Minimal Influential Node Set Detection in Social Networks

[1]M Venu Gopalachari, [2]P Hima Bindu, [3]S. Rakesh
[1]Associate Professor, [2]PG Student, [3]Assistant Professor
[1]Department of IT,
[1]Chaitanya Bharathi Institute of Technology, Hyderabad, India

Abstract : Social Networks became popular in recent times with the applications such as facebook, twitter, wechat etc, which provides a platform to share wide variety of information in turn involved in many business models like target marketing. Social influence analysis (SIA) is a vast research field on the social networks that includes topics such as community detection, influence maximization and minimization, flow of influence, and individual influence. This project focuses on influence maximization problem, which is to select a subset of nodes, so that the spread of information could be maximized. However, the existing works were more focused to define influence propagation but did not consider the size of the influence node set. The size of the influence node set matters if it overlaps neighbor node set. Also the existing works considered positive influence dominant set while the negative influence is ignored. This paper developed a methodology that detects minimum sized maximum influential nodes in social network which also considers the percentages of positive and negative influences of the users in the community.

IndexTerms **- Social Networks, Influential Node set, Minimum Influential Node, Greedy algorithm**

I. INTRODUCTION

A social network (e.g., Facebook, Google+, and Myspace) is composed of a set of nodes (such as individuals or organizations) theft share a similar interest or purpose. The social network is a powerful medium of communication for sharing, exchanging. and disseminating information, and for spreading influence beyond the traditional social interactions [1]. Since social networks emerged, they have significantly expanded our social circles and become a bridge to connect our daily physical life and the virtual web space. With the emergence of social applications (such as Flickr, Wikis, Netflix, and Twitter, etc.), a tremendous interest has focused on how social networks can be utilized effectively to spread ideas or information within a community. Capturing the dynamics of a social network is a complex problem, thus, it requires an approach to analyse the dynamics of positive and negative social influences that result from individual-to- individual and individual-to-group interactions. Individuals in a social network may have both positive and negative influences on each other.

For example, within the context of gambling, a gambling insulator has a positive influence on his friends/neighbours. Moreover, if many of an individual's friends are gambling insulators, then the aggregated positive influence is exacerbated [2]. However, an individual might become a gambler, who has a negative impact on his friends/neighbours. A social network is defined as a chain of individual and their personal connections. Expanding one's connections with other people is a technique that can be used both for personal or business reasons. It is based on the concept of "six degree of separation", where in any two people can make contact by a chain of most five mediators. Alternatively referred to as a virtual community or profile site, a social network is a website that brings people together share ideas and interest .this type of together to talk, share idea and interests this is knows as social media examples of social media is Discord, Quora, Vero, Flick etc. here comes about the opinion mining or sentiment analysis aims to determine the attitude of a speaker with respect to some topic or the overall contextual polarity of the document.

This opinion mining is important because opinions are the key influences of our behaviours, our beliefs and perceptions of reality are conditional how other see the world. Whenever we need to make a decision we often seek out the opinion from others.it is the area of research that attempts to make automatic system to determine human opinion from text written in natural language and the process of analysing the text about a topic written in a natural language classify them as positive negative or neutral based on the human sentimental emotions opinion expressed in it.

Social influence analysis is a vast research field that has attracted research interest in many areas we analyse social influence method including influence maximization and minimization flow of influence and individual influence. In a minimum sized positive individual domination set D, so that every other node has at least half of the its neighbours is considered, and while the negative influence is ignored. Most studies ignore negative influences among individuals and groups. For this a new optimization problem called the Minimum-sized Positive Influential Node Set (MPINS) selection problem is proposed to identify the minimum set of influential nodes such that every node in the network can be positively influenced by these selected nodes. In this paper a methodology is developed that detects minimum size maximum influential nodes that means in social networks which also consider the negative influences of the user in the community.

