Influence Maximization on Signed Social Networks with Integrated PageRank

Shubo Chen
School of Computer Science and Engineering, South China University of Technology
Guang Zhou, China
E-mail: mumu_311@163.com

Kejing He
School of Computer Science and Engineering, South China University of Technology
Guang Zhou, China
Corresponding Author. E-mail: kejinghe@ieee.org

***Abstract*—Online social networks (OSNs) have received a lot of attentions recently since they provide a new platform for product promotion and online viral marketing. Influence maximization problem has been extensively studied on some existing influence diffusion models in number of domains. However, most of the existing studies consider OSNs as friendly networks only containing friendship relationships, whereas the hostile relations do exist in many OSNs in real life, e.g., Epinions and Slashdot. In this paper, we integrate the PageRank on signed social networks and use the integrated PageRank to study influence maximization in OSNs with both friend and hostile relations which are respected as positive edges and negative edges on signed networks. In addition, we use the extended vote model to study the influence diffusion on signed networks. We then conducted comprehensive experiments on real social networks to select initial k seeds for influence maximization, and results indicate that our integrated PageRank method performs better than other heuristic algorithms.**

Keywords-PageRank; signed social networks; influence maximization; voter model

I. INTRODUCTION

A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of dyadic ties between these actors. With the fast development of the Internet and the growth of online social networks (OSNs), more and more people are active on OSNs, about two-thirds of online users are using social networks such as Facebook and Twitter and so on[1]. Thus the OSNs has become an important platform and propagation vehicle for the diffusion of opinions, message, innovation in the form of "word of mouth". It is a fundamental issue to find a small subset of influential individuals in a social network such that they can influence the largest number of people in the networks[2].

The influence maximization problem has raised a lot of research interests in recent years. The initial related research was conducted by Mark Granovetter et al. [3] in the 1970, known as the diffusion of innovations. Then Kempe et al. [4] formulated the problem as influence maximization problem and showed it is NP-hard under both LT model and IC model. Although the greedy algorithm always generates a seed set with high quality, the major limitation is the scalability. So many algorithms have proposed to solve the problem efficiently, such as CELF and CELF++. Even-Dar and Shapira[5] study the influence maximization problem in the voter model on simple unsigned and undirected graphs. Most of the existing studies consider social networks only with friend (i.e., positive or trust) relationship, regardless of the existing hostile (i.e., negative or distrust) relationship in real social networks. The signed networks with both positive and negative edges have gained attentions recently[6,7]. Some researches study the foe relationship on signed networks, e.g., Li et al. [7] does theoretical analysis on signed networks with hostile relationships. In this paper we consider both friend and hostile relationships on social networks. The friend relationship influence the user to adopt the opinion of his friends. In contrast, the user will choose the opposite value of the active neighbor with a hostile relationship.

In this paper, we develop the solution to the influence maximization problem under extended voter model on signed networks[8]. Our contributions are summarized as follows:

- We extended the classic voter model [9,10] to incorporate both negative and positive relationships for modeling the diffusion of opinions in a signed network. The basic voter model works in unsigned directed graph. In extended voter model, the node change its opinion according to the edge and the opinion of its neighbor it picks.
- We then study the influence maximization problem under the voter model for signed networks and analysed the PageRank algorithm and extended the algorithm to be available for digraph with negative edges. Then we use the Integrated PageRank method to choose seeds in signed networks for the influence maximization.
- Finally, we conduct massive simulations on real-world networks. The results demonstrate our the Integrated PageRank method performs better than other heuristic algorithms.

The remainder of this paper is organized as follows. In section II, we give a description of the problem with mathematical method and the extended vote model. Section III gives details of the Integrated PageRank method applied in signed networks. The experiment results on real-world social networks are presented in section IV. Section V concludes the paper together with future directions.

978-1-5090-1893-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SmartCity.2015.86

II. PRELIMINARIES AND PROBLEM STATEMENT

We introduce the extended vote information diffusion model, and then the problem statement.

