Journal of Web Engineering, Vol. 14, No. 1&2 (2015) 151–178

ANALYZING TOPOLOGICAL CHARACTERISTICS OF THE KOREAN BLOGOSPHERE

JIWOON HA, DUCK-HO BAE, MINSOO RYU, SANG-WOOK KIM*
Department of Computer and Software, Hanyang University
222 Wangsimni-ro, Seongdong-gu, Seoul, South Korea
jiwoonha@hanyang.ac.kr, dhbae@agape.hanyang.ac.kr, msryu@hanyang.ac.kr, wook@hanyang.ac.kr

SEOK-CHUL BAEK
Complex Network Analysis Team, NHN Corp
6 Buljeong-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, South Korea
spacentime@nhncorp.com

BYEONG-SOO JEONG
Department of Computer Engineering, Kyung Hee University
26 Kyungheedae-ro, Dongdaemun-gu, Seoul, South Korea
jeong@khu.ac.kr

JINSOO CHO
Department of Computer Engineering, Gachon University
1342 Seongnamdae-ro, Sujeong-gu, Seongnam-si, Gyeonggi-do, South Korea
jscho@gachon.ac.kr

Received August 27, 2013
Revised November 6, 2014

Due to their popularity and widespread use, blogs have become an important medium through which many people communicate and exchange information on the World Wide Web (WWW). The blogosphere has provided many opportunities for individuals and companies to establish new business models that investigate social relationships. In Korea, there are many blogospheres that appear to have characteristics that differ from other foreign blogospheres on the Internet. Consequently, it is inappropriate to apply the analysis methods used for the foreign blogosphere directly to the Korean blogospheres. To establish successful business policies for the Korean blogospheres, it is necessary to understand the characteristics of the Korean blogospheres and the behavioral patterns of the bloggers. In this paper, we analyze the characteristics of the Korean blog network, wherein each blogger forms a node and scraps by bloggers form edges. First, we demonstrate that the Korean blog network is a scale-free network, like the WWW. Second, we compare the bow-tie structure of the Korean blog network with that of the WWW. Lastly, we analyze the changes in the Korean blog network over time. Results of these analyses will be helpful in developing effective algorithms and establishing new business models targeted at the Korean blogosphere.

Keywords: Blogosphere, social network analysis, bow-tie, data mining, graph mining
Communicated by: G.-J. Houben & E.-P. Lim

1 Introduction

A *blog* is a type of personal homepage, where authors express their opinions, ideas, and information freely on the World Wide Web (WWW). A *blogger* refers to the individual author

*Corresponding author

of each blog, and a post comprises the document that the blogger publishes on his/her blog. Bloggers are able to perform various actions when using a blog. For example, *scrap* refers to the function of copying another person's blog post into his/her own blog. *Comments* enable bloggers to express their opinions on other blog posts. *Trackback* refers to a post that has a link to an existing post, and includes a back link to the newly written post. The *blogosphere* is made up of all of the blogs and their interconnections [1]. This term implies that blogs exist together as a collection of connected communities or as a social network, in which everyday authors can publish their opinions.

Blogs have become an important medium of communication and information on the WWW. Due to their accessible and timely nature, they are also an intuitive source for the spread of information and ideas. The advent of the blogosphere provides researchers with opportunities to establish a new business model that investigates the social relationships of the blogosphere. To establish a successful business policy, it is imperative that we understand the characteristics of the blogosphere and the behavioral patterns of the individual bloggers.

In this paper, we describe our methods for analyzing the characteristics of the Korean blogosphere, as well as our results from the analysis. The Korean blogosphere has the following diverse characteristics[1].

First, Korean blog service companies provide convenient features that enable each blogger to easily generate and update posts with little effort, unlike having to generate web pages via HTML coding.

Second, they provide several options for existing posts, including scraps, trackbacks, and comments. Other websites require cumbersome HTML coding to link other web pages on the WWW; however, Korean blog posts can be easily copied and linked with the click of a mouse.

Third, they provide optimized retrieval services by storing every action that takes place in the blogosphere to the server's storage, allowing a blogger to quickly and easily locate previous posts.

Fourth, they display the most popular and important posts on the blog service's start-up screen, so that bloggers can easily determine popular topics in the blogosphere.

Fifth, they also provide uni- and bi-directional neighbor functions, so that bloggers can register interesting bloggers and manage them. The neighbor function is identified as a bookmark in the web browser, although, unlike a bookmark, which is a list that contains web pages of interest, the neighbor function shows recently written posts and comments.

These attributes enable bloggers to easily express their opinions and ideas in the Korean blogosphere, as well as quickly determine other bloggers' activities, which, in turn, accelerates activity and reduces the distance between bloggers in the Korean blogosphere.

There has been research focusing on the characteristics of web space in recent years [2, 3, 4, 5, 6, 7]. The WWW has been simply modelled as a graph structure (i.e., web pages as nodes, hyperlinks between web pages as edges), and a variety of experiments have been conducted to determine the characteristics of the WWW. The blogosphere has also been analyzed to identify the characteristics of social relationships [8, 9, 10]. However, the existing research on the blogosphere focuses primarily on measuring the degree distribution and diameter of

[1]Naver blog, http://blog.naver.com/
Daum blog, http://blog.daum.net/
Egloos, http://www.egloos.com/
Tistory, http://www.tistory.com/

blog networks and modelling these temporal aspects [11, 12, 13, 14, 15]. In addition, there have been analyses on information diffusion in the blogosphere [16, 17, 18, 19, 20, 21]. In this paper, we describe a variety of experiments in order to examine the unique characteristics of the Korean blogosphere for the purpose of better understanding the current status of the Korean blogosphere.

The characteristics of the Korean blogosphere are applicable to a variety of applications [22, 23]. For example, they could be used in sampling blog networks. The main goal of network sampling is to create a smaller-sized sample, whose characteristics are similar to those of the original network [23]. While it is difficult to analyze a real-world social network in its entirety, due to the large amounts of data and the high complexity of the algorithms, it is possible to use the sampled network, as its analysis results are similar to those from the original network.

To understand the characteristics of the Korean blogosphere, we first analyze the connectivity of the Korean blog network by measuring the degree distribution, the sizes of the weakly connected components (WCC) and the strongly connected components (SCC), and the diameter. We subsequently analyze the internal structure of the Korean blog network using the Bow-tie structure [4], which was proposed as a method for analyzing the entire WWW structure from a macroscopic viewpoint. It divides the entire WWW structure into several components, based on the various relationships between nodes [4]. Finally, we analyze the internal structures of the Bow-tie components and examine the internal characteristics of the Korean blog network through our analysis of the relationships among the nodes in each component.

It is also important to analyze how the repeated eliminations of nodes or edges affect the topology of the entire blog network. Such an analysis can help blog service providers expect how much the successive withdrawals of bloggers influence their blogosphere, enabling them to make a plan on how to better manage them. To that end, we analyze the resilience of the Korean blog network in order to identify changes in the blog network if bloggers were iteratively excluded from the blog network. We do this by generating a blog network where bloggers were successively eliminated and then measure the changes in the ratio of Bow-tie components to WCC size.

Finally, it is possible to understand the evolution of the blogosphere by carefully analyzing changes in the blogosphere over time. This enables blog service providers to make plans for further expansions of facilities and to establish new business strategies. In this paper, we examine the changes that have taken place in the Korean blogosphere over time and analyze the behavioral patterns of blogosphere users.

The paper is organized as follows: In Section 2, we briefly survey the related literature. We describe ways to generate a blog network for experimental purposes in Section 3. We summarize the experimental results for Korean blog network connectivity in Section 4. In Section 5, we describe the characteristics of the Korean blogosphere, based on the Bow-tie structure. We analyze the resilience of the Korean blogosphere in Section 6. In Section 7, we analyze the characteristics of the Korean blogosphere over time. We analyze the scrapping and posting tendencies over time in Section 8 and in Section 9 we present our conclusions.

2 Related Work

In this section, we introduce the existing research related to the analysis of social networks targeting the WWW structure and blog networks. We summarize the degree distribution and diameter measurements as a mechanism for analyzing the connectivity between nodes. In addition, we explain the existing research that focuses on analyzing the whole WWW structure, based on the Bow-tie structure.

2.1 Connectivity Analysis

In this section, we summarize the previous research that has analyzed the connectivity between nodes in the WWW structure and blog networks. Several of the works utilize the distribution of in-degree and out-degree for each node in the graph structure, in which the web pages are considered nodes and the hyperlinks between web pages are considered edges [2, 3, 5, 6, 24, 25]. In [11, 15], the distributions of in-degree and out-degree for each node are measured using two different types of blog networks. One network used a blog as a node, considering the action between blogs as edges. The other network used a post as a node, considering actions between posts as edges.

The diameter has a popular use in representing the extent of connectivity between the nodes in a graph structure [2]. The diameter of a graph is the maximum distance of any vertex in the graph. That is, it is the greatest distance between any pair of vertices. To determine the diameter of a graph, one must first find the shortest path between each pair of vertices. The greatest length of any of these paths is the diameter of the graph. In this regard, diameter is highly sensitive to noise [12]. Thus, the effective diameter is widely used for solving noise problems [26]. The effective diameter is selected from the 90% level of the largest distance. Conversely, the general diameter is chosen as 100% of the largest distance (i.e., top-most value). In [27], the effective diameter of the blog network in which the post is considered to be a node was measured.

2.2 Bow-tie Structure Analysis

The Bow-tie structure was proposed for analyzing WWW structure from a macroscopic viewpoint [4]. Figure 1 demonstrates the Bow-tie structure of the WWW. The Bow-tie structure consists of four components: SCC, IN, OUT, and Tendrils. Every node in the SCC has a path that can access all of the other nodes in the SCC. Every node in the IN has a path to the SCC, but not out from the SCC. Conversely, every node in the OUT has a path out from the SCC, but not to the SCC. Nodes in the Tendrils do not have paths to or from the SCC. The nodes that have a uni-directional path from the IN to the OUT, without going through the SCC, are called Tubes.

In [6], the researchers analyzed the characteristics of the internal structures for each component in the Bow-tie structure of the WWW. Initially, they measured the distribution of degree, SCC, and WCC for each component. In order to determine the existing Bow-tie structures within the IN and OUT, they also measured the sizes of the WCC and SCC within the IN and OUT. To determine connectivity between the SCC and the IN/OUT, they analyzed the internal structure of the SCC by examining the nodes of the SCC that were directly connected with the IN and OUT. In [6], they analyzed the changes that occur when some of the nodes of the SCC have a high in-degree or out-degree and are eliminated by attacks and system errors from the outside.

