MINING COMMUNITIES OF BLOGGERS: A CASE STUDY ON CYBER-HATE

Jennifer Xu
Bentley College
Waltham, MA, USA
jxu@bentley.edu

Michael Chau
University of Hong Kong
Pokfulam, Hong Kong
mchau@business.hku.hk

Abstract

Blogs, or Weblogs, have become increasingly popular in recent years. Research has found that racists and hate groups exist in communities of bloggers. As these communities allow hate groups to spread their ideologies and even advocate hate crimes, it is important to study the structure and behavior of these communities. In this study, we analyzed the blogs of 28 anti-black hate groups on Xanga, a popular blog hosting site, using a semi-automated framework that includes blog spidering, information extraction, network analysis, and visualization. Our findings suggested that bloggers formed communities through subscription, comment, and group co-membership relationships. Subscription and commenting relationships facilitated the communication between bloggers and could help spread information, propagandas, and ideologies faster. In addition, we compared our findings with previous studies and found some interesting similarities and differences. Overall, we believe our research on online hate groups in the blogosphere is timely and important to the security of our society, and several future research directions are suggested in the paper.

Keywords: mining, data mining, network topology, data and knowledge visualization, security

Introduction

A blog is a Web-based journal often referred to as online personal diary. Blog tools combine personal Web site features for authors to publish their thoughts, add and edit content, and link to other pages, with features of Internet forums for readers to post comments and afterthoughts (Blood 2004). Because blogs basically allow anyone to express his/her ideas and thoughts, bloggers consider blogging as an outlet for their thoughts and emotions. As blogs have become increasingly popular, communities of racists, extremists, and hate groups have also emerged in the blogosphere (Franklin 2006; Chau and Xu, forthcoming). The consequences of the formation of such communities on the Internet cannot be underestimated. Young people, the major group of bloggers, are more likely to be affected and even "brainwashed" by ideas propagated through the Web as a global medium. Hatred, white supremacy, and extremist ideas could easily be embedded into their minds to make them become members of these hate communities or even conduct hate crimes. Therefore, it is important to study the phenomenon and the nature of these communities.

This paper reports our ongoing research and an extension of our previous analysis on hate groups in blogs (Chau and Xu, forthcoming). In our previous work, we studied the communities of hatred-related blogs and the relationships between bloggers. We proposed a framework for semi-automatic blog analysis that comprises blog spider, information extraction, social network analysis, and visualization. Our analysis results revealed some interesting demographical and topological characteristics. However, only two types of relationships, namely subscription and co-membership relationships (which will be discussed later in this article), were used in our previous study. One of the most important features in blogs – commenting – was not included. Commenting is the ability for any reader to write a comment on a blog entry. The ability to comment on blogs has facilitated the interaction between bloggers and readers. In the current study, we aim to study the linkage among these bloggers and the structure of the

communities they formed based on all three types of relationships. One of our main objectives is to ascertain whether the incorporation of comment relationship has an impact on the structure of the network formed.

The rest of the paper is organized as follows. Section 2 reviews the research background of hate group analysis and related research in Web mining and social network analysis. In Section 3, we pose our research questions. Section 4 presents our methodology of blogger network analysis. In Section 5 we present the study that we conducted on Xanga, a popular blog hosting site, and discuss our findings. We conclude our research and suggest some future research directions in Section 6.

Related Work

In this section, we review the basic characteristics of blogging and the blogosphere, the network analysis approach, and activities of racists and hate groups on the Internet.

The increasing number of blogs, connected by relationships such as reading, comments, hyperlinks, subscriptions, forms a network often known as the blogosphere (CACM 2004). Smaller cyber communities have also emerged in blogs. Communities in blogs can be categorized as explicit communities or implicit communities, like some other cyber communities on the Web (Kumar et al. 1999). Explicit communities in blogs are called blogrings that bloggers have explicitly formed and joined. Most blog hosting sites allow bloggers to form a new group or join any existing groups. On the other hand, implicit communities are not explicitly defined as groups or blogrings by bloggers. Instead, these communities are formed by interactions among bloggers, such as subscription, linking, or commenting.