II. RELATED WORK

Influence maximization is a classic optimization problem studied in the area of social network analysis and viral marketing [1]. Given a network, it is defined as the problem of finding k seed nodes so that the influence spread of the network can be optimized. Kempe et al. have proved that this problem is NP hard and the objective function is sub modular, based on which a greedy algorithm was proposed to give a near optimal solution [2]. However, this simple greedy algorithm is time consuming, which limits its application on large-scale networks. Heuristic algorithms generally cannot provide any performance guarantee [3, 4]. To solve this problem, in this paper they propose CoFIM, a community-based framework for influence maximization on large-scale networks. In our framework the influence propagation process is divided into two phases:(i) seeds expansion; and (ii)intra-community propagation. The first phase is the expansion of seed nodes among different communities at the beginning of diffusion. The second phase is the influence propagation within communities which are independent of each

other. Based on the framework, we derive a simple evaluation form of the total influence spread which is sub modular and can be efficiently computed. Then we further propose a fast algorithm to select the seed nodes [6].

Information propagation in network environment is a widely studied research topic, especially in Online Social Networks (OSNs), where the problem has gained significant popularity. Recent studies attempt to pick up the key nodes, who could maximize the network influence in OSNs. However, in addition to propagation in OSNs, another information propagation way is through words of mouth among people in the offline mobile network, which is an indispensable factor and is not considered in most cases [7, 8]. In [5], explored the Influence maximization for opinion formation IMOF as an extension of influence maximization, which aims at determining a set of influential nodes and making the average of opinions of network nodes maximized. They formulate the IMOF mathematically and propose a heuristic algorithm of the 3-hop measure based on the knowledge of node influence, i.e., RSPN3 to determine the top- m influential nodes.

III. MPINS TECHNIQUE

To spread information to entire community author is using MPINS (Minimum-sized Positive Influential Node Set) Greedy Algorithm. The main aim of this algorithm is to select or choose minimum sized nodes which can influence (spread information) entire community.

MPINS Greedy algorithm

Input: G(V,E,P(E)), a graph with vertices V, Edges E, Polarity P with predefined threshold t
Output: Influential node set I

- Initialize Infliuence set I = M
- While f(I) < |V| t do
    - Choose u which belongs to V over I maximizing f(I U {u})
    - I = I U u
- Return I

To implement this concept by means of following properties.

- Graph: A graph will form from communities' dataset (Social networking dataset which contains vertex id and edge id).
- Mset (minimum set M): We need to give this value as input in the form of source and destination vertex to get active nodes which can cover entire graph
- Positive influential node set: All nodes which contain edges to Mset will be consider as positive influential node set
- Active neighboring set or Neighboring set: All nodes which are neighbors of positive influential node set are called Neighboring set
- Non-active neighboring set: nodes which are not part of Positive influential node set are called inactive.
- Positive or Negative influence: if node is Positive node set then it will positively influence otherwise negative influence.

After building graph author is using edges weight and then calculating influence value and if influence value greater than theta threshold value then it will be considered as selected influential nodes. In below diagram we can see selected black cover nodes are the minimum selected node which can influence left over nodes.

In this paper as we are calculating total positive or negative influence percentage from nodes which are actively connecting or can cover max number of nodes. With this extension users can easily understand by selecting which node how much positive or negative influence will be there on selected social networks. To implement this extension work we are uploading WIKIVOTE dataset. In greedy method we added positive and negative influence percentage calculation. Below screen with code and comments help u in understanding how percentage is calculated.

IV. RESULTS

Sandford university there is a research lab for social Network there released the Data statistics. Wikipedia is a free encyclopaedia written collaboratively by volunteers around the world. Using the latest complete dump of Wikipedia page edit history we extracted all administrator elections and vote history data. This gave us 2,794 elections with 103,663 total votes and 7,006 users participating in the elections. out of these 1,235 elections resulted in a successful promotion, while 1,559 elections did not result in the promotion. About half of the votes existing admins, while the other half comes from ordinary Wikipedia users. Nodes In the network represent Wikipedia users and a directed edge from node I to node j represent that user I voted in user j Dataset The network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008.

