ISSN (Online): 0974-5645

ISSN (Print): 0974-6846

A Study on Influential Evaluation of Information hubs in Social Networks

A.N. Arularasan and S. Koteeswaran

¹Research Scholar, Department of Computer Science and Engineering, Vel Tech University, Chennai-62, TamilNadu, India. E- Mail: arularasan@live.com

²Associate Professor, Department of Computer Science and Engineering, Vel Tech University, Chennai-62, TamilNadu, India. E- Mail: s.koteeswaran@gmail.com

Abstract

Objectives: The objective of the work is to identify the most influential nodes and to classify the information hubs in the social network in promoting the viral marketing. **Methods:** N-gram model classification approach is used for categorizing the information hubs of different domains. **Findings:** The proposed method aims at estimating influence of each user in the network and categorizes the information hubs by using domain based classification to advertise a product in a social network. **Applications/Improvements:** The work can further be extended to slice the longer string into N characters to measure the similarity value depending on the set of characters between words and relationship among them.

Keywords: Social Network, Information hub, influence maximization, influential node, Viral Marketing.

1. Introduction

Rapid growth of information technology and web usage spawns various information sharing systems on internet. The most popular websites are Google, Facebook, YouTube and Twitter on the cyberspace. These website users can produce the social network that forms a popular platform for interaction^{1, 2}. Users publish their own profile and maintain the social relationship among users with similar interests. As the social network involves certain kinds of relationships among individuals, users can diffuse the information by following user-to-user links. Most social websites allow users to create a special group and it allows the users from other groups to join and view the information posted to a special group. Thus, the social networks play a vital role in personal and commercial online interaction^{3,4}.

The business model of social networks is based on advertising and it employs users' information for ad targeting purposes^{5, 6,7,8}. Oral communication or otherwise called viral marketing is simply the diffusion of business ideas or innovations in social networks to market business ⁹. An advertising sector offer free samples to

users randomly, but they are not equal in diffusing the information to others. Users differ widely in terms of the number of friends and interaction frequency. For a better result, an advertising sector offers free sample to information hubs only, instead of providing free samples to users randomly. A member or a node¹⁰ consists of many contacts with other nodes in a network is referred as an information hub or influential node^{11, 12}. A contact refers to the information sharing of a node with others. Thus, the information hub plays a fundamental role in diffusing information, including propaganda, ideology, and gossip in the social network, which considerably raise the effectiveness of oral communication and viral marketing¹³.

Generally, the users are interested in various topics, and hence the personalization of the user's profile becomes mandatory while selecting the information hubs for advertisement^{14,15}. It is essential to employ the classification approach for categorizing the information hubs under various domains based on hubs interest. For instance, an advertiser selects a small number of information hubs to issue sample movie tickets. Mostly, a single area of hubs' interest is considered in the entire network. However, in

^{*}Author for correspondence

practice it is not absolutely in a social network. Thus, the influence maximization model selects the information hubs according to their influence in spreading and relations, and then it classifies the information hubs based on their interest by observing their profiles and activities.

As the increased popularity and tremendous growth of information technology, the social networks are not only social platforms, but it is used as a marketing channel for business today. The marketing agents promote the products and service by posting business advertisements on social network sites. Even though business agents cannot contact users directly, they can reach millions of users by interfering with the highly information hubs. Information hub provides not only a channel to interact with customers, but also a source of valuable information about customers such as influential frequency and relationship. There is a relationship between hubs' interest and consumption choices. So the marketing agents are interested in monitoring the social interactions of information hubs to improve viral marketing. Classification of information hubs based on interest facilitates the advertisers to post the advertisement based on hubs' interests. The promotion of viral marketing can be realized with a constraint that sample products are sponsored only to the information hubs interested in a particular domain.

Applications of information hub

- Information hub mainly plays important role in business based applications in social network especially in marketing.
- Third parties make use of the specific interested area of the information hub in the social network for advertising purposes.
- Mining the logs of users on the social network has a wide range of applications such as improving search results, building user models, and understanding recent trends in the market
- Several organizations offer security consulting by identifying communities of users in social networks.