Traditionally the identification of communities on the Web has to a large extent relied on Web structure mining (Gibson et al. 1998; Kumar et al. 1999). Many Web community identification methods are rooted in the HITS algorithm (Kleinberg 1998). Web structure mining is also closely related to social network analysis (SNA). SNA is a sociological methodology for analyzing patterns of relationships and interactions between social actors in order to discover the underlying social structure (Wasserman and Faust 1994). When used to mine a network, SNA can help reveal structural patterns such as the central nodes that act as hubs, leaders, or gatekeepers; densely knit communities and groups; and patterns of interactions between the communities and groups. These patterns often have important implications for the functions of the network.

Moreover, recent advances in the statistical analysis of network topology (Albert and Barabási 2002) have brought new insights and research methodology to the study of network structure. Three models have been proposed to characterize the topologies of empirical networks, namely, *random model* (Bollobás 1985), *small-world model* (Watts and Strogatz 1998), and *scale-free model* (Barabási and Albert 1999). In a random network, each node has roughly the same number of links, and communities are not likely to exist. Small-world networks, in contrast, have a significantly high tendency to form groups and communities. In scale-free networks, a large percentage of nodes have just a few links, while a small percentage of the nodes are "central" and have a large number of links.

Hate groups have been increasingly using the Internet to express their ideas, spread their beliefs, and recruit new members (Lee and Leets 2002). Several studies have investigated Web sites that are related to racism or white supremacy (Burris et al. 2000; Douglas et al. 2005; Gerstenfeld et al. 2003; Zhou et al. 2005). For example, Gerstenfeld et al. (2003) conducted a manual analysis of 157 extremist Web sites and found that some hate Web sites were associated with hate groups while others were maintained by individuals. Many of these sites had links to other extremist sites or hate group sites, showing that some of these groups are linked to each other. Burris et al. (2000) systematically analyzed the networks of Web sites maintained by white supremacist groups and found that this network had a decentralized structure with several centers of influence. In addition, communities were present in this network in which groups sharing similar interests and ideologies tended to be closely connected. Recent years have seen the emergence of hate groups in blogs with a lot of high-narrative messages. This has made blogs an ideal medium for spreading hatred. Blogs have also made it possible for individuals to find others with similar beliefs and ideologies much more easily. As a result, hate groups have emerged in blogs. Content analysis and network analysis techniques are needed to analyze hate-group blogs in order to identify patterns and facilitate further analysis.

Research Questions

Based on our previous work, we pose the following research questions in the current study: (1) Do the networks of different types of relationships (subscription, comment, and co-membership networks) display different structural properties? (2) Are there bloggers who stand out as leaders in these groups by spreading their ideas through commenting on others' blogs, or attracting a lot of readers and comments to their blogs? (3) When incorporating the comment relationship, do communities emerge? Do these communities overlap with blogrings of hate groups? How do these communities look like and what implications do they have? Answering these questions will help us understand the social structure of the communities and the social movement of online hate groups.

Methodology

In this section, we present our methodology for collecting blogs, analyzing their relationships, and presenting analysis results. Our methodology is based on a semi-automated framework that we developed in our previous research (Chau and Xu, forthcoming). Figure 1 presents the framework, which consists of four main modules, namely *Blog Spider*, *Information Extraction*, *Network Analysis*, and *Visualization*.