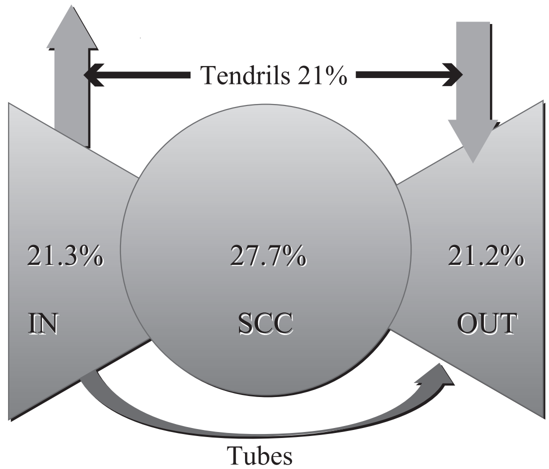


Fig. 1. Bow-tie structure of WWW [4].

Currently, there is a strong body of research that has focused on WWW structure using a Bow-tie structural analysis. However, there has been very little research specifically analyzing blog networks using a Bow-tie structural analysis. In this paper, we analyze blog networks in detail, while generating blog networks from actual blog data.

3 Generating Blog Network

To examine the characteristics of the Korean blogosphere, we used real blog data obtained from the Naver blog, which is one of the largest Korean blog service companies. Blog data were anonymously collected from the Naver blogosphere for several months, beginning in April 2006. The collected blog data consists of approximately 1.7 million bloggers, 1 billion posts, and 40 million scrapes. To analyze the changes over time, the entire duration in which the data were generated was divided into seven time periods. According to the experimental needs, the data from specific periods were used independently (fixed period) and the accumulated data from the first period to the specific periods were used collectively (accumulated period).

Previous analyses of the WWW structure considered web pages as nodes, and hyperlinks between nodes as edges; however, posts in the blogosphere do not have hyperlinks. Thus, we cannot generate a network that connects the entire blogosphere if we were to use hyperlinks to generate a network.

If one web page has a hyperlink to another web page, this indicates that the corresponding web page relates to the content of the linked web page in some way. Even though the blogosphere does not have hyperlinks, it has a variety of features that enable bloggers to demonstrate interest in specific posts. Such actions include scrap, trackback, and reply. In the Naver blogosphere, trackback is rarely used by bloggers, making its analysis as a component of the blog network difficult. Furthermore, the reply to a post feature is used more for acknowledgement or greeting than for demonstrating interest in a post. The scrap feature is primarily used for showing interest in a specific post.

In this study, we generate a blog network that recognizes bloggers as nodes and scraps by bloggers as edges. Figure 2 demonstrates the generation of a blog network. An arrow represents writing a post and a dotted arrow represents scrapping. Since blogger B scraps a post generated by blogger A in the left side of Figure 2, the directed edge from blogger B to blogger A is generated in the right side of Figure 2.

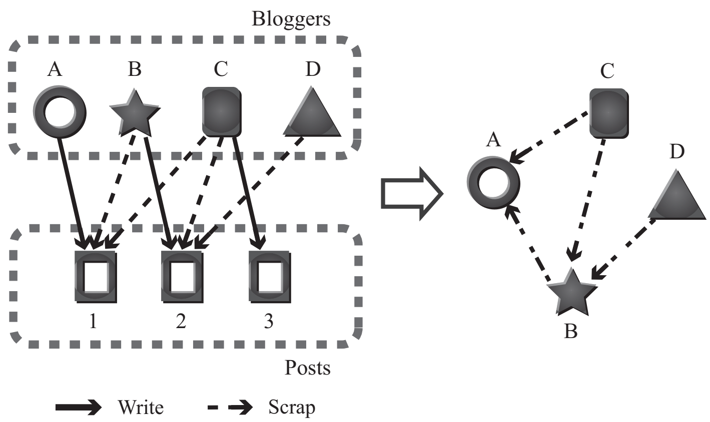


Fig. 2. Generating a blog network.

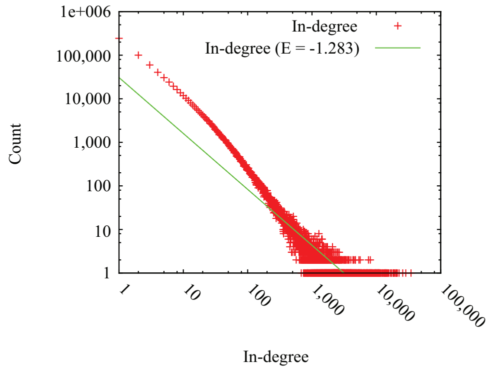
4 Analyzing the Connectivity of the Korean Blog Network

We observed whether or not power-law distribution holds for the Korean blog network, as it does in the WWW structure. Power-laws are expressions of the form, $y \propto x^a$, where a is a constant, x and y are the measures of interest. If the relationship between metrics x and the occurrence frequency of metrics y is equal to Eq. (1), the corresponding measure is defined so as to follow a power-law distribution [4]. When analyzing the WWW structure and blog network, x metrics represent in-degree or out-degree, while the occurrence frequency y represents the number of nodes that have corresponding in- or out-degrees.

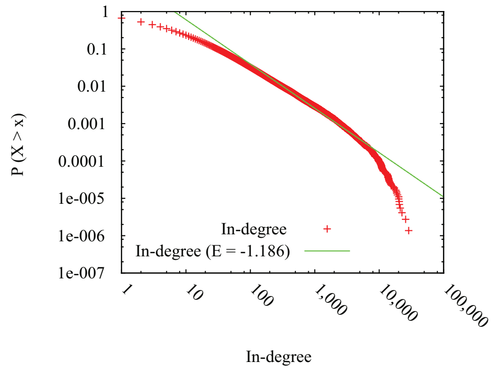
$$y = ax^e \tag{1}$$

We can determine the proper tendency of a corresponding graph by calculating the exponent e of Eq. (1). Generally, the graph for the distribution of degrees is represented by a log scale degree-count graph, which identifies the number of nodes that have a corresponding degree (Figure 3(a)). In this graph, we expect to obtain the value of the exponent e by using the least square method. However, if we use this approach, the trend line strays out of the whole data set, due to the very high value measured for the degree (Figure 3(a)) [28].

To overcome this problem, we drew the graph after transforming the distribution of the degree, using a complementary cumulative distribution function (CCDF), as demonstrated in Figure 3(b). The same approach was used in [28] to obtain the value of the exponent e using the least square method. The value, subtracted by one from the exponent e of the CCDF graph, is considered to be the value of the exponent e of the power-law distribution that best represents the corresponding data set [28].



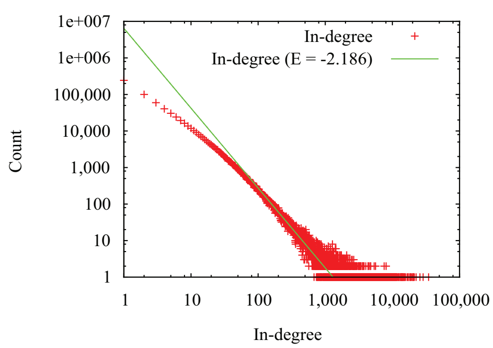
(a) Calculating from the degree - count graph.



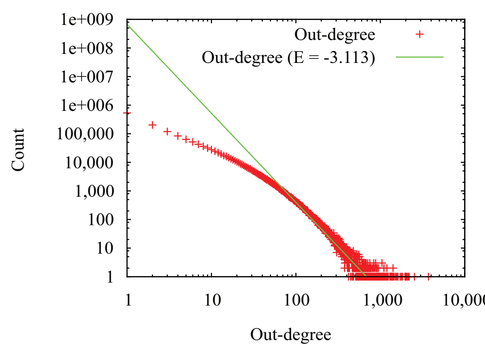
(b) Calculating from the CCDF graph.

Fig. 3. The degree distribution graph and trend line for exponents using different calculations.

A degree distribution graph is generally drawn on a log scale degree - count graph. Fixing the value of the exponent e obtained from the CCDF graph, we can obtain the value of a in Eq. (1) using the least square method, after drawing a trend line suitable for a degree - count graph. The value of a represents the y-intercept in the trend line, which demonstrates the power-law distribution in the log scale degree - count graph. The degree distribution in the Korean blog network is shown in Figure 4, demonstrating the trend line that was obtained using the aforementioned methods. As a result, the trend line from Figure 4(a) more properly represents the characteristics of the whole data, compared to Figure 3(a). The value of the exponent e, of the trend line for in-degree distribution in the Korean blog network, was approximately -2.2, while the value for the out-degree distribution was approximately -3.1. As demonstrated in Figure 4(b), the left part of the out-degree distribution diverged more from the whole trend line, compared to the in-degree distribution. Since this phenomenon also occurred in the out-degree distribution of the WWW network [4], we predicted that the tendency to link other nodes (posts or web pages) is very similar between the Korean blog network and the WWW network.



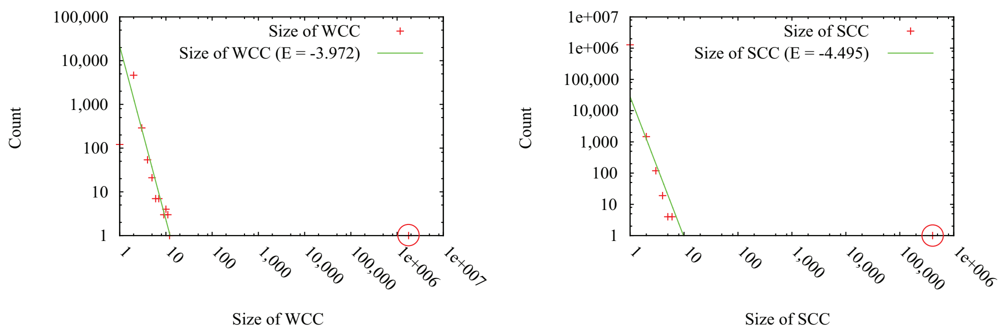
(a) In-degree distribution.



(b) Out-degree distribution.

Fig. 4. Node degree distribution in the Korean blog network.