2. Objective of the Study

An important platform for effective viral marketing practice is online social networks. In order to promote a product's sales, it is important to determine the node (individual) in the network with high influence. This information hub considerably leads to wider development in innovation than usual way of individual marketing. A decreasing cascade model in which the behavior spreads in a cascading manner is proposed and it is based on the perspective rule, starting with a group of originally "active" nodes. Decreasing cascade model is a greedy algorithm for expanding the influential extent.

A heuristic algorithm is proposed to achieve optimal result in most models16. It suggested two different debatable cascade models such as independent cascade (IC) as well as Linear Threshold (LT) model. Each active node in independent cascade gets only one chance to mobilize each of its neighbors independently with a certain possibility. Each edge and vertex of the linear threshold has a weight and threshold respectively that are chosen consistently at random, and if the threshold exceeds weighted sum of its active neighbors then the vertex becomes activated.

The basic model for the above techniques is a greedy approximation algorithm that suffers from the disadvantage that the under-lying objective function information diffusion cannot be computed exactly and effectively. In order to overcome this issue, a novel and innate method to find the most significant nodes with respect to a predefined metric in social networks is proposed using the Shapley value concept. The top-k influential nodes are identified in social network based on suitable information¹⁷.

The marginal role that the node causes the coalitional dynamics of a theoretic game is allotted by Shapley value meant for every node. As marginal contribution of the node is high, the Shapley value of that node will also become high, and the most important is that the node should be among the other nodes. It suggests the degree discount heuristic of the IC model, but the selection of the authority nodes is difficult in the social networks. It does not consider an edge between two nodes towards its degree when the authority node is already selected as the second node. This technique increases the expected number of nodes, but it does not increase the total flow of information in a network. It improves the existing greedy algorithms by selecting the authority nodes that is being defined to the information.

The considered portable social network extends the IC model to adapt contact frequency¹⁸. The extent of influence is moreover related to influence frequency and also connected with the user preferences. As the preferences for information increases, the more likely the influence occurs. An IC model is suggested for information diffusion in19 and each node activates each of its neighbors separately with a certain probability. The major disadvantage of this system is that it does not consider user preference.

To overcome the issues in^{20, 21} two-stage mining algorithm (GAUP) is proposed in discovering the important influential node on a particular subject. In the primary stage, GAUP exploits a collaborative filtering technique to estimate user preference on the subject. In the secondary stage, GAUP accepts a greedy algorithm to determine most influential nodes in a network. A new algorithm is suggested to determine the information hubs in²² and it can preserve user privacy. It identifies the information hubs without requiring a central entity to access entire community graph. It achieves the above objective by completely allotting the computation using the kempe-Mcsherry algorithm and also addresses user confidentiality concerns.

Most of the existing works have not taken into account the user preference for identifying the information hub. The existing schemes for identifying the influential node works, consider only a single area of interest in the network, i.e. they assume that the users on the entire network have the same area of interest. This fact does not suitable for real social networks. Therefore, it is necessary to determine several hubs in various topics of interest in a distributed fashion.

The existing models have been proposed to identify the contagion (information, disease) spread from one to other nodes²³. It selects the nodes to flood the information into the network randomly, but it neglects the external influencing factors. It is important to explicitly model the influencing spread of the external node in the diffusion framework. Information hubs are related to the high centrality nodes. Several influence maximization techniques have been proposed24,25 and the influential nodes are determined based on the steady state behavior of the network.

The next challenging problem is the influential maximization in social networks^{26, 27}. The target-set selection problem for influential spreading under the scenario of mobile data offloading is difficult. Maximizing the influential spread, i.e. the plenty of nodes receiving the message from one node through opportunistic communications is in need to investigate. This work introduces the overlapping community detection to determine the information hubs in the social networks. The overlapping community structure is made based on the history of interaction.

Based on the motivation for promoting the viral marketing, several works have been proposed on influence maximization in social networks^{28, 29}. It proposes the new expectation model for promoting the influence spread of a seed for improving the business marketing. Fast greedy based approximation method is proposed based on the expectation model. A new metric proposed for social interaction influencing is gravitation or user attraction. It takes into account the quantitative measurement of user interactions within the network. Further improvement in the social interaction includes topics and user profile such as gender, age, interests, and etc³⁰. However, the authority determination of interactions between users is not provided in the social networks and this is a much more challenging process in real scenarios. It is necessary to examine the influence optimization and the interaction tracing process.