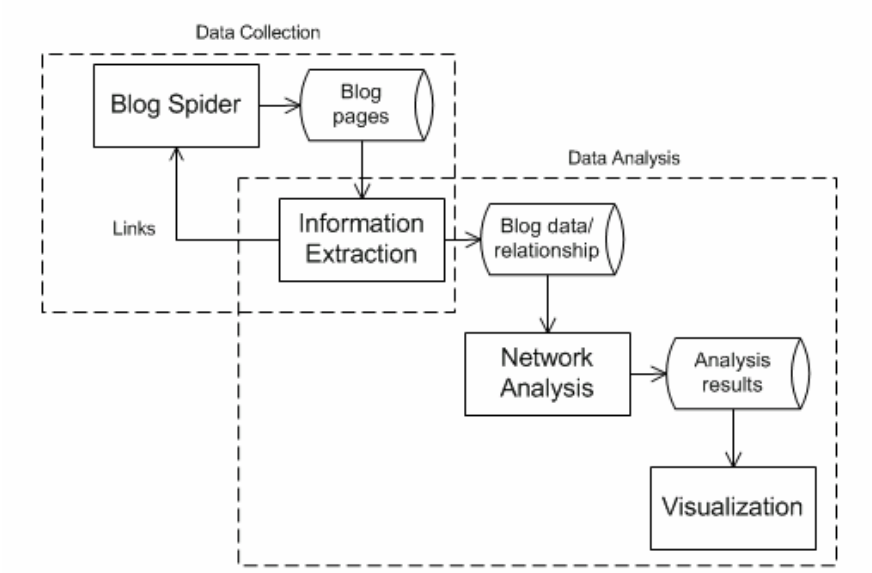


Figure 1. The Framework for Blog Collection and Analysis

The blog spider is designed to download the relevant pages from the blogs of interest in a way similar to general Web fetching. However, instead of following all extracted links, the blog spider should only follow links that are of interest, e.g. links to a group's members or other bloggers. In addition, the spider can use RSS (Really Simple Syndication) and get notification when a blog is updated. However, this is only necessary when monitoring or incremental analysis is desired and is not used in our current study.

After a blog page has been downloaded, it is processed in order to extract useful information from the page. This includes information related to the blog or the blogger such as user profiles and date of creation. This can also include relationship information between two bloggers, such as linkage, commenting, or subscription. Because different blogs may have different formats, it is not a trivial task to extract such information from blogs. Fortunately, some standard information such as name and location are oftentimes put into specific format (e.g. as a sidebar) in large blog hosting sites, and simple rules should suffice. In the current study, a pattern matching approach is employed.

The major module in our framework is network analysis including *topological analysis*, *centrality analysis*, and *community analysis*. We use four statistics that are widely used in topological studies to categorize the extracted network (Albert and Barabási 2002; Crucitti et al. 2003): *average shortest path length*, *efficiency, clustering coefficient*, and *degree distribution*. Average path length is the mean of all-pair shortest paths in a network.

Efficiency is defined as the average of the inverses of shortest path lengths over all pairs of nodes in a network (Crucitti et al. 2003). The most efficient network is a fully connected network and is often called a clique with efficiency value of 1. Clustering coefficient measures how likely nodes in a network form communities. The degree distribution, $p(k)$, is the probability that a node has exactly k links. A random network usually has a small average path length and is more efficient because an arbitrary node can reach any other node in a few steps. Small-world networks usually have significantly higher clustering coefficients than their random network counterparts of equal size. The degree distributions of random networks are bell-shaped Poisson distributions. However, scale-free networks are categorized by power-law degree distributions, which have long flat tails (Barabási and Albert 1999).

The goal of centrality analysis is to identify the key nodes in a network. Three traditional centrality measures can be used: *degree*, *betweenness*, and *closeness* (Freeman 1979). Degree measures how active a particular node is. It is defined as the number of links a node has. In a directed network,the in-degree refers to the number of in-links a node has and the out-degree refers to the number of out-links. Betweenness measures the extent to which a particular node lies between other nodes in a network. The betweenness of a node is defined as the number of geodesics (shortest paths between two nodes) passing through it. Nodes with high betweenness scores often serve as gatekeepers and brokers between different communities. They are important communication channels through which information, goods, and other resources are transmitted or exchanged (Wasserman and Faust 1994). Closeness is the sum of the length of geodesics between a particular node and all the other nodes in a network. A node with low closeness may find it very difficult to communicate with other nodes in the network.

Community analysis is to identify social groups in a network. In SNA a subset of nodes in an unweighted network is considered a community or a social group if nodes in this group have denser links with nodes within the group than with nodes outside of the group (Wasserman and Faust 1994). An unweighted network can be partitioned into groups by maximizing within-group link density while minimizing between-group link density. In this case, groups are densely-knit subsets of the network. Note that community and groups here do not refer to the explicit groups (blogrings). They refer to a subset of nodes that form implicit clusters through various relationships. In these communities, members subscribe to or post comments to each other's blogs frequently even though they may not belong to the same blogrings.

The extracted network and analysis results can be visualized using various types of network layout methods. Two examples are *multidimensional scaling* (MDS) (Kruskal and Wish 1978) and *graph layout* approaches (e.g. Davidson and Harel 1996).

The major factor that may affect the scalability of this framework lies in the community analysis part in the network analysis component. As community analysis relies on the clustering of network nodes, the low scalability of the clustering algorithm selected may become the bottleneck of the framework. Fortunately, some very efficient clustering techniques have been developed (e.g. Flake et al. 2000) that can help resolve this problem.

