Dynamic Selection of Activation Targets to Boost the Influence Spread in Social Networks

Cheng-Te Li[1], Man-Kwan Shan[2], Shou-De Lin[1]
[1] Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan
[2] Department of Computer Science, National Chengchi University, Taipei, Taiwan
{d98944005, sdlin}@csie.ntu.edu.tw, mkshan@cs.nccu.edu.tw

ABSTRACT

This paper aims to combine the viral marketing with the idea of direct selling to for influence maximization in a social network. In direct selling, producers can sell the products directly to the consumers without having to go through a cascade of wholesalers. Through direct selling, it is possible to sell the products in a more efficient and economic manner. Motivated by this idea, we propose a target-selecting independent cascade (TIC) model, in which during influence propagation each active node can give up to attempt to influence some neighboring nodes, named *victims*, who are hard to affect, and try to activate friends of its friends, termed *destinations*, who could have higher potential to increase the influence spread. The next question to ask is that given a social network and a set of seeds for influence propagation under TIC model, how to effectively select targets (i.e., victims and destinations) for the attempts of activation during propagation to boost the influence spread. We propose and evaluate three heuristics for the target selection. Experiments show that selecting targets based on influence probability between nodes have the highest boost of influence spread.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications – *Data mining*.

Keywords

Influence Propagation, Target Selection, Social Networks.

1. INTRODUCTION

Social structure plays a key role in the spread of information and influence for targeted marketing. The problem, *influence maximization* [3], is to find a subset of individuals (as *seeds*) such that they can eventually influence the largest number of people (i.e., *influence spread* is maximized) in a social network. Some greedy [6] and heuristic [1] methods have been proposed. Mathioudakis et al. [7] simplified the network to boost the time efficiency of finding seeds. Some variations are also proposed to tackle different real-world requirements. Leskovec et al. [6] selected a set of *sensors* such that their placements can efficiently detect the propagation of information or virus. Lappas et al. [4] found *effectors* which can cause an activation pattern very similar to the given active nodes. Li et al. [5] discovered *mediat*ors who control the bottlenecks of influence propagation when seeds want to activate some target nodes. In addition to find diverse roles of nodes under traditional influence propagation models, Cosley et al. [2] considered the temporal dynamics to devise new *sequential propagation* models. Yang and Leskovec [8] proposed new *linear influence* model for propagations in implicit networks by predicting the newly-infected nodes.

In this paper, we propose to boost the influence spread under the settings of influence maximization, by integrating the concepts of *direct selling* into *viral marketing*. Direct selling is a marketing strategy that sells products or services directly to customers through one-to-one presentation, party plan, and personal contact arrangements, instead of relying on some mediators such as wholesalers. In direct selling, salespersons are often paid not only for the sales they made, but also for the sales made by people they recruits and sold products to. In real-world scenarios, marketers tend to utilize both viral marketing and direct selling together to promote their products. Based on existing influence propagation models (e.g. *independent cascade* and *linear threshold* models) which realize the viral marketing, we add the idea of direct selling by imposing a *dynamic target selection* mechanism with the goal to maximize the influence spread under the new influence propagation model.

2. DYNAMIC TARGET SELECTION

To combine the viral marketing and direct selling, we propose a dynamic target selection mechanism and impose it on an existing influence propagation model. Note that although in this paper we focus on the *independent cascade* (IC) model [3], the target-selecting mechanism can easily be applied to other propagation models. We first describe the *Target-selecting Independent Cascade* (TIC) model, and then define the problem of influence maximization under TIC model.

