International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
ISSN (Online) 2349-6967 *Volume 2 , Issue 1(Jan-Feb 2015), PP1-5*

A Greedy Algorithm Approach for Mobile Social Network: Review

Smita Bhosale [1], Dhanshree Kulkarni [2]
[12] *(Dr.D.Y.Patil College Of Engineering, Ambi, Pune, Maharashtra, India)*

Abstract:- With the proliferation of mobile devices and wireless technologies, mobile social network systems square measure more and more out there. A mobile social network plays an important role because the unfold of data and influence within the sort of "word-of-mouth". it's a elementary issue to find a set of influential people in an exceedingly mobile social network such targeting them ab initio (e.g. to adopt a replacement product) can maximize the unfold of the influence (further adoptions of the new product). The matter of finding the foremost influential nodes are sadly NP-hard. It's been shown that a Greedy rule with demonstrable approximation guarantees will provide smart approximation; but, it's computationally valuable, if not preventative, to run the greedy rule on an outsized mobile social network. During this paper, a divide-and-conquer strategy with parallel computing mechanism has been adopted. We have a tendency to first propose a rule referred to as Community-based Greedy rule for mining top-K influential nodes. It encompasses 2 components: dividing the large- scale mobile social network into many communities by taking under consideration data diffusion and choosing communities to find influential nodes by a dynamic programming. Then, to additional improve the performance, we have a tendency to place the influence propagation supported communities and take into account the influence propagation crossing communities. Also, we have a tendency to provide preciseness analysis to point out approximation guarantees of our models. Experiments on real large-scale mobile social networks show that the planned strategies square measure a lot of quicker than previous algorithms, meanwhile, with high accuracy.
Keywords: - mobile social network, influence maximization, greedy algorithm, divide-and-conquer

I. INTRODUCTION

Today mobile social network systems are more and more offered. Mobile social network is enjoying necessary role because the unfold of data and influence. It's a main issue to search out a set of cogent people in a very mobile social network. Cogent nodes target 1st. it'll maximize the unfold of the influence. Think about a social network alongside the estimates for the extent to that individual influence each other, and therefore the network performs because the platform for selling. an organization would really like to promote a replacement product, hoping it'll be adopted by an outsized fraction of the network. the corporate plans to at first target atiny low variety of "Influential" people of the network by giving them free samples of the merchandise (the product is dear or the corporate has restricted budge in order that they'll solely select atiny low variety of people).

The corporate hopes that at first elect users can advocate the product to their friends; their friends can influence their friends'. Friends and then on, therefore several people can ultimately adopt the new product through the powerful spoken impact (or known as infectious agent marketing). Similar things may apply to the promotion of concepts and opinions, like political candidates attempting to find early supporters for his or her political proposals and agendas and rewriting the influence of them to induce a lot of supporters, government authorities or firms associate degree attempt attempting} to win public support by finding and convincing an initial set of early adopters to their concepts. The matter here is to decide on a group of people to send the free samples such they eventually influence the most important variety of individuals within the network

The most cogent nodes are referred to as NP-hard. Greedy algorithmic rule with demonstrable approximation guarantees will offer sensible approximation; this Greedy algorithmic rule is dear. During this project, a divide-and-conquer strategy with parallel computing mechanism has been used. I 1st used associate degree algorithmic rule known as Community-based Greedy algorithmic rule for mining top-K cogent nodes. It encompasses 2 components: dividing the large-scale mobile social network into many communities by taking into consideration data diffusion and choosing communities to search out cogent nodes by a dynamic programming.. Experiments on real large-scale mobile social networks show that the projected ways ar abundant quicker than previous algorithms, meanwhile, with high accuracy. Connecting people or organizations, social network seems a system, like email networks, on-line Facebook, and scientific collaboration networks etc.

With the shape of spoken data or influence spreads upon the social network, that may be a elementary issue in social network analysis.

An organization would really like to promote a replacement product, hoping it'll be adopted by an outsized fraction of the network. At first target atiny low variety of “influential” people of the network by giving they free samples of the merchandise. The corporate hopes that the at first elect users can advocate for product to their friends.& because of their friends can influence their friends ‘friend. Adopt the new product through the powerful spoken impact or known as infectious agent selling. During this to decide on a group of people to send the free samples such they influence the most important variety of individuals.

“Divide-and-conquer” strategy and “Parallelized computation”. First used associate degree algorithmic rule for mining top-K cogent nodes, known as Community-based Greedy algorithmic rule CGA, These networks ar weighted directed graphs. giant social network seems with native densely connected subsets of nodes whereas solely thin links exist among completely different native targeted regions. Communities are social teams. And therefore people in a very community can influence one another within the variety of “words”. During this Project parallel mechanism for the discovered communities is employed. It’s wont to calculate the influence in every community at the same time, known as Parallelized Community-based algorithmic rule (PCA).

