Scalable Dynamic Networks with Influential Node Tracking Under An Interchange Greedy Algorithm

R.Leelavathi[1], R.Indra[2], P.Kiruthika[3]
[1, 2, 3] Department of Computer Science
[1, 2, 3] Shrimati Indira Gandhi College, Trichy, Tamilnadu, India.

Abstract- *Real world marketing campaign utilizing the world-of-mouth effect usually lasts a long time, where multiple sets of influential users need to be mined and targeted at different times to fully utilize the power of viral marketing. As both social network structure and strength of influence between individuals evolve constantly, it requires tracking the influential nodes under a dynamic setting. To address the above problem, we explore the Influential Node Tracking (INT) problem as an extension to the traditional Influence Maximization [6], problem under dynamic social networks. While Influence Maximization problem aims at identifying a set of k nodes to maximize the joint influence under one static network, INT problem focuses on tracking a set of influential nodes that keeps maximizing the influence as the network evolves. Utilizing the smoothness of the evolution of the network structure, we propose an efficient algorithm, Upper Bound Interchange Greedy(UBI) to solve the INT problem. Instead of constructing the seed set from the ground, we start from the influential seed set we find previously and implement node replacement to improve the influence coverage. Furthermore, by using a fast update method to maintain an upper bound on the node replacing gain; our algorithm can scale to dynamic social networks with millions of nodes.*

Keywords- Influence maximization, influential nodes tracking, social network, Scalable algorithm.

I. INTRODUCTION

The processes and dynamics by which information and behaviors spread through social networks have long interested scientists within many areas. Understanding such processes have the potential to shed light on the human social structure, and to impact the strategies used to promote behaviors or products. While the interest in the subject is long-standing, recent increased availability of social networks and information diffusion data (through sites such as Face book, Twitter, and LinkedIn) has raised the prospect of applying social network analysis at a large scale to positive effect. Influence maximization, is the problem of selecting a small set of seed nodes in a social networks, such that their overall influence on other nodes in the network, defined according to particular models of diffusion, is maximized. For example, links appear and disappear when users follow/unfollow others in Twitter or friend/unfriend others in face book. Moreover, The strength of influence also keeps changing, as you are more influenced by your friends who you contact frequently, while the influence from a friend a friend usually dies down as time may lead to poor influence coverage after the evolution of social network, which suggests that using one static set as seeds across time could lead to unsatisfactory performance. It turns out that targeting at different nodes at different time becomes essential for the success of viral marketing. We proceed to illustrate the idea of considering the dynamic perspect in influence maximization [7].

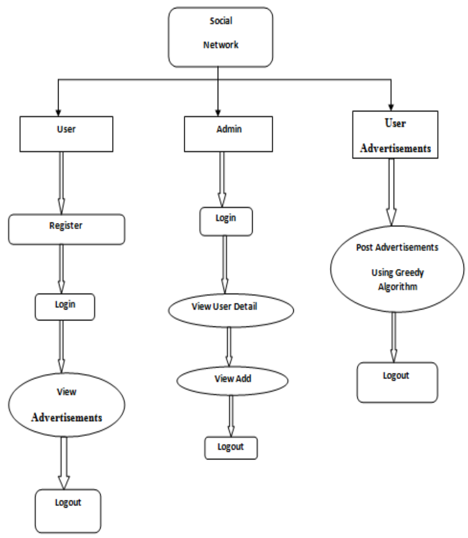

Figure 2.

II. BACKGROUND AND RELATED WORK

A. Temporal analysis of telecom call graphs:

Real world graphs like call graphs, email communication graphs [1], are temporal in nature in which edges between nodes exist only for a limited span of time. Temporal analysis can lead to new insights such as densification laws and shrinking diameters. In this paper we have analyzed temporal properties like diameter, clustering coefficient, number of calls and other properties of Call Detail

Records of more than 1 billion calls. To analyze the data we used moving windows at multiple time scales, such as day-night windows, weekday-weekend windows, etc. We also analyzed the number of unique calls with respect to days of week which lead to the rather surprising conclusion that no day of week dominates other days in terms of highest number of unique calls. To best of our knowledge, this is the first study of temporal properties on telecom call graphs with this particular set of splits.

B. Learning and Predicting the Evolution of Social Networks:

With the increasing availability of large social network data, there is also an increasing interest in analyzing how those networks evolve over time.1 Traditionally, the analysis of social networks has focused only on a single snapshot of a network. Researchers have already verified that social networks follow power-law degree distribution, 2 have a small diameter, and exhibit small-world structure3 and community structure.4 Attempts to explain the properties of social networks have led to dynamic models [2], inspired by the preferential attachment model, 5 which assumes that new network nodes have a higher probability of forming links with high-degree nodes, creating a “rich-get-richer” effect. Recently several researchers have turned their attention to the evolution of social networks at a global scale. For example, Jure Leskovec and his colleagues empirically observed that networks [5], become denser over time, in the sense that the number of edges grows super linearly with the number of nodes.6 Moreover, this densification follows a power-law pattern. They reported that the network diameter often shrinks over time, in contrast to the conventional wisdom that such distance measures should increase slowly as a function of the number of nodes. Although some effort has been devoted to analyzing global properties of social network evolution, not much has been done to study graph evolution at a microscopic level. A first step in this direction investigated a variety of network formation strategies,7 showing that edge locality plays a critical role in network evolution.

