Determining Content Power Users in a Blog Network

Seung-Hwan Lim  
Dept. of Electronics and  
Computer Engineering  
Hanyang University  
Seoul 133-791, Korea  
shlim@zion.hanyang.ac.kr

Sang-Wook Kim  
Dept. of Electronics and  
Computer Engineering  
Hanyang University  
Seoul 133-791, Korea  
wook@hanyang.ac.kr

Sunju Park  
School of Business  
Yonsei University  
Seoul 120-749, Korea  
boxenju@yonsei.ac.kr

Joon Ho Lee  
NHN Corp.  
Seongnam 266-1, Korea  
joonho@nhn.com

ABSTRACT

In a blog network, there are special users who induce other users to actively utilize blog services. In this paper, these users whose contents exhibit large influence over other users are defined as Content Power Users (CPUs). Accurately determining who content power users are in a blog network is important in order to establish an effective business policy that stimulates usage of blog services. In this paper, we propose a new method of determining content power users. First, we measure the influence of each document owned by individual users. Since the influence of a document tends to be high with long exposure, we adjust the value of influence of a document based on exposure duration. Then, by adding up the values of influence of all the documents owned by each user, we calculate the influence of the corresponding user. We analyze the performance of our method, by applying the proposed method to an actual blog network and comparing its results to those of preexisting methods for determining power users. The experimental results demonstrate that our method is well suited for the dynamic nature of the blog network.

Categories and Subject Descriptors

J.4 [Computer Applications]: SOCIAL AND BEHAVIORAL SCIENCES—Sociology; H.3.3 [Information Systems]: INFORMATION STORAGE AND RETRIEVAL—Retrieval models

General Terms

Human Factors, Algorithm, Experimentation, Economic Activities

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The 3rd SNA-KDD Workshop '09 (SNA-KDD'09), June 28, 2009, Paris, France.

Copyright 2009 ACM 978-1-59593-848-0 ...$5.00.

1. INTRODUCTION

Due to modern-era individualism and recent developments in Internet technology, self-expression and networking together 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. Blogosphere is a primary example of such online social networks. A blog is a user-generated website where an individual publishes her own documents. The service provided to users that facilitate 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.

As the number of blog users increases, companies have become interested in providing products and services that utilize blogs [4][7][8][17][18]. To ensure the success of a business geared towards blog networks, encouraging users within a blog network to actively utilize blog services is an essential prerequisite [26]. Within a blog network, there are special users who contribute inducing other users to actively utilize blog services. If users such as these can be identified, it is possible to establish diverse business policies centered on these users that will increase usage of blog services more effectively. In this paper, we define Content Power Users (CPUs) as those users whose contents exhibit large influential power and thus induce a large amount of activities of other blog users within a blog network. CPUs are more precisely defined in Section 3.

In early studies, power users were determined based on the topology [3][20][22] and characteristics of a social network [2][14][19][25]. 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 [9][11][12][15][21][23]. Typical examples are the independent cascade model [11],
the linear threshold model [12], the general model that combines the independent cascade model and the linear threshold model together [15], mining of network values of customers [9][21], and information diffusion in blogspace [15]. They employed different definitions of power users, and proposed methods for determining their own power users. As explained in Section 2, however, they have limitations on identifying the content power users since they do not properly consider whole activities of various types performed by users in a blog world.

If we are able to utilize the degree and frequencies of various activities between users in addition to topology when analyzing a blog network, properly determining power users may become possible. In this paper, we propose a new approach for determining power users using user activities and usage patterns of services in a blog network, which are readily available in the database of the blog service provider.

Our method of determining content power users 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.

By applying the proposed method to actual blog networks and comparing its results to those of preexisting methods, we analyze different methods of determining power users. The experimental results reveal that the power users selected by a topology-based method and those by our method are very different, and that the content power users selected by our method reflect well the dynamic nature of a blog network.

