Influence Maximization on Mobile Social Network using Location based Community Greedy Algorithm

Smita Bhosale
Department of Computer Engineering
Dr.D.Y.Patil College of Engineering, Ambi, Pune

Dhanshree Kulkarni
Department of Computer Engineering
Dr.D.Y.Patil College of Engineering, Ambi, Pune.

ABSTRACT

A mobile social network plays an important role as the spread of information and influence in the form of "word-of-mouth". It is basic thing to find small set of influential people in a mobile social network such that targeting them initially. It will increase the spread of the influence .The problem of finding the most influential nodes in network is NP-hard. It has been shown that a Greedy algorithm with provable approximation guarantees can give good approximation. Community based Greedy algorithm is used for mining top-K influential nodes. It has two components: dividing the mobile social network into several communities by taking into account information diffusion and selecting communities to find influential nodes by a dynamic programming. Location Based community Greedy algorithm is used to find the influence node based on Location and consider the influence propagation within Particular area. Experiments result on real large-scale mobile social networks show that the proposed location based greedy algorithm has higher efficiency than previous community greedy algorithm.

General Terms

Mobile social network, Influence maximization, community greedy algorithm, Location based community greedy algotithm.

Keywords

CGA - Community-based Greedy Algorithm
LCGA – Location Based Community Greedy Algorithm

1. INTRODUCTION

Mobile social network plays an important role in social Network. It's a main issue to search out a set of influential people in a mobile social network. It will maximize the influence. 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 small variety of "Influential" people of the network by giving them free samples of the product. The corporate hopes that first selected 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. Similar things may apply to the promotion of concepts and opinions, like political candidates attempt 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 to win public support by finding and convincing an initial set of early adopters to their concepts. The most cogent nodes are referred to as NP-hard. During this project a divide and conquer strategy has been used. Community based Greedy algorithmic rule is used for mining top-K influential nodes. It has two 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 proposed algorithm is quicker than previous algorithms with high accuracy.

The new algorithm is known as Location based community Greedy algorithm. Here we have to find most influential node in Particular location. The people in same area are more influence as compare to the people in different area or state. Persons in same area always have more contact than persons in different area. Communication Time between persons & location of person these two parameters are considered in Location Based Community Greedy algorithm.Experimently shows that Location Based community greedy algorithm have higher accuracy and efficiency than existing community based Greedy algorithm.

2. RELATED WORK

Independent cascade model and linear threshold model are two extensively studied influence diffusions models originally summarized by Kempe et al .It prove that the generalized versions of these two models are same. Based on the IC and LT model, Kempe et al propose a greedy algorithm to solve the influence maximization problem (brought about by Richardson) to maximize the spreading of ideas & innovations under these two models. Many follow up studies propose alternative heuristics and try to solve the influence maximization problem more efficiently. The idea in of finding efficient local graph structures is used to speed up the computation. CLDAG algorithm is similar to the LDAG algorithm, which is also based on the DAG algorithm. Our CLDAG algorithm is novel in dealing with competitive influence diffusion using the dynamic programming method. Number of studies on competitive influence diffusion is done. Bharathi et al, extend the IC model to model competitive influence, but they only provide a polynomial approximation algorithm for trees.

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.

The proposed analysis is scaled to a real large social network in following ways. For example, the academic community of Computer Science has more than 1 million researchers and more than 10 million 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 proposed method. Data Input: Two inputs are required to our social influence analysis: 1) networks 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 categories 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 networking 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.

3. METHODOLOGY

Our community greedy algorithm uses community detection algorithm to fine community. Community detection algorithm consists of 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 unweighted graph. It 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
- Top-K Influential Nodes Mining
- Precision Analysis of CGA

LCGA Implementation
- Location based Mechanism
- Top-K Influential Nodes Mining
- Precision Analysis of LCGA
- Regional Factor-

I) Location Based community Greedy Algorithm consider region Factor.

- Node in same region has more contact.
- Region factor is denoted as SR=2

II) Node in different region has less contact with each other. Node in Different regional factor is SR=1.

4. ALGORITHM AND PARAMETERS

In the experiments, we compare our LCGA algorithms with existing CGA algorithm for influence maximization. Algorithm & parameter are listed as follow:

Independent Cascade Model-
The model is originally proposed by Lopez Pintado . It is the most common dynamic model in information diffusion. It is widely used in influence maximization problem. We assign two states to nodes: *active* and *inactive* in the model. Active nodes are those that are influenced by other active nodes. These active nodes are able to influence their inactive neighbors; inactive nodes are those that are not influenced by their active neighbors. The state of a node can be switched from being inactive to being active, but not reverse.

Diffusion Speed-
The algorithm has an important parameter called diffusion speed λ. When an active node vi contacts an inactive node vj, the inactive node becomes active at a probability (rate) λ. In a viral marketing, diffusion speed models the tendency of individuals to accept a product. Thus the diffusion is affected by diffusion speed, node degree, and the number of initial active nodes.

Influence Degree-
Let A be the initial set of active nodes. The influence degree of set A is computed as:

R(A) = VA/N

VA is the number of nodes influenced by A during information diffusion process.

CGA:
It is known as community-based greedy algorithm.

The CGA algorithm is to divide a network into communities, and then choose communities to find top- K influential nodes within communities. Individuals within a community have frequent contact and thus are more likely to influence each other; In contrast, individuals across communities have much less contact with each other and thus are less likely to influence each other. This property suggests that it might be a good approximation to choose influential nodes within communities instead of the whole network. This algorithm considers into account the diffusion model.