Vote Model for Signed Networks The voter model is proposed in [9,10], and the diffusion model is suitable for modeling opinion diffusion in which people may switch opinions back and forth from time to time, because of the interactions with other people in the social network. The voter model is a simple mathematical model of opinion formation in which voters are located at the nodes of a network denoted as a directed weighted graph $G = (V, E, W)$. Initially each voter has an opinion (in the simplest case, 0 or 1), and randomly chooses voter assumes the opinion of one of its neighbors. At step $t \geqslant 1$, every user v randomly chooses one outgoing neighbor u with the probability to the weight of (u,v), namely $W_{uv}/\sum_p W_{up}$, and changes its opinion to the opinion of v.

We model a signed network as a directed graph $G = (V, E, A)$, where V, E, and A represent nodes, edges, and adjacency relation matrix, respectively. A directed edge from node v to node u indicate v is influenced by u. In extended model, the seeds have positive opinion (denoted by s(i)=1) while other nodes are inactive (denoted by s(i)=0). Then at step $t \geqslant 1$, every user v randomly choose one outgoing neighbor u that have been activated before step t with the same probability, if $A_{uv} = 1$ (the relationship is friend) the user u changes its opinion to the same with user v, otherwise $A_{uv} = -1$ (the relationship is hostile) ,then user u changes its opinion to the opposite of user v and the user u becomes active.

Problem Statement Given a signed social network, in which the relationships are friend or hostile. We aim to mine a set of k(the budget) influential nodes on the network to maximum the influence (the number of people who have positive opinion) using not friend relationships but hostile relationships under the extended voter model. The PageRank algorithm can be used to identity influential users, but the algorithm can not be applied in a graph with negative edges, e.g., signed networks. So we present the Integrated PageRank method to select seed nodes in signed networks which consider negative relationships.

III. THE INTEGRATED PAGERANK

Larry's PageRank Citation Ranking[12] which could be applied in social networks is used to rank the quality of pages that have hyperlinks to some web pages. In PageRank Citation calculation, they model a directed graph G such that web pages are represented as nodes and the hyperlinks are represented as directed edges. As for hanging nodes, they assume that the node links to the other nodes to make the PR converge. Let i denotes each node ($1 \leqslant i \leqslant$ n), PR_{it} denotes the PR score of node i after t rounds calculation. Initially, every node has a non-zero value, and then we calculate the PR scores according to the connection between nodes. Let d_i denotes the number of outgoing edges of node i, $f(i,j)$denotes the edge from i point to j. We can define the row stochastic matrix $P = (p_{ij}) \in R_{n\times n}$, in the form

$$p_{ij} = \begin{cases} d_i^{-1} & f(i,j) \in E \\ 0 & otherwise \end{cases} \quad 1 \le i, j \le n \qquad (1)$$

For each node i, we calculate the PR score as (2). The α (0< α< 1) is the so-called damping factor (that we use as α= 0.85). This value is favorable in terms of computational performance and is also often considered as the default value for PageRank calculations in literature[8].

$$PR_{it} = \frac{1-\alpha}{n} + \alpha \sum_{j=1}^{n} p_{ji} PR_{j,t-1} \qquad (2)$$

We introduce the famous PageRank method in Algorithm 1 used to get PR score of nodes. In Algorithm 1, the **1** and U are the all one vector with |V|-dimensional and |V|×|V|-dimensional respectively. In online social networks, there are usually many hanger points that are the nodes without outgoing edges because there are many people have no friends. The hanger points will affect the PageRank algorithm. To eliminate the hanger points, we should make the value of the elements in the row of P equal to 1/|V| when the sum of a row is zero which means the node with the row number has no outgoing edges, otherwise the matrix G will can not converge.

Algorithm 1 The PageRank Algorithm: PageRank(P)

1: **Input:** Unsigned transition matrix P,
2: **Output:** Rank of nodes PR.
3: PR_0=**1**, $G=\alpha * P^T+(1-\alpha)U$, t=1, $PR_1=G*PR_0$;
4: **while** $\max\|PR_t-PR_{t-1}\| < \theta$ **do**
5: $t=t+1$;
6: $PR_t=G*PR_{t-1}$;
7: **end while**;
8: Return PR_t.