Analyses and Results

Dataset

Following our previous study, we chose to study the hate communities against blacks and limited our data to the blogs on Xanga (www.xanga.com). According to statistics provided by Alexa (2005), Xanga is the second most popular blog hosting site, only after the Google-owned Blogger (www.blogger.com). We used a set of keywords, such as "KKK", "niggers", "white pride", to search for groups (blogrings) on Xanga that have any of these words in their group name or description. We then manually checked these groups and filtered out those not related to anti-black, resulting in a list of 28 groups.

We then sent spiders to download the description page and member list of each of these groups, obtaining a list of 925 bloggers. The extraction program was executed to extract some basic information of each blogger, including user ID, real name, date of birth, city, state, country, and date of registration. The date of registration is the date when a blogger first registered to the blog hosting site (Xanga) and created his/her blog.

Three types of relationships between bloggers were also extracted: (a) subscription relationship, which occurs when one blogger subscribes to another blog; (b) comment relationship, which is established when one blogger makes a

comment on another one's blog; and (c) group co-membership, which exists between two bloggers if they belong to the same blogring.

Results

Bloggers in Hate Groups

In our dataset, 700 bloggers indicated their genders, 600 bloggers provided their dates of birth, and 573 bloggers reported their countries. We found that 64.6% of these bloggers are male and 35.4% are female; and 76.7% of them are between 15 and 25 years old. The majority of the bloggers are from the United States (82.5%), and the rest are from Germany (2.8%) and 42 other countries (19.3% in total).

We recognize that the use of self-reported demographical data has a problem — low reliability. A blogger may hide his/her demographics or even provide misleading information. Nonetheless, as blogs are considered personal diary, many bloggers choose to release partially true information. The statistics thus can provide a rough (although not accurate) picture of the sample of bloggers in hate groups. It shows that young males from the United States are dominant bloggers in black-hatred groups. In general, youths who were born in 1980s and early 1990s are the primary users of blogs. We see that some of them are also actively participating in hate groups. Blogs offer them an ideal platform to learn the ideologies of extremisms and hate groups, express and disseminate their adopted believes, and find supporters, collaborators, and friends who share the similar anti-black views.

We also analyzed the growth pattern of the hate group blogs over the years. Because the server on the blog hosting site (Xanga) records the time when each blogger registers to and joins a hate group, the data about bloggers' dates of registration are rather accurate and reliable. We found that there have been an increasing number of people joining these anti-black blogrings since their inception in 2002. This implies that hate groups have been gaining popularity in blogosphere over years as more and more bloggers joined these groups.

Structural Properties of Blogger Networks

We performed topological analysis on three different networks among bloggers: (a) network of subscriptions, (b) network of comments, and (c) network of all three types of relationships. We did not consider the network of only group co-membership because the resulting network was rather dense and consisted of several fully-connected cliques. Each clique corresponded to a blogring, in which nodes were connected with all other nodes in the same clique. Such a network is less valuable than the other three networks from the structural analysis perspective because nodes are almost indistinguishable in their structural roles. The same problem existed when we constructed network (c). We thus included in this network only co-membership links whose weights were greater than one. That is, we considered a co-membership link between two bloggers a valid link only if they shared memberships of at least two common groups.

When analyzing the topological properties of these networks, we ignored the direction and weight of relationships. As a result, there could be at most one link between a pair of nodes. The resulting networks do not include all the 925 bloggers. Around 38% the bloggers are left out. These isolated bloggers have not subscribed to or received subscription from any other blogger; and they have never commented on and received any comment from others. They have joined only one blogring (thus their co-membership relationship weights were less than two).

In addition, these networks are not connected graphs. They consist of several disjoint components, between which no link exists. In each network, there are several very tiny components with only 2 or 3 nodes and a single large connected component. This largest connected component is often called giant component in graph theory (Bollobás 1985). Table 1 reports the statistics of the three networks regarding their sizes, giant component sizes, and the average and maximum degrees in the giant components.

Table 1. The Statistics of the Three Networks

Network	(a) Subscription	(b) Comment	(c) All Relation Types
Number of nodes, *n*	351	490	567
Number of links, *m*	566	1048	2034
n (Giant Component)	247	452	532
m (Giant Component)	489	1020	2014
Average Degree (Giant Component)	3.96	4.51	7.57
Max Degree (Giant Component)	32	76	87

We studied topological properties (average shortest path length, efficiency, and clustering coefficient) of the giant component in each network. The results are reported in the "Data" columns in Table 2. To compare the topology of a network with its random graph counterparts, we generated 30 random networks consisting of the same number of nodes (e.g. 247 for the subscription network) and number of links (e.g. 489 for the subscription network) with those in the giant component. The averages and standard deviations for these measures are listed in the "Random" columns labeled in Table 2.