We also measured WCC and SCC distribution of the Korean blog network and drew a trend line to identify that the WCC and SCC distribution of the Korean blog network follows power-law distribution. Excluding the GWCC (giant WCC) and the GSCC (giant SCC), which are represented by circles in the graph from data set, we drew a power-law trend line for the remaining data. The results demonstrate that the WCC and SCC distribution of the Korean blog network also follow the same power-law distribution as the WWW network. However, the Korean blog network has an exponential trend line of -4 and -4.5, while both the WCC and SCC distribution of the WWW network have an exponential trend line of about -2.54. This indicates that the size of the WCC and SCC, excluding the GWCC and GSCC, is generally small, and the GWCC and GSCC are a greater part of the whole Korean blog network. From this, we determined that the nodes in the Korean blogosphere are more closely connected with each other, as compared to the WWW network.



(a) WCC distribution.

(b) SCC distribution.

Fig. 5. WCC and SCC distribution of the Korean blog network.

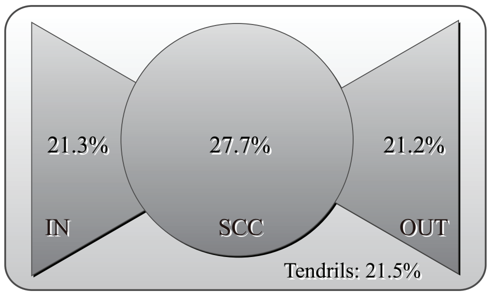
In order to identify connectivity between members in the Korean blog network, we compared the diameter of the Korean blog network with that of the WWW network. Previous studies indicate that the diameter of the whole WWW network is measured at 19 [2], with the SCC being 28 [4], and the effective diameter of the blog post network has been measured to be approximately 14 [27]. Our study demonstrates that the effective diameter of the Korean blog network is approximately 4.5. This implies that nodes in the Korean blogosphere are closer to each other than nodes in the WWW network.

The Korean blogosphere is functional in terms of high quality retrieval. Furthermore, it has a blog portal, which refers to an entry to a blog service. In a blog portal, the most popular and important posts are selected by the blog service company using a diverse set of criteria and are then registered. Using this retrieval function and blog portal service, each blogger can directly access original posts. Arguably, this contributes to the relatively small diameter of the Korean blog networks, compared with WWW networks.

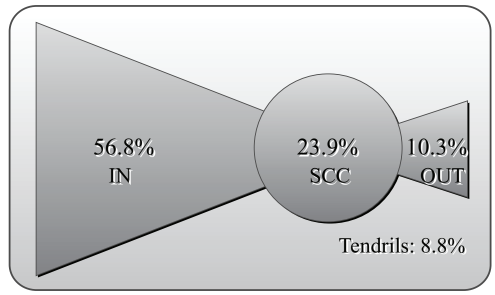
5 Bow-tie Structural Analysis of the Korean Blog Network

In this section, we generated a Bow-tie structure of the Korean blogosphere so as to analyze it from a macroscopic viewpoint. We then compared the Korean blogosphere to WWW networks. Figure 6 demonstrates the Bow-tie structures of the WWW network and the

Korean blog network, respectively. As shown in Figure 6, the ratios of the IN, SCC, and OUT in the WWW network are very similar. However, the Bow-tie structure of the Korean blog network demonstrates that the IN is nearly 5.5 times larger than the OUT.



(a) Web network [4].



(b) Korean blog network.

Fig. 6. Comparison of Bow-tie structure.

Bloggers who belong to the OUT represent the type of blogger who primarily writes his own posts. Bloggers of the IN represent a blogger who primarily scraps other blogger's posts. Bloggers belonging to the SCC write their own posts and scrap other blogger's posts.

Because blogs are popular for putting necessary information in one place, the activity of scrapping other bloggers' posts makes up a large portion of blogging activity. As a result, the amount of the IN appears to be much larger. At the same time, convenient blogging services and facilities, (i.e., scrapping) provided by blog service companies, also appear to impact the growth of the IN.

Another difference between Korean blogs and WWW networks arises from how they originate. Because every Korean blog site is managed by a blog service company, all of the posts can be accessed and retrieved, even if the corresponding posts have not been scraped by other bloggers. On the other hand, web pages are managed independently by different web servers and are only identified via hyperlinks. As a result, there may be many web pages that are not hyperlinked at all, which prevents us from accessing them. Since these web pages belong to the IN of the Bow-tie structure, the portion of the IN for the WWW networks is lower than that of the Korean blog networks.

We used the characteristics of the Bow-tie components to determine the probability that a path exists between any two nodes. In order for a path to exist between any two nodes, the two nodes should be directly connected or there should be at least one node to link them. The following four conditions are cases in which the previous situation is satisfied.

(1) A case in which one node exists in the IN and the other node exists in the OUT
(2) A case in which one node exists in the IN and the other node exists in the SCC
(3) A case in which one node exists in the SCC and the other node exists in the SCC
(4) A case in which one node exists in the SCC and the other node exists in the OUT

To summarize the previous cases, we can express Eq. (2)[2]. Let $n(total)$ be the total

[2] Additionally, even though there can be a path between nodes among IN and OUT respectively, we ignore them in this paper since the portion of them is very small in whole blog network [4].

number of nodes in the blog network. Let $n(SCC)$, $n(IN)$, and $n(OUT)$ be the number of nodes in the SCC, IN, and OUT, respectively.

$$\frac{n(IN)+n(SCC)}{n(total)} \times \frac{n(SCC)+n(OUT)}{n(total)} \tag{2}$$

In the WWW network, the probability that a path exists between any two nodes is approximately 24% [4]. Our study demonstrates that the probability of the Korean blog network is approximately 25%. This indicates that connectivity between nodes in the WWW network and the Korean blog network are quite similar.

Table 1. Internal structure of IN and OUT in WWW and the Korean blog network.

	Web [6]		Korean blogosphere	
The IN component				
nodes in IN	14.4 M	11%	962,019	56.80%
max SCC	5,876		2	
number of WCCs	3.68 M		40,163	
max WCC	197,500	1.30%	766	0.08%
The OUT component				
nodes in OUT	53.3 M	39%	176,073	10.39%
max SCC	9,349		7	
number of WCCs	25.4 M		5,764	
max WCC	14.94 M	28.01%	5,955	3.38%

We also analyzed the internal structure of the Bow-tie components. To verify that another Bow-tie structure was created within the IN and OUT, we measured the distribution of the SCC and WCC within the IN and OUT[3]. In Table 1, we show the comparison between the internal structures of the IN and OUT for the WWW and the Korean blog network.

As demonstrated in Table 1, the max WCC of the IN was 0.08%, and the max WCC of the OUT was 3.38%. These values indicate that the Korean blog network does not have a large enough WCC or SCC to generate a Bow-tie structure within the IN and OUT, unlike the WWW network. These values also indicate that the max WCC of the IN and OUT in the Korean blog network is very small compared with the WWW network, but that the Korean blog network has a relatively large portion of IN, compared to the WWW network. From this, we infer that members within the IN and OUT in the Korean blog network do not actively communicate with each other.

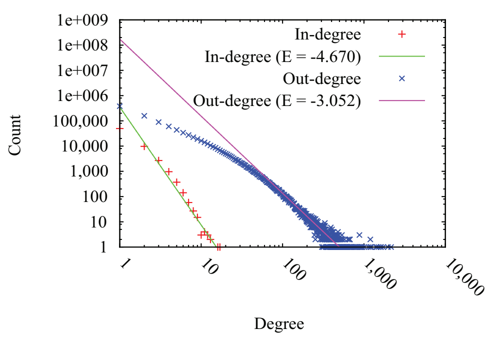
Table 2. Internal structure of SCC in WWW and the Korean network.

	Web [6]		Korean blogosphere	
The SCC				
nodes in SCC	44.7 M	33%	427,101	24.07%
entry points	2.6 M	5.87%	285,671	66.89%
exit points	29.6 M	72.03%	292,050	68.38%
bridges	2 M	4.58%	204,127	47.79%
connectors	2.96 M	6.63%	14,213	3.33%
petals	1.4 M	3.14%	940	0.22%

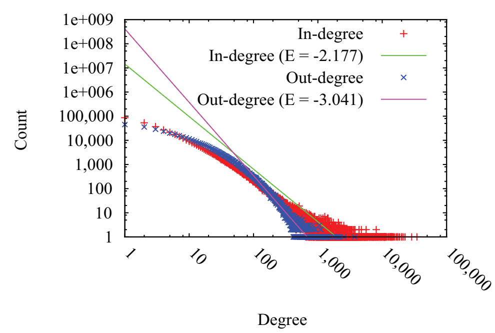
[3] We refer to [6] in doing experimentation.

In this study, we identified the internal structure of the SCC using several SCC nodes, which have specific relationships with the IN and OUT. The specific SCC nodes include: the entry point, exit point, bridges, connectors, and petals. The entry point indicates an SCC node where more than one node of the IN has a link. The exit point indicates an SCC node in which there is a link to more than one node of the OUT. Bridges are the nodes that act as both an entry point and an exit point. A connector refers to an SCC node that has only one link and is linked by only one node. If a connector node has a link to a node and is also linked by same node, then it is referred to as a petal. Table 2 describes our comparison of the internal structures within the WWW and Korean blog networks.

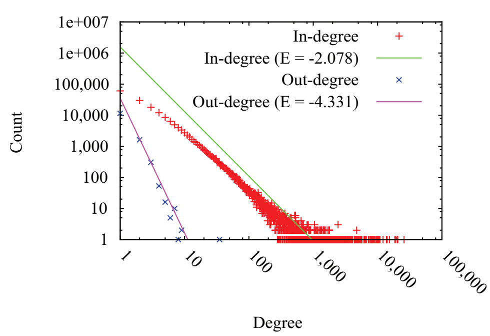
Our results from this study indicate that the ratio of entry points and bridges in the Korean network is 10 times higher than that of the WWW network. While the ratio of exit points is slightly lower in the Korean blog network, it maintains a relatively high ratio given that the ratio of the OUT is approximately 50%, compared to the WWW network. This indicates that most SCC nodes are directly connected to the IN and OUT. We infer that the connectivity between the SCC and IN/OUT increases due to the functionalities provided by a blog service company, including retrieval, blog portal, and neighboring.



(a) Degree distribution of IN.



(b) Degree distribution of SCC.



(c) Degree distribution of OUT.

Fig. 7. Degree distribution of each Bow-tie component.

We measured the degree distribution of each Bow-tie component in order to determine the behavioral tendencies of bloggers for each component. Figure 7 shows the distribution of the in-degree and out-degree within each Bow-tie component of the Korean blog network.

The results indicate that the degree distribution of each component follows a power-law distribution. Among all of the components, the degree distribution of the OUT was most similar to the power-law trend line. In the degree distribution of the SCC, the low degree part diverged from the power-law trend line to some degree for both the in-degree and out-degree.