2. RELATED WORK

The problem of identifying power users in social networks has been discussed 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 researches focused on measuring centrality of a user in a social network based on its topological structure [3][20][22]. Depending on how to measure centrality, they are classified into degree centrality, closeness centrality, and betweenness centrality. Degree centrality measures the number of relationships of a user. That is, it determines users with many relationships as power users. Closeness centrality measures the power of a user by adding up the shortest distances to other users in a network and taking the inverse of it. It determines as power users the users who have a low value of closeness centrality. Betweenness centrality measures the frequency that a user is on the shortest paths between other users. Using betweenness centrality, power users are determined as the ones who are on the shortest paths between other users more frequently.

The topology-based methods mentioned above, however, are not appropriate for identifying power users in blog networks. In a blog network, users having many relationships with others are not necessarily the ones with high influence. Sometimes, users with a small number of relationships can exhibit great influence over other users, thereby inducing more activities in the blog world. The methods proposed in recent researches tried to quantify the influential power of users directly or indirectly on the other users by using the strengths of relationships among users other than the topological information [9][11][12][15][21][23]. The users can mutually give and receive influence, and due to this influence, the tendency of a specific user can become similar to that of other users that influence him. In this case, the user is said to be assimilated by those who influenced him. Information diffusion is a phenomenon that a document in a blog is disseminated over a blog world by those assimilations through blog services such as a trackback.

In a linear threshold model [12], when a user’s accumulated value of the weighted influence received from surrounding users is bigger than her threshold value, that user is regarded as having been assimilated by the users who influenced her. In an independent cascade model [11], when influence occurs between users, it decides whether or not assimilation occurred according to a probability. In [15], a general framework was proposed to explain both features of the independent cascade and linear threshold models. In [9][21], the influential power of a customer is modeled as a network value, which is the expected profit from sales to other customers she may influence to buy. In [23], a new concept of a diffusion rate was introduced to describe how rapidly the information is diffused in the network.

Now, let us consider the applicability of the previous models and approaches to identifying CPUs in a blog network. In a blog world, a user is assimilated with some document through her independent relationships with other users. Thus, the models and approaches proposed in [12][15][9][21], which consider the sum of influences received from multiple surrounding users, are inappropriate in blog networks. The independent cascade model in [11] is inappropriate to model information diffusion in blog networks. For determining power users, it requires the assimilation probability between every pair of users. It is extremely difficult, however, to compute the probability accurately in real applications. In [13], a method was proposed to determine diffusion power users in a blog world, based on the independent cascade model, and assigns assimilation probability by considering users’ actions. This model (as that proposed in [11]) determines a power user by the number of other users assimilated from her. The CPUs we are seeking in this research, however, are those users that, within their blogs, contain quality contents of large influential power and thus induce a large amount of activities of other users in the blog world. Thus, in order to indentify such CPUs properly, we should consider all the activities of various types performed by users in a blog network.

3. PROPOSED METHOD

In this section, we propose a method to determine CPUs. First, we discuss two ways of constructing a blog network by capturing different influence relationships: one based on (more static) bookmark and the other based on (more dynamic) user activities. We argue that a user’s action on another user’s document may be viewed as the former being influenced by the latter and that user actions, in particular reproductive actions like trackbacking, are better suited for capturing the propagation of influence in a blog network. Second, we propose a method to compute the content power of a document. The influence power of a document is greater with more activities of other users on the document and with more activities of other users on the documents that are re-produced from the original document. Therefore, the
document content power is computed from both direct and indirect user activities. Third, the content power of a user is computed by adding the content power of all the documents owned by the user. Since a document with longer exposure tends to receive more user activities, the raw value of the document content power is inversely adjusted with exposure duration. Finally, when determining a content power, users in a blog network, we select top n users from the highest user content power.

3.1 Construction of a Blog Network

A blog user is often provided with a functionality that enables her to keep track of the blogs of her interest, which makes it easy and convenient for her to visit those blogs [4][7][8][17][18]. Such functionality is called bookmark or neighbor. A blog user performs actions on a document in someone else’s blog, such as read, comment, trackback, scrap, etc. Read and comment are reading and putting comments on someone else’s document, respectively. Trackback is writing a new document related to someone else’s document putting a link to the original document in one’s own blog. Unique to blog sites in Korea, scrap is an action of copying someone else’s document to one’s own blog. Trackback and scrap actions can be viewed as a way of reproducing the original document. The reproduced document may in turn trigger someone else to perform actions of read, comment, trackback, and scrap on it.

