Finding Influential Seed Successors in Social Networks

Cheng-Te Li[1], Hsun-Ping Hsieh[1], Shou-De Lin[1], Man-Kwan Shan[2]
[1] Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan
[2] Department of Computer Science, National Chengchi University, Taipei, Taiwan
{d98944005, d98944006, sdlin}@csie.ntu.edu.tw, mkshan@cs.nccu.edu.tw

ABSTRACT
In a dynamic social network, nodes can be removed from the network for some reasons, and consequently affect the behaviors of the network. In this paper, we tackle the challenge of finding a successor node for each removed seed node to maintain the influence spread in the network. Given a social network and a set of seed nodes for influence maximization, the problem is to effectively choose successors to inherit the jobs of initial influence propagation when some seeds are removed from the network. To tackle this problem, we present and discuss five neighborhood-based selection heuristics, including degree, degree discount, overlapping, community bridge, and community degree. Experiments on DBLP co-authorship network show the effectiveness of devised heuristics.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications – *Data mining*.

Keywords
Influence Maximization, Influential Successors, Social Networks.

1. INTRODUCTION
Social network plays a key role in the spread of information and influence for target marketing and immunization setting. The problem, *influence maximization* [4], is to find a subset of influential individuals (as *seeds*) such that they can eventually influence the largest number of people in a social network. Some greedy [7][9] and heuristic [1][2] methods are proposed to effectively and efficiently solve this problem. Mathioudakis et al. [8] simplified the network to boost the time efficiency of finding seeds. Goyal et al. [3] mined the past propagation traces to identify seeds instead of simulation-based methods. Some variations are also proposed to tackle different real-world requirements. Leskovec et al. [7] selected a set of social *sensors* such that their placements can efficiently detect the propagation of information or virus in a social network. Lappas et al. [5] found a set of *effectors* who can cause an activation pattern as similar as possible to the given active nodes in a social network. Li et al. [6] discovered a set *of mediat*ors which control the bottlenecks of influence propagation when seed nodes want to activate some target nodes.

In this paper, we reveal another crucial problem for influence maximization in finding influential *successors* in a social network. Considering a node v, which is one of the seeds for the spread of influence, is removed from the network due to some reason. To maintain the effectiveness of propagation, one usually has to choose another seed as a replacement. This research tries to identify a neighboring node of v as seed, which has the best chance to maintain the overall degree of propagation, In other words, when one seed node along with its incident links disappear, in order to maintain the influence spread, we would need a strategy to select the successor node from its neighbors as a new seed. . Note that in this paper we focus on finding successors from the neighbors of the removed seed, rather than from all the remaining nodes in the network. It is reasonable because people tend to transfer the jobs to somebody who is close to them.

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

There are potential applications in real world. For example, in viral marketing, if a salesman retires or is laid off, the manager might need to find a colleague of this person to transfer the duty of propagation can to maximize the influence spread. In blogosphere, if the favorite blog of a user is closed, the user tends to obtain needed information from one of the hyperlinked blogs.

Preliminary. First, we adopt the *Independent Cascade* (IC) model [4] for information propagation. In IC model, at time step t each active node u has a single chance to activate its inactive neighbors v with a pre-determined probability $p(u,v)$. If u succeeds, v will become active in step $t+1$; otherwise, u will not attempt to activate v again. Second, based on IC model, the original influence maximization problem focuses on discovering a set of k nodes (i.e., seeds) S such that the expected number of activated nodes incurred by S, referred to as $R(S)$, is maximized. Third, we define a *neighbor-selection function* $f(v)$, which is a one-to-one mapping from one node to another, to choose the successor node of a removed node v. Note that the neighbor set of a node v is defined as the set of one-step linked nodes from v, denoted by $N(v)$.

Problem Definition (*Influential Successors Problem*)**.** 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 $p(u,v)\in[0,1]$ as weight, (b) the IC model, (c) a set of k seed nodes S and the corresponding influence spread $R(S)$, and (d) a subset of S, denoted by T ($T\subseteq S$), which is removed from G, the goal is to find a new set T' ($T'\cap T=\emptyset, and\ |T'|=|T|$) as successor nodes such that $R(S')$ is maximized, where $S'=T'\cup(S\backslash T)$, and each $t'\in T'$ needs to be determined by $f(t)=t', t\in T$.

Our goal aims to devise the neighbor-selection function f to find the successor set T' mapped from T such that $R(S')$ is maximized. We present and examine a series of heuristic methods to develop the neighbor-selection function f. Each heuristic method possesses its own physical meaning and is able to determine the successor for removed node efficiently. We will conduct experiments to show and compare the effectiveness of the discussed methods.

2. METHODOLOGY
We present five neighbor-selection heuristics, including Degree, Degree Discount, Overlapping, Community Bridge, and Community Degree, as discussed in the following.

Degree is a commonly used heuristic to find seeds for influence maximization. Kempt et al. [4] reported that high-degree nodes can outperform other centrality-based heuristics. Therefore, we select the neighboring node with the highest degree value as the successor for the removed seed t, i.e., $f(t)=max_{u\in N(t)}d(u)$.

