Discovering Influential Users in Social Media to Enhance Effective Advertisement

Mohammad J. Tarokh[1], Hosniyeh S. Arian[2] and Omid R. B. Speily[3]

[1] **Associate professor, IT Group - Faculty of Industrial Engineering, K. N. Toosi University of Technology**
Tehran, Iran
mjtarokh@kntu.ac.ir
[2] **Department of Computer Engineering and Information Technology, Islamic Azad University**
Qazvin, Iran.
hosniyeh.safiarian@gmail.com
[3] **Department of Computer Engineering and Information Technology , Urmia University of Technology**
Urmia, Iran
Speily@uut.ac.ir

Abstract

Recently, companies have attempted to take advantage of social advertising to deliver their advertisements to appropriate Customer. Social network advertising, also social media targeting is a group of terms that are used to describe forms of online advertising that focus on social networking services. One of the major benefit of this type of advertising is that advertisers can take advantage of the users' demographic information and target their ads appropriately. In social advertising ,identifying influential users is important issues because companies can reduce cost of advertising by them.in this paper, we try to proposed a method based propagation structure , advertising content and behavioral of users until identifying influential users. Our experimental results showed that the proposed method could be useful for advertising in social media.

Keywords: *social network, content, trust, interest, influential users.*

1. Introduction

Social network includes a set of individuals and ties among them, such as relationships and interactions [1]. Online social networks such as Facebook and LinkedIn have been made friends and finding jobs, and other types of social networks [2]. Recent researches indicate a great usage of social network which lead to a dramatic increase in the capacity of business materials on the web [3]. they give marketers a great opportunity to diffuse information [4], plays an important role for the spread of information, ideas, and influence among its members[4].

Also Social network analysis has received increasing interest in many different areas in recent years, including community detection [5,6], role detection [7,8], etc .Social advertising is a kind of recommendation system, of sharing information between friends.

In 2010, e-Marketer reported that 90% of consumers depend on recommendations from people that they can trust. The efficiency of social advertising is more than the traditional advertising. It seems that social advertising has become an important advertisings for marketers [9].

Some users can influence on the other users that called active users and another users called passive. this influencing can be positive or negative when a user posts a message it would be directly transmitted to its followers then the followers will decide whether to repost this message or not .there are many factors for influencing ,same as propagation models, social network structure ,friends and so on[10].in this paper we devise a method for discovering influential users in social media and improve advertising in social media. The rest of this paper is organized as follows. In the next section, some of the most important related works are reviewed. In section 3, the structure of proposed method for discovering influential users in social media to enhance effective advertisement. The evaluation of this method is discussed in section 4. Finally, last section presents conclusion and directions for future works.

2. Related Works

In [11] studied on choosing set of good customers in viral marketing for optimization problem .they proved that the optimization problem was NP-hard for both LT and IC Models. They presented the greedy algorithm. The greedy solution would been produced result which was optimal , but scalability was low in this method. In [12] studied on sub modular property of influential function in kempe's greedy algorithm and proposed CELF algorithm based lazy forward. in scalability, CELF was better than greedy algorithm and memory usage of CELF was low than greedy algorithm .In [13] paid attention on greedy algorithm by kemp and it's improvement by Leskovec they tried to reduce time run greedy algorithm.they combined

ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 5, No.17 , September 2015
ISSN : 2322-5157
www.ACSIJ.org

greedy algorithm with CELF algorithm called mixed greedy algorithm. Influential spread in mixed greedy algorithm was exactly that of Kemp greedy algorithm but running time was less than CELF. In[14] presented CELF++ algorithm based CELF algorithm. CELF++ algorithm was avoided unnecessary re computations of marginal gains incurred by CELF. Experiment was showed CELF++ algorithm was significant improvement in running time and number of node look up but the memory usage CELF++ was more than CELF. In [15] presented SIMPATH algorithm for influence maximization under the LT model. SIMPATH was optimized computation and improved quality of seed selection by Vertex Cover Optimization and Look Ahead Optimization SIMPATH can active nodes more than another algorithms and it's run time ,memory use age and quality of seed sets was better than LDAG[16].In[17] proposed a new algorithm for mining top-K influential nodes, called Community-based Greedy algorithm (CGA). This algorithm had community structure property of social networks. CGA was faster than mixed greedy but the number active nodes were near to mixed greedy and new greedy. In [18] tried to solve scalability problem with sparsification of network called SPINE. The goal of sparsification detected a sub network with its properties. SPINE had two phases, first selected a set of finite likelihood edges and second it greedily seek a solution of maximum. SPINE can reduce time running of influence maximization problem and number of active nodes in sparsification of network close to full of network. In [19] proposed a method based action logs that consist of relation Actions (User, Action, Time) for detecting influential nodes and learning influential probability. The propagation of influence was modeled by edges connecting them in direction of propagation. In [18] proposed a method for discovering leaders in social network based frequent pattern of Influence and association rules mining.