The Method and technique along with the key features are provided in the Literature survey Table 1.

3. Proposed Solution

Most research in viral marketing over social networks has been focused on selecting the information hubs for improving the influence maximization. User's information plays a fundamental task in the influence spread from one user to others. However, most of the existing works have not taken into account the user preference for identifying the information hub and interest or influence of users in a particular domain to reach effective advertisement. The users on the entire network have the same area of interest, but it is not absolute in a real social network. Therefore, it is necessary to identify the information hubs in various topics of interest in a distributed fashion. As in case of advertisement, long standing challenges are to identify the highly influential users or information hubs in large scale social networks and observe hub's interest on a particular domain. Moreover, the existing works lacks techniques for classifying the information hubs in various topics of interest for spreading information in a social network.

To remedy this research gap, this paper proposes a new technique which takes into account the influence spread and various topics of interest for identifying the information hubs more effectively over social networks.

Overview

The main objective of this work is to identify the information hub successfully to promote the viral marketing in social networks.

Table 1. Survey Table

Sl. No.	Year	Authors	Method / Technique	Key features
1.	2014	Bo Liu, Gao Cong, Yifeng Zeng, Dong Xu, Yeow Meng Chee	Time step based simulation algorithm	Explicit incorporation of the time factor improves influential maximization
2.	2014	Yeslam Al-Saggaf, Md Zahidul Islam	Privacy - Content analysis	Masking the individual's data carefully helps to preserve privacy
3.	2013	Bo Han, Jian Li, Aravind srinivasan	Random walk sampling	The targeted immunization using iWander achieves a comparable performance with a degree-based immunization policy
4.	2013	Jong-Ryul Lee, Chin-Wan Chung	Fast greedy-based approximation method	Identifying local influencing regions makes the basic greedy algorithm faster
5.	2013	Guojun Wang, Wenjun Jiang, Jie Wu, Zhengli Xiong	Feature based influence model	It can distinguish the social influence of different users that reduces the duplication
6.	2013	Muhammad U. Ilyas, Zubair Shafiq M, Alex X. Liu, Hayder Radha	Solution using eigen-vector centrality	It eliminates the requirement of a central authority for identifying hubs
7.	2013	Mario Cataldi, Mittal Nupur, Marie-Aude Aufaure	N-Gram model classification approach	Studying the relationships among users based on the topics they share improves the information diffusion
8.	2013	J. Cruz, R. Therón, E. Pizarro, Francisco J.	Virtual Worlds	Provides the opportunity to meet a number of hidden data
9.	2013	Sukhpreet Dhaliwal, Hikmat Jahangirli, Riad Aljomai, Abdullah Sarhan, Wadhah Almansoori, Reda Alhajj	Clustering, association rules	Used to overcome the problem of sparsity. Finds out preferred items and sort them from most recent to oldest
10.	2011	Yunlong Zhang, Jingyu Zhou, Jia Cheng	Two-stage mining algorithm	Latent semantic indexing model efficiently adopts the greedy algorithm to find most influential nodes
11.	2010	M.Kimura, K.Saito, R.Nakano	Greedy algorithm	Bond percolation and graph theory were used to compare the proposed method with the conventional method in terms of computational complexity
12.	2010	Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, Xiangyang Lan, Siddharth Suri	Independent cascade Model	It reduces the duplication
13.	2010	Yu Wang, Gao Cong, Guojie Song, Kunqing Xie	Community based greedy algorithm	High speed in information diffusion
14.	2010	Wei Chen, Chi Wang, Yajun Wang	Two-stage mining algorithm	Collaborative filtering technique accurately estimates user preference on a topic
15.	2010	Ramasuri Narayanam, Narahari Y	Independent cascade Model	Improves the information diffusion by identifying influence frequency
16.	2010	Wei Chen, Yifei Yuan, Li Zhang	Linear threshold model	Easily scale up to networks of millions of users and it maximizes the expected number of nodes
17.	2009	Wei Chen, Yajun Wang, Siyu Yang	Heuristic algorithm	Heuristic algorithms are orders of magnitude faster than all greedy algorithms
18.	2008	Rama Suri N, Narahari Y	Shapley Value Based Algorithms	Distributed influential hubs identification and also addresses user privacy concerns

- To classify the information hubs under various domains using N-Gram model classification approach.
- To expand the tweets that are often quite short using term-similarity based tweet expansion method.
- To classify (label) the interest of information hubs by analyzing the profile information using machine learning technique and to use these labeled-information hubs to advertise a product or service in a social network.