In generally, they calculate the first iterative ranking of each node based on the initial PageRank values, and then calculate the second rank according to the first iteration. The process continues until the PageRank estimator converges to its practical value. The PageRank algorithm can not be directly applicable to signed networks since there exist some negative edges. So we develop the Integrated PageRank algorithm to solve the influence maximization problem on signed networks under voter model, as shown in algorithm 2.

As for a given signed network graph $G = (V, E, A)$, let $G^+ = (V, E^+, A^+)$ and $G^- = (V, E^-, A^-)$ denote the unsigned sub-graphs consisting of all positive edges E^+ and all negative edges E^-, respectively. And the both the adjacency matrices in both G^+ and G^- are non-negative. Now we can apply the PageRank algorithm to the two sub-graphs and get the PR scores of each node indicated as PR^+ and PR^- respectively. We use the Integrated PageRank method (IPR) which is defined as (3) to identify the influence of a node.

$$IPR_i = PR_i^+ - PR_i^- \qquad (3)$$

Algorithm 2 Integrated PageRank Algorithm

Input: Singed network $G = (V, E, A)$, budget k
Output: A seeds set S
Compute the positive adjacent matrix P^+ and negative adjacent matrix P^-;
PR^+ = PageRank (P^+), PR^- = PageRank (P^-)
$IPR = PR^+ - PR^-$;
return the top k nodes with highest IPR_i for influence maximization.

The PR^+ calculated from the sub-graph G^+ constituted of all friend relationships indicates the positive influence of a node. As the foe relationships makes the influenced user to choose the opposite opinion of the influenced one, the PR^- calculated from the sub-graph G^- constituted of all hostile relationships indicates the negative influence of a node.

Therefore, we use the difference between the positive influence and negative influence that is IPR to represent the real influence of nodes on the social network. Then we choose the top-k (the number of k is according to the budget) nodes with the highest IPR to maximize the number of people who have positive opinion on signed social network.

IV. EXPERIMENT

In this section, we use two real-world signed networks to experiment, and their basic statistics are summarized in Table 1, including: (1) Epinions. This is who-trust-whom online social network of a general consumer review site Epinions.com [13]. (2) Slashdot. It is a technology-related news website known for its specific user community[14].

In our experiments, we compare our Integrated PageRank method (denoted by IPR) with four other heuristics listed as follows: (1) selecting top-k nodes with the original PageRank algorithm which only consider the friend relationships (denoted by PR), (2) selecting seed nodes with the highest in-going degrees (denoted by d^++d^-), (3) selecting seed nodes with the method SVML-S in [6] with the t=60 (denoted by SSVIM), and (4) selecting seed nodes randomly (denoted by rand). In our evaluation, we compare the number of positive influenced people between our method and other heuristics to indicate the performance.

Fig. 1, and Fig. 2 shows the result of experiment on Slashdot and Fig. 3 shows the result of experiment on Epinions. From the result of Fig. 1 and Fig. 2, we can find that the curve of pagerank is almost constant below in the carve of IPR. The pagerank only consider the friend relationships in online social networks while IPR consider both friend and hostile relationships. Without hostile relationships, some information would become missed parts such that they could not help to represent the real situation on social networks, since there are distrust between users in real life. The result illustrates that the foe relationship is also as important as friend relationship , and it is necessary to utilize both friend and foe relationships in influence maximization[15] on social networks.

TABLE I. STATISTIC OF EPINIONS AND SLASHDOT DATASETS

Dataset	Epinions	Slashdot
Nodes	131828	82144
Edges	841372	549202
Nodes in largest SCC	41441	27382
Edges in largest SCC	693737	346652
Average clustering coefficient	0.1279	0.0588
90-percentile effective diameter	4.9	4.7