5. LOCATION BASED COMMUNITY 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 Independent Cascade information diffusion model. It has been proved that the optimization problem is NP- hard. However, the community s greedy algorithm is used in whole network for solving the influence maximization problem on a large-scale network. We propose Location Based community greedy algorithm which mine the Influential nodes in each community rather than the whole network considering location factor. Location area is denoted with location no. This algorithm is used to mine a set of top-k influential nodes in particular area.
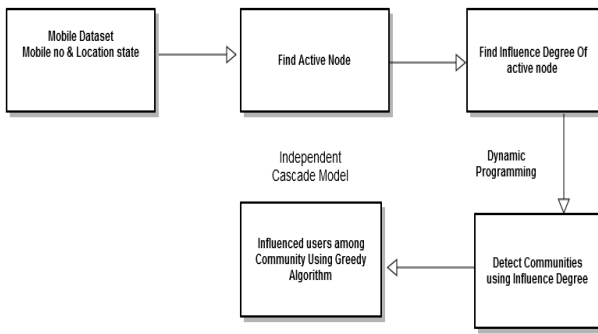
Fig 1: Location Based community Greedy Algorithm Illustration

Algorithm LCGA

1) Network $G = (V, E, W)$, size of result k, influence speed, size of result, Location_Id

2. Detect communities & Find out influence degree using dynamic Programming algorithm

3. Calculate maximal increase of influence degree with regard to community.

4. Sort according to location degree.

5. Choose community that which yields the largest increase of influence degree among all communities.

6. Select community from first m communities to mine influential node.

6. EXPERIMENTAL SETUP

Data Set
We have a CDR (call detailed record) data of Mobile. We extract a Mobile Social Network from the CDR data using the CGA & LCGA presented in this Paper. We evaluate algorithm in Mobile Social Networks datasets having 1000 records.

Experimental Results
This experiment is to evaluate the performance of the proposed Location Based community Greedy algorithm on a Mobile Social Network.

We have to consider following parameter.

1. Varying K
This experiment is to evaluate the difference of the parameter K on the influence degree and efficiency of CGA & LCGA algorithms. We vary K from 1 to 6. The results are shown in Figure 2. It can be seen in Figure 2 that 1) LCGA method is much better than CGA method. This is Because of Location parameter.
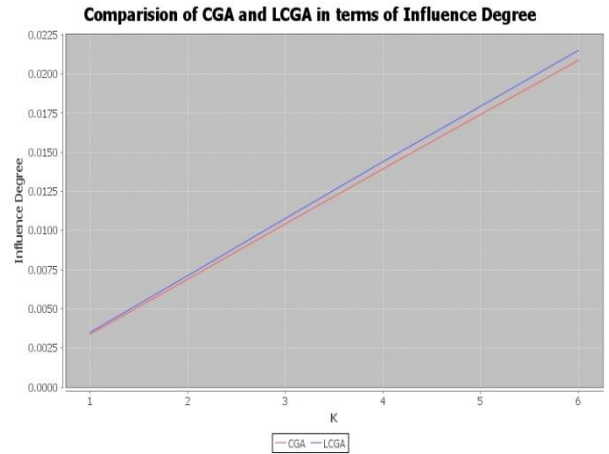

Fig 2: Comparison of CGA and LCGA in terms of Influence Degree.

2. Varying the diffusion speed

This experiment is to compare CGA algorithm with LCGA in terms of influence degree and diffusion speed. We vary the average diffusion speed λ. We vary λ from 0 to 0.1. From the influence spread and efficiency shown in Figure 3. The influence Degree of LCGA is more to the influence degree of CGA.This is because Location Based Community greedy algorithm consider particular area. Influence speed in particular area is more than different area.
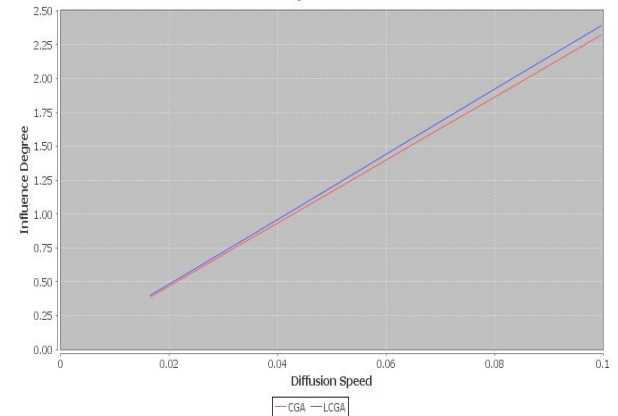

Fig 3: Comparison of CGA and LCGA in terms of Influence Degree & Diffusion Speed.

7. CONCLUSION
In this paper, Location Based Community Greedy algorithm (LCGA) is proposed. It is used for mining top-K authoritative nodes in an exceedingly MSN. Independent Cascade model is required to consider weight fringe of Mobile social network. LCGA has two main elements; AN formula is used for detective work communities by taking under consideration data diffusion. Dynamic programming formula is used for choosing communities to find authoritative nodes. LCGA algorithm considers both influential time & Location Factor. Empirical studies on mobile social network show that LCGA algorithms have great improvement on both efficiency and accuracy compared with CGA methods.

8. REFERENCES

[1] 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

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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

[11] 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