II. RELATED WORK

Independent cascade model and linear threshold model are two extensively studied influence diffusions models originally summarized by Kempe et al. based on earlier works of Kempe et al. prove that the generalized versions of these two models are equivalent. Based on the IC and LT model, Kempe et.al pro- pose a greedy algorithm to solve the influence maximization problem (brought about by Richardson) to maximize the spreading of a single piece of ideas, innovations, etc. under these two models. Many follow-up studies propose alter- native heuristics and try to solve the influence maximization problem more efficiently. In terms of efficient algorithm design, our work follows the idea in of finding efficient local graph structures to speed up the com- putation.

In particular, our CLDAG algorithm is similar to the LDAG algorithm, which is also based on the DAG structure, but our CLDAG algorithm is novel in dealing with competitive influence diffusion using the dynamic programming method. Recently, there are a number of studies on competitive influence diffusion. Bharathi et al, extend the IC model to model competitive influence, but they only provide a polynomial approximation algorithm for trees. Kostka et al. study competitive rumor spreading on a more restricted model than IC and LT, and focused on showing the hardness of computing the optimal solution for the two competing parties. Pathak et al. study a model of multiple cascades, which is an extension of a different diffusion model called the voter model, even though they claim it to be a generalization of the linear threshold model. They only study model dynamics and do not address the influence maximization problem.

The social influence analysis problem poses a unique set of challenges: First, how to leverage both node-specific topic distribution and network structure to quantify social influence? In another word, a user’s influence on others not only depends on their own topic distribution, but also relies on what kinds of social relationships they have with others. The goal is to design a unified approach to utilize both the local attributes (topic distribution) and the global structure (network information) for social influence analysis.

Second, how to scale the proposed analysis to a real large social network? For example, the academic community of Computer Science has more than one million researchers and more than 10 mil- lion coauthor relations; Facebook has more than 50 million users and hundreds of millions of different social ties. How to efficiently identify the topic-based influential strength for each social tie is really a challenging problem. Next we discuss the data input and the main intuition of the pro- posed method. Data Input: Two inputs are required to our social influence analysis: 1) net- works and 2) topic distribution on all nodes. The first input is the network backbone obtained by any social networks, such as online social networks like Facebook and MySpace. The second input is the topic distribution for all nodes. In general, the topic information can be obtained in many different ways. For example, in a social network, one can use the predefined cat egories as the topic information, or use user-assigned tags as the topic information. In addition, we can use statistical topic modeling to automatically extract topics from the social net- working data. In this paper, we use the topic modeling approach to initialize the topic distribution of each node.

Topical Affinity Propagation (TAP): Based on the input network and topic distribution on the nodes, we formalize the social influence problem in a topical factor graph model and propose at opical affinity propagation on the fact or graph to automatically identify the topic-specific social influence. Our main idea is to leverage affinity propagation at the topic- level for social influence identification. The approach is based on the theory of factor graph, in which the observation data are cohesive on both local attributes and relationships. In our setting, the node corresponds to the observation data in the factor graph and the social relationship corresponds to edge between the observation data in the graph. Finally, we propose two different propagation rules: one based on message passing on graphical models, the other one is a parallel update rule that is suitable for Map-Reduce framework.

III. PROJECT METHODOLOGY

Our community detection algorithm consists of two steps, partition and combination.

1) Partition. We extend the algorithm with the information influence mechanism based on Independent Cascade model. The algorithm, a nearly linear algorithm for community detection, is designed for undirected and un weighted graph, and thus is not directly applicable.
2) Combination. The generated communities by the partition step are very small and dispersed; we develop a method to combine communities such that the difference between influence degree of a node in its community and its influence degree in the whole network is restricted.

CGA Implementation

- Mining TOP-L Influential Nodes
- Top-K Influential Nodes Mining
- Precision Analysis of CGA

PCA Implementation

- Community-based Parallelization Mechanism
- Influence Propagation Crossing Communities

Precision Analysis of PCA

ALGORITHM AND PARAMETERS

In the experiments, we compare our algorithms (CGA, PCA) with existing representative algorithms for influence maximization. We listed as follow:

• **Mix Greedy**: We take Mix Greedy as the benchmark to evaluate the proposed algorithm CGA for two reasons. First, MixGreedy is the state-of-the-art Greedy Algorithm for influence maximization, and it is shown that MixGreedy outperforms previously proposed Greedy Algorithms, such as GA and CELF. Second, CGA adopts MixGreedy to find influential nodes within communities, and thus a performance comparison between them will reveal their pros and cons.
• **Degree Discount**: The degree discount heuristic of Chen et al. Developed for the uniform IC model.
• **SA**: Unlike previous algorithms, it takes a totally different approach by using artificial intelligence

• **MIA**: The most representative algorithm is proposed by Chen et at., which uses local arborescence structures of each node to approximate the influence propagation.
• **Random**: simply select K random nodes in the graph.
• **CGA**: It is the community-based greedy algorithm pro- posed in this paper.
• **PCGA**: PCGA is CGA in parallelization. It only uses LTP algorithm to balance the computation.
• **PCA**: It is the algorithm that mining Top-K influential nodes with considering the propagation of crossing com munities based on parallelism in this paper.

