



Comparative Study on Various Methods to find Content Power User

Shashank Khare

*M.Tech. Research Scholar
Shri Ram Group of Institutions, Jabalpur
Jabalpur, (M.P.) [INDIA]
Email: ishashank29@gmail.com*

Sapna Choudhary

*Assistant Professor, Head of the Department
Department of Computer Science & Engineering
Shri Ram Group of Institutions, Jabalpur
Jabalpur, (M.P.) [INDIA]
Email: choudharysapnajain@gmail.com*

Abstract—In today's world weblogs are exclusive social and web portent affecting perceptions and behaviors of their users and helping them to maintain/create social presence. It is remarkable to analyze the growth and evolution of weblogs where either we consider the marketing strategy or new services being offered, also from a scientific viewpoint, since their evolution may share similarities with tangible social network. Several techniques for analyzing social networks have been developed, to gauge quantitative properties (e.g., defining metrics and measures of fundamental characteristics of the networks) or qualitative aspects (e.g., studying the attachment model for the network evolution and the link prediction problem). However, online social network analysis poses novel challenges both to computer and Social scientists. Influential user is a unique web and social phenomenon affecting tastes and behaviors of their users and helping them to maintain/create friendships. There can be many different ways to calculate the power of user. In this section, we are describing novel method to calculate the influence power of users in online social network. In our Research we focusing on user activity based influence power. The influence power of a document is greater when activities like Read, Comment, Trackback and Scrap are more of other users on the document and with more activities of other users on the documents that are reproduced (i.e. Trackback or scrap) from the original document and also the degree of

friends of friend. Therefore, to the computational point of view we are going to calculate document power as well as user's power in existing network.

Keywords:— Blog Network, Blogs, Content Power User (CPU), Power User, Document Content Power.

1. INTRODUCTION

During the last years, we are experiencing an unprecedented growth of the use of Internet for both scientific and commercial purposes; especially in the commerce sector. More importantly, the Internet has changed the way people communicate. As in previously information was shared on face-to-face communication or paper, fax. The digitalized new world makes it easier to communicate and float information. Initially Internet brought us e-mail, but as the Internet is maturing the growth of the Blog services day by day. Blog services connects millions of people whom build relationships online, and is growing amazingly. These relationships are represented using various kinds of activities which indicates the user behavior that can be stored in database files specially created for this purpose.

There is a need, therefore, once the data is there to find ways of extracting information out of them; a way to dig into the large files for patterns of users' behavior must be found. In this paper we review the previous methods to

identify power users in a social network, the methods to measure influential power of a user. In particular, the problem of identifying power users in a social network has been studied for a long time in the field of viral marketing. Its primary goal is to determine a small group of customers that can produce the maximum marketing effect.

2. LITERATURE REVIEW

In this section we will review previous methods for the identification of existing Influence in a network.

2.1 Topology Based Method:

Most early studies focused on the topological structure of a social network and measured the *centrality*[1] of the user. Structural location of the node is advantageous to find the relative significance in the graph. There are various types of centrality measures of a node used to find the importance in the social network structure.

Degree Centrality: Degree Centrality measures the number of neighbor relationships a user has. That is, it determines those with many relationships as power user.

Closeness Centrality: Closeness Centrality measures the power of a user by adding up the shortest distance to everyone in the network. It determines as power users those with a low sum of shortest distance.

Between Centrality: Between Centrality measures the frequency that a user is on the shortest between any pair of user. Based on Between Centrality, power users are determined as those who appear most often on the shortest path between any pair of two users.

2.2 Link Topological Ranking methods:

PageRank: PageRank determines the authority score of a page based on the authority scores of those pages pointing to the page. A page is ranked high if it is pointed to by a number of pages of high authority score.

HITS: HITS calculates the hub score and the authority score for a web page, both are which are used to determine the rank of the page. A page has a high authority if it is pointed to by a number of pages of high hub score; A page has high hub score when it points to a number of pages of high authority score. As a result, pages of high authority score get high ranked.

2.3 Technique based on the Diffusion Model:

Linear Threshold Model: designates a threshold value to each user and a weight to a relationship between users. When a user's accumulated value of the weighted influence received from surrounding user is bigger than the threshold value, that user is regarded as having been influenced. In determining a user's power, the number of users who have been directly or indirectly influenced by her is measured.

Independent Cascade Model: designates a probability value to a relationship between users. Whether influence occurs between them is determined according to the probability.

2.4 Method based on Community based Mining:

Kempel[6] et al. state that optimization problem of influence maximization is NP-Hard. Accordingly to solve this type of problem, Greedy algorithms with provable approximation may give better outcome. But Greedy methods are expensive in computation, so as a result it is not practicable to social network. Yu Wang et al. proposed novel method called "Community based Greedy algorithm for mining top-K influential nodes" which divides the network into a number of communities, and then selects individual community to find top-K influential nodes. The community structure is a main property of social network features: Individuals within a community have frequent contact; in contrast, individuals across communities has much less contact with each other and thus is less likely to influence each

other. This property suggests that it might be a good approximation to identify influential nodes within communities instead of the whole network. This work gives several directions to expand research of location based social network to find influential over time.

