Running Time Improvement of Influence Maximization in Social Network Using Graph Reducing Technique (GRT)

Ashis Talukder, Anupam Kumar Bairagi, VanDung Nguyen, and Choong Seon Hong

Department of Computer Science & Engineering, Kyung Hee University.

Email: {ashis, anupam, ngvandung85, cshong}@khu.ac.kr

Abstract

Influence maximization is to find a small seed set of most influence people in a domain. With the increased popularity of social network, influence maximization in the social network has become a potential problem and the result has been begun to be applied in various applications like viral marketing, domain expert search, recommending network etc. The basic influence maximization problem is NP-hard and many researchers has developed different approximation algorithms. In this research we have proposed an algorithm named Graph Reduced Technique (GRT) which estimates influence based on heuristics. The algorithm offers better running time and with feasible amount of spread of influence.

1. Introduction

Similar to the real life where people exchange ideas, thoughts, information and even rumor from person to person by the effect of *word of mouth* (also called *viral marketing*), social network, now a days, has become very popular medium for spreading ideas, thoughts and influence among the users [1], [2], [3], [4], [8]. Facebook, Twitter, Linkedin, Youtube, Google+, Netflix etc are the prominent social networks using direct or referral marketing [1], [2].

Analogous to real life, one friend (followee) influences or motivates other friends (followers) to some action (viz. buy any product and/or service) in the social network. The *Influence Maximization (IM)* problem is to find an optimal seed set so that expected number of activated node is maximum in a graph when an initial small seed set size is given. Generally the number of activated nodes are considered to be the amount of spread of influence in the network [1], [2], [3], [8]. The scope of influence maximization covers diverse fields of applications including product/service marketing, target marketing, public or community awareness program, discovering community leader [7], domain expert search [3], recommending (product, service, movie, music) [2], and isolating rumor with its offenders [9] etc.

In this paper we have proposed an approximation algorithm based on heuristic. By heuristic approximation we find and delete all the inactive nodes and associated links to get the Reduced Social Graph (RSG). Then from the reduced graph, find the k top influential nodes to be seed set. The main contribution of this paper is:

- To formulate a technique to reduce the social graph by probabilistic estimation and heuristics.
- This will decrease running time remarkably.

The rest of the paper is organized as follows: in section two a study of the state-of-the-art has been provided and the problem formulation is given in section three. System model and the algorithm are described in the section four and five respectively. Evaluation and conclusion have been described in the next consecutive sections.

2. Related works

The pioneer of the IM research community is Kempe et al. [8]. They proposed two classical greedy algorithms: Linear Threshold (LT) model and Independent Cascade (IC) model. In general IM problem is NP hard. So they used submodular function and got approximation with factor of (1- 1/e) i.e. 63%.

H Nguyen et al. [1] used Directed Acyclic Graph (DAG) and then efficient heuristic algorithm and got an approximation of (1-1/√e). In [2] the researchers have solved the problem by Continuous Markov process they also devised a spread prediction technique. Chen et al. [4] used quality factor and quality sensitivity ratio on tree structured social network. It has the same approximation (1-1/e) factor with standard greedy method but faster. Many authors have tried to extend the Linear Threshold and Independent Cascade Model.

Goel et al. [7] proposed Simple Path (SIMPATH) Algorithm using vertex cover and look ahead optimization to extend LT model [11] and also used his previous algorithm Cost Effective Lazy Forwarding (CELF) [10]. The SIMPATH algorithm is faster but do not guaranty any approximation. Our GRT algorithm gives a reasonable amount of influence value while it is faster.

3. Problem formulation

Consider a scenario that there are n users in a social network and they have friendship relation. The directed graph $G(V,E)$ represents social network where nodes are users and edges represent their social relationship. The relationship in G is an adjacency matrix and an activation threshold vector

$t = [t_1, t_2, ..., t_n]^T$ is given to decide whether the node will be activated or not. There is another matrix p called influence probabilities with entries

$$p_{i,j} = \begin{cases} x & 0 < x \le 1, \text{ x is the influence probability of node i to node j if } g_{i,j} \ne 0 \\ 0 & otherwise \ (if \ \ g_{i,j} = 0) \end{cases}$$

