Chen, Wei Lu, Ning Zhang.: Time-Critical Influence Maximization in Social Networks with Time-Delayed Diffusion Process, http://arxiv.org/abs/1204.3074, 2015.

[15] Manuel Gomez-Rodriguez, Le Song, Nan Du, Hongyuan Zha, and Bernhard Scholkopf. Inuence estimation and maximization in continuous-time diffusion networks. ACM Trans. Inf. Syst, Volume 34(2), DOI: http://dx.doi.org/10.1145/2824253, 2016.

[16] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameter, ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 2007.