

ANALYSIS OF WEBLOG LINK STRUCTURE – A COMMUNITY PERSPECTIVE

Ying Zhou, Joseph Davis

School of Information Technology, The University of Sydney, Australia

Keywords: Weblog, community, social network, social tie.

Abstract: In this paper, we report a two level study on weblog link structures. At the micro level, we carried out an in-depth investigation of individual weblogs. Our goal was to obtain some preliminary understanding of the different types of links that might indicate underlying communities of bloggers. Complete and detailed link data was collected from eight weblogs followed by a variety of analyses. The result shows that both incoming and outgoings follow Zipf like distribution in terms of the sources of those links. These suggest clustering patterns (communities) within the whole blogspace. We also examine the temporal aspects of weblogs. The average life span of a weblog entry is fairly long in most of our sample cases. In addition, analysis on individual comment authors shows that in average, active comment authors maintain a rather long relationship with a certain weblog. It provides evidence that historical data may be useful in understanding weblog communities. On a larger scale, we developed a program to collect complete link data from large number of interconnected weblogs and performed cluster analysis on it. Communities with common topics are successfully extracted using those link data.

1 INTRODUCTION

Weblogs are web pages with several dated entries usually arranged in reverse chronological order (Kuma et al. 2004). This new form of online diary has become an influential web application, with thousands of blogs added on the web everyday. In general, blog sites cover a wide range of topics. Sites devoted to politics or technology-related topics usually receive thousands of hits per day.

Blogs usually contain a large number of links to other pages (Barabasi et al. 2001). This could be links to regular webpages or links to other blog entries. Each entry of a blog has its own “*permalink*” (permanent link). It could be an individual webpage or a section in a webpage. Blogs with similar topic are usually interwoven into a network of communities referred to as blogspace (Searls & Sifry 2003). Within a blogspace, bloggers list blogs they read and other links on the sidebar, sometimes referred to as *blogrolls*. They comment on each other’s postings, generating periods of bursty activities around interesting topics. Such bursty activities are usually exemplified by heavy linkage amongst the blogs involved with in a time interval (Kumar et al. 2003). Blogs have evolved into both link magnets and sources of links on Internet. The

result, as illustrated by Searls and Sifry (2003), is both striking and to be expected, They state “name a topic with a community of interest around it. Now go to Google and look it up. There is a good chance one or more of the top results will include somebody’s weblog (aka blog)”. The example given by them are: 802.11b, Segway and weblog. For each search term, Google listed a weblog among the top three results, which still holds at the time of writing of this paper. There are also reports that people prefer to have the latest news or development trends through blogs rather than through traditional media.

Blogspace represents a new form of online community as well as a new form of online knowledge repository connected by hyperlinks. Yet it is not an easy task for a newcomer to discover the virtual communities if they are interested in tracking particular community discussions. Google recently released a beta version of blog search services (<http://blogsearch.google.com/>). It indexes a huge collection of blog and news feeds. This search service is based on the feed data. Most query results point to individual entry of a weblog, with a few highlighted matched entire weblogs appear on the top. Beside the latest Google blog search, a few specialized blog/news search engine has also

emerged, including bloglines.com, daypop.com and blogdex.com. Each has its own copy of a collection of weblogs and can perform content related search for weblogs and news. Most of them are focused on page-level search rather than community oriented search. Hence the results are usually a mixture of regular news pages and some blog pages based on conventional static page rank algorithm. Weblogs, compared the regular webpages are more dynamic and evolves very quickly. Besides, People typically search search weblogs for a different purpose. Most of the time, they are trying to find a weblog that they can keep reading regularly. A desirable weblog search and ranking service requires different methods to model and organize the whole blog space. Understanding of individual blogs as well as the structure of linkage interaction among weblogs is important in the design of a weblog search service.

We carried out a two level study on weblog search. We first conduct a case study on a number of weblogs and investigated the feature of link component within each of them. This case study tries addresses the following questions: what kind of data should we collect to study the community of weblogs? what is the average life span of a weblog entry? what are the general interaction patterns of a weblog with other weblogs? Based on the findings of the case study we conducted a larger scale experiment to extract weblog communities by exploiting their link structure.