C. UBLF: An Upper Bound Based Approach to Discover Influential Nodes in Social Networks:

Influence maximization, defined as finding a small subset of nodes that maximizes spread of influence in social networks, is NP-hard under both Linear Threshold (LT) and Independent Cascade (IC) models, where a line of greedy/heuristic algorithms have been proposed. The simple greedy algorithm achieves an approximation ratio of $1-1/e$. The advanced CELF algorithm, by exploiting the sub modular property of the spread function, runs 700 times faster than the simple greedy algorithm on average. However, CELF is still inefficient [4], as the first iteration calls for N times of spread estimations (N is the number of nodes in networks), which is computationally expensive especially for large networks. To this end, in this paper we derive an upper bound function for the spread function [3]. The bound can be used to reduce the number of Monte-Carlo simulation calls in greedy algorithms, especially in the first iteration of initialization. Based on the upper bound, we propose an efficient Upper Bound based Lazy Forward algorithm (UBLF in short), by incorporating the bound into the CELF algorithm. We test and compare our algorithm with prior algorithms on real world data sets. Experimental results demonstrate that UBLF, compared with CELF, reduces more than 95% Monte-Carlo Simulations and achieves at least 2−5 times speed-raising when the seed set is small.

D. Using Crowd sourced Data in Location-based Social Networks to Explore Influence Maximization:

Online social networks have gained significant popularity recently. The problem of influence maximization in online social networks has been extensively studied. However, in prior works, influence propagation in the physical world, which is also an indispensable factor, is not considered. The Location- Based Social Networks (LBSNs) are a special kind of online social networks in which people can share location-embedded information [4]. In this paper, we make use of mobile crowd sourced data obtained from location-based social network services to study influence maximization in LBSNs. A novel network model and an influence propagation model taking influence propagation in both online social networks and the physical world into consideration are proposed. An event activation position selection problem is formalized and a corresponding solution is provided. The experimental results indicate that the proposed influence propagation model is meaningful and the activation position selection algorithm has high performance.

III. ALGORITHM

GREEDY ALGORITHM

A greedy algorithm is an algorithmic paradigm that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum. In many problems, a greedy strategy does not in general produce an optimal solution, but nonetheless a greedy heuristic may yield locally optimal solutions that approximate a global optimal solution in a reasonable time.

Algorithm 1. Greedy$(G = (V, E), k)$

initialize $S = \emptyset$
for $i = 1$ to k do
  $v^* = \arg\max_{v \in V - S} \{\sigma(S + \{v\}) - \sigma(S)\}$
  $S = S + \{v^*\}$
end for
Output S

Figure 2.

The exact computation of the marginal gain has shown to be #P-hard, though approximate estimation can be achieved via multiple times of Monte-Carlo simulations, which are extremely inefficient for large networks. To tackle the inefficiency of the above greedy algorithm, numerous methods are proposed. Though with much better efficiency, the algorithms may still spend at least minutes on a network with millions of nodes.

SENTIMENT ANALYSIS

Sentiment analysis (also known as semantic inference) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

CLASSIFICATION

There are two major forms of data analysis that can be used for extracting models describing important classes or to predict future data trends. These two major forms are as follows

- Classification
- Prediction

models predict categorical class and prediction models predict continuous valued functions. For example, we can build classification model to categorize bank loan applications as either safe or risky, or a prediction models to predict the expenditures in dollars of prospective customers on computer equipment given their income and occupation. Following are the examples of cases where the data analysis task is Classification.

Examples

- A bank loan officer wants to analyze the data in order to know which customer (loan applicant) is risky or which are safe.
- A marketing manager at a company needs to analyze a customer with a given profile, who will buy a new computer.
- In both of the above example, a model or classifier is constructed to predict the categorical labels. These labels are unsafe or safe for loan application data and sure or not for marketing data.

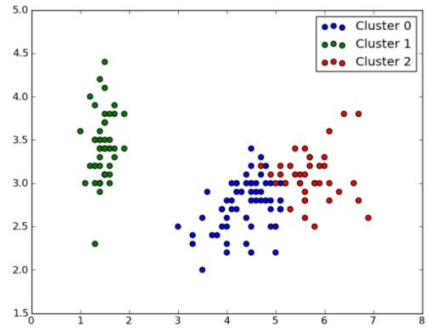

Figure 3.

IV. CONCLUSION

In this paper, we explore a novel problem, namely Influential Node Tracking problem, as an extension of Influence Maximization problem to dynamic networks, which aims at tracking a set of influential nodes dynamically such that the influence spread is maximized at any moment. We propose an efficient algorithm greedy to solve the problem based idea of the Interchange Greedy method. We utilize the upper bound on node replacement gain to accelerate the process. Moreover, an efficient method for updating the upper bound is proposed to handle the evolution of the network structure. Extensive experiments on three real social networks show that our method outperforms state-of-the-art baselines in terms of both influence coverage and running time. Then we propose greedy algorithm that improves the computation of the upper bound and achieves better influence spread. In future work, we would like to generalize our greedy algorithm to track influential nodes under the other widely adopted diffusion model, Linear Threshold model under dynamic networks. Moreover, it will be interesting if we can combine our work. That is to track a series of influential nodes where the diffusion process is also carried out under a dynamic network instead of the static snapshot graph.

REFERENCES

[1] J. Leskovec, J. M. Kleinberg, and C. Faloutsos, “Graph evolution: Densification and shrinking diameters”, 2007.

[2] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, “Microscopic evolution of social networks,” Knowledge Discovery Data Mining, 2008.

[3] C. Zhou, P. Zhang, J. Guo, X. Zhu, and L. Guo, “UBLF: An upper bound based approach to discover influential nodes in social networks,” Conference in Data Mining, 2013.

[4] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in Process of ACM SIGKDD 2002.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance, “Cost-effective outbreak detection in networks ”Conference in Knowledge Discovery Data Mining, 2007.

[6] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, and K. Xie, “Simulated annealing based influence maximization in social network,” 2011.

[7] K. Jung, W. Heo, and W. Chen, “IRIE: Scalable and robust influence maximization in social networks,” Conference in Data Mining, 2012.