Table 2 shows that networks (a) and (b) are roughly as efficient as their random graph counterparts. Network (c) is significantly less efficient than its random graph counterparts (3.72 is greater than the upper bound of the 97% confidence interval for the mean of the random counterparts). However, although the difference is statistically significant, the value of the difference is less than half of a step. It takes a blogger about only four steps (three intermediate bloggers) to reach another arbitrary blogger in the networks. Therefore propagandas, beliefs, opinions, and ideas can be quickly distributed among the bloggers through subscription and comment links, making it easier for bloggers to influence one another. Moreover, these networks all have significantly higher clustering coefficients, which are at least 10 times more than those of their random graph counterparts. This implies that these networks are small worlds, in which densely-knit communities are very likely to exist.

Table 2. The Topological Properties of the Giant Components in the Three Networks.

Network	Average Path Length		Efficiency		Clustering Coefficient	
	Data	Random	Data	Random	Data	Random
(a) Subscription	4.09	3.96 (0.150)	0.28	0.26 (0.007)	0.14	0.013 (0.005)
(b) Comment	4.07	3.96 (0.073)	0.28	0.27 (0.004)	0.15	0.009 (0.002)
(c) All Relation Types	3.72	3.33 (0.037)	0.30	0.32 (0.003)	0.27	0.015 (0.002)

The values in the "Random" columns are averages over 30 generated networks. Numbers in parentheses are standard deviations.

In terms of degree distribution, all these networks displayed scale-free topology. We plotted the cumulative degree distributions of these networks on Figure 2. The cumulative degree distribution, P(*k*), is defined as the probability that an arbitrary node in the network has at least *k* links (Amaral et al. 2000). If the degree distribution follows a power law, the cumulative distribution should also follow a power law, appearing as a straight line on a scatter plot in logarithmic scales (Amaral et al. 2000). The curves for networks (a) and (b) are roughly straight, with exponents of 1.60 and 1.50, respectively. However, the network (c) shows a bump in the curve, displaying characters of two-regime power-law distribution (Barabási et al. 2002). In general, a two-regime power law distribution may result from the new links that are formed between existing nodes when the network grows (Barabási et al. 2002). In our network, these internal links are the group co-memberships added to the subscription and comment networks. That is, group co-memberships connect bloggers who do not have any subscription and comment relationships.

To summarize, these three networks can be categorized as small-world networks because they all have short average distances and a high tendency for forming clusters. In addition, they also have scale-free properties with power-law

degree distributions. These findings are consistent with those for many other empirical networks that display both small-world and scale-free properties (Albert and Barabási 2002).

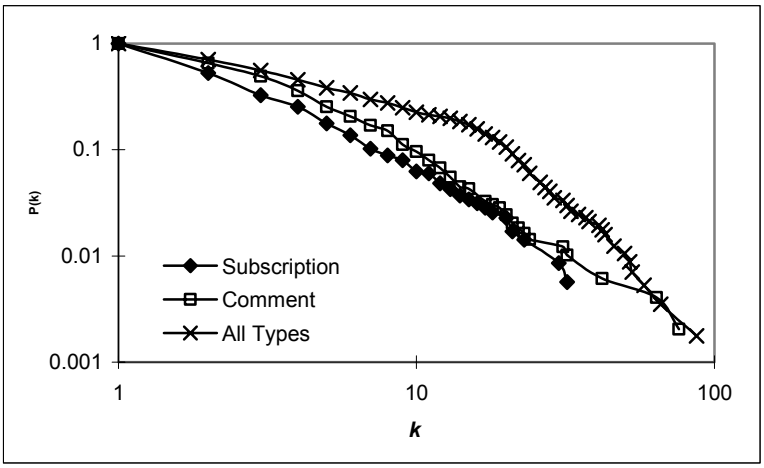
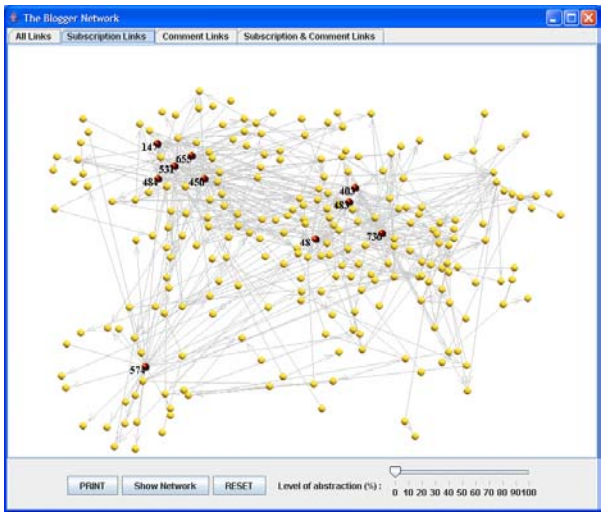


Figure 2. Cumulative Degree Distributions of the Three Networks.