3. Proposed Method

Users communicate with another person and join to different groups on social network. Also they do some activities same as like, share, reply [20]. All of these data are stored in DB. Proposed method uses these logs for its input data that includes of user's interest and activities. As illustrated in Fig.1 the proposed method generally. consists of :user's log file , computing trust detecting interest ,clustering based on interest and trust, influence check ,detecting leader.

3.1. Computing Trust

users tend to relate with another persons who are similar to themselves[21].users trust to these persons and fallow them

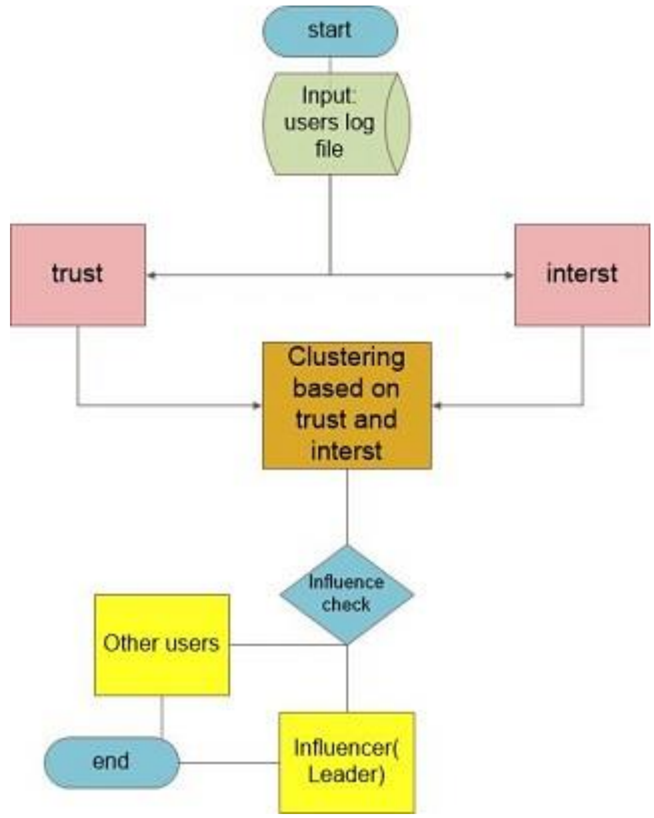

Figure 1: proposed Method

on social network[22]. We Consider on an inactive user u and the set of its active neighbors S (that is all nodes in S already performed a certain task). Trust probability is gained by equation (1)[19].

$$p_u(S)=1-\prod_{v\in S}(1-p_{v,u}) \tag{1}$$

3.2. Computing user's interest

Users share information and fallow their interest content. When a user u does an interest action under influence of its neighbors, who already performed the action before u, all the active neighbors of u are assigned or shares partial 'credit' for influencing u to fallow this action. Partial 'credit' is gained by equation (2)[19].

$$Credit_{v,u}(a)=\frac{1}{\sum_{w\in S} I(0<t_u(a)-t_w(a)\le \tau_{v,u}} \tag{2}$$

In equation [2] if user u performs an action a is 1 and otherwise returns 0. $\tau_{v,u}$ is computed in equation (3)[23], A_{v2u} is the number of action propagated from v to u.

$$\tau_{v,u}=\frac{\sum_{a\in A}(t_u(a)-t_u(a))}{A_{v2u}} \tag{3}$$

3.3. Clustering based on interest and trust

in part 3-2 and 3-3 similarity was detected by content and interest .so users join to comminutes that similar to themselves . The algorithm works in a greedy manner starting with each vertex in its own cluster and then repeatedly joins the two communities whose amalgamation provides the maximum increase in the modularity. Community is gained by equation (4)[17] as the fraction of edges that join vertices in community v to vertices in community w in equation(5)[17] and which is the fraction of edges that are attached to vertices in community v in equation (6)[17].