We analyzed the distribution of the edges between two components in order to analyze the connectivity between two Bow-tie components of the Korean blog network (Figure 8). Our results indicate that the edges from the SCC to SCC take up the largest portions, due to the highly active blogging of the SCC, while edges from the IN to SCC occupy the second largest portion. It is interesting to note that the number of edges from the SCC to OUT and from the IN and OUT appear to be similar. This suggests that the bloggers of the IN primarily access the posts of the OUT, with no connection with bloggers of the SCC. We argue that this is a result of the retrieval, blog portal, and neighboring features that are provided by each blog service company, which provide direct access to original posts.

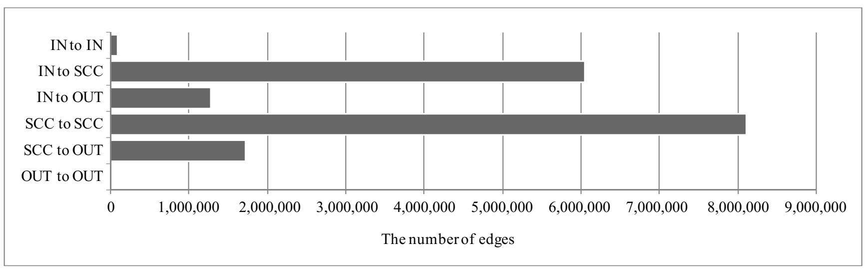


Fig. 8. Distribution of edges between two components of the Bow-tie structure.

We summarize the characteristics of the Korean blogosphere's Bow-tie structure, as compared with that of the WWW, as follows:

(1) The ratio of the IN and OUT in the Korean blog network differs dramatically from the WWW network.
(2) Connectivity between the SCC and IN/OUT is very high.
(3) The number of direct edges from the IN to OUT is very large.

These results indicate that the macroscopic structure of the Korean blog network differs significantly from the WWW network. Figure 9 demonstrates the expected Korean blogosphere's macroscopic structure, based on the aforementioned analysis.

6 Resilience Analysis of the Korean Blog Network

Next, we analyzed the resilience of the Korean blog network. To determine the resilience of the WWW network in [6], the researchers measured the size change of the updated SCC, while eliminating the SCC nodes with a high in-degree or high out-degree. In this study, we analyzed the resilience of the Korean blogosphere by eliminating power bloggers[4]. Power bloggers are considered highly influential bloggers in the blogosphere. We measured the size

[4]In this work, the in-degree of a blogger was used for the selection criteria defining power bloggers.

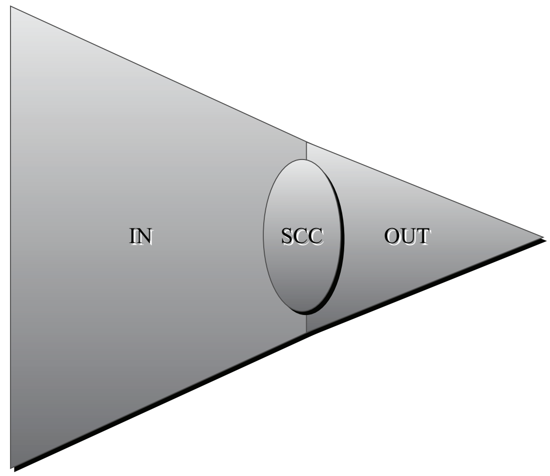


Fig. 9. Macroscopic structure of the Korean blogosphere.

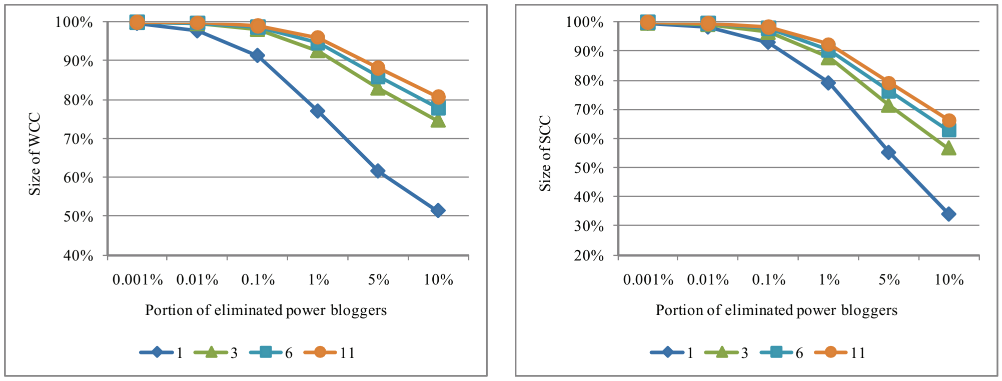
of the WCC and SCC after eliminating the power bloggers and scraps from their posts in the Korean blog network.

To analyze the resilience of the Korean blog network in more detail, we generated several blog networks by varying edge acceptance criteria. The edge acceptance criteria are applied when we generate an edge of a blog network. In section 3, we generated an edge from A to B if blog A scrapped at least one post from B. By applying the edge acceptance criteria, we generated an edge from A to B if blog A scrapped posts of blog B more than a given number. We refer to the given number as edge acceptance criteria.

Applying different edge acceptance criteria in the generation of a blog network is reflective of two things. First, generating a blog network with a low value of edge acceptance criteria eliminates noise. For example, if we accept an edge with only one scrap, then many edges might result in noise. By applying the low edge acceptance criteria, we can eliminate noisy edges. Second, when applying high edge acceptance criteria, we can generate strongly connected blog networks. Edges, in this circumstance, imply a very strong relationship between nodes.

The resilience of a blog network consisting of relatively active bloggers can be determined by eliminating power bloggers, which is done using different edge acceptance criteria. We measured the size of the WCC and SCC while varying the edge acceptance criteria (i.e., 3, 6, and 11) and the portion of the eliminated power bloggers (i.e., 0.001%, 0.01%, 0.1%, 1%, and 5%) (Figure 10). Each line on the graph represents a differently applied edge acceptance criteria and each point on the graph represents the size ratio (%) of the WCC and SCC among the blog networks with and without the power bloggers.

Our results indicate that the size reduction of the WCC and SCC, related to the elimination of power bloggers, tends to decreased with increasing edge acceptance criteria. Therefore, we suggest that bloggers in reduced blog networks are not dependent on power bloggers. Rather, they are more evenly connected with other bloggers via common interests.



(a) Changes in the size of the WCC.

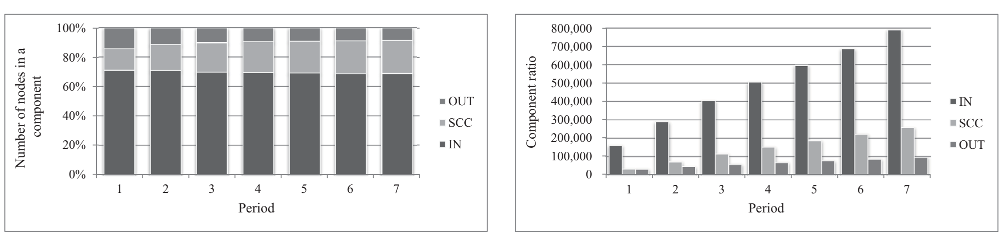
(b) Changes in the size of the SCC.

Fig. 10. Changes in the blog network related to the elimination of power bloggers.

7 Characteristics of the Korean Blogosphere over Time

In this section we analyzed the characteristics of the blogosphere over time. In order to do so, we used data from each period to analyze the blogosphere's transformation process over time. In addition, we analyzed how the unique properties of the Korean blogosphere affected the transformation.

First, we analyzed the transformation of the Bow-tie structure of the Korean Blogosphere. In order to assess how the Bow-tie structure has changed, we accumulated data from each period and measured the changes in the number of nodes of the bowtie component and the relative ratio as time elapsed. We conducted experiments using three key elements that account for approximately 90% of the Bow-tie structure: IN, SCC, and OUT. Figure 11(a) shows the number of nodes in the Bow-tie component of the blog network over time. The horizontal axis represents the periods and the vertical axis represents the number of nodes in each component of the Bow-tie structure. Figure 11(b) displays the changes in the relative ratio of the Bow-tie components of the blog network. The horizontal axis represents the period over time, and the vertical axis represents the percentage of each component in the Bow-tie structure.



(a) Number of nodes in a Bow-tie component over time.

(b) Ratio of Bow-tie components over time.

Fig. 11. Bow-tie structure of the blog network over time.

As shown in Figure 11(a), the number of nodes in the IN rapidly increases, the number of nodes in the SCC increases gradually, and the number of nodes in the OUT increases very slowly over time. The nodes in the IN represent bloggers who mostly scrap other bloggers' posts, which can be viewed as one of the main characteristics of new bloggers. These findings suggest that convenient blog creating and scrapping features, which are provided by Korean blog service providers, provide opportunities for newcomers to join the blogosphere.

Unlike the number of nodes in the IN, which appears to be increasing at a rapid rate, the ratio of the SCC has gradually increasing in terms of the Bow-tie component ratios (Figure 11(b)). Nodes that belong to the SCC represent bloggers who are actively engaged in the blogosphere and influence each other. Accordingly, as the blogosphere expands, the ratio of active bloggers increases, while the percentage of bloggers who merely scrap others' posts, or whose posts are scrapped by others, decreases. This suggests that, in addition to quantitative growth, the Korean blogosphere is showing signs of development in terms of quality.

In order to analyze the connectivity among the nodes in the Korean blogosphere over time, we measured the average connected distance, which refers to the average shortest distance between all of the connected node pairs [4]. The in-links and out-links are the average connected distances, calculated using the reverse and forward edges, respectively. "Undirected" refers to the average connected distance calculated after removing directionality from the edges. Figure 12 displays the experimental results, which indicate that the average connected distance gradually decreases over time. This suggests that the average distance between bloggers slowly reduces as time elapses. It appears that the increase in the average connected distance during Period 2 was caused by a temporary spike in the number of new bloggers entering the IN, who had little to no relationships with other bloggers.