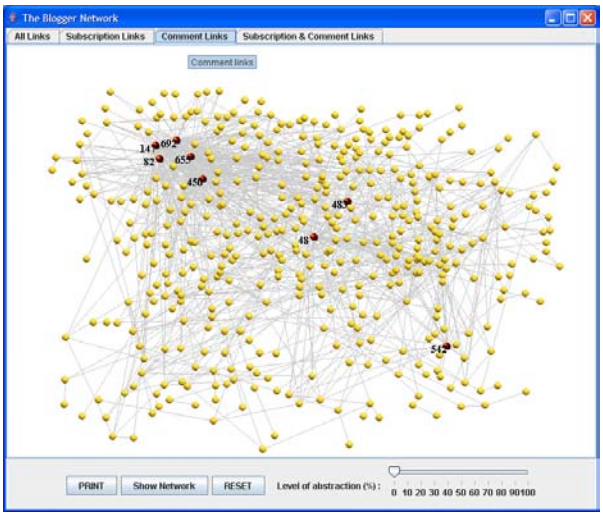
Key Bloggers and Communities

We used a prototype system we developed (Xu and Chen 2005) to find the central bloggers and to identify the communities. In order not to confuse between explicit and implicit communities, we use community to refer to the implicit community found by the system and blogring to refer to the explicit community. The three networks are shown in Figure 3, in which a node represents a blogger and a straight line represents a link. In the subscription network, some links are directed with the heads pointing to the bloggers whose blogs receive subscriptions from others. An undirected link implies a mutual subscription. The layout of the network was determined using the MDS method. We fixed the positions of the nodes across the layouts of the networks to make the comparison of networks easier.

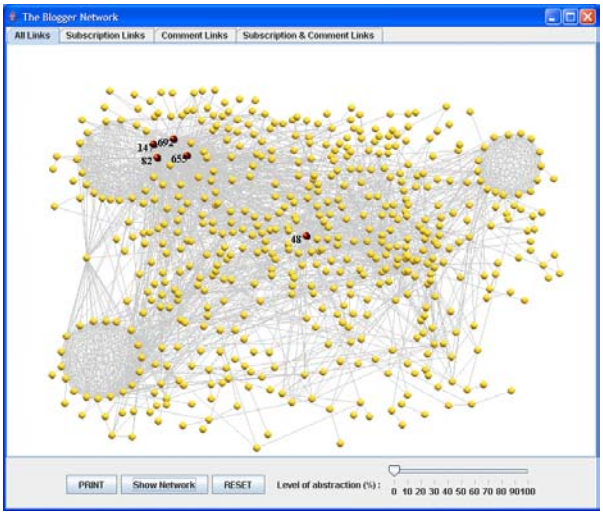
In Figure 3, bloggers who have the highest degrees (in-degrees and out-degrees in the subscription network and comment network) are highlighted and labeled with their system generated ID numbers. We did not show the real login names of these bloggers to protect their privacy. These bloggers are those who may participate in multiple blogrings or have many subscription or comment relationships with other bloggers.



(a)



(b)



(c)

Figure 3. The Giant Components of the Three Networks: (a) Subscription Network, (b) Comment Network, and (c) Network of All Types of Relationships. The highlighted nodes are those who have large degrees.

In general, it is worth a closer look at the bloggers with high in-degrees in the subscription or comment networks. A content analysis on their blog entries may reveal whether they actually are popular leaders who express extreme beliefs, opinions, and ideologies. Because such information may easily be spread among subscribers and readers, these blogs need to be closely monitored. On the other hand, the hubs with a high out-degree should also be paid attention to since they may be intermediate channels pointing to the leaders' blogs.