A GREEDY ALGORITHM APPROACH

Given a mobile social network G = (V, E, W), we aim to mine a set of top-K Influential nodes I on the network such that R (I) is maximized using the extended Independent Cascade information diffusion model. It has been proved that the optimization problem is NP- hard [5]. A greedy algorithm can approximate the optimum to within a factor of (1− 1 e). However, the greedy algorithm is expensive for solving the influence maximization problem on a large-scale network. So we propose a community based greedy algorithm which mine the Influential nodes in each community rather than the whole network.

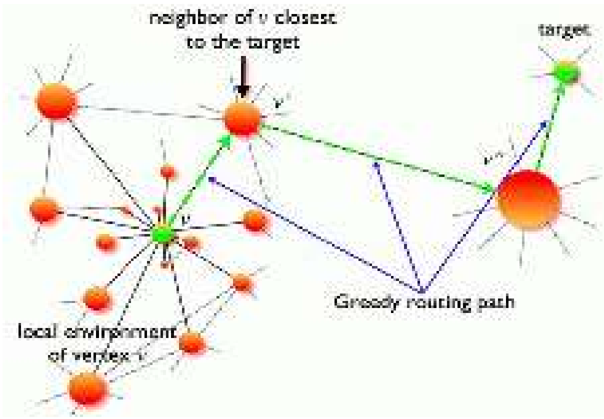

Fig.1

Algorithm: PCA algorithm

Input: G = (V,ε,⊒), M communities;
Output: I, Top-K influential nodes;

```
I = I1 = I2 = ··· = IM = ∅ ◃ I and Ij are the set of K influential
        nodes in the whole network and community Cj respectively
for k = 1 to K do
        for m = 1 to M do
                Cm.W = Cm.N ∗β1 + Cm.E ∗β2
        end for
        Sort(C, 1, M)
        for m = 1 to M do
                i = argminpi∈P (workload[pi])  ◃ pi is ith loaded processor
                threadwork[i].push back (Cm)
        end for ◃ The following is the parallelization step
        omp set num threads (P) ◃ The number of threads
        ♯ pragma omp parallel ◃ Parallel mark of OpenMP
                compute ΔRm
        vmax = argmaxvj∈Cj (Rj(Ij∪{vi})−Rj(Ij))
        Ij = Ij∪{vmax}
        I = I∪vmax
end for
```

In Algorithm, line 3-5 computes the computational weight of each community, and assign them to each loaded pro- cessor(line 7-10); Line 11-13 parallelize the influence spread by computing ΔRm of each community m. Specifically, we compute ΔRm with consideration of influence propagation crossing communities, which will be detailed in Section 4.2.2. In line 14-16, we find the node vmax that maximizes Rj(Ij∪{vi})−R(Ij) in community Cj. In addition, we optimize the above algorithm by the fol- lowing technique: if the kth influential node and its influence spread are not in community Cm, the weight of this community is not computed repeatedly.

IV. CONCLUSION

In this paper, we have a tendency to propose the economical algorithms known as Community-based Greedy algorithm(CGA) and Parallelized Community-based algorithm(PCA) for mining top-K authoritative nodes in an exceedingly MSN. We have a tendency to initial extend the essential free lance Cascade model to require weight fringe of MSN into thought. CGA has 2 main elements, AN formula for detective work communities by taking under consideration data diffusion, and a dynamic programming formula for choosing communities to find authoritative nodes. Then, to any improve the efficiency and accuracy, we have a tendency to set the influence unfold in every community with a balance computation assignment and take into account the influence propagation crossing communities by removing less authoritative nodes through a rigorous established regulation. We conjointly offer demonstrable approximation guarantees for CGA and PCA. Empirical studies on an outsized real world mobile social network show that these algorithms have nice improvement on each potency and accuracy

REFERENCES

[1] F.Bass, "A new product growth model for consumer durables," Manage- ment Science, vol. 15, pp. 215–227, 1969.

[2] J.Brown and P.Reinegen, "Social ties and word-of-mouth referral be- havior," Journal of Consumer research, vol. 14, no. 3, pp. 350–362, 1987.

[3] J.Goldenberg, B.Libai, and E.Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," Marketing Letters, vol. 12, no. 3, pp. 211–223, 2001

[4] V .Mahajan, E.Muller, and F.Bass, "New product diffusion models in marketing: A review and directions for research," Journal of Marketing, vol. 54, no. 1, pp. 1–26, 1999.

[5] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 137–146.

[6] D.Kempe, J.Kleinberg, and E.Tardos, "Influential nodes in a diffusion model for social networks," In ternational colloquium on automata, languages and programming, no. 32, pp. 112–1138, 2005.

[7] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 199– 208.

[8] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 420–429.

[9] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.

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

[11] P. Domingos and M. Richardson, "Mining the network value of cus- tomers," in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 57–66.