2 RELATED WORKS

At a macro scale there are Gruhl et al (2004)'s information diffusion study and Kumar et. Al (2003)'s bursty evolution study. Both studies draw on a large collection of weblogs over time. Gruhl et.al (2004) utilizes the epidemic model of disease-propagation to investigate the dynamics of information propagation through networks of bloggers. Their work can help to identify the hottest topics and predict the diffusion of certain piece of information in the community of bloggers. Kumar et. al adopted the notion of time graph and use that as a basis to extract temporal communities and studied bursty behavior. They argued that blog communities have striking temporal characteristics; in particular, communities only formed when an interesting topic arose and it faded away after a certain period.

On a micro scale, Judit Bar-Ilan (2004) monitored 15 sample weblogs for 2 months and generated statistics related to the blogs and posts such as, average postings per day, number of posting day in a period of time, average links per post and so

on. The majority of her sample weblogs are technology or research oriented. Her results show that the topics of most postings in the blogs were typically closely related to the declared topics. Since the data collected by Bar-Ilan were over a two-month period, it cannot give a complete picture of long running blogs and their evolution.

3 BLOG COMMUNITIES

We define a blog community as a network of weblogs with similar topics connected through hyperlinks. Within the community, bloggers may read, write and comment on other bloggers' entries. These are reflected by the three types of links indicating certain interactions between two weblogs. They are candidate links we need to consider for discovering and constructing the blogspace.

Many bloggers list a few other blogs they read on the sidebar. The list is called a blogroll. The blogroll links indicate a *read* relationship between this blogger and others.

It is very common for a blogger to cite or reference another blogger's writing on his/her own post. This indicates a response relationship between the two bloggers.

Bloggers may respond to one another's writing through other channels. They may leave a comment on the posting site. Sometimes, a commenter identifies self by a link to his or her own weblog in the comment body. This enables a crawler to build the connection between the two weblogs. In addition, some blog authoring tools provide "trackback" feature. This is a protocol that a responder can use to notify the author that he/she has cited a particular article of the author in his/her own blog. The notification is achieved by sending a ping message to the original blog entry, which will then update its trackback list to include the senders' URL. Trackback links make it very easy to get all incoming links of a particular blog entry. However, this feature is not used extensively at present.

4 METHODOLOGY

In this study, we first take a micro scale approach and closely investigated links in a few individual weblogs and the three different types of associations between blogs. The micro level is more appropriate here since we are not going to study the general structure of the whole blogspace. Rather, we are more interested in the individual weblogs, and their communication pattern with other weblogs.

The sample we traced consists of eight Weblogs. The main criterion we use to choose the sample cases is that the weblogs should have been running actively for some time to make sure there are enough data. In addition, the sample weblog should have extensive comments to ensure the activeness of the interactions. The cases chosen include two cooperate weblogs, four pure technology weblogs and two weblogs on personal opinions. These are:

1. MSN Search's Weblog (<http://blogs.msdn.com/msnsearch/>): an official weblog on MSN's search related products and discussion since November 2004.
2. Yahoo! search weblog (<http://ysearchblog.com/>): a weblog written by Yahoo! staff on Yahoo! search product since August 2004.
3. Fabulous Adventures In Coding (<http://blogs.msdn.com/ericlippert/>): a weblog written by Eric Lippert discussing all sorts of coding issues, .NET technology and a few other things. It is started on September, 2003.
4. Sorting it all out (<http://blogs.msdn.com/michkap/>): a weblog written by Michael Kaplan mainly about locales, keyboards, Unicode and other language related techniques. It is started on November 2004.
5. Micro persuasion (<http://www.micropersuasion.com/>): a weblog written by Steve Rubel on "how new technologies are transforming marketing, media and public relations". The weblog started since since April 2004. There are many articles regarding blogging news, practices and systems in micropersuasion.
6. Schneier on Security (<http://www.schneier.com/blog/>): Bruce Schneier's weblog on security and security technology since October, 2004.
7. BuzzMachine(<http://www.buzzmachine.com/>): started since July 2005, buzzmachine is a weblog written by Jeff Jarvis, currently the president and creative director of Advance.net. Jeff Jarvis is a high profile media people and long time supporter of weblogs. Buzzmachine has a wide coverage on many topics including weblogs, newspapers, open source and politics.
8. Hot Points with BOB Parsons (<http://www.bobparsons.com/>): a weblog written by Bob Parsons, the president and founder of GoDaddy.com, a company that provide Internet domain name registration, web hosting, email accounting and lots of other Internet related services. The weblog contains his thoughts and opinions on Internet innovations and lots of other things. It is started on December 2004