Degree Discount is proposed by Chen et al. [2], 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 neighboring nodes which have been selected as seeds. Applying to devise the neighbor-selection function with the IC model, the successor of the removed seed t will be $f(t)=max_{u\in N(t)}dd(u)$, where the degree discount value of node u is $dd(u) = d_u - 2s_u - (d_u - s_u) \times s_u \times p$, where s_u is the number of neighbors of u which have been selected as seeds, and p is the influence probability on the edge between u and its neighboring node.

Overlapping. We consider the social acquaintance of two individuals in this heuristic. The idea is that it is more reasonable and reliable to transfer the jobs to the most familiar friend. The familiarity can be measured by the number of common friends of two individuals. From the other perspective, if the position of a person is critical in the network structure, it is intuitive to find the substitute who plays the similar topological role if such person disappears. Therefore, we find the successor of a removed seed t based on the number of overlapping neighbors, i.e., $f(t) = max_{u \in N(t)} intersect(N(u), N(t))$, where $intersect()$ counts the number of common neighbors.

Community Bridge. We assume that a node with connections to diverse communities, called *bridge*, can benefit the spread of influence, comparing to the node in the central position within a certain community. Thus, we aim to estimate the expected number of communities to which a node is attached. We consider that the extent of a node being a community bridge is determined by the connectivity of its neighbors. If the neighbors of a node connect with each other, they tend to become a clique and belong to the same community. On the contrary, if the neighbors have few links to one another, the node could be the bridge connecting communities. We select the successor of a removed seed t by $f(t) = max_{u \in N(t)} CB(u)$, where $CB(u) = \sum_{v \in N(u), u \neq v} 1/(1 + I(u, v))$, $I(u, v) = 1$ if there is a link between u and v and 0 otherwise.

Community Degree. Based on the community bridge, we further assume a node with higher degree and higher community-bridging power could produce more influence spread. Thus, we devise the neighbor-selection function as $f(t) = max_{u \in N(t)}\{nor(d(u)) + nor(CB(u))\}$, where $nor()$ is a normalization function.

3. EXPERIMENTAL RESULTS

We conduct the experiments to demonstrate the effectiveness and efficiency of our method. We compile the DBLP bibliography data to a connected co-authorship network, which contains 22,285 nodes and 49,365 edges in some recent premier conferences of data mining (i.e., KDD, ICDM, SDM, PAKDD, PKDD, SIGIR, WWW, and CIKM). The probabilistic weights on edges are determined by the number of co-authorship between two persons. If #coauthor is higher than 100, the edge weight is as 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 Decay of Influence Spread* $(R(S) - R(S'))/R(S)$. We compare the effectiveness of the presented five methods while using a random successor-selection method as the baseline. The results are derived by averaging 20,000 simulations of influence propagation of the IC model, in which the original seed nodes (the number of seeds k is set to be 100) are selected using the greedy method [4] to ensure the quality of seeds. we conduct the experiments under two different strategies of seed removals: (a) *random* removal, and (b) *descending* removal based on the influence spread of seed. The former is closer to real-world scenarios while the latter is experimental one.

By varying the number of removed seeds (i.e., the number of successors) from 1 to 50, the results are shown in Figure 1 and 2. For the random removal of seeds, we can find the community degree outperforms others, while the traditional wisdom of transferring jobs to well-acquainted persons (i.e., overlapping) performs the worst. For the descending removal of seeds, similar results are reported except for the two degree-based methods are competitive with the community degree. We think this is due to the fact that seeds removed in the descending order are really influential, and thus their neighbors have higher potential to connect with other influential nodes. Consequently, neighbor-selection in itself is easy to find nodes connecting to diverse communities. In short, selecting successors with community-based methods is effective, especially for the realistic case of random removals of seeds.

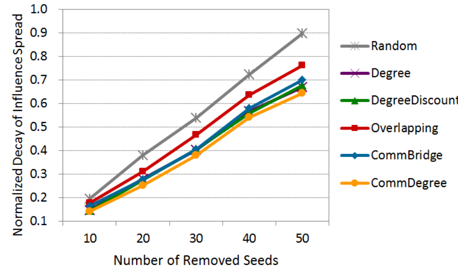

Figure 1. Results of random removals of seeds.

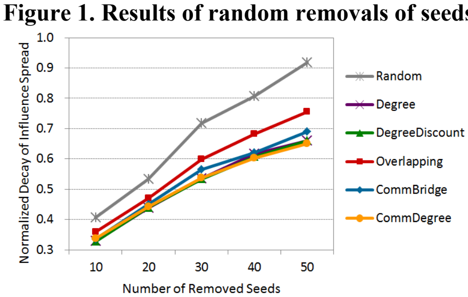

Figure 2. Results of descending removals of seeds.

4. CONCLUSIONS

This paper introduces and defines the problem of finding successors while achieving influence maximization when some seed nodes are removed from the social network. We present and discuss five heuristics to efficiently identify the best successors for removed seeds. Ongoing work focuses on finding global successors from all remaining nodes in the network, instead of the local neighbors in this paper.

5. 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

6. REFERENCES

[1] W. Chen and C. Wang. Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks. In *KDD* 2010.

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

[3] A. Goyal, F. Bonchi, and L.V.S. Lakshmanan. A Data-Based Approach to Social Influence Maximization. In *VLDB* 2012.

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

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

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

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

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

[9] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based Greedy Algorithm for Mining Top-k Influential Nodes in Mobile Social Networks. In *KDD* 2010.