Target-selecting Independent Cascade Model. In the original IC model, let S_t be the set of activated nodes at time t, each active node $u \in S_t$ has a single chance to activate its inactive neighbors v with an influence probability $p(u,v)$. If u succeeds, v will become active in round $t+1$ and be added into S_{t+1}; otherwise, u will not attempt to activate v again in the future. The new TIC model is similar to the IC model except for the step of choosing neighboring nodes for activation. The dynamic target selection mechanism is applied when choosing target neighboring nodes. The idea is that rather than attempting to activate all neighbors, we consider some neighbors are negligible because they are either un-influential or nearly impossible to be activated. Some neighbors of neighbors, which might be more influential or easier to influence, are more worthwhile to be chosen. This idea corresponds to the intuition of direct selling that sales usually target a few unacquainted *friends of friends* whose interests are similar to him, in addition to some close friends. The TIC model works as follows. For each active node $u \in S_t$, we virtually change some of its incident edges: *hiding* a subset of links to its neighbor set $N_1(u)$, and *adding* a set of *virtual* links to the set of neighbors of neighbors $N_2(u)$, $N_2(u) = \{\cup_{v \in N_1(u)} N_1(v)\} \backslash \{N_1(u) \cup \{u\}\}$. In other words, the node u will select m victim nodes to ignore from $N_1(u)$ and select m destination nodes from $N_2(u)$ to create virtual connections for activation, where $m = \delta \times |N_1(u)|$, $\delta \in [0,1]$ is the *selecting ratio*. Since the links virtually added or hidden, the network structure will not be changed. To determine the influence probability $p(u,w)$ between u and each potential destination node w, we compute $p(u,w)$ using Jaccard coefficient of the labels between u and w: $p(u,w) = |L_u \cap L_w| / |L_u \cup L_w|$, where L_u is the label set (e.g. interests, skills) associated with node u. That says, if u and w has more common attributes, u will have higher potential to affect w.

Problem Definition. Given (a) a social network $G = (V,E,P)$, where V stands for individuals and each undirected edge $(u,v) \in E$ is associated with an influence probability $p(u,v) \in [0,1]$, (b) the TIC model, (c) a set of k seed nodes S, (d) the influence

WWW 2012 Companion, April 16 – 20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

spread $R_{IC}(S)$ under the original IC model, the goal is to select *targets* (*victim* nodes to hide links and *destination* nodes to add virtual links) for each active node during the propagation such that the *boost* of influence spread between the original and new models $R_{TIC}(S) - R_{IC}(S)$ is maximized.

3. TARGET SELECTION STRATEGY

We discuss three efficient heuristics for target (victim and destination) selection, including Degree Discount, Outward Probability and Between Probability, as presented below.

Degree Discount is proposed by Chen et al. [1], which improves the degree heuristic and nearly approaches the performance of the greedy method. The degree discount heuristic considers that the degree of a node should be discounted by the number of neighbors which have been selected as seeds. We propose to select nodes with few connections to be victims and nodes with more friends to be destinations. For an active node u, we find m victims with $min_{v \in N_1(u)} dd(v)$ and m destinations with $max_{v \in N_2(u)} dd(v)$, where $dd(v) = d_v - 2s_v - (d_v - s_v) \times s_v \times p$, where s_v is the number of v's neighbors which have been activated, and p is the influence probability of the edge between v and its neighbor.

Outward Probability. An alternative to measure the influence potential of a node is to consider the influence probabilities of its incident edges toward inactive nodes. Thus, we define the average outward probability as $p_{out}(v) = (\sum_{w \in N'(v)} p(v,w))/|N'(v)|$, where $N'(v) = \{x \mid x \in N_1(v) \text{ and } x \text{ is inactive}\}$. For an active node u, we select m victims with $min_{v \in N_1(u)} p_{out}(v)$ and m destinations with $max_{v \in N_2(u)} p_{out}(v)$.

Between Probability. To select victims and destinations, the previous two methods consider the influence abilities of neighboring nodes. However, in real-world marketing scenario, one tends to give up the attempts to influence the individuals who are either too hard to activate or can hardly provide benefit to the ultimate goal. Instead, one could turn to promote the products to those prone to be activated. Therefore, we determine the targets to virtually hide and add links based on the difficulty of activation, which is controlled by the influence probability. Specifically, for an active node u, we select m victims with $min_{v \in N_1(u)} p(v,u)$ and m destinations with $max_{v \in N_2(u)} p(v,u)$, where $p(v,u)$ is computed by their common attributes, as described in the previous section.