All the chosen weblogs have different layouts. A few specialized web crawlers were developed to extract and collect the required information from

them. The customized crawlers can accurately extract specified information, which is different from the simple heuristic based crawling algorithm used in Kumar et al. (2003) study. Information collected include blogroll links, links appeared in each entry, and comments made on each entry. Each entry's publishing date and time as well each comment's publishing date and time are also collected. The links referenced in entries are placed in two broad categories: *blog* link and *other* link. *Blog* link points to another weblog or weblog entry. We use the algorithm proposed by Ceglowski (2003) to judge if a link point to a blog entry. The data collection was carried out in October 2005. All data were current up to that month.

5 RESULTS

5.1 Basic Demographic Information

Table 1 gives a summary of the basic information based on the target weblogs. For all weblogs except Schneier, we collected all data from the very first posting. The number of occurrences for each type of interactions is given to provide an overall picture of the intensity of the communication coming in and going out of a particular weblog. All of our target weblogs have significant levels of communication with other sites or weblogs. They referenced many web pages and also attracted lots of discussions on their own sites.

Table 1: Basic demographic information.

	Blogroll Link	No. of Entries	Entry Link	Comments
MSN Search	9	77	468 (6)**	1545 (20)
YSearchblog	21	55	320 (5.9)	634 (11.7)
MichKap	0	887	4398 (4.9)	4272 (4.8)
EricLippert	12	605	889 (1.5)	4995 (8.3)
Micropersuasion	0	2548	7252 (2.8)	4672 (1.8)
Schneier*	0	86	200 (2.3)	3354 (39)
buzzmachine	1	580	4211 (7.3)	8485 (14.6)
BobParsons	0	68	197 (2.9)	5924 (87)

* Only 3 months data of this weblog is collected

** average number

5.2 Tie Strength Distribution

The blog community can be viewed as a special type of social network. A social network is a set of people or groups connected with each other under a particular relationship (Wasserman & Faust 1994). Examples of typical social networks include the friendship network of high school students, email network of employees, and scientific co-authorship network of academics (c.f. Barabasi et al. 2001, Newman 2001). In social network terminology, the people or groups are called "actors" and the connections are called "ties". In a blog community context, the bloggers who write weblogs are the actors and the hyperlinks between weblogs are the ties between two bloggers. If we consider each link as a communication instance between two bloggers, we can measure the strength of the tie according to the number of links between two particular weblogs. In this section, we will use links appearing in blog entries and links created by comment authors to study the distribution of tie strength for each particular weblog.

Figure 1 shows the tie strength distribution for the outgoing links found in weblog entry. The x-axis represents the number of times a weblog has been referenced (strength score of tie) and the y-axis represents the number of weblogs that have been references for a particular number of times (number of ties with a certain strength score). For instance, in Michkap weblog, we have 15 other weblogs being referenced by Michkap once (15 weak ties), 1 weblog being referenced by MichKap 10 times (1 strong tie with strength score 10) and so on.

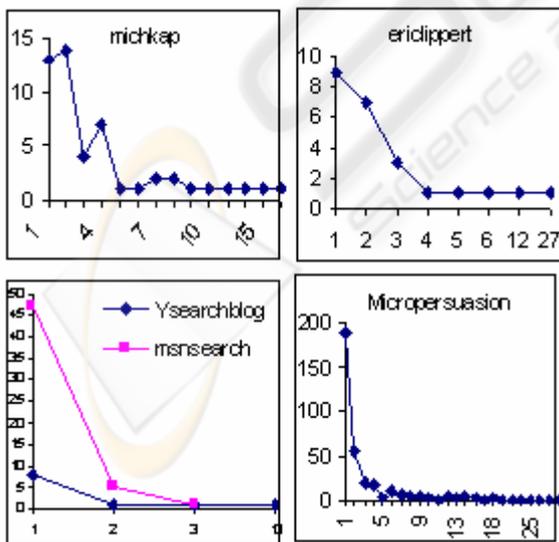


Figure 1: Outgoing tie strength distribution.