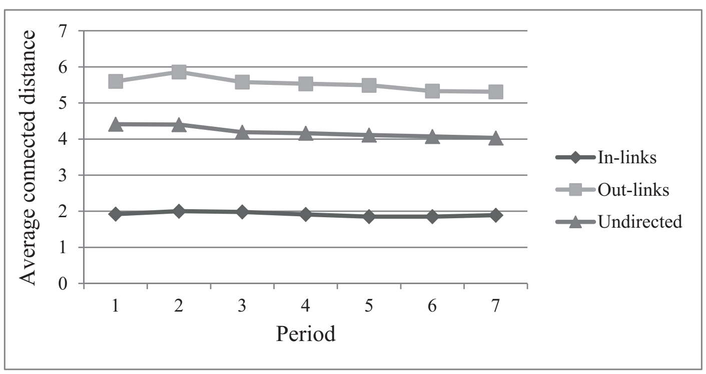


Fig. 12. Average connected distance of the blog network over time.

The effective diameter [26] is another index that is widely used to assess the connectivity among the nodes in a network. Figure 13 displays the changes in the effective diameter over time. Our experimental results indicate that the effective diameter of the blog network slowly decreases as time elapses. This suggests that, as with the decreasing average connected distance, interactions among bloggers increases over time, creating diverse relationships and shortening the distance among the nodes in the blog network.

Based on the changes in the Bow-tie component ratio, the average connected distance, and the effective diameter over time, we determined that the connectivity among the nodes in

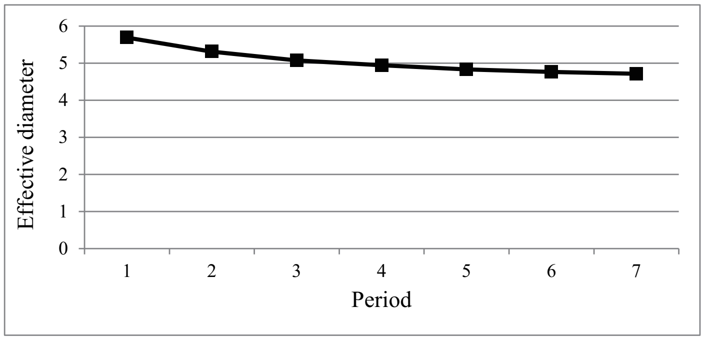


Fig. 13. Effective diameter of the blog network over time.

the blog network increases with time. Korean blog service providers allow bloggers to scrap and express interest in other people's posts, facilitating searching for and easily finding the desired information. Moreover, they publicize, on their main home page, popular and high quality posts that are likely to draw the attention of other bloggers. Arguably, these functions provide bloggers with easy access to blog services, easy recognition of the activities of other bloggers, and a convenient means of expressing their own opinions, all of which contributes to enhancing connectivity in the blog network.

8 Posting and Scrapping Tendencies

In order to assess the general tendencies pertaining to the creation of posts in the Korean blogosphere, we measured the changes in the total number of posts for each period. Figure 14 displays the number of posts created each day, which showed a gradual increase over time. This indicates that the overall level of activity has been increasing as the blogosphere has expanded. Marked with a circle in Figure 14, there were an exceptionally high number of posts created on Day 122; however, this was noise created by an automatic robot. Based on a detailed analysis, we determined that three bloggers had written approximately 20% of all of the posts created that day. An analysis of the posting times verifies that the three bloggers engaged in abnormal blogging behavior, such as writing 40 posts in a minute.

We also measured the average number of posts created on each day of the week. Figure 15(a) shows the average number of posts for each day of the week; the horizontal axis represents the days of the week, and the vertical axis is the average number of posts per day. Our experimental results indicate that approximately 190,000 posts are created on a daily basis, with negligible differences among the days of the week. In other words, there was no difference between weekday and weekend patterns in other countries' blogospheres [11].

Figure 15(b) displays the average number of posts during each hour of the day; the horizontal axis represents the hour of the day, and the vertical axis is the average number of posts during each hour. Our experimental results indicate that there is hardly any blogging activity in the early morning hours, between 3 and 8 AM. These results also show that blogging activity slows down between 6 and 7 PM, when most people are travelling home from work.

In [13, 15, 27], the burstiness of blog networks was measured using an entropy plot [29]. Burstiness is a measure of the level of concentration regarding a particular phenomenon that

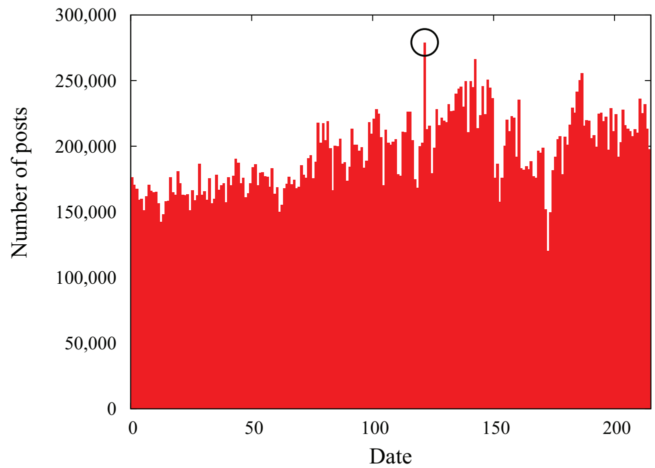


Fig. 14. Number of posts created on each day.

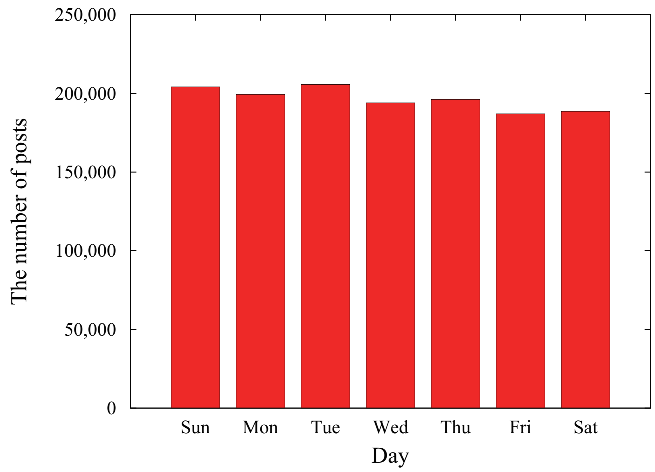
occurs in a specific section. The entropy plot is expressed using the entropy $H(W)$ distribution of a particular phenomenon for a resolution of W. The resolution refers to how finely the entire section is divided into subsections, and entropy $H(W)$ at resolution W is defined as Eq. (3) [29], where $p_{i,w}$ denotes the ratio between the number of occurrence in the entire section i. That is, $p_{i,w} = n_{i,w}/n$.

$$H(W) = -\Sigma_i p_{i,w} log_2(p_{i,w}) \tag{3}$$

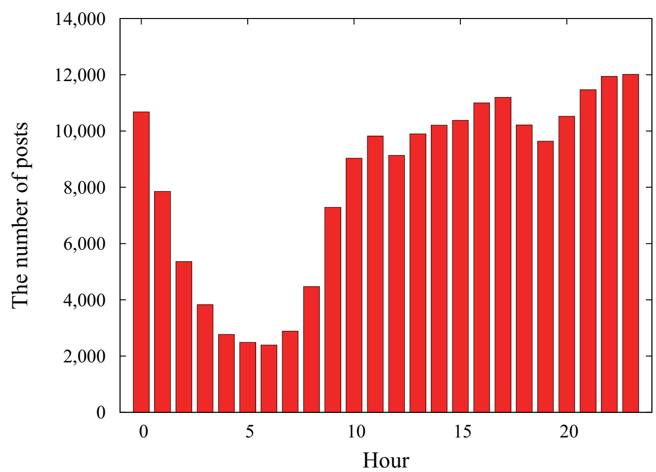
In an entropy plot, the vertical axis is $H(W)$ and the horizontal axis is $log_2(W)$. When data on an entropy plot displays a linear function, the phenomenon is described as self-similar [29]. When a phenomenon is self-similar, a pattern found in the entire section is also displayed in the subsections. Self-similarity is a representative characteristic of fractals.

When a linear regression line is drawn for the elements on the entropy plot, the slope f of the linear function is referred to as the information fractal dimension [29], which provides a measure of burstiness for the corresponding phenomenon; $f = 1$ indicates that the phenomena occurred uniformly, and $f = 0$ indicates that the phenomena occurred simultaneously. In order to examine the burstiness of posting in the Korean blogosphere, we constructed an entropy plot for the number of posts created and measured the corresponding information's fractal dimension. The entropy plot of the number of posts created appears as a linear function (Figure 16). This indicates that posting in the blogosphere has self-similar characteristics. The information fractal dimension of the number of posts created was 0.997, which is close to 1. In comparison, the information fractal dimensions calculated in studies focusing on other countries' blogospheres were 0.88 [15] and 0.72 [13], respectively. The fact that the information fractal dimension of the Korean blogosphere is closer to 1 indicates that posting occurs more uniformly in Korea than in other countries.

Higher uniformity in posting suggests that the number of posts created is similar each day. Unlike a situation wherein the number of posts would drastically increase due to a particular social issue, our findings indicate that the number of posts created each day remains almost



(a) Number of posts created each day of the week.



(b) Average number of posts created each hour of the day.

Fig. 15. Average number of posts created at various time periods.

constant, regardless of various social issues. This indicates that Korean bloggers write posts in a series format about topics they are interested in or about their daily lives and, that the activities in the Korean blogosphere are not significantly affected by external factors.

In addition, we analyzed the patterns showing when posts created in the Korean blogosphere were disseminated by other bloggers. A blogger scrapping a post written by another blogger signifies that the content of the corresponding post has been disseminated [16]. As seen in Figure 17, the horizontal axis represents the months, and the vertical axis shows the amount of scrapping that occurs each month. Our experimental results indicate that the amount of scrapping, i.e., the level of disseminated information, has increased over time.

We also charted the number of active relationships, with the horizontal axis indicating the months, and the vertical axis the number of active relationships formed each month (Figure

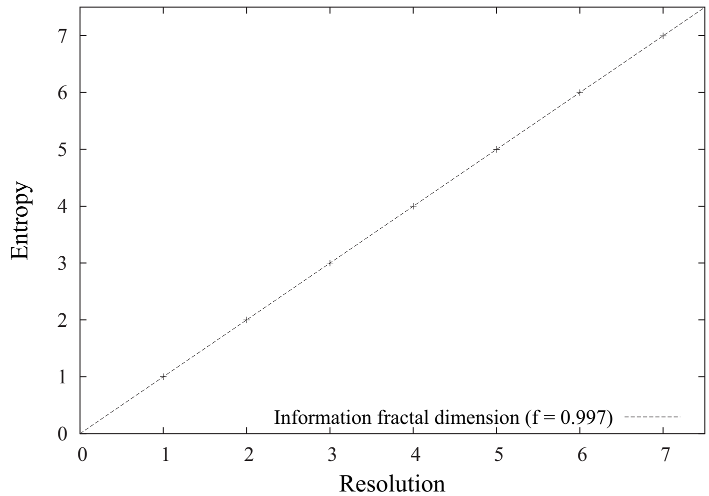


Fig. 16. Entropy plot of the number of posts created.