Table 1: Key Features of Various Influential Identification Techniques

Name of the Technique	Year launched	Key Features
Centrality Measures[1][2]	1966	Identify the relative importance of Nodes in Network.
HITS Algorithms	1998	Identify Authority Pages and Hub Pages in Network.
PageRank Algorithms[3]	1998	Maintain a single metric for information of all Web Pages
Linear Threshold Model[4]	1978	Focuses on Threshold behavior of nodes
Independent Cascade Model [5]		Focuses on Individual's Interaction in Network
Community Modelling[6]	2010	Efficient over Greedy method and orthogonal to existing algorithms of Influential detection

3. PROBLEM DEFINITION

In this section we will discuss some of the drawbacks of previous work. First is topological based method which focuses on centrality measures i.e degree centrality, closeness centrality and between centrality. The topological based method is unable to capture the dynamic nature of the Blog network. Since Closeness centrality and between centrality could not handle a huge network where the number of nodes reached ten of millions of users. so the topology based methods mentioned in the above section, however, are not appropriate for identifying

power users in a blog network. Due to the computation complexity, closeness centrality and between centrality are applicable only to small size network. A blog network in general is too big to compute closeness and between centralities. Although computationally tractable, degree centrality does not correctly identify Content Power Users in a blog network. In a blog network a user having many relationships with others are not necessarily the one with high influence. Sometimes, users with a small number of relationships may exhibit great influence over other users, thereby inducing more activities in a blog network.

Most recent studies have tried to qualify the influential power of users directly or indirectly on other users by using the strengths of relationship among users. Although not originally proposed for selecting power users, linear threshold model and independent cascade model be used in identifying power users in a blog network. They both measures a users power by the number of users who were directly or indirectly influenced by the user. But unlike linear threshold model, the independent cascade model assumes a user is independently influenced by her surrounding users. Therefore, the independent cascade model is more appropriate to model influential power in a blog network. There are two weaknesses in applying the independent cascade model, however. First, accurate assigning of probability between every pair of users is a prerequisite in producing correct power users with this model. It is extremely difficult, however, top compute the assimilation probability accurately in real application. Second, the power users defined in the independent cascade model do not capture the importance of the quality of contents of Content Power Users.

Although originally proposed for the use in information retrieval, both PageRank and HITS two link based ranking algorithms that rate web pages can also be adapted in selecting power users. As an indicator of an

individual web page's value PageRank determines the authority score of a page based on the authority score of those pages pointing on the page. A page is ranked high if it is pointed to by a number of pages of high authority score. HITS calculates the hub score and the authority score for a web page, both of which are used to determine the rank of the page. Both PageRank and HITS calculate their rank based on reproductive action based method also known as T&S method used the total number of actions performed on the document in one's blog. But these actions are trackback and scrap only. So this method ignore actions like read and comment.

Now there comes diffusion methods to calculate power user in a blog network. It includes linear threshold model and linear cascade model. They focuses on threshold behavior of nodes and individual interaction in network. But it assigns a probability by considering read, write, reproduce action in a blogosphere and influence occurs between them id determined according to this probability. There are two weaknesses in applying the independent cascade model, however. First, accurate assigning of probability between every pair of users is a prerequisite in producing correct power users with this model. It is extremely difficult, however to compute the assimilation probability accurately in real applications. Second, the power users defined in the independent cascade model do not capture the importance of the quality of contents of the content power user.

4. CONCLUSION

This paper presents different approaches used for the identification of influential users in a blog network. Some of the approaches are hybrid form of structural properties and behavior of the user in a blog network. Structural properties mainly include topological features along with links between users inside social boundaries. Behavioral aspects, fully dependent on user's activities inside network, published contents and actions of users depicting the behavior of

individual in online social network. The future direction to this work is to development of approaches by considering temporal aspects. Another major area for extending works by basis of temporal and location based methodologies for identification of influencers in online society, additionally not only calculates influence power but to type of influence (negative or positive)

REFERENCES:

- [1] G. Sabidussi: The centrality index of a graph. *Psychometrika* 31 (1966) 581–603.
- [2] L.C. Freemann: A set of measures of centrality based on betweenness. *Sociometry* 40 (1977) 35–41
- [3] S. Brin, L. Page: The anatomy of a large scale hyper textual web search engine. WWW Conference, Australia, 1998
- [4] M. Granovetter: Threshold models of collective behaviour. *American Journal of Sociology*, vol. 83, no. 6, pp. 1420-1443, November 1978
- [5] 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*, pp. 211-223, August 2001.
- [6] Yu Wang, Gao Cong, Guojie Song, Kunqing Xie: Community-based Greedy Algorithm for Mining Top-K Influential Nodes in Mobile Social Networks. *KDD'10*, July 25–28, 2010, Washington.
- [7] Brown, Jo, Amanda J. Broderick, and Nick Lee (2007), "Word of Mouth Communication within Online Communities: Conceptualizing the Online Social Network," *Journal of Interactive Marketing*, 21 (July), 2-20