All links appearing in the blog entry of *blog* type are extracted. Those links usually point to a blog entry of another weblog. To construct the tie between two weblogs, we need to find weblog home page URL of those links. This is achieved through the RSS or ATOM feed of a weblog. The <link> value of the <channel> element in RSS usually gives the homepage of a weblog. By replacing the actual links with its home page URL we can easily count the number of links between two weblogs and discover how many strong and weak ties exist for a weblog.

Five weblogs are included in the diagram. The rest do not have enough links pointing to other blog entries. The chart on the bottom left panel contains data for two weblogs. The data from all five weblogs follows Zipf-like distribution. From the social network point of view, the observation can be interpreted as a weblog has many weak ties (being referred once or twice), and only a few strong ties (being referred many times) to other weblogs.

Similar procedures are performed on the incoming ties reconstructed from the comment section of each weblog entries. Each unique comment author is considered as an agent in a social network and one comment is considered as a communication instance between the comment author (usually a blogger) and the blogger. We use the number of comments left by the same author to measure the strength of the tie between comment author and blogger. Figure 2 shows the tie strength distribution for the incoming ties. The x-axis represents the number of comments a reader left (strength score of tie) and the y-axis represents the number of readers that have left a particular number of comments (number of ties with a certain strength score). All weblogs show clear Zipf-like distributions. All weblogs have large number of occasional visitors who only made one or two comments and a few frequent visitors who made lots of comments.

The Zipf like distribution of the tie strength is consistent with Granovetter (1983) view of social world. In his description, social world is structured with highly connected clusters (strong ties) with many external weak tie connecting these clusters. The observation implies the existence of clusters consisting of weblogs with strong ties to each other in the blogspace. These clusters are actually weblog communities of interest. It also indicates that we can discover these blog communities by measuring the link strength between weblogs.

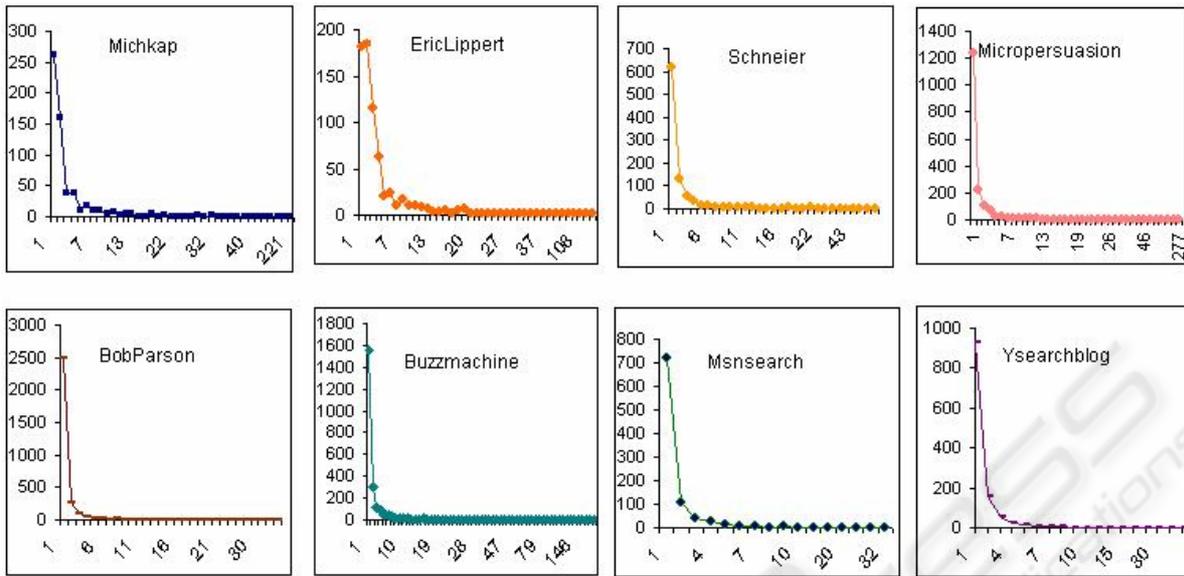


Figure 2: Incoming tie strength distribution.