Based on either bookmark or user actions, two different influence relationships may be defined. First, an influence relationship can be defined between a user and the blogs in her bookmark. This is based on the assumption that a user keeps a blog in her bookmark only when she has been influenced by it. Bookmark, however, may not reflect the actual influence relationship between users since it becomes outdated quite easily. Even when she does not visit a blog in her bookmark anymore, for example, the user often keeps it in her bookmark for a long time. Using bookmark to construct a blog network, therefore, may fail to capture the dynamic and fast-changing influence relationships in a blog network.

Second, an influence relationship can be defined between a user who has performed an action over a document and the owner of the original document. This is based on the assumption that an action indicates that the user has been influenced by the original document, and scrap actions reflect the propagation of influence, and they seem appropriate for capturing indirect influence relationship in a blog network.

Figure 1 depicts an example of a blog network. Figure 1(a) shows the actions of users in a blog world. Figure 1(b) shows the blog network representation based on bookmark. Figure 1(c) shows the blog network based on reproductive actions. In Figure 1(a), each rectangle, labeled A through E, represents a blog, whereas each small rounded rectangle represents a document in a blog. The arrow represents an action either from a document to another (Trackback/Scrap), from a blog to a document (Comment), or from a blog to a blog (Bookmark). The documents in the same shade represent they are related through reproductive actions. For example, user A put comments on document 1 of blog B and document 2 of blog D, respectively, while user C put trackback to document 3 of blog A in document 1 within her blog. In Figure 1(b) and Figure 1(c), each circle represents a blog, and the arrow represents the influence relationship between

Figure 1: An example of a blog network.

3.2 Definitions

We define terminologies and symbols used in this paper. $U_i$ represents user $i$. $D_i$ represents the set of documents owned by $U_i$, and $D_{ij}$ represents document $j$ of user $i$. Document Content Power (DCP) is defined as the content power of a document, and $DCP(D_{ij})$ represents DCP of document $D_{ij}$. Similarly, User Content Power (UCP) is defined as the content power of a user, and $UCP(U_i)$ represents UCP of user $i$. Action Type (AT) represents the types of actions (i.e., comment, trackback, scrap, etc.) a user can perform in a blog network. An action of type $k$ is denoted as $A_k$. When computing the content power of a document, different weights may be assigned depending on the types of actions. The weight for $A_k$ is denoted as $W_{A_k}$.

3.3 Document Content Power

The document content power (DCP) is defined as a degree of influence of a document on other users in the blog world. When a user performs an action to a particular document, 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.

A document can have either direct influence on other users who access the document directly, or indirect influence on others who access the document in someone else’s blog that is reproduced through trackback or scrap of the original document. In this paper, we call the former as DirectDocument-ContentPower (D_DCP) and the latter IndirectDocument-ContentPower (I_DCP). DCP is the sum of the weighted values of D_DCP and I_DCP. Depending on applications, one may adjust the weights, \( w_D \) and \( w_I \), which will change the degree of relative importance of D_DCP and I_DCP.

\[
\begin{align*}
& DCP(D_{i,j}) = w_D \times D\_DCP(D_{i,j}) + w_I \times I\_DCP(D_{i,j}) \\
& D\_DCP(D_{i,j}) = \sum_{A_k \in \{D, T, S\}} W_{A_k} \times Count(D_{i,j}, A_k) \\
& Count(D_{i,j}, A_k) = \text{Frequencies of } A_k \text{ on } D_{i,j} \\
& I\_DCP(D_{i,j}) = \sum_j IED(D_{i,j}) \times DCP(D_{i,j}) \\
& \text{where } D_{i,j} \text{ represents documents reproduced by other blog users from } D_{i,j}
\end{align*}
\]

\[D_{i,j} = (4, 1, 2, 3)\]

Table 1: Computation of DCP of a document.

