An Improved Approach to Find Content Power User in Blog

Abhay Kumar Mishra[1], Dr. Tasneem Bano Rehman[2]
[2]*Associate Professor*
[1,2]*Department of Computer Science, SRIST Jabalpur, M.P. , India*

Abstract — In a blog network, there are special users who induce other users to actively utilize blogs. Identifying such influential users is important when establishing business policy and business models for the blog network. This paper defines the users whose contents exhibit significant influence over other users as content power users (CPUs) and proposes a method of identifying them. We analyze the performance of the proposed method by applying it to an actual blog network and comparing its results with those of preexisting methods for determining power users. The experimental results demonstrate that the definition of CPUs is adequate to address the dynamic nature of the blogosphere and the main concerns of the blog industry. We also discuss the business models based on CPUs that could be used to stimulate user activities in a blog network.

Keywords: Blog network, blogs, business models, content power users (CPUs), power users.

I. INTRODUCTION

Due to modern-era individualism and recent developments in Internet technology, self-expression and networking are rapidly mobilizing content development from offline to online. As a result, online social networks have emerged where individuals write documents, exchange information, and form relationships online. The blogosphere is a primary example of such online social networks. A blog is a user generated website where individuals publish their own *documents*. The service provided to users that facilitates the production and maintenance of a blog is referred to as *blog service*. Using blog services, users may establish relationships with other users, from which social networks are created. This paper defines a *blog network* as the social network established through blogs and their relationships. Examples include facebook.com, myspace.com, linkedin.com, blogger.com, cyworld.com, and blog.naver.com.

As the number of blog users increases, companies have become interested in providing products and services that utilize blogs. To ensure the success of a business geared toward blog networks, encouraging users within a blog network to actively utilize blog services is an essential prerequisite [13]. Within a blog network, there are special users who contribute inducing other users to actively utilize blog services. If such influential users can be identified, it is possible to establish diverse business policies centered on these users that will increase the usage of blog services more effectively.

In early studies, power users in a social network are determined based on the topology [3] and characteristics of a social network [2], [11]. In a blog network, however, topology does not seem to reflect the actual influence relationship between users, and thus topology-based selection may not correctly identify power users. There have been excellent research results on identifying special users with influential power greater than others in social networks [7], Typical examples are the independent cascade model [8], the linear threshold model [9], the general model that combines the independent cascade model and the linear threshold model together [12], mining of network values of customers [7], and information diffusion in blogosphere [10]. They employed different definitions of power users and proposed methods for determining their own power users. As explained in Section II, however, they have limitations on identifying *content power users* (CPUs) since they do not properly consider various types of activities performed by users in a blogosphere.

The proposed method of determining CPUs is based on the *influence relationship* between users. First, we measure the influence of each document owned by individual users. Since the degree of influence of a document tends to increase with the increase in the time of exposure, we adjust the measured value of influence of each document based on the time of exposure. Then, by adding up the values of influence of all the documents owned by an individual user, we determine the degree of influence of the user.

II. RELATED WORK

In this section, we review the previous methods to identify power users in a social network, the methods to measure

influential power of a user, and two link-based ranking algorithms. We also discuss their applicability to identifying CPUs in a blog network.

Previous studies on social networks have proposed various methods to identify power users in a social network. 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 [5], [6].

Most early studies focused on the topological structure of a social network, and measured the centrality of a user [3]. Depending on how to measure centrality, they are classified into *degree centrality*, *closeness centrality*, and *betweenness centrality*. *Degree centrality* measures the number of neighbor relationships a user has. That is, it determines those with many relationships as power users. *Closeness centrality* measures the power of a user by adding up the shortest distance to everyone in a network. It determines as power users those with a low sum of shortest distances. *Betweenness centrality* measures the frequency that a user is on the shortest path between any pair of users. Based on betweenness centrality, power users are determined as those who appear most often on the shortest path between any pair of two users.

The topology-based methods mentioned above, however, are not appropriate for identifying power users in blog networks. Due to computation complexity, closeness centrality and betweenness centrality are applicable only to a small-size network. A blog network in general is too big to compute closeness and betweenness centralities. Although computationally tractable, degree centrality does not correctly identify CPUs in a blog network. In a blog network, the users having many relationships with others are not necessarily the ones with high influence. Sometimes, users with a small number of relationships

may exhibit great influence over other users, thereby inducing more activities in the blogosphere. We justify our claim in Section IV, by comparing our proposed method with the degree-centrality based method

III. PROPOSED METHODOLOGY

In this section, we propose a method to determine content power users or CPUs. First, we build a blog network composed of bloggers and their actions. We discuss two ways of constructing a blog network by capturing different influence relationships: one based on (more static) bookmarks and the other based on (more dynamic) user activities.