5.3 Blog Entry Life Span

The study of weblog community would involve both spatial and temporal dimensions. In the previous section, we investigate the dynamics of weblog interactions and discovered that around each weblog, there is a small circle with highly frequent interaction. These would form the basis for closely knit weblog communities. Yet the questions regarding how stable these small circles might be along a weblog's life span and how frequently those communities might evolve remain unanswered.

Weblogs, nicknamed as online diary, share many features of news sites. They both focus on current events, be it a social, political event or a technical problem recently raised. They are both updated frequently and have some established readership base. This raises the problem of the value of back issues. If we are going to study the community around a particular topic, how much historical data do we need to collect to ensure the accuracy of our community information.

To develop a better understanding of the problem, we used the publishing date and time information recorded on each blog entry and comment to get some preliminary idea of the average life span of a blog entry. We used the latest time a comment was made by somebody other than the author and the blog posting time to estimate the average life span of a blog entry. The life spans of entries vary significantly with most posts having a one day life span and a few having life spans as high as nearly two years. For instance, one entry in Eric Lippert's blog (<http://blogs.msdn.com/ericlippert/>

<http://blogs.msdn.com/ericlippert/archive/2003/10/06/53150.aspx>) originally posted in November, 2003 achieved some recent discussion around two years later. We discovered quite a few posts in Eric's weblog with more than 600 days' of life span. Table 2 give the descriptive statistics of entry life span for each weblog in the sample. Although varying enormously in terms of average life span, a consistent message coming through all those cases is that historical entries are not completely ignored by blog readers. This suggests that in analyzing weblog communities, we do have to include some historical data.

Table 2: Weblog entry life span.

Unit: day	Mean	S.D	Range
MSN Search	31.4	50.1	210 (330)*
YSearchblog	60.3	86.5	343 (420)
MichKap	25.7	54.1	282 (330)
EricLippert	75.0	168.9	704 (760)
Micropersuasion	6.0	36.7	482 (550)
Schneier	7.8	9.3	47
buzzmachine	7.2	13.3	87 (110)
BobParsons	84.9	71.6	295(300)

* Weblog running days to October, 2005

5.4 Temporal Aspects

The concept of time graph and temporal community introduced in (Kumar et al. 2003) suggests certain bursty patterns along a weblog's life time. These may be generated by totally different groups of people or the same group of people. On the one hand, if bursts are generated by different groups, a natural conclusion would be that a weblog may be in

different communities from time to time. In that case, it would be difficult to measure the stability of a community and predict any future interaction among different weblogs. On the other hand, if most bursts are generated by relatively stable groups of members, this suggests that communities around a weblogs are fairly stable and it is reasonable to use historical data to predict future trends.

We took a few weblogs in our sample with relatively long life time and extract the most active five comment authors along with all instances of their communications. The purpose was to see if there was clustering pattern along the time line, that is, comment authors tend to comment with in a short duration and may never come back. Figure 3 shows the monthly communication intensity distribution of the top five comment authors in MichKap, EricLippert and Micropersuasion weblogs. The size of the area is proportional to the number of communications in that month. We do see peaks from time to time which support findings in (Kumar et al. 2003). In Figure 3, different colors represent different authors. The numbers in the legend indicate the total number of communications made by that author. A consistent pattern is that majority of authors interact with the target weblog for a relatively long period of time. Except for a few cases, many authors maintain the commenting relationship for more than one year. However, some may be more active in the early stage, for instance, Dan Shappire in EricLippert Weblog. Others may be more active in late stage, for instance, Nicolas Allan in EricLippert Weblog. Both covered a period of more than one-and-a-half year. This suggests that the communities of readership for most weblogs in our sample are quite stable for a certain period of time.