Figure 2 shows an example of the computation of content power of a document. In this example, the weights for both D_DCP and I_DCP are assumed to be 1, and the values of D_DCPs are arbitrarily given. The three-value tuple for each document represents \((D\_DCP, I\_DCP)\). In Figure 3, document 2 of user D \((D_{2,2})\) does not have any diffusion history, so I_DCP is 0 and DCP is the same as D_DCP of 5. Therefore, I_DCP of \(D_{C,3}\) is 5, and adding the values of I_DCP and D_DCP results in DCP of 8. I_DCP of document \(D_{A,1}\) is 12, by adding DCP of \(D_{C,3}\) (which is 8) and DCP of \(D_{B,2}\) (which is 4). DCP of document \(D_{A,1}\) is computed to 16 by adding D_DCP of 4 and I_DCP of 12.

### 3.4 User Content Power

The user content power (UCP) is defined as a degree of influence of a user on other users in the blog world. UCP is computed by adding up DCPS of all the documents in the blog of a particular user. DCP of a document tends to increase as it is exposed to 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 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 exposed duration to DCP, we correct the distortion of DCP values, if any. Equation (1) computes UCP of a user by adding up the exposure-duration-adjusted DCPS of her documents. Here, \(IED(D_{i,j})\) is the inverse of the relative exposure duration of \(D_{i,j}\).

\[
UCP(U_i) = \sum_j IED(D_{i,j}) \times DCP(D_{i,j})
\]

Figure 3 shows computing of UCP\((U_A)\). The arrow indicates the creation time of document \(D_{i,j}\). In this example, UCP\((U_A)\) = 50*0.25 + 35*0.4 + 35*0.8 + 10*0.95 = 64.

4. EXPERIMENTS AND ANALYSIS

4.1 Experimental Setup

For experimental analysis, this paper used anonymized data which was collected from blog.naver.com, one of the largest blogospheres in Korea, for several months starting from July 2006. The number of documents created during the analysis period was about 100 million. When constructing the model of the blog network, nodes represented blog users, and links represented the fact the relationships between two users were above a predefined level. The analysis were performed based on two types of influence relationships: influence based on actions (trackbacks and scraps) and influence based on bookmarks. Static bookmarks were used to represent the influence relationship only when employing the previous topology-based method; Actions were used to represent the influence relationship when employing our proposed method.

To use our method of determining CPUs, we needed to determine the weights of user actions and those of direct and indirect content powers. Different combinations of weights could be used based on the properties of power users that need to be selected in a target application. If one wanted to find power users who made others to reproduce, for example, one might assign the weight of reproductive actions (Trackback & Scrap, T&S in short from now on), \(W_{T&S}\), to be higher than that of Comment, \(W_C\). Similarly, to find the power users with high direct influence, one might assign the weight of direct content power, \(w_D\), to be higher than that of indirect power, \(w_I\).

The different combinations of weights used in our experiments are as follows.

![Figure 3: An example of computation of UCP.](image-url)
• Ratio of weights of Comments and T&S : \((W_C:W_{T&S})\) = (1:1), (1:1.33), (1:1.5)

• Ratio of weights of Direct and Indirect content powers : \(\{w_D:w_I\} = \{1:0\}, \{1:1\}, \{1:2\}, \{2:1\}, \{0:1\}\)

Note that we use parentheses to represent the ratio of weights of user actions and curly brackets to represent the ratio of weights of direct and indirect content powers.

Three different weight ratios were used for user actions. Note that the weight ratio of (1:1.33) reflects the frequency ratio between Comments and T&S exhibited in the actual blog network. Five different weight ratios were used for direct and indirect content powers. Note that \{1,0\} reflects the direct content power only, while \{0,1\} reflects the indirect content power only. We denote the method for determining content power users with different weight combinations as \(CP(W_C:W_{T&S}) \{w_D:w_I\}\).