The influence $\sigma(S)$ is measured by the number of activated nodes by all the member of the S. The formulation of the problem is restricted on a given an initial seed size k. The influence maximization problems is to find such a set S of fixed size k such that spread of influence, $\sigma(S)$ is maximized *i.e.*

$$\begin{aligned} &\max \quad \sigma(S) \\ &s.t. \quad |S| = k \end{aligned} \qquad (1)$$

4. System Modeling

A) Find the reduced Influence graph:

The term $n(i)$ is defined as the set of the out neighbors of a node i and the term *outdegree(i)* = $|n(i)|$. The relative out degree which is named as *influence index* is defined as follows:

$$\beta_{index}(i) = \frac{out\deg ree(i)}{\max\limits_{v \in V} \ out\deg ree(v)} \qquad (2)$$

The idea is that the person having more friends (who have more followers in the follower-followee graph) will have higher probability of being more influential.

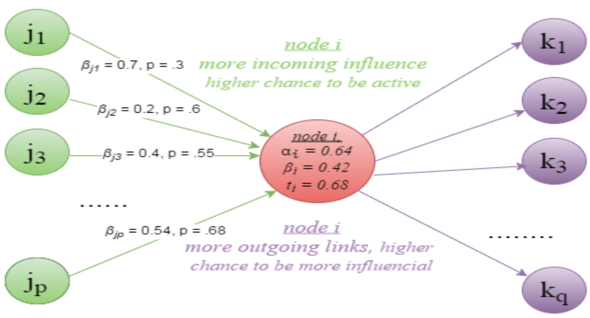

Figure 1: *Node i with activation threshold t_i = 0.68, influence index β = 0.63, activation index α = 0.64 with in degree p=11, max in degree = 17 out degree q = 8, max out degree = 19*

The term $n^{-1}(i)$ is defined as the set of in neighbors of node i and the term *indegree(i)* = $|n^{-1}(i)|$. The relative in degree which is named as *activation index* is defined as follows:

$$\alpha_{index}(i) = \frac{in\deg ree(i)}{\max\limits_{v \in V} \ in\deg ree(v)} \qquad (3)$$

The logic is that the person (who is following more people in the follower-followee graph) being friend of more people will have higher probability of being influenced.

Now in calculation the activation probability we will sum up all the weighted probabilities to a node *i* as

$$P_i = \frac{1}{|n^{-1}(i)|+1} \left\{ \sum_{j \in n^{-1}(i)} p_{j,i} * \beta_j + \alpha_i \right\} \qquad (4)$$

If P_i is no less than t_i then node *i* is activated otherwise stays inactive. All the inactive nodes and associated edges are deleted to get the reduced social graph $G'(V', E')$.

B) Influence Estimation within Heuristic limit:

Now influence value for each node $\sigma(v')$ is estimated by traversing the reduced social graph $G'(V', E')$ using simple breadth-first algorithm starting from each node $v' \in V'$. Finally k nodes with maximum influence value are selected to be in the seed set S.

5. Algorithm

Algorithm 1 Graph Reduced Technique(GRT)

Calculate out neighbor set $n(i)$ and in neighbor set $n^{-1}(i)$ of each node $i \in G$
Calculate $outdegree(i) = |n(i)|$ and $indegree(i) = |n^{-1}(i)|$ of each node $i \in G$
Calculate maximum in degree $inmax$ and maximum out degree $outmax$
Calculate $\beta_{index}(i) = \frac{outdegree(i)}{outmax}$ for all $i \in G$
Calculate $\alpha_{index}(i) = \frac{indegree(i)}{inmax}$ for all $i \in G$
Initialize $m = 0$ {Number of active nodes}
for $1 \le i \le n$ **do**
  $P_i = \frac{1}{|n^{-1}(i)|+1} [\sum_{j \in n^{-1}(i)} \beta_{index}(j) \times p_{j,i} + \alpha_{index}(i)]$
  if $P_i \ge t[i]$ **then**
    $active[i] = 1$ //Node i is activated
    $m = m + 1$
  else
    $active[i] = 0$ //Node i is NOT activated
  end if
end for
Delete all node with $active[i] = 0$ to get $G'(V', E')$
for $1 \le i \le m$ **do**
  $influence[i]$ = the number of nodes traversable from i
end for
Sort The array$influence[i]$ is descending order with index of the node
Output first k indices as seed set S

6. Evaluation

We have simulated the algorithm by a program in C on a machine featuring Core i7 2GHz, 2.5GHz processor, 4GB RAM, Windows 8 with synthesized data.