We also performed community analysis[1] and found two communities in network (a), two communities in network (b), and three communities in network (c). Bloggers interact frequently with each other in their communities, where they read other community members' blogs, share their ideas and opinions, and sometimes make comments on these thoughts. They may also interact with bloggers from other communities, passing information to other communities or bringing new information from outside. Communities found in these three networks are different from each other. The two communities in network (a) differ greatly in size with the bigger community consisting of more than 80% of the nodes in the network and the other less than 20%. The network (b) is divided into two relatively even communities of similar sizes. The three communities in network (c) are clearly formed around the three densely-knit cliques that resulted from group co-membership. The two communities found in network (b) can be roughly mapped to two communities in this network. Comparing cross the communities from the three networks, we found that two bloggers being in the same blogrings (as represented by the cliques) may not necessarily lead to more interactions (subscription and comment) between them.

Discussion

In this section we discuss our findings and answer the research questions regarding the topological properties, key bloggers, and community structures in networks of bloggers in online hate groups. Specifically, we extend the findings from our previous study which was based only on subscription and co-membership relationships.

- The topologies of blogger networks: *Do the networks of different types of relationships display different structural properties*? In this study, we analyzed three types of relationships: subscription, comment, and group co-membership. As mentioned previously, unlike the subscription network and the comment network, the network consisting of only group co-membership links contained several fully connected cliques, which corresponded to blogrings. Such a structure provided little information from the structural analysis point of view. Comparing the networks of subscription, comments, and all three types of relationships, we found that these networks are very similar as they are roughly equally efficient in information diffusion and

[1] Communities found are not shown in the figure.

communication, and they all have high tendency to form communities in the networks. The major difference among the networks based on these relationships lies in the degree distributions. It remains an open question whether group co-memberships actually facilitate communication and information diffusion among bloggers as the subscription and comment relationships do.

- Key bloggers in online hate groups: *Are there bloggers who stand out as leaders in these groups by spreading their ideas through commenting on others' blogs, or attracting a lot of readers and comments to their blogs*? Our findings suggest there are some bloggers who are connected with many others through subscription, comment, and group co-membership links. These bloggers may be either leaders of opinions and ideologies or hubs of communication. Both types of center require closer analysis and examination. For example, we found that bloggers who receive many subscription links may not necessarily be the ones who receive many comment links. Only two bloggers were identified in both the top five in-degree list for network (a) and the top five in-degree list for network (b). No overlap was found between the top five out-degree lists for these two networks. This means that bloggers who subscribe to many others' blogs may not comment frequently on others' blogs.
- Community structure in blogger networks:
  - *Do communities emerge in these networks after comment relationships are incorporated*? We found densely-knit clusters of bloggers in all networks under study. Inside these communities, bloggers frequently interact with each other. The incorporation of comment links clearly helped form the communities of bloggers sharing similar interests.
  - *Do these communities overlap with blogrings of hate groups*? It is interesting that the communities formed based on bloggers' interactions (subscription and comment) do not necessarily overlap with blogrings. Bloggers have their own "cliques" in which they have their own interests, beliefs, and even leaders. What is more meaningful, we think, is these implicit communities, since they are the environment in which bloggers actually exchange and spread their racist and extremist messages.

Conclusions and Future Directions

In this paper, we report our research in progress on the phenomenon of hate in blogs and the communities formed by these bloggers. Using the semi-automated framework, we investigated the anti-black bloggers on a popular blog host and applied network analysis techniques to study their network structure. Specifically, we extended our previous study by incorporating a very important type of relationship between bloggers — commenting. Overall, the findings are interesting and some previously unattended patterns have been revealed.

Our research has several contributions. First, our framework consists of a set of blog collection and network analysis techniques that can be applied to the study of blogosphere. With the increasing popularity of blogs, we believe that the framework can be used for other applications (such as cybercrime analysis) or even other domains (such as marketing analysis). Second, we found that comment relations in blogs are an important feature for network analysis. Our results showed that including comment relations does have a significant impact on the blogger network. Third, we believe our research is timely and important to the security of society by studying the activities of online hate groups in the blogosphere. Law enforcement and other intelligence agencies can use the methodology to monitor the activities of hate groups in blogs, identify the key bloggers, and analyze the changes in the communities. Other organizations such as the Hate Directory (Franklin 2006) can also use the methodology to periodically compile lists of hate blogs to warn the public of the online hate groups.

We have several directions for our future research. First, the current study only investigated the hate group activities on one single blog site, Xanga. In the future we will include other blog hosts for a more comprehensive study. Second, we plan to apply more content analysis and text mining techniques on blogs.