When choosing a weight combination for the following sets of experiments, we tried to use a general method that did not exhibit any particular bias. Since \(CP(1:1.33)\{1:1\} \) included the highest number of power users in common with the other methods, (1:1.33)\{1:1\} seemed to be a good candidate for weight combination in generic experiments. Therefore, we used \(1:1.33\{1:1\}\) in the following, unless noted otherwise. However, we note that, in real situations, the weight combination is normally set by the domain expert so that it can meet the goals of a target application.

When analyzing the performance, we compared four different methods of selecting power users: topology-based, comment-based, reproductive-action-based, and content-power-based. As a method that used network topology, a degree centrality based method (DEG) was used. Among the preexisting methodologies, closeness centrality and betweenness centrality were excluded in our experiments since they could not handle a huge network where the number of nodes reached tens of millions of users as in our case.

In determining power users, the content-power-based method (CP) used the user content power of one’s blog as computed in Section 3. The comment-based method (COM) used the total number of comments made by others on the documents in one’s blog. The reproductive-action-based method (T&S) used the total number of reproductive actions performed on the documents in one’s blog. The COM and T&S methods were selected for comparison since the CP method also used comments and reproductive actions. The COM and T&S methods were actually the same as \(CP^*(1:0)\{1:1\}\), \(CP^*(0:1)\{1:1\}\), where \(CP^*\) denotes the CP method not adjusted for exposure duration.

### 4.2 Analysis

We performed three sets of experiments. The first set of experiments computed power of users based on different methods and analyzed their distributions. The second set of experiments compared the similarity between the power users selected by different methods. The third set of experiments compared the survival rates of the power users selected by different methods.

Some readers may wonder about the accuracy of our approach and thus may expect the experimental results on the accuracy. In this paper, however, we are not saying that our approach is more accurate than others in identifying power users in the blog world. Rather, we are saying the followings: (1) In some blog-related businesses, it is fairly important to find users whose documents make other users perform a lot of activities in the blog world; (2) The definition of the content power user is adequate to these situations; (3) Our approach utilizing diffusion history structures is suitable for identifying these content power users effectively. In particular, we note that, by its definition, the user content power successfully quantifies the activities performed in the blog network. Thus, we do not deal with any accuracy-related results in this paper.

#### 4.2.1 Distribution of Power of Users

In the first set of experiments, we computed power of users using different methods and analyzed their distributions. Figure 4 shows the distributions of power of users. The x-axis represents users ranked by their power, and the y-axis represents the value of power. To make it easy to read, the graphs are presented in log-log scale. The distributions of user powers all resemble power functions, no matter what methods are used. The result is similar to that of the existing research [10][16]. This implies that, in a blog network, a small number of users with high power have a big impact while most users exercise little power. This result supports our claim that, to stimulate the activities in a blog network, it is better to set up a business policy that targets a small number of power users rather than the whole community in the network.

#### 4.2.2 Similarity of Power User Groups

In the following two experiments, we compared the CP method with pre-existing methods of selecting power users. The experiments were divided into two groups. In experiment 2, we compared the similarity of the power users selected by different methods. Figure 5 depicts the similarity of top 200 power users while fixing user-action weights to be (1:1.33) and changing the direct/indirect weight combination.

The similarity of power users ranges from 3.5% to 99%. The DEG method shows the similarity of 5.93 on average, which is significantly lower than other methods. This means that the power users identified by the DEG method, which is only based on the topology of social networks, are quite different from those who incur a large amount of activities
COM DEG T&S CP(1:1.33) Avg.
---
DEG 8 3.5 6 6 6 6 6 5.93
COM 8 44.5 60 59.5 60 60 40 47.43
T&S 3.5 44.5 83 82.5 79.5 82.5 49.5 60.71

Figure 5: Similarity of power users (while changing the $W_D:W_I$ of CP).

<table>
<thead>
<tr>
<th>DEG</th>
<th>COM</th>
<th>T&amp;S</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1:1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1:3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1:2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2:1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0:1)</td>
</tr>
</tbody>
</table>

Figure 6: Similarity of power users (while changing the $W_C:W_{T&S}$ of CP).

performed by other users in the blog world. In particular, the similarity between the DEG and T&S methods is mere 3.5, which means the number of power users in common between two methods are only 7 out of 200 selected. This implies that the power users selected by the DEG method have little impact on diffusing and propagating documents. Compared to the COM method, the power users selected by the T&S method are more in common with the CP method.

