Journal of Engineering Sciences
Vol 10, Issue 11, Nov /2019
ISSN NO:0377-9254

UPPER BOUND GREEDY METHOD TO INFLUENTIAL MAXIMIZATION IN DYNAMIC ENVIRONMENTS

[1]P.ESWARAIAH, [2]ALAPTI ROHITH
[1]Associate Professor, Dept. of CSE, PBR VITS, Kavali, AP, India.
[2]M.Tech, Dept. of CSE, PBR VITS, Kavali, AP, India

Abstract – 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 this problem, this project explores the Influential Node Tracking (INT) problem as an extension to the traditional Influence Maximization problem (IM) 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, this project proposes an efficient algorithm, Upper Bound Interchange Greedy (UBI) and a variant, UBI+. Instead of constructing the seed set from the ground, this project 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 by calculating the marginal gain of nodes, our algorithm can scale to dynamic social networks with millions of nodes. Empirical experiments on three real large-scale dynamic social networks show that our UBI and its variants, UBI+ achieves better performance in terms of both influence coverage and running time.

Keywords – influential node tracking, social Network, Influential Maximization.

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 network and information diffusion data (through sites such as Facebook, Twitter, and LinkedIn) has raised the prospect of applying social network analysis at a large scale to positive effect.

One particular application that has been receiving interest in enterprises is to use word-of-mouth effects as a tool for viral marketing. Motivated by the marketing goal, mathematical formalizations of influence maximization have been proposed and extensively studied by many researchers [1], [2], [3], [4], [5], [6], [7], [8], [9]. Influence maximization is the problem of selecting a small set of seed nodes in a social network, such that their overall influence on other nodes in the network, defined according to particular models of diffusion, is maximized.

Marketing campaign is usually not a one-time deal; instead enterprises carry out a sustaining campaign to promote their products by seeding influential nodes continuously. Often, a marketing campaign may last for months or years, where the company periodically

allocates budgets to the selected influential users to utilize the power of the word-of-mouth effect. Under this situation, it is natural and important to realize that social or information networks are always dynamics, and their topology evolves constantly over time [10]. For example, links appear and disappear when users follow/unfollow others in Twitter or friend/unfriend others in Facebook. 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 usually dies down as time elapses if you do not contact with each other. As a result, a set of nodes influential at one 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.

II. PROPOSED SYSTEM

For real dynamic social network, it is unlikely to have abrupt and drastic changes in graph structure in a short period of time. As a result, the similarity in structure of graphs from two consecutive snapshots could lead to similar seed sets that maximize the influence under each graph. Based on the above idea, we propose UBI algorithm for the INT problem, in which we find the seed set that maximizes the influence under Gt+1 based on the seed set St we have already found for graph Gt. Instead of constructing the seed set for graph Gt+1 from the ground, we start with St and continually update by replacing the nodes in St to improve the influence coverage. Our algorithm first uses an initial set and several rounds of interchange heuristic to maximize the influence, as mentioned in the paper. So the interchange heuristic obviously works on a snapshot graph. When extended to the dynamic graph, our algorithm only needs to interchange for a few more rounds after each time window and can achieve a faster update.

a) Interchange Heuristic

We use the Interchange Heuristic proposed in as our strategy to replace the nodes inS^t. Starting from an arbitrary set$S \subseteq V$, Interchange Heuristic means to find a subset $S' \subseteq V$ that differs from S by one node and has the same cardinality.

It has been shown by Nemhauser et al. in that applying Interchange Heuristic to monotone sub modular function until no further improvement is possible leads to a solution with approximation guarantee 1/2.