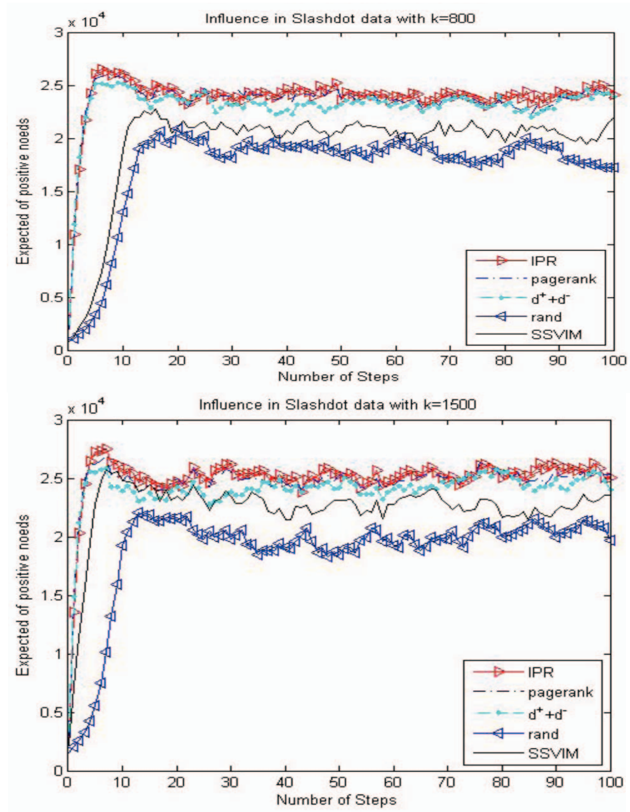

Fig. 1. Influence in Slashdot with seed set size k=800 and 1500.

We can observe that the SSVIM perform better than d^++d^- when k=5000 in Fig. 2. That because the method in [7] is based on an assume that every node is active at first. The more seeds, the better performance. It may perform better than IPR when the k is up to a number close to |V|, but it increases the budget dramatically. Above all, the IPR performs better than other heuristics. We also conducted massive experiments on Epinions. Because of the structure of the network, all the heuristics perform almost. As showed in Fig. 3, the final number of positive nodes is around 40000 no matter what the value of k is. Then we analyses the structure of the network and find that the nodes in SCC have nearly influence and the diffusion mostly happened in SCC. Therefore we can not distinguish these methods due to the special structure very well. However, compared to d^++d^-, our method still performs better as Fig.3 shows. That is due to the d^++d^- seeks the seeds almost in SCC.

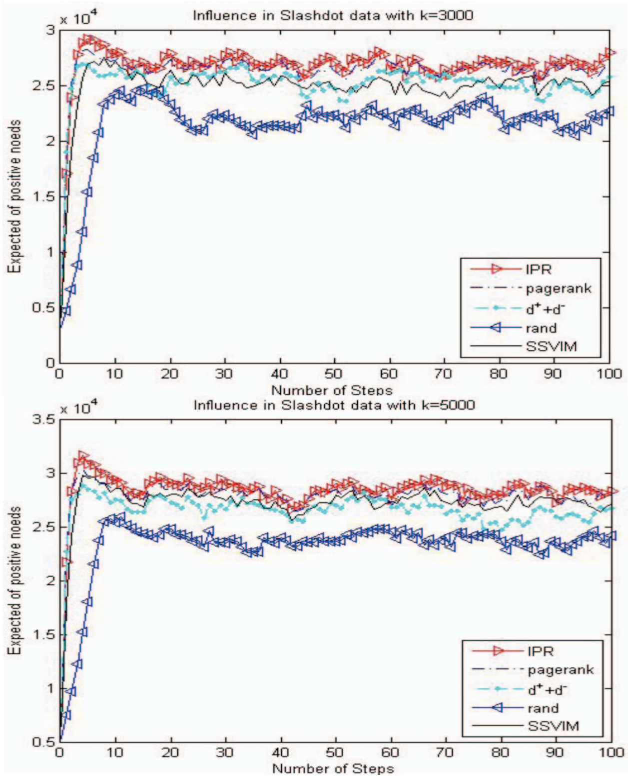

Fig. 2. Influence in Slashdot with seed set size k=3000 and 5000.