6 MACRO-SCALE APPLICATION

The above case study indicates the existence of community of bloggers and the types of link data that need to be collected to discover the weblog communities. To test those findings on a much larger scale, a web crawler was developed to collect link data from large collection of weblogs. Clustering is then applied on the data to identify the community structure.

The weblog crawler takes a weblog URL as seed and incrementally adds linked pages in the collection. Table 1 illustrates the main crawling algorithm. This weblog crawler can extract complete set of links from a weblog. This includes links on sidebar, links in blog entry and links in comment section. If any of those links points to another

weblog, a complete link set from that weblog will be extracted as well. The crawling depth is controlled by the *maxDepth* variable and is currently set to 6 to reflect the “six degrees of separation” rule (Barabasi

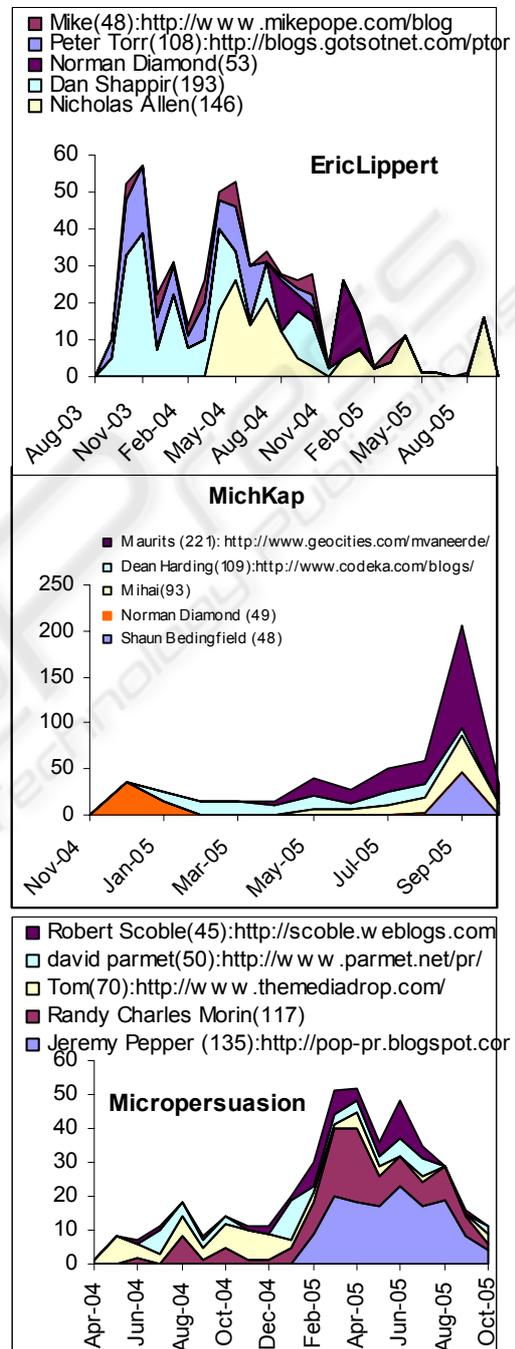


Figure 3: Communication Strength Distribution.

2002). Internet was reported to have 19 degree of separation between any individual webpages (Barabasi 2002). Since our unit of analysis is the weblog and not the individual page, and each

weblog represents a person who writes it, we think it is more appropriate to follow the six degrees of separation concept in our research.

The crawling result is a collection of relational records with two fields, the source weblog url and the target weblog url. Each record indicates a link from source to target, with a value of 1.