Note that the average similarity of CP(1:1.33){1:0} is quite low (43.93). This is based on the fact that the original documents created by power users receive many comments, but the documents reproduced from the original do not. That is, content power users primarily receive a high value of direct content power in general. CP(1:1.33){1:0}, however, completely ignores the direct content power, which results in significantly different from other methods.

Figure 6 depicts the similarity of the power users while fixing the direct/indirect weights of the CP method to be {1:1} and changing the user-action weight combination. Similar to Figure 5, the average similarity of the DEG method is the lowest, and the power users selected by the T&S method are in common with those of the CP method more than the COM method. The CP method shows a higher level of similarity with the T&S method and a lower level of similarity with the DEG and COM methods, respectively as the weight ratio for indirect actions increases. As explained in Figure 5, this is because the power users selected by the DEG method will not be able to diffuse and propagate documents well.

4.2.3 Survival rates of power user groups

In the third set of experiments, we compared the survival rate of a group of power users selected by each method. We divided the whole analysis period into 7 units, and selected a group of top 200 power users in each time unit. The survival rate of a power user group is defined as the ratio that the power users selected in time unit 1 are included in the power users in subsequent units. Figure 7 depicts the experimental results. The x-axis represents the time unit, and the y-axis represents the survival rate. Regardless of the methods for determining power users, the survival rate decreases. The DEG method shows the survival rate significantly higher than those from the activity-based COM, T&S, and CP methods. This is because the addition and deletion of blogs in a bookmark does not happen as frequently as user actions in a bookmark does not happen as frequently as user actions of Comments and T&S. The T&S method, CP(1:1.33) shows the lowest rate of survival. The low survival rate of CP(1:1.33) can be explained similar to the explanation given in experiment 2. The reproduced documents do not receive as many comments as CP(1:1.33) does.

Figure 8 depicts the survival rate of the power users while fixing the direct/indirect weights of the CP method to be {1:1} and changing the user-action weight combination. The results show trends similar to those in Figure 7. The survival rate of the DEG method is the highest, the T&S and all CP methods are similar, and the COM method shows a lower survival rate. The results of experiment 3 can be used to predict the probability that the power users selected on a particular time would still be power users on a later time. Furthermore, they can be used to determine how long the selected power users are admitted as such. The experimental results indicate that the power users are selected by the activity-based methods frequently changing in a blog network. We observe that more than 40% of the power users selected in the time unit 1 have been changed to others in the time unit 2 with the activity-based methods. It also suggests that one may need a business policy that sets up
an expiration date of power users. For example, 35 percent of the power users selected by CP{1:1}(1:1.5) will still be the power users in 120 days later. If one wants to establish a business model that requires the survival rate of power users to be 40 percent or above, on the other hand, the power users selected by CP{1:1}(1:1.5) should be admitted as power users only for 90 days.

5. CONCLUSIONS

In this paper, we discussed and proposed a new method of determining CPUs within a blog network. Due to the lack of information regarding the depth of relationships within a social network, most prior studies identified power users based on the structural topology and characteristics of social networks. By utilizing the information on user activities and usage patterns, we analyzed the influence relationships between users and proposed a new approach in determining CPUs who contribute to the increase in blog service usage.

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. Comparing the proposed method and preexisting methods, we conducted the performance evaluation. The experimental results demonstrate that our method is well suited for the dynamic nature of the blog network.

6. ACKNOWLEDGMENT

This work was supported by NHN Corp. Any opinions, findings, and conclusions or recommendations expressed in this material are the authors’ and do not necessarily reflect those of the sponsor. Also, this work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. R01-2008-000-20872-0).

7. REFERENCES

[23] X. Song et al., “Mining in Social Networks Information Flow Modeling based on Diffusion Rate for Prediction

Figure 8: Survival rates of power users (while changing the W_C:W_{T&S} of CP).