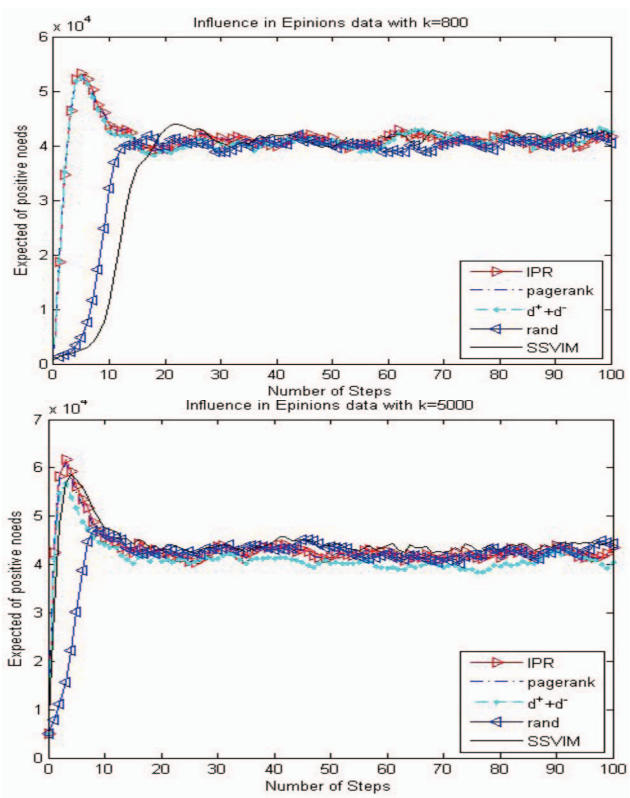

Fig. 3. Influence in Epinions with seed set size k=800 and 5000.

V. CONCLUSION

There are many studies about influence maximization on social networks, but most of them do not incorporate with hostile relationships. In this paper, we first extended the basic voter model for signed network with the negative edges, and then propose the Integrated PageRank method used in singed networks to select seeds to maximum influence. The result demonstrates the efficiency of our method.

There are still several open problems and the future directions. One problem is to decide the value of *k*, it does not mean the higher value of k is set, the better result we can get. Sometimes the higher value of k does not bring the more positive nodes but increases the burden of the budget.

ACKNOWLEDGMENT

This research was jointly supported by the National Natural Science Foundation of China (NSFC) (61272200), the Program for Excellent Young Teachers in Higher Education of Guangdong, China (Yq2013012), and the Fundamental Research Funds for the Central Universities (D2153440).

REFERENCES

[1] Facebook statistics.[Online]. Available: http://www.facebook.com/press/info.php?statistic

[2] D.Kempe, J.Kleinberg, and E.Tardos, “Influential nodes in a diffusion model for social networks,” in International colloquium on automata, languages and programming No32, pages 1127-1138, 2005.

[3] M. Granovetter, “Threshold models for competitive influence in social networks,” American Journal of Sociology, 1978.

[4] D.Kempe, J.Kleinberg, and E.Tardos, “Maximizing the spread of influence through a social network,” in KDD, 2003.

[5] E.Even-Dar and A.Shapira, “A note on maximizing the spread of influence in social network,” in WINE,2007.

[6] Jure Leskovec, Daniel Huttenlocher, Jon Kleinberg, “Signed networks in social media,” in CHI, 2010.

[7] Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang, “Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships,” in WSDM’13, 2013.

[8] Boldi, P., Santini, M., and Vigna, S, “ PageRank as a Function of the Damping Factor,” in Proceedings of Fourteenth International Conference on World Wide Web, Chiba, Japan, May 2005.

[9] P.Clifford and A.Sudbury, “A model for spatial conflict,” Biometrika, 60(3):581,1973.

[10] R. Holley and T. Liggett, “Ergodic theorem for weakly interacting infinite systems and the voter model,” The annals of probability, 1975.

[11] J. Leskovec, D. Huttenlocher, J. Kleinberg, “Signed Networks in Social Media,” 28th ACM Conference on Human Factors in Computing Systems (CHI), 2010.

[12] L. Page, S. Brin, R. Motwani, T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” Stanford Digital Library Technologies Project, 1999.

[13] Soc-sign epinions network. [Online]. Available: http://snap.stanford.edu/data/soc-sign-epinions.html

[14] Soc-sign-Slashdot090221 network. [Online]. Available: http://snap.stanford.edu/data/soc-sign-Slashdot090221.html

[15] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In ACM KDD, 2010.