However, it remains to specify how we should choose set S' in the Interchange Heuristic. In this work, we choose S' in order to maximize the gain achieved via the replacement for any fixed$v_s \in S$. Let $\delta_{v,v_S}(S)$ be the replacing gain by changing from $v_s \in S$ to $v \in V - S$. Let $v^* = \arg\max_v \delta_{v,v_S}(S)$, we choose $S' = S - v_s + v^*$.This strategy needs to evaluate the gain by replacing v_s with any node in $v \in V - S$, which calls for $|V - S|$times of influence estimation. The calculation by running Monte- Carlo simulations is unaffordable even for network with moderate size. Inspired by the UBLF optimization proposed, we use the upper bound on replacing gain to reduce a large number of influence estimations.

Assume that we have already calculated the upper bound on replacing gain $\delta_{v,v_S}(S)$ for any node $v \in V - S$. Let the upper bound on the replacement gain be $\delta_{u,v_S}(S)$, then if for node u such that $\delta_{u,v_S}(S) \leq \delta_{v,v_S}(S)$, the expensive computation of replacing gain for node u becomes unnecessary as its gain is guaranteed to be less or equal than that for node v. The

computation of $\delta_{u,v_S}(S)$ will be presented in next section.

We use the subroutine in Algorithm 1 to carry out the Interchange Heuristic for any fixed $v_s \in S$. If the largest replacing gain δ_{v,v_S} is less than a given threshold with $\gamma \geq 0$, we stop to find another vs for interchange (line 5-7). This reduces the computations for the case of insignificant improvements and accelerates the process of interchange. It remains to show how v_s is selected to complete the description of our method. It turns out that we utilize the derived bounds to choose v_s with the largest possible replacing gain, namely, $v_s = \arg\max_{v_s \in S}\{\max_{v \in V-S} \bar{\delta}_{v,v_s}(S)\}$.

Algorithm 1 Interchange($G = (V, E), S, v_s, \bar{\delta}_{\cdot,v_s}(S)$)

1: Set $\delta_{v,v_s} \leftarrow \bar{\delta}_{v,v_s}(S), v \in V - S$
2: Set $cur_v \leftarrow$ false, $v \in V - S$
3: while true do
4: $\quad v^* = \arg\max_{v \in V-S}\{\delta_{v,v_s}\}$
5: $\quad$ if $\delta_{v^*,v_s} \leq \gamma\sigma(S)$ then
6: $\quad\quad$ break
7: $\quad$ end if
8: $\quad$ if cur_{v^*} then
9: $\quad\quad S \leftarrow S - v_s + v^*$
10: $\quad\quad$ break
11: $\quad$ else
12: $\quad\quad \delta_{v^*,v_s} \leftarrow \sigma(S - v_s + v^*) - \sigma(S)$
13: $\quad\quad cur_{v^*} \leftarrow$ true
14: $\quad$ end if
15: end while
16: Output S

With the interchange strategy defined above, we present our Upper Bound Interchange Greedy, in short UBI as Algorithm 2.

Algorithm 2 UBI($G = (V, E), S$)

1: Compute $\bar{\delta}_{v,v_s}(S)$ for $v \in V - S, v_s \in S$
2: for $i = 1$ to $|S|$ do
3: $\quad v_s^* = \arg\max_{v_s \in S}\{\bar{\delta}_{\cdot,v_s}(S)\}$
4: $\quad S \leftarrow$ Interchange($G, S, v_s^*, \bar{\delta}_{\cdot,v_s}(S)$)
5: $\quad$ Update $\bar{\delta}_{v,v_s}(S)$ for any $v \in V - S, v_s \in S$ according to the interchange result
6: end for
7: Output S

It should be noticed that instead of carrying out node replacement until no further improvement is possible, we apply at most jSj rounds of replacement in our implementation.

While sacrificing the theoretical guarantee, we significantly improve the efficiency of our method, as it may take an exponential number of replacements until no improvement exists. As we will illustrate in the empirical experiments, the proposed method achieves comparable results as the hill-climbing greedy algorithm where the 1□1=e approximation is guaranteed.