Most of the work lacks technique for considering the domain of interest of information hubs for promoting the viral marketing. The proposed methodology taking into account the user's profile and influence spread of a user based on their exchange of information about a particular domain. It generates a directed multi-graph in which the nodes represent the individual user and directed links represents the relation of being in pairwise contest. Every link is associated with a timestamp representing the previous time of contact. The contact graph assists to determine the information hubs or highest degree nodes, having more links.

The proposed methodology attempts to classify the information hubs by observing the profile information and analyzing each tweet and re-tweet of an information hub with respect to the domain of interest. A tweet refers to the messages sent and read by the users and retweet refers to the re-post of a tweet from another user. The proposed methodology integrates a N-Gram model classification approach for classifying the information hubs under various domains. The main challenge associated with the tweets is the messages that are often quite short and the words posted in tweets often lack context. This work employs term similarity based tweet expansion method. It estimates the cosine similarity between words in two tweets, posted by a information hub and it simply combines similar tweets according to the distance of their tweet vectors. Moreover, it estimates the appearance frequency of each word of an expanded tweet in a corpus of documents and profile information about the information hub. For each tweet and profile information, the system classifies the information hubs in descending order of similarity value in various domains of interest. Finally, the proposed methodology aims at reckoning the influence of each user in the social network and classifies the information hubs by using the interest based domain classification.

4. Results and Discussion

The existing works identifies the information hubs, by building the contact graph or pattern of a network based on the user's interaction and friends link. However, none of the works analyzed how the information hub gets influenced by specific domain or a company's product. Identifying different sets of influential nodes of various area of interest are important, yet utilizing the similarities among the area of interest. The most challenging feature of social networking is to support the scalability. The social website normally scales up to several millions, and it is difficult to locate the functionality of crawling, user's area of interest, and statistic maintenance across multiple servers. Classifying the area of interest of the particular user among hundreds of million users is a difficult task in a social network. Only a few works focused on using machine learning techniques in social network analysis.

5. Conclusion

In this paper, we presented a notion of identifying the most influential nodes and classifying the information hubs based on their interest in the social network. The integrated N-Gram model classification approach classifies the information hubs under various domains that improve the viral marketing in social networks. This paper has provided a brief description about existing algorithms and various models of detection techniques that are used in social networks to analyze and visualize information hubs.

6. Future Work

One of our future works is to slice the longer string into N characters to measure the similarity value depending on the set of characters between words and relationship among them. Further work can be done to classify the short messages most precisely.

7. References

- 1. Hanneman R A, Riddle M. Introduction to social network methods. University of California: Riverside, 2005.
- Breiger R L. The Analysis of Social Networks Handbook of Data Analysis, Sage Publications Ltd.: London 2004, 505-526.
- Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge University Press: USA, 1994
- Koteeswaran S, Visu P, Kannan E, Enhancing JS-MR Based Data Visualisation using YARN, Indian Journal of Science and Technology, 2015, 8(11), pp. 1-5.