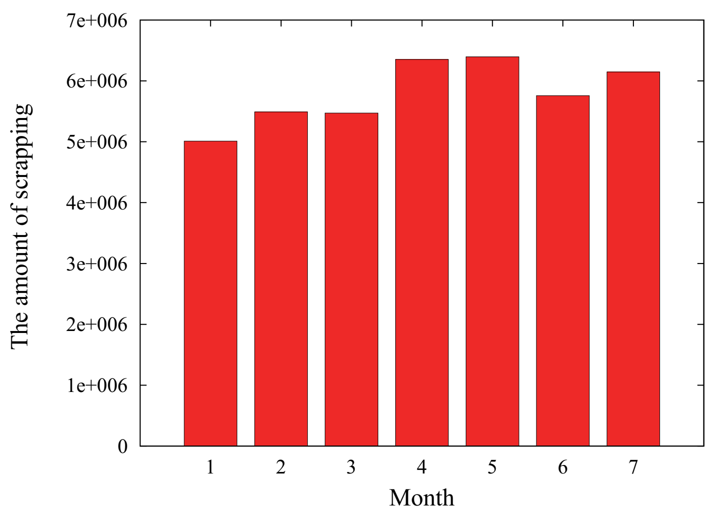


Fig. 17. Amount of scrapping each month.

18). A relationship is represented by the edge between the nodes in a blog network and, according to the conditions of the blog network creation, an edge between two bloggers is formed when a blogger scraps a post written by another blogger more than once. Scrapping that occurs between a pair of bloggers is expressed as a single edge, regardless of the amount of scrapping. In this paper, we defined an active relationship as a newly formed relationship within a particular time period, or when scrapping occurs in an old relationship during a particular time period. Our experimental results indicate that the number of active relationships also increases over time, as does the amount of scrapping.

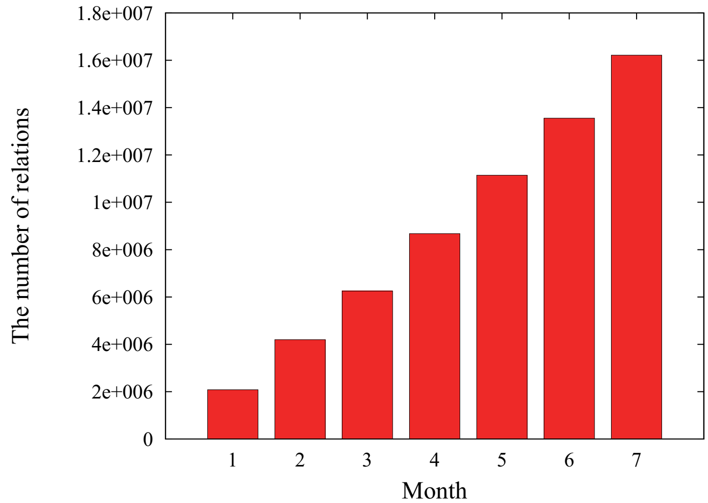


Fig. 18. Number of active relationships formed each month.

Similar to our process for constructing the entropy plot for post creation and measuring the information fractal dimension, we constructed an entropy plot and measured the information fractal dimension for scrapping in the Korean blogosphere. As with posting, the entropy plot for scrapping also displayed a linear function (Figure 19). This indicates that scrapping has self-similar characteristics in the blogosphere. The information fractal dimension of scrapping was very close to 1, at 0.999, which is higher than the information fractal dimension of the in-link generation of 0.72 for another country's blogosphere [13]. These findings suggest that scrapping occurs with stronger uniformity in Korea than in the blogospheres of other countries.

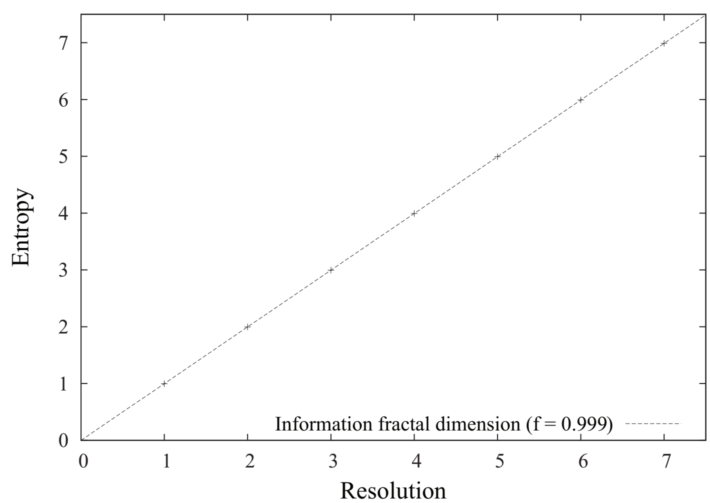


Fig. 19. Entropy plot of the amount of scrapping.

In order to analyze bloggers' scrapping tendencies, we measured the time between new postings and scrapping. Figure 20 displays the number of days from posting to scrapping; the horizontal axis represents the number of days from posting to scrapping, and the vertical axis is the amount of scrapping that occurs on a particular hour drawn on the logarithmic scale. Unlike most in-links in other blogospheres, which occur within 24 hours of posting [11], our experimental results indicate that only 15% of scrapping in the Korean blogosphere occurred during the first 24 hours of posting. Although the amount of scrapping quickly declines after the first 24 hours, as reported in [11], some scrapping does continue to occur for approximately 10 days.

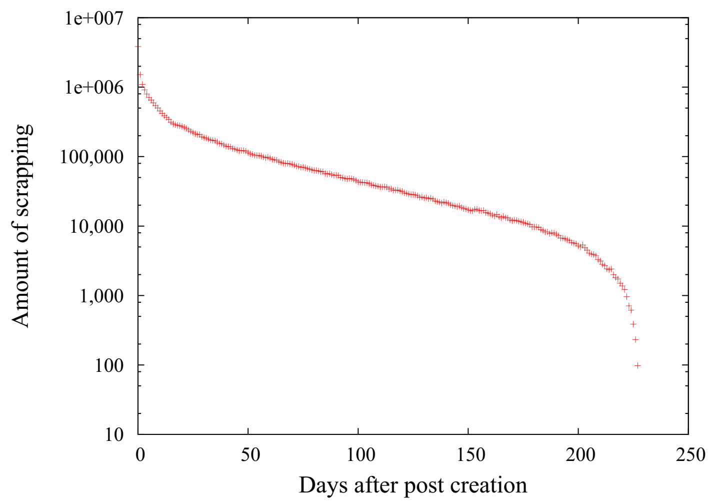


Fig. 20. Time between posting and scrapping.

The correlation between each blogger's in-degree and out-degree was analyzed based on a scatter plot (Figure 21). Each point represents a blogger, and the horizontal and vertical axes drawn to logarithmic scales represent the bloggers' in-degree and out-degree, respectively. Our experimental results indicate that bloggers are widely distributed throughout the entire space, and do not display any particular pattern.

Although it can be assumed that the level of interest that a blogger receives (in-degree) is associated with the blogger's level of activity (out-degree), one study [11] determined that there is actually a very low correlation between bloggers' in-degrees and out-degrees. Similarly, the correlation between a bloggers' in-degree and out-degree turned out to be very low in the Korean blogosphere, where scrapping plays the role of hyperlinks in other blogospheres. This means that the amount of scrapping done by a blogger (out-degree) does not affect the level of attention he or she receives from other bloggers (in-degree). There were rare cases of bloggers with more than 1,000 out-degrees and 10,000 in-degrees; however, in these situations, we believe that they scrapped many high quality posts created by other bloggers and categorized them well, which were then re-scrapped by other bloggers.

In addition, in order to examine whether these characteristics change over time, we conducted identical experiments using a network divided into monthly units. These results

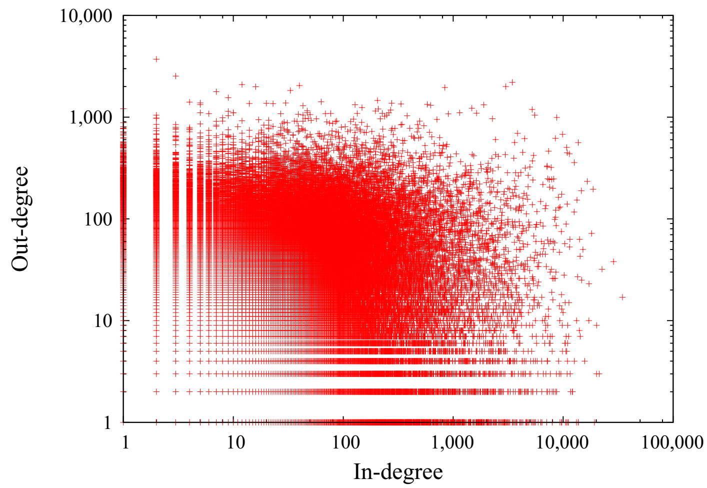


Fig. 21. Relationship distribution between bloggers' in-degree and out-degree.

showed the same tendencies in each network, indicating that there is no correlation between the amount of scrapping done by a blogger and the level of interest he or she receives from other bloggers for any of the monthly periods examined.

We also conducted an experiment to assess the correlation between bloggers' in-degrees and out-degrees, as well as the number of bloggers with a corresponding number of degrees. Figure 22 shows a 3D representation of bloggers' in-degrees and out-degrees, with the number of bloggers with a specific number of in-degrees and out-degrees, on a logarithmic scale, which was mapped onto a 2D plane.

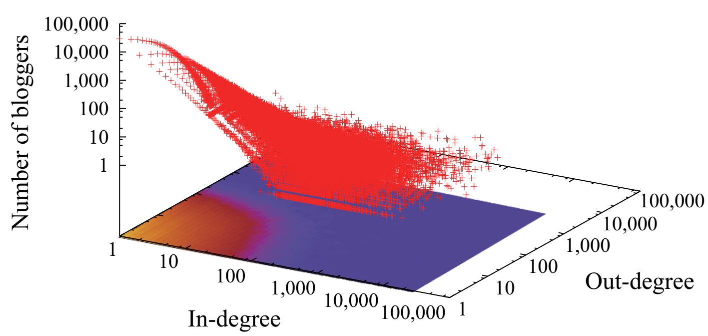


Fig. 22. Relationship distribution between in-degree, out-degree, and the number of bloggers.