$$Q = \sum_{v} (e_{vw} - a^2) \tag{4}$$

$$e_{vw} = \frac{1}{2m} \sum_{ij} (A_{ij}) \partial(C_i , v) \partial(C_j , w) \tag{5}$$

$$(a)^2 = \frac{1}{2m} \sum_{k=0}^{n} (K_i) \partial(C_i , v) \tag{6}$$

3.4. detecting influential users

The goal of this stage identifies influential users in a social network. Influential users would maximize the influence spread function. There are three constraints in finding influential users[24]:

- For an action, should influence sufficiently large number of users ($>\psi$)
- For an action, should influence these users in a reasonable amount of time ($<\pi$)
- Should act as a leader in sufficiently large number of actions ($>\sigma$)

Influential users are detected by (7)[23].

$$\exists S \subseteq I, | S | > \sigma : \forall a \in S. size(\inf_{\pi}(v, a)) \geq \psi \tag{7}$$

So a user should do actions lager than a given action threshold (σ) and Inf(π) is number of users performed the action a after v within time threshold π .

4. Evaluation

In order to evaluate the proposed method on Twitter offer application program interface (API) [1] and Yahoo Meme application program interface (API) [2] enabling data collection easily. we collect detailed information from users and followers. 1500 users selected randomly, they fallow their friends and share information and feedbacks same as like, share, comment on social networks. This information includes of news, linked, events and other diversity topics. Influential users can maximize the influence spread information.

The statistical information of this experiment are depicted in table 1 and table 2 .there are some groups and they show their interests based their interest and activities same as like, share, comment, mention.

Table 1 Summary of properties of Twitter dataset

Parameters name	Value of parameters
Number of users	1500
Number of contents	13500
average number of like	39150
average number of Share	22950
average number of comment	35100
average number of mention	28650
Average number of friends	33615

Table 2 Summary of properties of Yahoo Meme dataset

Parameters name	Value of parameters
Number of users	1500
Number of contents	9450
average number of like	26649
average number of Share	16065
average number of comment	25420
average number of mention	19264
Average number of friends	28815

4.1. Performance Analysis

The goal of our experiments is to show that influence spread achieved by our proposed method improves influence spreads that can be achieved by standard approaches like [24][25].We compared influence spread, number of nodes activated by seed set discovered, achieved by our proposed method and [24][25]. With the following approaches.Figure2 shows the influence spreads of proposed method and other two methods[24][25] on social network generated from twitter data set. Our proposed method very closely to method [24].It also outperforms method [25] for all seed set sizes. Also in Yahoo Meme dataset the spread achieved by proposed method outperforms both methods [24][25] based solutions figure 3. As expected method[24] performs inconsistently in both datasets and in some cases it even performs below the method[25].

[1] http://snap.stanford.edu/data/index.html

[2] http://snap.stanford.edu/data/index.html

4.2. Runtime of APG

To compare runtime of proposed method with [24] and [25] we recorded time required to select Influential nodes of different sizes. Figure 4 reports the runtime comparison on twitter dataset and figure 5 reports the same on Yahoo Meme dataset. In both datasets performs almost in constant time. Proposed method longer than [24] as the size of the required set of influential nodes increases in both datasets.

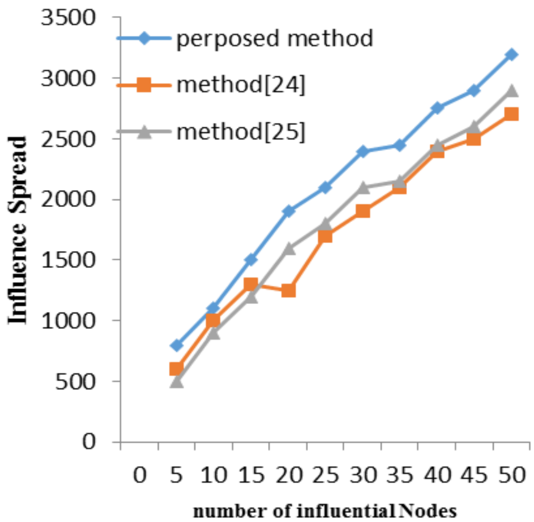

Figure 2: Influence spreads of different algorithms on twitter Dataset under proposed method