Time and space complexities Let n (resp. m) be the number of nodes (resp. edges) in social network G. The first lines of Algorithm 3 take O(n) time. For the entire for loop, the dominant cost is on interchanging the nodes in seed set. In the worst case, the algorithm 2 needs O(n) to explore all nodes in the graph. Thus the running time is O(kn) for the for loop, which is also the time complexity of Algorithm 3. In addition to the input social graph, Algorithm 2 only needs to store bounds and replacement gain for each node, the space needed by which is O(n). Thus the space complexity of Algorithm 2 is O(n+m), which is dominated by the input of social network.

III. RESULT AND DISCUSSION

A) Experiment Results of UBI

Influence coverage and running time on real dynamic networks. We first present our main result on comparing our UBI algorithm to other baseline methods on three real world dynamic networks. For Mobile network, we set the window size to one hour while the time difference is set to two minutes. For both HepPh and HepTh network, we set the window size to three years and the time difference to one month. Moreover, we choose the seed size k as 30.

The results on influence coverage of the selected seed sets for each snapshot graph are shown in Figure 6.1 and Figure 6.2. As Greedy is too slow to finish within a reasonable time, we do not include Greedy on Mobile dataset.

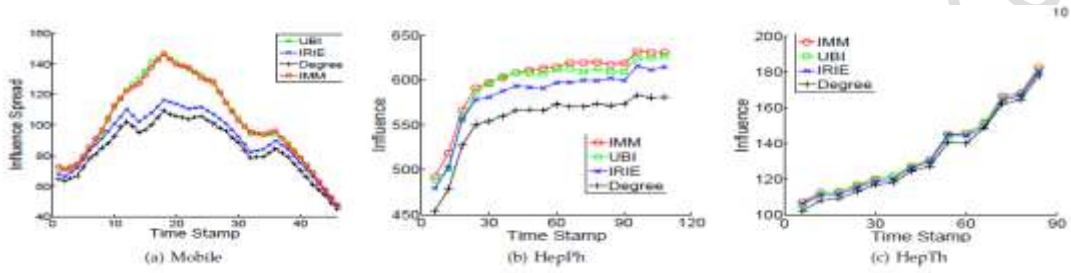

Fig 6.1: Influence Tracking Results under UA model with k = 30.

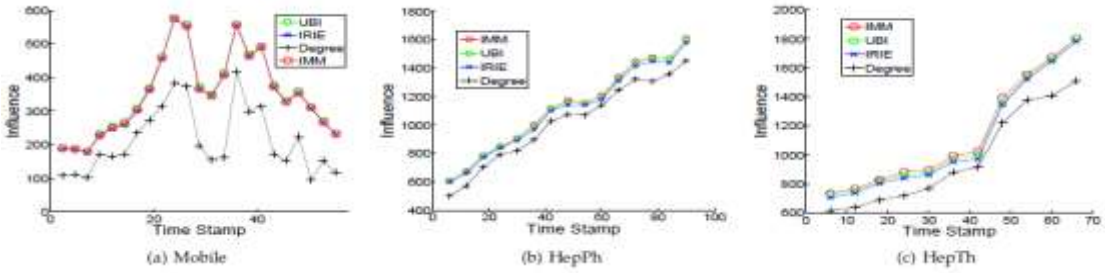

Fig 6.2: Influence Tracking Results under DWA model with k = 30.

We also calculate the average influence spread over all snapshot graphs for all three networks and present the results in Table 1 and Table 2 for better comparison. For the above results, we can easily find that UBI algorithm results in better influence coverage compared with IRIE averaged over all datasets. As our method has a little loss of accuracy on influence to achieve fast tracking, UBI achieves slightly lower influence compared to IMM and Greedy

.