Acknowledgements

We thank Porsche Lam from the University of Hong Kong for their help in preprocessing the data.

References

Albert, R. and Barabási, A.-L. "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* (74:1) 2002, pp. 47-97.

Alexa. Top English language sites. [Online] Retrieved from http://www.alexa.com/site/ds/top_sites?ts_mode=langandlang=en on October 7, 2005.

Amaral, L.A.N., Scala, A., Barthelemy, M., and Stanley, H.E. "Classes of Small-World Networks," *Proceedings of the National Academy of Science of the United States of America* (97:21) 2000, pp. 11149-11152.

Barabási, A.-L. and Albert, R. "Emergence of Scaling in Random Networks," *Science* (286) 1999, pp. 509-512.

Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. "Evolution of the Social Network of Scientific Collaborations," *Physica* (311) 2002, pp. 590-614.

Blood, R. "How Blogging Software Reshapes the Online Community," *Communications of the ACM* (47:12) 2004, pp. 53-55.

Bollobás, B. *Random Graphs*, London, Academic, 1985.

Burris, V., Smith, E., and Strahm, A. "White Supremacist Networks on the Internet," *Sociological Focus* (33:2) 2000, pp. 215-235.

CACM, *Communications of the ACM, Special Issue on "The Blogosphere"*, (47:12), 2004.

Chau, M. and Xu, J. J. "Mining Communities and Their Relationships in Blogs: A Study of Online Hate Groups," *International Journal of Human Computer Studies*, forthcoming.

Crucitti, P., Latora, V., Marchiori, M., and Rapisarda A. "Efficiency of Scale-Free Networks: Error and Attack Tolerance," *Physica A* (320) 2003, pp. 622-642.

Davidson, R. and Harel, D. "Drawing Graphs Nicely Using Simulated Annealing," *ACM Transactions on Graphics* (15:4) 1996, pp. 301-331.

Douglas, K. M., McGarty, C., Bliuc, A., and Lala, G. "Understanding Cyberhate: Social Competition and Social Creativity in Online White Supremacist Groups," *Social Science Computer Review* (23:1) 2005, pp. 68-76.

Flake, G. W., Lawrence, S. and Giles, C. L. "Efficient Identification of Web Communities," in *Proceeding of the 6th International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD 2000),* ACM, Boston, MA, 2000, pp. 150-160.

Franklin, R. A. "The Hate Directory." [Online] Retrieved from http://www.bcpl.net/~rfrankli/hatedir.htm on April 28, 2006.

Freeman, L. C. "Centrality in Social Networks: Conceptual Clarification," *Social Networks* (1) 1979, pp. 215-240.

Gerstenfeld, P. B., Grant, D. R., and Chiang, C. P. "Hate Online: A Content Analysis of Extremist Internet Sites," *Analyses of Social Issues and Public Policy* (3) 2003, pp. 29-44.

Gibson, D., J. Kleinberg, and Raghavan, P. "Inferring Web Communities from Link Topology," in *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*, Pittsburgh, PA, 1998.

Glaser, J., Dixit, J., and Green, D. P. "Studying Hate Crime with the Internet: What Makes Racists Advocate Racial Violence?" *Journal of Social Issues* (58:1) 2002, pp. 177-193.

Kleinberg, J. "Authoritative Sources in a Hyperlinked Environment," in *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, California, January 1998, pp. 668-677.

Kruskal, J. B. and Wish, M. *Multidimensional Scaling*. Beverly Hills, CA, Sage Publications, 1978.

Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. "Trawling the Web for Emerging Cyber-Communities," *Computer Networks* (31:11-16) 1999, pp. 1481-1493.

Lee, E., and Leets, L. "Persuasive Storytelling by Hate Groups Online: Examining its Effects on Adolescents," *American Behavioral Scientist* (45) 2002, pp. 927-957.

Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge, Cambridge University Press, 1994.

Watts, D. J. and Strogatz, S. H. "Collective Dynamics of 'Small-World' Networks," *Nature* (393) 1998, pp. 440-442.

Xu, J. J. and Chen, H. CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery," *ACM Transactions on Information Systems* (23:2) 2005, pp. 201-226.

Zhou, Y., Reid, E., Qin, J., Chen, H., and Lai, G. US Domestic Extremist Groups on the Web: Link and Content Analysis," *IEEE Intelligent Systems* (20:5) 2005, pp. 44-51.