U.S. Domestic Extremist Groups on the Web: Link and Content Analysis

Yilu Zhou[1] , Edna Reid[1], Jialun Qin[1], Hsinchun Chen[1] and Guanpi Lai[2]

[1]*Artificial Intelligence Lab, Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721*

{yiluz, ednareid, qin, hchen}@eller.arizona.edu

[2]*Department of Systems and Industry Engineering, The University of Arizona, Tucson, AZ 85721, USA*

guanpi@email.arizona.edu

Abstract

U.S. Domestic extremist groups have increased in numbers and are intensively utilizing the Internet as an effective tool to share resources and members with limited regard for geographic, legal, or other obstacles. Researchers find that monitoring extremist and hate groups' Web sites, and analyzing their usage and content have become time consuming and challenging. In response, this study describes the development of automated or semi-automated methodologies for capturing, classifying, and organizing domestic extremist Web site data and using them for analysis. We found that by analyzing the hyperlink structures and content of domestic extremist Web sites and constructing social network maps, their inter-organizational structure and cluster affinities could be identified. Such analysis results could help experts in terrorism, law-reinforcement, intelligence, and policy-making domains better understand the domestic extremist phenomena and eventually boost our national security.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications - *Data and knowledge visualization, Text mining, Web mining*

H.4.2 [**Information Technology and Systems Applications**]: Types of Systems - *Decision support*

I.2.1 [Artificial Intelligence]: Applications and Expert Knowledge-Intensive Systems - *Decision support*

Keywords: terrorism, Dark Web, Web harvesting, Web content analysis, Web link analysis, visualization

1. Introduction

Increasingly, extremist and hate groups are using the Internet as a powerful tool for facilitating recruitment, reaching global audiences, linking with other extremist groups, and spreading hate materials that help to persuade others to violence and terrorism. Although U.S. extremist and hate groups may not be as well-known as some of the international extremist organizations, they are considered as domestic extremist organizations that are based and operate entirely from within the continental United States and Puerto Rico [2] and pose a significant threat to U.S. homeland security.

According to the Southern Poverty Law Center (SPLC), the number of active extremist and hate groups operating in the U.S. was 708 in 2002 [14]. Their Web sites increased from 443 in 2002 to 497 in 2003, a 12 percent increase. Researchers and watchdog organizations, such as SPLC, the Simon Wiesenthal Center, and SurfControl, that monitor and analyze these Web sites are finding that keeping track of existing and new Web sites, and exploring their usage and content have become time consuming and challenging [6,8]. Since such content on the Internet is expanding, it is important that tools that allow researchers to monitor, analyze, and predict changes and developments in extremist and hate groups' use of the Web and their influences be developed [8].

The objectives of this paper are twofold. First, we propose the development of automated or semi-automated procedures and systematic methodologies for capturing extremist groups' Web site data and that the data be used for subsequent analyses. By analyzing the Web sites' content and visualizing the hyperlinks at the collection level, our methodology formalizes the process of knowledge discovery. Second, we seek to broaden our understanding of how domestic extremist groups utilize the Web infrastructure so that we can develop a comprehensive understanding of the extremists themselves. Because the groups are volatile and often associated with illegal activities and violence, they pose great

difficulties for researchers seeking to understand their structure and dynamics of their movements [4].

Since these groups are active in using the Internet, Web-based research on domestic extremist groups should prove valuable for supplementing and modifying earlier findings. In this paper, we present the related literature, proposed methodology, results, and implications of this investigation.

2. Previous Research

2.1 Social Movements Research on Extremists and Internet

Research on social movements organizations, such as extremist and hate groups use of the Internet, is in its early stages [4,10]. Researchers have identified a wide variety of different extremist groups such as White Supremacist, Black Separatist, and Militia, and how they are using the Internet to support their resource mobilization strategies. Resource mobilization is a process of securing control over resources needed for collective action such as communication, money, information, human assets, and specialized skills [10]. Clandestine groups are constantly seeking ways to improve the effectiveness of their communication [15], information operations [4], and to facilitate collective identity, solidarity, and leaderless resistance [8,15].

American extremist and hate groups have continuously exploited technology to enhance their operations and were among the early adopters of computer bulletin boards that eventually evolved into the Internet [8]. Stormfront.org, a neo-Nazi's Web site set up in 1995, is considered the first major domestic "hate site" on the World Wide Web because of its depth of content and its presentation style which represented a new period for online right-wing extremism [15]. The neo-Nazis groups share a hatred for Jews and other minorities, and a love for Adolf Hitler and Nazi Germany. A social network analysis of extremist Web sites revealed that the Stormfornt.org served as a central node that occupied a prominent position within the White Supremacist network [4].

In addition to Web sites, extremists use the Internet to access private message boards, email, research, listservs, and sell merchandise such as the Web site of Resistance

Records, the e-commerce music site of the National Alliance which is estimated to have had about $1 million in sales revenue in 2001 [8]. White supremacist groups have a significant presence on the Internet with several hundred sites ranging in complexity from single one-page sites to those that contain extensive documentations, discussion groups, music collections [10]. In the literature, the white supremacist movement is depicted as a fragmented, decentralized, and often sectarian network of organizations that can be grouped into three categories: Ku Klux Klan, neo-Nazi, and the Racist Skinheads [4]. An important unifying aspect of the movement is the Christian Identity theology that teaches that Whites are the only true children of God [4].