Table 3: Web crawler algorithm.

```

1  Url = seed url;
2  depth = 0;
3  method crawl (Url, depth)
4    if (depth < maxDepth)
5      for all hyper links link in Url
6        if link belongs to the same weblog
7          crawl (link, depth)
8        else if link is a weblog post
9          find the home page of link as link.home
10         add a record Url.home and link.home
11         crawl(link.home, depth + 1)
12       end if
13     end for
14   end if
15 end method

```

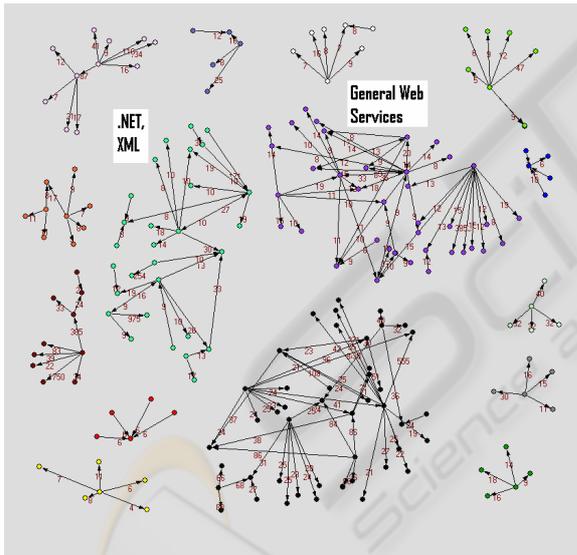


Figure 4: Web Services blog communities.

We take Savas Parastadist's weblog (Savas.parastadist.name) as seed in the experiment to run the web crawler. The key theme of Savas Parastadist's weblog is web services standards and products. The result contains around 3800 unique weblogs and over 33000 links. Multiple links between two weblogs are removed by summing up the number of links and use it as the link value between two weblogs. We then use Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) to perform clustering on the collection. The clustering

algorithm tries to discover clusters within a large graph so that each node in a cluster should have more communication with other nodes inside the cluster than with nodes outside the cluster.

Figure 4 gives a visual display of the communities extracted from the data and the structure within each cluster. In total, 15 communities are identified from the data.

We inspected members in those communities. Table 4 give a list of members in the community that contains the original seed. The topics of those member weblogs include .NET, XML and general web services. Some communities are more focused on general web services while others focused on certain web services-related technologies such as XML and .NET. The result shows that it is possible to identify communities based on complete link data of weblogs.

However, we can collect more recent data to build communities based on clustering technique as well. From the result on individual comment author, not all comment authors are active throughout the life span of a weblog. Some may be more active in the early stage, while others may be more active in late stages. The clustering algorithm can be applied on feed data rather than on entire weblog pages (as what Google BlogSearch does). A sliding time window can be used to update the blogspace with new feeds and to remove obsolete data. However, we believe that there will not be a standard window size that can fit all sorts of weblogs. Different weblogs usually have very different average life spans.

7 CONCLUSIONS

This research examined eight weblogs as special cases to study the link structure within weblogs. Two different types of links were examined: links embedded on blog entry and links created by comment authors. The result shows that most weblogs cited a wide range of other weblogs with approximate Zipf-like distributions. Many weblogs have large number of readers commenting on their writings. The distribution of commenting communication intensity also follows Zipf like curve. Majority of the readers left one or two comments on the average with a small number of readers left large number of comments. These Zipf like distributions observed with respect to entry links and comment authors suggest clustering patterns (communities) within the whole blogspace. We also examine the temporal features of weblogs. The average life span of a weblog entry is fairly long in most our sample cases. In addition, analysis of

individual comment authors shows that in average, some active comment authors maintain a rather long relationship with a certain weblog. This suggests that historical data may be useful in understanding weblog communities. To test the above findings, we developed a program to collect complete link data from large number of interconnected weblogs and performed clustering analysis on it. Communities with common topics are successfully extracted from these link data.

Table 4: Community member list.