We next attempted to determine the relationship between the amounts of scrapping a blogger receives and the number of posts he or she creates. Figure 23 displays the relationship between the number of posts created by bloggers and the amount of scrapping they received on a 2D plane; each point represents a blogger, with the horizontal axis signifying the number of posts, and the vertical axis showing the amount of scrapping. The general shape of the distribution is a rhombus, with a small top-right protrusion. The Pearson correlation coefficient was 0.21, indicating that there is a very weak correlation between the number of posts created by a blogger and the amount of scrapping he or she receives.

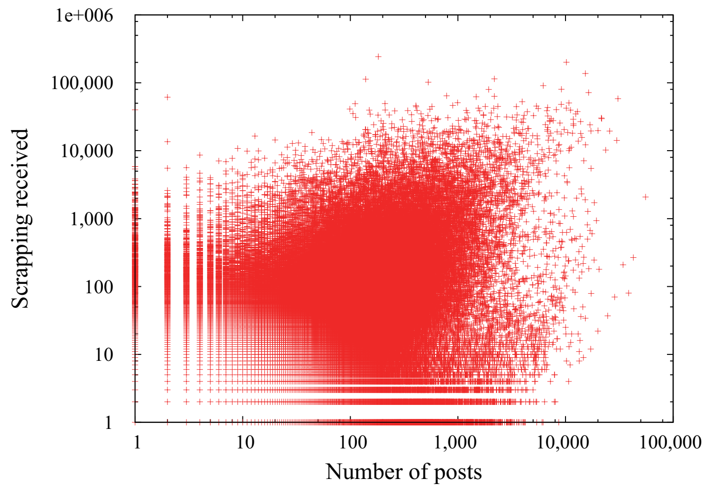
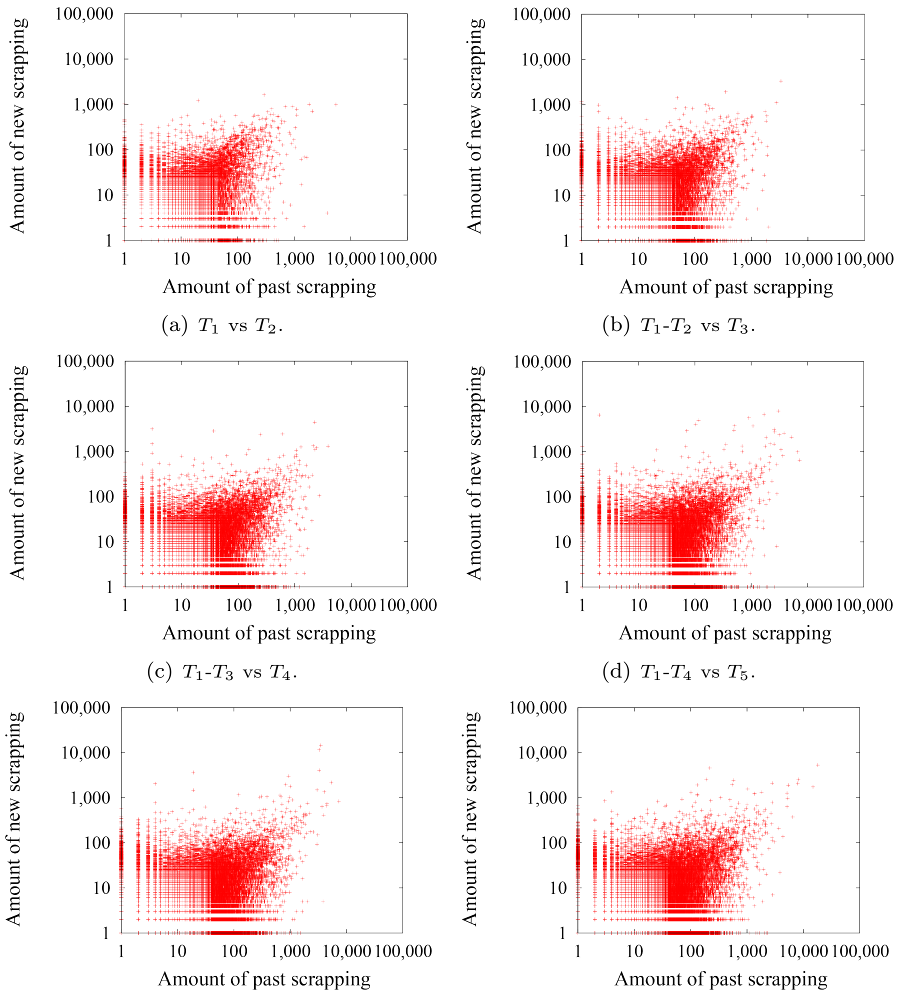


Fig. 23. Relationship distribution between the number of posts created and the amount of scrapping.

Next, we conducted an experiment to examine whether there is a tendency for bloggers to scrap posts from blogs whose content they had previous scrapped. Figure 24 displays the correlation between the amount of new scrapping and the amount of previous (cumulative) scrapping in each relationship. Each point represents the relationship between the two bloggers (an edge in the blog network), the horizontal axis shows the total amount of scrapping from T_1 to T_{n-1}, and the vertical axis represents the amount of new scrapping in the corresponding relationship during T_n. Figures 24(a)-(f) depict the distribution in each period. In order to calculate the total amount of scrapping in a particular relationship, we used accumulated data from each time period for this experiment.

Our experimental results yielded a rhombus shape distribution with a small top-right protrusion, which did not vary significantly as time elapsed. The average Pearson correlation coefficient was 0.45, and there were minor discrepancies among the periods. These results indicate that there is a tendency for bloggers to scrap from blogs where they had previous scrapped, and this tendency continues over time.

We conducted experiments to assess the influence of the total in-degree on the behavior of other bloggers. Figure 25 displays the relationship distribution between the bloggers' accumulated in-degrees and the amount of new scrapping they received. Each point represents a blogger, with the horizontal axis showing the in-degree from T1 to Tn-1, and the vertical axis representing the distribution of the amount of new scrapping they received during Tn. Figures 25(a)-(f) shows the distribution during each time period. As with the previous experiments, we compiled the data accumulated for each time period to measure the total in-degree of each particular blogger.



(a) T_1 vs T_2.
(b) T_1-T_2 vs T_3.
(c) T_1-T_3 vs T_4.
(d) T_1-T_4 vs T_5.
(e) T_1-T_5 vs T_6.
(f) T_1-T_6 vs T_7.

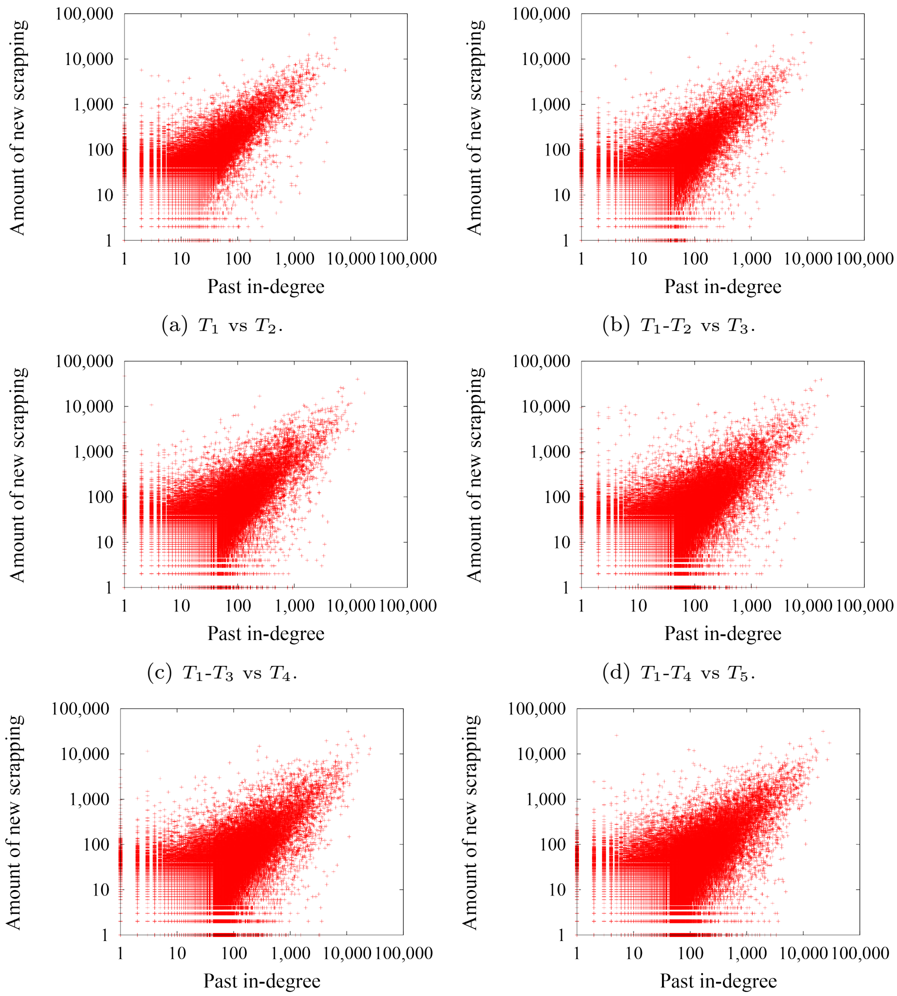
Fig. 24. Correlation between the amount of new scrapping and the amount of previous (cumulative) scrapping in each relationship.

9 Conclusions

In this paper, we analyzed the unique characteristics of the Korean blogosphere and generated a blog network by identifying bloggers as nodes and identifying scraps by bloggers as edges.

First, we analyzed the structural characteristics of the Korean blogosphere. We identified the characteristics of this scale-free blogosphere using degree distribution, and we measured the diameter of the Korean blog network. Our analysis indicates that nodes in the Korean blog network are more tightly bound together than in the web or foreign blog networks. We concluded that the features (retrieval function, collected interesting posts in a blog start screen, etc.) provided by Korean blog service companies may impact these results.

Second, we generated a Bow-tie structure for the Korean blog network and identified



(a) T_1 vs T_2.

(b) T_1-T_2 vs T_3.

(c) T_1-T_3 vs T_4.

(d) T_1-T_4 vs T_5.