Table 1. Average influence spread in UA Model

Dataset	Mobile	Hepph	Hepth
Greedy		71.49	65.49
IMM	95.42	71.24	65.43
IRIE	87.99	70.64	64.88
UBI	94.36	71.02	64.36

Table 2. Average influence spread in DWA model

Dataset	Mobile	Hepph	Hepth
Greedy		124.35	74.81
IMM	1053.78	124.33	74.48
IRIE	943.69	122.94	74.32
UBI	1033.74	123.79	74.35

We can easily find that Greedy is extremely slow that it even fails to finish on the largest Mobile network. Though performing well in influence coverage, IRIE performs well in running time on Hepth and Hepph but bad on Mobile with million nodes and edges. IMM performs better than IRIE on Mobile. However, our method, UBI, achieves consistent lowest running time on all the three networks with comparable influence coverage compared with IRIE, IMM and Greedy algorithm. UBI is about 30 times faster than IRIE and 2 times faster than IMM. Notice that UBI achieves insignificant improvement compared with IRIE and IMM under the last two dataset. This is because they are in relatively small size. At the same time, as the size of networks grows, UBI can scale to networks like Mobile with million nodes and edges as shown in the following experiment.

B) Experiment Results of UBI+

Influence coverage on dynamic networks. We present our result on comparing our improved UBI algorithm, UBI+ to UBI on three real-world dynamic networks. For Mobile network, we set the window size to one hour while the time difference is set to two minutes. For both HepPh and HepTh network, we set the window size to three years and the time difference to one month. Moreover, we choose the seed size k as 30. We calculate the average influence spread over all snapshot graphs for all three networks and present the results in Table 3 and Table 4. For the above results, we can easily find that our UBI+ algorithm achieves a better influence spread than UBI. Notice that UBI+ merely reaches about 2% and 1% better on the Hepph and Hepth dataset, this is because that UBI already performances very close to the influence spread upper bound(which is also the Greedy algorithm's result), so UBI+ only reaches an influence much closer to the theoretically influence bound. However, UBI+ get a 10% improvement in Mobile dataset and this shows that our new algorithm significantly improves the result in large datasets. Similar to the experiment results of UBI, the average influential users coverage of UBI+ is are 154, 119, 143 for Mobile, Hepph and Hepth dataset.

Table 3. Average influence spread in UA Model

Dataset	Mobile	Hepph	Hepth
IMM	95.42	71.24	65.43
UBI	94.36	71.02	64.36
UBI+	95.01	71.15	65.04

Table 4. Average influence spread in DWA model

Dataset	Mobile	Hepph	Hepth
IMM	1053.78	124.33	74.48
UBI	1033.74	123.79	74.35
UBI+	1045.15	124.01	74.42

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 UBI to solve the INT 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 UBI+ algorithm that improves the computation of the upper bound and achieves better influence spread.

V. FUTURE ENHANCEMENT

As a direct future work, we would like to generalize our UBI 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] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social network," in KDD, 2009, pp. 199–208.

[2] P. Domingos and M. Richardson, "Mining the network value of customers," in KDD, 2001, pp. 57–66.

[3] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of inffluence through a social network," in KDD, 2003, pp. 137–146.

[4] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in PKDD, 2006, pp. 259–271.

[5] W.Yu, G.Cong, G.Song, and K.Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in KDD, 2010, pp. 1039–1048.

[6] W.Chen, C.Wang, and Y.Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in KDD, 2010, pp. 1029–1038.

[7] W. Chen,W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," in AAAI, 2012.

[8] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in Data Mining (ICDM), 2010 IEEE 10th International Conference on. IEEE, 2010, pp. 88–97.

[9] N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha, "Scalable influence estimation in continuous-time diffusion networks," in Advances in neural information processing systems, 2013, pp. 3147–3155.

[10] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in KDD, 2005, pp. 177–187.

AUTHORS

P.Eswaraiah working as Associate Professor in the department of CSE, PBR Visvodaya Institute of Technology & Science, Kavali with total teaching experience of 21 years. Now He is pursuing Ph.D at SV University in DATA WAREHOUSING. He has guided 35 P.G. and 110 U.G. students.

Alapti Rohith has received her B.Tech degree in CSE from JNTU, Anantapur in 2014 and M.Tech degree in CSE from PBR VITS, affiliated JNTU, Anantapur in 2018.