A. Definitions

Table I summarizes the terminologies and symbols used in this paper. *Ui* represents user *i*. *Di* represents the set of documents owned by *Ui*, and *Di,j* represents document *j* of user *i*. *Document Content Power (DCP)* is defined as the content power of a document, and *DCP*(*Di,j*) represents the *DCP* of document *Di,j* . Similarly, *user content power* (*UCP*) is defined as the content power of a user, and the *UCP* (*Ui*) represents *UCP* of user *i*.

TABLE I : SUMMARY OF TERMINOLOGIES

- U_i: User i
- $D_{i,j}$: Document j of user i
- $D_i = \{D_{i,1}, D_{i,2}, ...\}$: Set of documents owned by U_i
- $DCP(D_{i,j})$: Document content power of $D_{i,j}$
- $UCP(U_i)$: User Content Power of U_i
- $AT = \{A_1, A_2, ...\}$: Types of user actions
- A_k: Action of type k
- W_{A_k}: Weight of A_k

B. Document Content Power

The *DCP* is defined as a degree of influence of a document on other users in the blogosphere. When a user performs an action to a particular document (such as writing a comment), it indicates the document has some influence on the user. This observation leads to a new method of quantifying the *DCP*. The main idea is that the *DCP* can be computed by adding up the weighted frequencies of other users' activities induced by that particular document

C. User Content Power

UCP is defined as a degree of influence of a user on other users in the blogosphere. The *UCP* is computed by adding up the *DCP*s of all the documents in the blog of a particular user. The *DCP* of a document tends to increase as it is exposed to an audience longer, which may result in an older document having a higher *DCP* than a newer document.

We resolve this potential problem by taking into account the exposure duration of a document (i.e., the exposure duration since the creation of a document). While assuming the entire analysis period as 1, we compute the relative exposure duration of a particular document. By multiplying

the inverse of this relative exposure duration to the *DCP*, we correct the distortion of *DCP* values, if any.

IV. EXPERIMENTS AND ANALYSIS

We compare the proposed method and existing methods of determining power users through experiments on a real-world blog network. Section IV-A explains experimental setup, and Section IV-B analyzes experimental results.

We performed five sets of experiments. The first set of experiments (experiment 1) compared the similarity of power users identified by the CP method with different weight combinations. The results were used to select the most "representative" parameter setting to be used in subsequent experiments. The second set of experiments (experiment 2) compared the similarity between the power users selected by different methods. The power users identified by the DEG method were quite different from those by the proposed method, and the DEG method did not seem to select the users who had big influence on diffusing and propagating contents. The third set of experiments (experiment 3) compared the survival rates of the power users selected by different methods. The power users selected by our activity-based methods were frequently changing, and the proposed method seemed to capture the dynamic nature of a blog network well. Although experiments 2 and 3 indicate that CPUs are different from other power users,

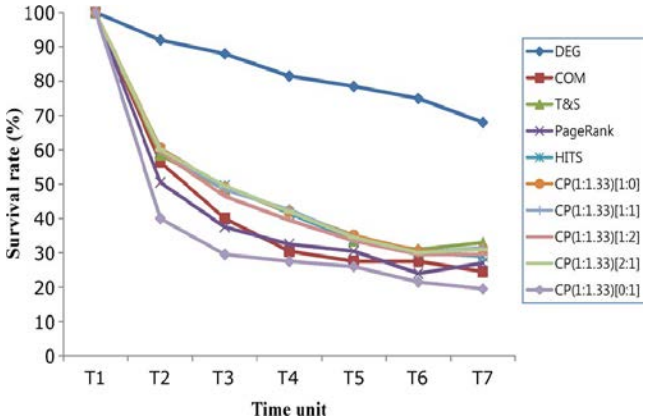

Fig. 1. Survival rates of power users (while changing the*WD* : *WI* of CP).

They do not show which methods are better. Two sets of user studies were performed to demonstrate the superiority of the CP method. In experiment 4, human experts evaluated various methods based on the power users identified by them. In experiment 5, human experts ranked the power users. In both experiments, the results show that the CP method has selected the power users most appropriate for the blog network.