Nodes	50	60	70	80	90	100
LT	7.14	9.34	13.19	15.93	19.78	24.18
GRT	4.40	6.04	8.24	10.44	13.33	16.71

Table 1: *Running time comparison.*

We have compared the running time of our GRHA algorithm with classic Linear Threshold (LT) Method. The result is shown in the ***Table 1 & Figure 2***. We have found that the graph reduce technique (GRT) has better running time.

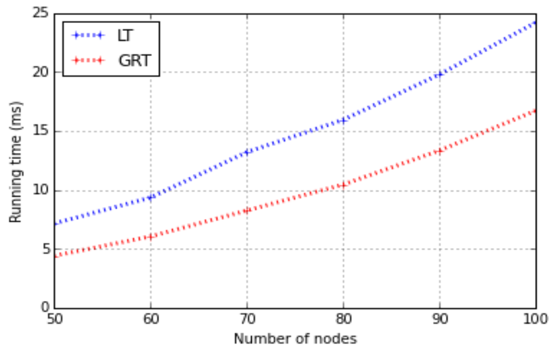

Figure 2: *Running time comparison.*

7. Conclusion

In this research we have seen that our model Graph Reduces Technique (GRT) estimates the reasonable amount of influence while two levels of heuristic approximation gives better performance in terms of running time.

We had approximation heuristic in calculating active nodes to reduce the running time but the amount of influence is approximated by the heuristic. This has scope for future research to get better amount of influence.

Acknowledgement: *This work was supported by the Industrial Core Technology Development Program(10049079, Development of Mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) *Dr. C S Hong is the corresponding author.*

8. References

[1]. Nguyen, Huy, and Rong Zheng. ***"On budgeted influence maximization in social networks."*** Selected Areas in Communications, IEEE Journal on 31.6 (2013): 1084-1094.

[2]. Li, Jingxuan, Wei Peng, Tao Li, Tong Sun, Qianmu Li, and Jian Xu. ***"Social network user influence sense-making and dynamics prediction."*** Expert Systems with Applications 41, no. 11 (2014): 5115-5124.

[3]. Zhang, Yunlong, Jingyu Zhou, and Jia Cheng. ***"Preference-based top-k influential nodes mining in social networks."*** Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on. IEEE, 2011.

[4]. Chen, Wei, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. ***"Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate."*** SIAM International Conference on Data Mining (SDM), vol. 11, pp. 379-390. 2011.

[5]. Goyal, Amit, Francesco Bonchi, Laks VS Lakshmanan, and Suresh Venkatasubramanian. ***"On minimizing budget and time in influence propagation over social networks."*** **Social Network Analysis and Mining 3, no. 2 (2013): 179-192.**

[6]. Goyal, Amit, Wei Lu, and Laks VS Lakshmanan. ***"SIMPATH: An efficient algorithm for influence maximization under the linear threshold model."*** In Data Mining (ICDM), 2011 IEEE 11th International Conference on, pp. 211-220. IEEE, 2011.

[7]. Goyal, Amit, Francesco Bonchi, and Laks VS Lakshmanan. ***"Discovering leaders from community actions."*** In Proceedings of the 17th ACM conference on Information and knowledge management, pp. 499-508. ACM, 2008.

[8]. D. Kempe, J. Kleinberg, and E. Tardos., ***"Maximizing the spread of influence through a social network"***, In Proc. of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003.

[9]. A. Talukder, R. Kamal, A. K. Bairagi, M. G. R. Alam, S. F. Abedin, Md A. Layek, N. H. Tran and C. S. Hong, ***"Rumors in the Social Network: Finding the Offenders Using Influence Maximization,"*** The Conference of Korea Computer Congress (KCC), pp. 1214-1216, 2015.

[10]. Goyal, Amit, Wei Lu, and Laks VS Lakshmanan. ***"CELF++: optimizing the greedy algorithm for influence maximization in social networks."*** Proceedings of the 20th international conference companion on World Wide Web. ACM, 2011.

[11]. Zhou, Shengfu, Kun Yue, Qiyu Fang, Yunlei Zhu, and Weiyi Liu. ***"An efficient algorithm for influence maximization under linear threshold model."*** The 26th Control and Decision Conference (2014 CCDC), China, pp. 5352-5357. IEEE, 2014.