4. EXPERIMENTAL RESULTS

We conduct experiments to demonstrate the effectiveness of our method. We compile the DBLP bibliography data to a co-authorship network containing 22,285 nodes and 49,365 edges in some premier conferences of data mining (i.e., KDD, ICDM, SDM, PAKDD, PKDD, SIGIR, WWW, and CIKM). The influence probabilities on edges are determined by the number of co-authorships between two persons. If #coauthor is higher than 100, the edge weight is 1; otherwise, it is set to #coauthor/100. One will have higher potential to activate its neighbor if they have published more co-works. We measure the effectiveness by the *Normalized Boost of Influence Spread* $(R_{TIC}(S) - R_{IC}(S))/R_{IC}(S)$. We compare the effectiveness of the presented three target selection methods while using a random method as the baseline. The results are derived by averaging 20,000 simulations of influence propagation of the IC and TIC models, in which the original seed nodes (the number of seeds k is set to be 50) are selected using the greedy method [3] to ensure the quality.

The evaluation plan consists of two parts: (1) testing the performance for the combinations of different target selection methods (selecting ratio δ=0.3), and (2) presenting the effectiveness by varying the selecting ratio δ=0.1,0.2,...,0.6, and fixing the victim selection method as the between probability. The experimental results are shown in Figure 1 and Figure 2, respectively. Note that the influence spread $R_{IC}(S)$ under the original IC model is about 310. In Figure 1, we can find that using the between probability to select victims and destinations outperforms other combinations. In Figure 2, we find that as the selecting ratio increases, the effects of boosting influence spread become more evident. These results indicates that when allowing direct selling in propagating influence, selecting individuals who are easy to activate is more effective than choosing targets based on their influence potentials.

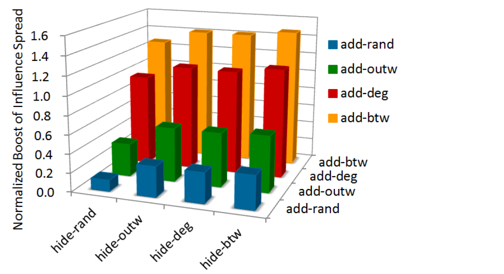

Figure 1. Results of combinations of different target selection strategies (victims to hide links and destinations to add links).

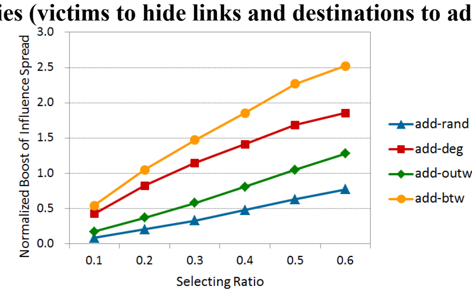

Figure 2. Results of varying the selecting ratio under different target selection methods to find destinations to add links.

5. CONCLUSION

This paper proposes to boost the influence spread by devising a dynamic target selection mechanism and imposing it on the independent cascade influence propagation model. The target-selecting mechanism realizes the idea of direct selling. Experimental results show that selecting victims and destinations can effectively maximize the boost of influence spread. Ongoing work aims at finding seed nodes for influence maximization under TIC model with the investigation of the submodular property.

6. ACKNOWLEDGEMENT

This work was supported by National Science Council, National Taiwan University, and Intel Corporation under Grants: NSC 100-2911-I-002-001, and 10R70501

7. REFERENCES

[1] W. Chen, Y. Wang, and S. Yang. Efficient Influence Maximization in Social Networks. In *KDD* 2009.

[2] D. Cosley, D. Huttenlocher, J. Kleinberg, X. Lan, and S. Suri. Sequential Influence Models in Social Networks. In *ICWSM* 2010.

[3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the Spread of Influence through a Social Network. In *KDD* 2003.

[4] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila. Finding Effectors in Social Networks. In *KDD* 2010.

[5] C.-T. Li, S.-D. Lin, and M.-K. Shan. Finding Influential Mediators in Social Networks. In *WWW* 2011.

[6] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective Outbreak Detection in Networks. In *KDD* 2007.

[7] M. Mathioudakis, F. Bonchi, C. Castillo, A. Gionis, and A. Ukkonen. Sparsification of Influence Networks. In *KDD* 2011.

[8] J. Yang and J. Leskovec. Modeling Information Diffusion in Implicit Networks. In *ICDM* 2010.