1	http://savas.parastatidis.name
2	http://research.microsoft.com/news/msrnews/
3	http://pluralsight.com/blogs/dbox/
4	http://pluralsight.com/blogs/craig/
5	http://pluralsight.com/blogs/tewald/
6	http://pluralsight.com/blogs/aaron/ ***
7	http://pluralsight.com/blogs/keith/
8	http://pluralsight.com/blogs/fritz/
9	http://pluralsight.com/blogs/mgudgin/
10	http://msdn.microsoft.com/msdnmag/
11	http://unboxedolutions.com/sean/
12	http://glazkov.com/blog/
13	http://devauthority.com/blogs/csteen *
14	http://weblogs.asp.net/ericjsmith/
15	http://blogs.msdn.com/smguest
16	http://samgentile.com/blog/
17	http://weblogs.asp.net/rhurlbut/
18	http://jcooney.net/
19	http://blogs.msdn.com/yassers
20	http://weblogs.asp.net/cweyer/
21	http://www.innoq.com/blog/st/
22	http://www.jonfancey.com/
23	http://codebetter.com/blogs/jeffrey.palermo
24	http://blog.whatfettle.com/
25	http://blogs.msdn.com/brada
26	http://blogs.msdn.com/mpowell
27	http://weblogs.asp.net/jgaylord/
28	http://blogs.msdn.com/robcaron **
29	http://weblogs.asp.net/despos/
30	http://www.theserverside.net
31	http://blogs.msdn.com/mfussell
32	http://weblogs.asp.net/mnissen/
33	http://blogs.msdn.com/trobbins
34	http://blogs.msdn.com/tomholl
35	http://weblogs.asp.net/wallym/
36	http://dotnetjunkies.com/WebLog/barblog

REFERENCES

- Ada, E. & Zhang, L (2004) Implicit Structure and the Dynamics of Blogspace. In *Workshop on the Weblogging Ecosystem, WWW2004*, New York City.
- Adamic, A.L and Huberman, B.A. (2002), Zipf's law and the Internet. *Glottometrics* 3, 143-250
- Barabasi, A.L., Jeong, H., Neda, Z., Ravasz E., Schubert A., & Vicsek T. (2001, April), Evolution of the social network of scientific collaborations. *arXiv:cond-mat/0104162* v1. 10.
- Barabasi A.L. (2002) Linked: the new science of networks. Perseus Books Group
- Bar-Ilan Judit. An outsider's view on "topic-oriented" blogging. *WWW2004*, May 2004, New York, USA
- M.Ceglowski. (2003) Wwww::identify-identify blogging tools based on url and content. Retrieved from <http://search.cpan.org/~mceglows/WWW-Blog-Identify-0.06/Identify.pm>,
- Downes. S. (2003 July/August) Web logs at Harvard Law. *The Technology source*. Retrieved from <http://ts.mivu.org/default.asp?show=article&id=2019>
- Flake G.W., Lawrence S, & Giles C.L. (2000) Efficient identification of web communities. In *Proc. 6th ACM SIGKDD Intel. Conf. On Knowledge Discovery and Data Mining*, page 150-160
- Granovetter, M. (1983) The Strength of weak ties:a network theory revisited. *Sociological Theory*, Vol. 1, 201-233
- Gruhl, D., Guha, R., Liben-Novell D. & Tomkins A. (2004 May) Information diffusion through blogspace. *WWW2004*, , New York, USA
- Kumar, R. Novak, J., Raghavan, P. & Tomkins, A. On the bursty evolution of blogspace. (2003 May) *WWW2003*, Budapest, Hungary
- Kumar, R., Novak J., Raghavan, P., & Tomkins, A. (2004 Dec) Structure and evolution of blogspace, *Communications of the ACM*, Vol.47/No.12
- Nanno, T., Fujiki, T., Suzuki, Y. & Okumura, M. (2004 May) Automatically collecting, monitoring and mining Japanese Weblogs. *WWW2004*, New York City
- Newman, M.E.J. (2001) Scientific collaboration networks. I. Network construction and fundamental results, *Physical Review* Vol. 64, 016131,
- Rubel, S. (2004, Oct) Finding Influential Blogs That Reach Your Key Audiences, Retrieved from http://www.micropersuasion.com/2004/10/finding_influen.html
- Searls, D. & Sifry, D. (2003 Mar), Building with Blogs, *Linux Journal*, Issue 107
- Tedeschi, B. (2005 Jul. 4) Blogging while browsing, but not buying", *The New York Times*.
- Wasserman, S. & Faust, K. (1994) *Social Network Analysis*, Cambridge University Press, Cambridge
- Winer, D. (2002 May) History of weblogs, Retrieved from <http://newhome.weblogs.com/historyOfWeblogs>.