- 5. Brown J J, Reingen P H. Social ties and word-of-mouth referral behavior. The Journal of Consumer Research. 1987, 14(3), pp. 350–62.
- Andrew Yang T, Kim D J, Dhalwani V. Social networking as a new trend in e-marketing. Research and Practical Issues of Enterprise Information Systems II. Springer: US, 2008, 847-856.
- 7. Falch M, Henten A, Tadayoni R, Windekilde I. Business Models in Social Networking. CMI International Conference on Social Networking and Communities, Denmark, 2009, pp. 1–23.
- 8. Hu X. Social media business model analysis-Case Tencent, Facebook, and Myspace. Aalto University: Finland, 2011, pp. 108.
- 9. Misner R. The World's best known marketing secret: Building your business with word-of-mouth marketing. Bard Press: USA, 2nd edition, 1999.
- Koteeswaran S, Srinivasan R, VarunKumar K.A., Region Specific Election Routing Protocol for Wireless Sensor Networks, Indian Journal of Science and Technology, 2014, 7(12), pp. 2083–87.
- 11. Kempe1 D, Kleinberg J, Tardos E. Influential nodes in a diffusion model for social networks, Proceedings of the 32nd International Colloquium on Automata Languages and Programming, 2005, pp. 1127–38.
- 12. Kimura M, Saito K, Nakano R, Motoda H. Extracting influential nodes on a social network for information diffusion. Data Mining and Knowledge Discovery. 2010, 20(1), pp. 70–97.
- 13. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, 137–146.
- Cataldi M, Nupur M, Aufaure M-A. Estimating domainbased user influence in social networks. Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC), Portugal, 2013, 1957–1962.
- 15. Weng J, Lim E-P, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers. ACM International Conference on Web Search and Data Mining, 2010, pp. 261–70.
- Chen W, Yuan Y, Zhang L. Scalable influence maximization in social networks under the linear threshold model. IEEE 10th IEEE International Conference on Data Mining, Sydney, NSW, 2010, pp. 88–97.
- 17. Rama Suri N, Narahari Y. Determining the top-k nodes in social networks using the Shapley value. Proceeding of 7th International Conference on Autonomous Agents and Multi agent Systems, Vol.3, 2008, pp. 1509–12.
- 18. Narayanam R, Narahari Y. A shapley value based approach to discover influential nodes in social networks. IEEE trans-

- actions on Automation Science and Engineering, 2010, pp. 1–18.
- 19. Wang Y, Cong G, Song G, Xie K. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. KDD'10, 2010, pp. 1039–48.
- Chen W, Wang C, Wang Y. Scalable influence maximization for prevalent viral marketing in large scale social networks. ACM KDD Conference, 2010, pp. 1029–38.
- Zhang Y, Zhou J, Cheng J. Preference-based Top-K Influential Nodes Mining in Social Networks. International Joint Conference of IEEE Trust Com Conference, 2011, pp. 1512–18.
- 22. Ilyas M U, Zubair Shafiq M, Liu A X, Radha H. A Distributed Algorithm for Identifying Information Hubs in Social Networks. IEEE journal on selected areas in communications, 2012, 31(9), pp. 629–40.
- Cosley D, Huttenlocher D, Kleinberg J, Lan X, Suri S. Sequential influence models in social networks. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010.
- 24. Chen W, Wang Y, Yang S. Efficient Influence Maximization in Social Networks. KDD Proceedings of the 15th ACM SIGKDD international Conference on Knowldege discovery and data mining, 2009, pp. 199–208.
- 25. Kempe D, Kleinberg J, Tardos E. Maximizing the Spread of Influence in a Social Network. ACM SIGKDD Conference of the 9th ACM SIKDD International Conference on Knowledge Discovery and data minig, 2003, pp. 137–46.
- 26. Liu B, Cong G, Zeng Y, Xu D, Chee Y M. Influence Spreading Path and Its Application to the Time Constrained Social Influence Maximization Problem and Beyond. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8), pp. 1904–17.
- Han B, Li J, Srinivasan A. Your Friends Have More Friends Than You Do: Identifying Influential Mobile Users Through Random-Walk Sampling. IEEE/ACM Transactions on Networking, 2013, 22(5), pp. 1389–400.
- 28. Wang G, Jiang W, Wu J, Xiong Z. Fine-grained Feature-based Social Influence Evaluation in Online Social Networks. IEEE Transactions on parallel and distributed systems, 2013, 25(9), pp. 2286–96.
- 29. Lee J R, Chung C W. A Query Approach for Influence Maximization on Specific Users in Social Networks. IEEE Transactions on knowledge and data engineering, 2013, 27(2), pp. 340–53.
- Tang J, Sun J, Wang C, Yang Z. Social influence analysis in large–scale networks, Proceeding of the ACM international Conference on Knowledge discovery, KDD'09, 2009, pp. 807–16.