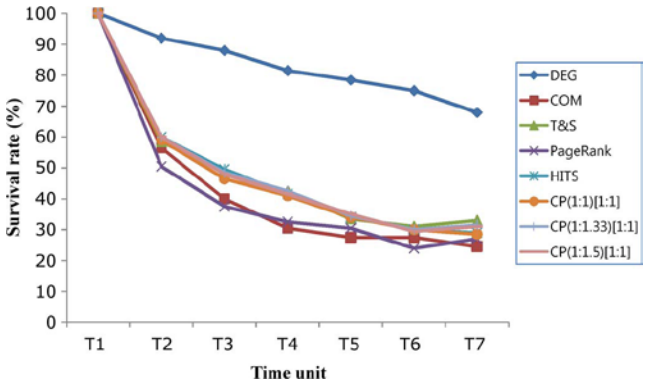

Fig. 2. Survival rates of power users (while changing the*WC* : *WT*&*S* of CP).

V. CONCLUSION

The contributions of this paper are at both construct and measure levels. Primary is the construct-level contribution of defining power users for a blogosphere, and the measure-level contribution of implementing the algorithm for identifying CPUs follows naturally.

The following are main contributions of this paper:

- Unlike the concept of power users derived from the structural topology and characteristics of social networks, the new concept of "content power user" was proposed, which reflected the influence of a user within a blog network more appropriately.
- We proposed a method of measuring the content power of a document and that of an individual user. In order to measure content power correctly, we proposed a normalization method based on the exposure time of a document.
- By soliciting domain experts for user study, we revealed that the proposed method performs best in finding those users who actually contribute to revitalizing the blog network.
- We proposed several business models that use the concept of CPUs to stimulate activities in a blog network.

In this paper, we proposed a new definition of power users. Note that this definition counts the activities of other users and does not count the number of users influenced by the

user. In some business applications, knowing the number of influenced users may be more valuable, which may require power users to be determined based on a user's ability to influence *the greatest number of other users*. As an extension of the research reported in this paper, we are in the process of developing a method to identify power users who persuade the largest number of users to take actions in a blogosphere.

REFERENCES

[1] Seung-Hwan Lim, Sang-Wook Kim, Sunju Park, and Joon Ho Lee, "Determining Content Power Users in a Blog Network: An Approach and its Applications," IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Volume 41, No.5 September - pp. 853-862, 2011.

[2] S. Bagchi, G. Biswas, and K. Kawamura, "Task planning under uncertainty using a spreading activation network," *IEEE Trans. Syst., Man,Cybern. A, Syst., Humans*, vol. 30, no. 6, pp. 639–650, Nov. 2000.

[3] Seung-Hwan Lim, "Determining Content Power Users in a Blog Network: An Approach and Its Applications," VOL. 41, NO. 5, pp. 855–862, SEP.2011.

[4] S. Brin and L. Page, "The anatomy of large-scale hypertextual web search engine," in *Proc. Int. Conf. WWW*, 1998, pp. 107–117.

[5] J. Brown and P. Reinegen, "Social ties and word-of-mouth referral behavior," *J. Consum. Res.*, vol. 14, no. 3, pp. 350–362, Dec. 1987.

[6] J. Coleman, H. Menzel, and E. Katz, *Medical Innovations: A Diffusion Study*. Indianapolis, IN: Bobbs-Merrill, 1966.

[7] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. ACM Int. Conf. Knowl. Discov. Data Mining, SIGKDD*,2001, pp. 57–66.

[8] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complexsystems look at the underlying process of word-of-mouth," *Market. Lett.*,vol. 12, no. 3, pp. 211–223, Aug. 2001.

[9] M. Granovetter, "Threshold models of collective behavior," *Amer. J.Sociol.*, vol. 83, no. 6, pp. 1420–1443, May 1978.

[10] D. Gruhl, R. Guha, D. Nowell, and A. Tomkins, "Information diffusionthrough blogspace," in *Proc. Int. Conf. WWW*, 2004, pp. 491–501.

[11] Karmeshu and D. Goswami, "Transient bimodality and catastrophicjumps in innovation diffusion," *IEEE Trans. Syst., Man, Cybern. A, Syst.,Humans*, vol. 38, no. 3, pp. 644–654, May 2008.

[12] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. ACM Int. Conf. Knowl. Discov.Data Mining, SIGKDD*, 2003, pp. 137–146.

[13] WegoNet, *Brand Strategy in Communities*. Seoul, Korea: E-Design, 2004.

[14] A. Bavelas, "Communication patterns in task-oriented groups," *J. Acoust. Soc. Amer.*, vol. 22, no. 6, pp. 725–730, Nov. 1950.