This paper describes concept to find minimum number of peoples in a community which can spread positive or negative influence to maximum number of users in entire community. Sometime we need to spread information to community to advertise new products or advertise smoking quit events or any other event. To do so earlier peoples were using 'mouths to words' to spread such information but now-a-days social networking sites contains similar groups as communities and we can use that group to spread information. To spread positive or negative information we need to select nodes which are active and can cover or spread data to maximum or entire community.

TABLE 1: STATISTICS OF WIKIPEDIA DATA SET

Nodes	7115
Edges	103659
Nodes in Largest WCC	7066 (0.993)
Edges in Largest WCC	103663(1.000)
Nodes in Largest SCC	1300(0.183)
Edges in Largest SCC	39456(0.381)
Average clustering coefficient	0.1409
Number of Triangles	608389
Fraction of closed Traingles	004564
Diameter (Longest Shortest Path)	7

Similar to other state-of-the-art algorithms we use two metrics to evaluate the performance of our algorithm. Influence spread Given seed set, influence spread is defined as the number of expected active nodes with the Monte-Carlo simulation in IC and LT models. It is used to evaluate the accuracy of an influence maximization algorithm. Algorithm with Higher influence spread value has more accurate than others. We repeat 1000 Monte-Carlo simulations to compute the influence spread for any given seed set with k seed nodes.

TABLE 2: IMPACT OF CRITERION ON THE INFLUENCE SPREAD AND RUNTIME OF THE MNIPS TECHNIQUE ON WIKIPEDIA DATASET

Spread Threshold	**0.4**	**0.6**	**0.8**	**0.9**
K=1	3	3.10	3.2	3.02
K=10	24.71	25.09	24.90	24.91
K=20	46.32	46.20	46.21	46.27
K=30	66.13	65.96	64.06	65.63

V. CONCLUSION

In this paper, the MPINS selection problem in social networks, which has useful commercial applications. Through reduction, it shows that MPINS is APX-hard under the independent cascade model. Subsequently, a greedy algorithm called MPINSGREEDY is proposed to solve the problem. I'm validating our proposed algorithm through simulations on random graphs and experiments on seven different real-world data sets. Simulation and experimental results indicate that MPINS-GREEDY can construct smaller satisfied initial active node sets than the latest related work PIDS. Moreover, for small-scale networks, the performance of MPINS-GREEDY similar to that of the optimal solution of MPINS.

REFERENCES

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network", in Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137– 146.

[2] K. Saito, M. Kimura, and H. Motoda, "Discovering influential nodes for sis models in social networks", in Proc. International Conference on Discovery Science, 2009, pp. 302–316.

[3] Y. Li, W. Chen, Y. Wang, and Z. Zhang, "Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships", in Proc. of the Sixth ACM International Conference on Web Search and Data Mining, 2013, pp. 657–666.

[4] M. Han, M. Yan, Z. Cai, Y. Li, X. Cai, and J. Yu, "Influence maximization by probing partial communities in dynamic online social networks", Transactions on Emerging Telecommunications Technologies, vol. 28, no. 4, p. e3054, 2016.

[5] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model", in Proc. of the 2012 SIAM International Conference on Data Mining, 2012, pp. 463– 474.

[6] W. Lu, F. Bonchi, A. Goyal, and L. V. Lakshmanan, "The bang for the buck: Fair competitive viral marketing from the host perspective", in Proc. of the 19th SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 928–936.

[7] F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan, "On positive influence dominating sets in social networks", Theoretical Computer Science, vol. 412, no. 3, pp. 265– 269, 2011.

[8] M. Han and Y. Li, "Influence analysis: A survey of the state of-the-art", in International Symposium on Bioinformatics Research and Applications, 2018, pp. 259–264.

[9] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks", in Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, pp. 807–816.

[10] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks", in Proc. of the Third ACM International Conference on Web Search and Data Mining, ch2010, pp. 241–250.