(e) T_1-T_5 vs T_6.

(f) T_1-T_6 vs T_7.

Fig. 25. Correlation between bloggers' accumulated in-degree and the amount of new scrapping.

various differences between the Korean blogosphere and the WWW network. Our analysis also demonstrated that the ratio of IN and OUT in Korean blog networks differs significantly from that of the WWW network. An analysis of each component of the Bow-tie structure and its internal structures determined that most SCC nodes are directly connected with the IN and OUT nodes. We also determined that bloggers from the IN directly access the posts from the OUT to a large degree, without going through bloggers from the SCC. The Bow-tie structural analysis of the two networks demonstrates fairly large differences, due to the various purposes uses have for blog services and web pages, as well as the different service environments available to these networks.

Third, we analyzed the resilience of the Korean blog network, where relatively active bloggers exist by eliminating power bloggers from a blog network generated by edge acceptance criteria. The results from this experiment indicate that the impact of power bloggers tends to weaken with increases in the edge acceptance criteria. We conclude that, in a reduced blog network generated by a high edge acceptance criteria, most bloggers appear to be connected by common interests, rather than by power bloggers.

Fourth, we analyzed the changes in the characteristics of the blogosphere over time. By analyzing the changes in the number and the ratio of nodes of the Bow-tie components in the blog network, the average connected distance, and the effective diameter, we showed that connectivity in the blog network increases over time. This corresponds to the findings in [12], which argue that densification occurs over time in social networks. We believe that the positive environmental factors created by Korean blog service providers, such as scrapping and search functions, as well as exposing popular and high quality posts on their main home pages, contribute to the enhanced connectivity in the blog network.

Fifth, we analyzed the tendencies associated with the creating and scrapping of posts in the Korean blogosphere. We analyzed the related patterns, constructing entropy plots. Based on our analytical results, we were able to prove that posting and scrapping patterns in the Korean blogosphere are more uniform than in other blogging communities. This suggests that bloggers are very active in the blogosphere, even when there are no particular social issues or events.

Finally, we analyzed the correlation between the amount of new scrapping and previous scrapping, as well as the correlation between the amount of new scrapping and accumulated in-degree. These results indicate that there is a tendency for bloggers to scrap posts from blogs that they had previously scrapped, and popular blogs receive increasing amounts of scrapping. We believe that these patterns are affected by the Korean blog service providers' search functions and the high exposure that popular posts receive.

Our findings can be useful in establishing business strategies, in developing efficient algorithms, and in making plans to vitalize the blogosphere. For example, blog service providers can utilize our results in order to better manage the connectivity of the blogosphere. Furthermore, our resilience analysis showed that many bloggers are connected to other bloggers (independently of power bloggers) due to their common interests.

By vitalizing communities comprised of these bloggers in order to increase the connectivity of the blogosphere, the resilience of the whole blogosphere could be significantly improved. Furthermore, if it were possible to identify the communities consisting of bloggers with common interests, using a resilience analysis, there could be successful target marketing campaigns aimed at these communities.

These findings also provide direction for various diverse strategies aimed at vitalizing communities. Bloggers could be made aware of other bloggers with similar common interests using a resilience analysis. In addition, popular or high quality posts could be introduced to other members or communities who would likely be interested in the content.

Our results can be used to revitalize the whole blogosphere. For example, the results of Section 8 proved that the popular bloggers consistently receive a lot of scraps. There might be new bloggers who are not widely recognized, despite producing high quality posts. If blog service providers are able to expose these bloggers via the main home page of the blog service provider, they could become popular bloggers soon. As a result, a number of popular bloggers could contribute to the vitalization of the entire blogosphere.

Using our results, blog service providers can establish a plan for hardware extensions for blogosphere maintenance. Our results showing the changes in the blogosphere and the number of posts created over time enables providers to establish effective maintenance strategies. We believe that our contributions could be the basis for successful blog-related businesses.

Acknowledgements

This work was supported by (1) Basic Science Research Program through National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (No. 2012R1A1A2007817), (2) the ICT R&D program of MSIP/IITP (14-824-09-001, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis), and (3) MSIP (Ministry of Science, ICT, and Future Planning), under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1022) supervised by the NIPA (National IT Industry Promotion Agency).

References

1. S.C. Herring, I. Kouper, J.C. Paolillo, L.A. Scheidt, M. Tyworth, P. Welsch, E. Wright, and N. Yu (2005), *Conversation in the Blogsphere: An Analysis "From the Bottom Up"*, in Proc. of Annual Hawaii Int'l Conf. on System Sciences, pp. 107b-107b.
2. R. Albert, H. Jeong, and A. Barabasi (1999), *Diameter of the World Wide Web*, Nature, vol. 401, no.6749, pp. 130-131.
3. A. Barabasi, R. Albert, and H. Jeong (2000), *Scale-free Characteristics of Random Networks: the Topology of the World-Wide-Web*, Journal of Physica A: Statistical Mechanics and its Applications, vol. 281, no. 1, pp. 69-77.
4. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener (2000), *Graph Structure in the Web*, Computer Networks, vol. 33, no. 1, pp. 309-320.
5. D. Donato, L. Laura, S. Leonardi, and S. Millozzi (2007), *The Web as a Graph: How Far We Are*, ACM Trans. on Internet Technology, vol. 7, no. 1, pp. 4.
6. D. Donato, S. Leonardi, S. Millozzi, and P. Tsaparas (2008), *Mining the Inner Structure of the Web Graph*, Journal of Physics A: Mathematical and Theoretical, vol. 41, no. 22, pp.224017.
7. A. Clauser, C. Shalizi, and M. Newman (2009), *Power-law Distribution in Empirical Data*, SIAM Reviews, vol. 51, no. 4, pp. 661-703.
8. S. Yoon, S. Kim, and S. Park (2009), *Determining the Strength of the Propensities of a Blog Network*, in Proc. IEEE Int'l Symposium on Computational Intelligence and Data Mining, pp. 140-145.
9. S. Kim, K. Kim, C. Faloutsos, J. Lee (2011), *Spectral Analysis of a Blogosphere*, in Proc. of ACM Conf. on Information and Knowledge Management, pp. 2145-2148.
10. S. Yoon, J. Shin, S. Kim, S. Park, and J. Lee (2012), *Subject-Based Extraction of a Latent Blog Community*, Information Sciences, vol. 184, no. 1, pp. 215-229.
11. J. Leskovec, M. McGlohon, C. Falooutsos, N. Glance, and M. Hurst (2007), *Patterns of Cascading Behavior in Large Blog Graphs*, in Proc. of SIAM Int'l Conf. on Data Mining, pp. 551-556.
12. J. Leskovec, J. Kleinberg, and C. Faloutsos (2007), *Graph Evolution: Densification and Shrinking Diameters*, ACM Trans. on Knowledge Discovery from Data, vol. 1, no. 1, pp. 2.
13. M. McGlohon, J. Leskovec, C. Faloutsos, M. Hurst, and N. Glance (2007), *Finding Patterns in Blog Shapes and Blog Evolution*, in Proc. of Int'l Conf. on Weblogs and Social Media.
14. L. Akoglu, M. McGlohon, and C. Faloutsos (2008), *RMT: Laws and a Recursive Generator for Weighted Time-Evolving Graphs*, in Proc. of Int'l Conf. on Data Mining, pp. 701-706.
15. M. Gotz, J. Leskovec, M. McGlohon, and C. Faloutsos (2009), *Modeling Blog Dynamics*, in Proc. of Int'l Conf. on Weblogs and Social Media, pp.18-25.

16. Y. Kwon, S. Kim, S. Park, S. Lim, and J. Lee (2009), *The Information Diffusion Model in the Blog World*, in Proc. of ACM KDD Workshop on Social Network Mining and Analysis, pp. 4.
17. Y. Kwon, S. Kim, and S. Park (2009), *An Analysis of Information Diffusion in the Blog World*, in Proc. of ACM CIKM Workshop on Complex Network in Information Knowledge Management, pp. 27-30.
18. S. Lim, S. Kim, S. Park, and J. Lee (2009), *Determining Content Power Users in a Blog Network*, in Proc. of ACM KDD Workshop on Social Network Mining and Analysis, pp. 5.
19. S. Lim, S. Kim, S. Kim, and S. Park (2011), *Construction of a Blog Network Based on Information Diffusion*, in Proc. ACM Symposium on Applied Computing, pp. 937-941.
20. S. Lim, S. Kim, S. Park, and J. Lee (2011), *Determining Content Power Users in a Blog Network: An Approach and Its Applications*, IEEE Trans. on Systems, Man, and Cybernetics PART A, vol. 41, no. 5, pp. 853-862.
21. J. Ha, S. Kim, S. Kim, C. Faloutsos, and S. Park (2014), *An Analysis on Information Diffusion through BlogCast in a Blogosphere*, Information Sciences, vol. 290, no. 1, pp.45-62.
22. S. Yoon, J. Kim, J. Ha, S. Kim, M. Ryu, and H. Choi (2014), *A Novel Approach for Link-Based Similarity Measures Using Reachability Vectors*, The Scientific World Journal, vol. 2014, pp. 1-13.
23. S. Yoon, K. Kim, S. Kim, and S. Park (2014), *Sampling in Online Social Networks*, in Proc. the ACM Symposium on Applied Computing, pp. 845-849.
24. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal (2000), *The Web as a Graph*, in Proc. of ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Data Systems, pp. 1-10.
25. I. Han and S. Lee (2007), *Graph Structure and Evolution of the Korea Web*, Journal of Korea Information Processing Society, vol. 14-D, no. 3, pp. 293-302.
26. C.R. Palmer, P.B. Gibbons, and C. Faloutsos (2002), *ANF: a Fast and Scalable Tool for Data Mining in Massive Graphs*, in Proc. ACM Int'l Conf. on Knowledge Discovery and Data Mining, pp. 81-90.
27. M. McGlohon, L. Akoglu, and C. Faloutsos (2008), *Weighted Graphs and Disconnected Components: Patterns and a Generator*, in Proc. ACM Int'l Conf. on Knowledge Discovery and Data Mining, pp. 524-532.
28. J. Leskovec (2009), *Modeling Large Social and Information Networks*, Tutorial at Int'l Conf. on Machine Learning, 2009.
29. M. Wang, T. Madhyastha, N. Chan, S. Papadimitriou, and C. Faloutsos (2002), *Data Mining Meets Performance Evaluation: Fast Algorithms for Modeling Bursty Traffic*, in Proc. IEEE Int'l Conf. on Data Engineering, pp. 507-516.