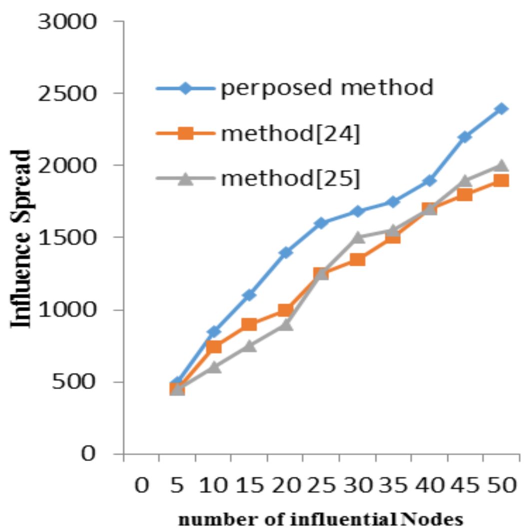

Figure 3: Influence spreads of different algorithms on Yahoo Meme Dataset under proposed method

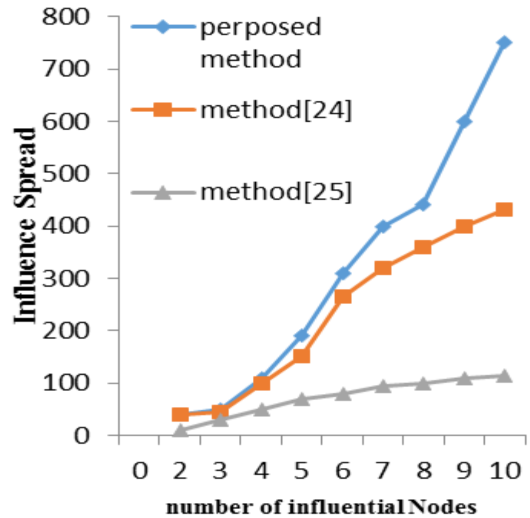

Figure 4: Running time of different algorithms on twitter Dataset under proposed method

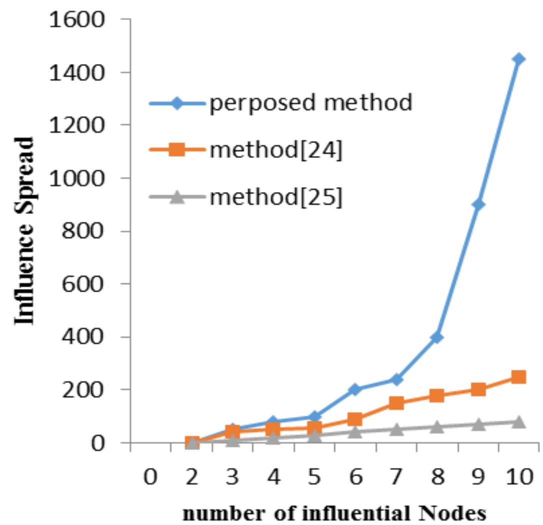

Figure 5: Running time of different algorithms on Yahoo Meme Dataset under proposed method

5. Conclusion

The contributions of this paper are summarized as follows. Firstly, from the perspective of system innovation, while discover influential users on social network .

Secondly, from the perspective of methodology, we consider the user’s interest and also propagate (social activeness, social interactions, and social similarity) factors in the evaluation of nodes' diffusion capabilities to discover the people who can spread the adverting messages widely.

Thirdly, from the perspective of performance, that our method can raise the visibility of advertising information. Our proposed method can widely extend the spreading coverage of advertisements.

References

[1] Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks Science, 286, 509–512

[2] Tsuyoshi Murata and Sakiko Moriyasu, "Link Prediction of Social Networks Based on Weighted Proximity Measures," in International Conference on Web Intelligence, 2007, pp. 85-88.

[3] D. Watts, J. Peretti, Viral marketing for the real world, Harvard Business Review (2007) 22–23.

[4] M.E. Newman, J. Park, Why social networks are different from other types of networks, Phys. Rev. E 68 (3) (2003) 036122.

[5] P.A. Grabowicz, J.J. Ramasco, V.M. Eguíluz, Dynamics in online social networks, Dynamics on and of Complex Networks, vol. 2, Springer, 2013, pp. 3–17.

[6] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of online social networks, in: Link Mining: Models, Algorithms, and Applications, Springer, 2010, pp. 337–357.

[7] J. Jiang, C. Wilson, X. Wang, W. Sha, P. Huang, Y. Dai, B.Y. Zhao, Understanding latent interactions in online social networks, ACM Trans. Web (TWEB) 7 (4) (2013) 18.

[8] R. Xiang, J. Neville, M. Rogati, Modeling relationship strength in online social networks, in: Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, pp. 981–990.

[9] Agarwal, N., Liu, H., Tang, L., and Yu, P. S. Identifying the influential bloggers in a community. In *Proceedings of the international conference on Web search and web data mining*. WSDM ’08. ACM, New York, NY, USA, 207–218. 2008.

[10] B.N. Hou, Y.P. Yao, D.S. Liao, Identifying all-around nodes for spreading dynamics in complex networks, Physica A 391 (2012) 4012–4017.

[11] Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. KDD ’03. ACM, New York, USA, 137–146. 2003.

[12] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. KDD ’07. ACM, New York, NY, USA, 420–429. 2007.

[13] Chen, W., Wang, Y., and Yang, S. 2009. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. KDD ’09. ACM, New York, NY, USA, 199–208.

[14] Goyal, A., Lu, W., and Lakshmanan, L. V. Celf++: optimizing the greedyalgorithm for influence maximization in social networks. In Proceedings of the 20thinternational conference companion on World wide web. WWW ’11. ACM, NewYork, NY, USA, 47–48. 2011.

[15] Goyal, A., Bonchi, F., and Lakshmanan, L. V. S. A data-based approach tosocial influence maximization. Proc. VLDB Endow. 5, 73–84. 2011.

[16] Chen, W., Yuan, Y., and Zhang, L. 2010. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. In Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10). IEEE Computer Society, Washington, DC, USA, 88-97

[17] Wang, Y., Cong, G., Song, G., and Xie, K. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In Proceedings of the۲thE ACM SIGKDD international conference on Knowledge discovery and datamining. KDD ’10. ACM, New York, NY, USA, 1039–1048©. 2010.

[18] Mathioudakis, M., Bonchi, F., Castillo, C., Gionis, A., and Ukkonen, A. 2011Sparsification of influence networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. KDD ’11. ACM,

[19] Goyal, A., Bonchi, F., and Lakshmanan, L. V. 2010. Learning influence probabilities in social networks. In Proceedings of the third ACM international conference on Web search and data mining. WSDM ’10. ACM, New York, NY, USA, 241–250

[20] Gartner 2013. Gartner’s Social Marketing Survey 2013.Accessedfrom.http://blogs.gartner.com/adam-sarner/2013/04/01/gartners-social-marketing-survey-2013.

[21] U. Brandes and T. Erlebach (Eds.), Network Analysis - Methodological Foundations, LNCS Tutorial 3418, Springer Verlag, 2005.

[22] Golbeck, J. Parsia, B. Trust network-based filtering of aggregated claims, Int. J. metadata, Semantics and Ontologies, Vol. 1, No. 1, 2006

[23] Goyal, A., Bonchi, F., and Lakshmanan, L. V. 2008. Discovering leaders from community actions. In Proceeding of the 17th ACM conference on Information and knowledge management. CIKM ’08. ACM, New York, NY, USA, 499–508.

[24] Yen-Liang Chen, Kwei Tang, Chia-Chi Wu, Ru-Yun Jheng,2014, Predicting the influence of users' posted information for eWOM advertising in social networks.in Electronic Commerce Research and Applications.
[25] Yu, B., Chen, M., Kwok, L., 2011. Toward predicting popularity of social marketing messages, LectureNotes in Computer Science 6589, 317–324.

Mohammad J. Tarokh is an associate professor in the department of Industrial Engineering at K.N. Toosi University of Technology, Tehran, Iran. He received his B. Sc. from Sharif University of Technology in Tehran, M.Sc. from University of Dundee in UK and Ph.D. from University of Bradford, UK. His main research interests are in knowledge management, business intelligence, customer relationship management and supply chain management.

Hosniyeh S. Arian is a M.Sc. student in Electronic Commerce at Qazvin Islamic Azad University, Qazvin, Iran. She received her bachelor's degree in information technology engineering from PNU University, Hamedan, Iran. Her current primary research interests include Social network, Data Mining and e-Marketing.

Omid R. B. Speily received his B.S. degree in computer engineering from Urmia University (2006) and his M.S. degree in Information Technology from AmirKabir University of Technology (2009). He is currently Ph.D. Student in AmirKabir University of Technology. Since 2009 he joined Urmia University of Technology as faculty member. He granted Iran minister of science national fellowship since 2011. His research interest includes the dynamics complex networks, graph theory, intelligent system.