Besides the White Supremacists, the leftist environmental and animal liberation groups also use the Web as a tool for propaganda and violent leaderless resistance [8]. Table 1 identifies several studies that use systematic methodologies such as Web content and link analysis to explore a range of research questions about domestic extremist groups' exploitation of Internet technology.

Table 1: Summary of Research on Extremist and Hate Groups' Use of Internet

Methodology	**Finding**
Observation	Tracing of the early usages of the Internet by extremists identified patterns of usages of racial computer games, USENET, bulletin boards, and Web sites [15].
Content analysis (157 Web sites)	Majority sites contained external links to other extremist sites, half included multimedia content, and half contained racist symbols. Used Web sites to expand their reach to international audiences, link to diverse extremists groups, and allow the groups to have maximum image control [8].
Network & content analysis (80 Web sites)	Internet hyperlinks appeared to provide a reasonable accurate representation of interorganizational structure of the movement. Use of the Internet assisted in the creation of an international virtual extremist community [4].
Egocentric network & content analysis (226 Web sites)	Selection of Aryan Nations Web site as ego was effective and different from previous network studies. Factions within White Supremacy movement engaged in coalition building [10].

Most of the studies identified in Table 1 involved manual processes for gathering the Web sites, classifying them, coding the Web sites, and visualizing the patterns. From a post-retrieval analysis perspectives, existing research tools to gather and explore the interpretations of Web sites' content and usage patterns are limited as yet [5]. Existing

tools provided by Web search engines and watchdog organizations' Web sites offer limited capabilities for integrating the resources and supporting information fusion.

2.2 Web Harvesting Approaches

The first step towards studying the terrorism Web infrastructure is to harvest extremist Web sites back to a local repository for further analysis. Web harvesting is the process of gathering and organizing unstructured information from pages and data on the Web [11]. Previous studies have suggested three types of approaches to harvesting Web contents in specific domains: manual, automatic, and semi-automatic.

In order to gather samples of extremist and hate groups' Web sites on analysis, all previous extremist Web content studies used a manual approach [4,8,10,15]. For example, Burris, Smith, and Strahm used a manual approach to collect and download seed URLs for a two-week period in 1997 [4]. The seeds were identified using seven watchdog organizations that monitor hate and extremist groups such as Net Hate and HateWatch. Hate Watch was under SPLC. The limitation of such a manual approach is that it is time-consuming and inefficient.

In some other relevant domains such as e-Government domain, automatic collection building methods were used. Albertsen used an automatic approach in the "Paradigma" project [1]. The goal of Paradigma is to archive Norwegian legal deposit documents on the Web. It employed a focused Web crawler, an automatic program that discovers and downloads Web sites in particular domains by following Web links found in the HTML pages of a starting set of WebPages. Metadata was then extracted and used to rank the Web sites in terms of relevance. The automatic approach is more efficient than the manual approach; however, due to the limitations of current focused crawling techniques, automatic approaches often introduce noise (off-topic Web pages) into the harvest results.

The "Political Communications Web Archiving" group employed a semi-automatic approach to harvesting domain-specific Web sites [13]. Domain experts provided seed URLs as well as typologies for constructing metadata that can be used in the crawling process. Their project's goal is to develop a methodology for constructing an archive of

broad-spectrum political communications over the Web. Based on our review, we believe that the semi-automatic approach is the most suitable approach in harvesting terrorism Web sites, because it combines the high accuracy and high efficiency of manual and automatic approaches.

2.3 Web Link and Content Analysis

Once the extremist Web sites are harvested, two types of analysis methods can be applied to study the extremists' use of the Web: Web link analysis and Web content analysis.

Web link analysis is based on hyperlink structure and has been previously used to discover hidden relationships among communities [9,12]. Borgman [3] defines two classes of Web link analysis studies: relational and evaluative. Relational analysis gives insight into the strength of relations between Web entities, in particular Web sites, while evaluative analysis reveals the popularity or quality level of a Web entity. Terrorism research utilizes relational analysis because it provides us with insights into the nature of relations between extremist Web sites and extremist organizations. Relational link analysis approach has been used in various domains outside terrorism research. For example, Gibson [9] describes an automated methodology for discerning Web communities on the WWW. His work is based on Hyperlink-Induced Topic Search (HITS), a tool that searches for authoritative hypermedia on a given broad topic. Reid [12] made use of hyperlink-based topologies to uncover companies' non-customer online communities. However, her approach was based on manual categorization. With the vast amount of information on the Web, this qualitative methodology is difficult to apply to large-scale studies.

In order to reach an understanding of the various facets of the extremist and hate groups' Web usage and communications, a systematic analysis of the Web site content is required. Demchak and Friis' [7] work provides a well-defined methodology for analyzing communicative content in government Web sites. Their work focuses on measuring "openness" of government Web sites. To achieve this goal they developed a Web site Attribute System tool that is basically composed of a set of high level attributes such as transparency and interactivity. Each high level attribute is associated with a second layer of attributes at a more refined level of granularity. For example, the right "operational

information" and "responses" on a given Web page can induce an increase in the interactivity level of a government Web site. Demchak and Friis' work, an example of a well-structured and systematic content analysis exercise, provides guidance for the present study.

3. Proposed Approach

This study is part of a Dark Web Portal project [5] that builds on our system development experience. The goals are to understand how U.S. domestic extremists are using the Internet and identify appropriate techniques for collecting high-quality Web pages of extremist and hate groups and automating systematic procedures for analyzing and visualizing the content of individual Web sites. As illustrated in Figure 1, our proposed approach consists of three components: 1) collection building, 2) content analysis, and 3) link analysis.

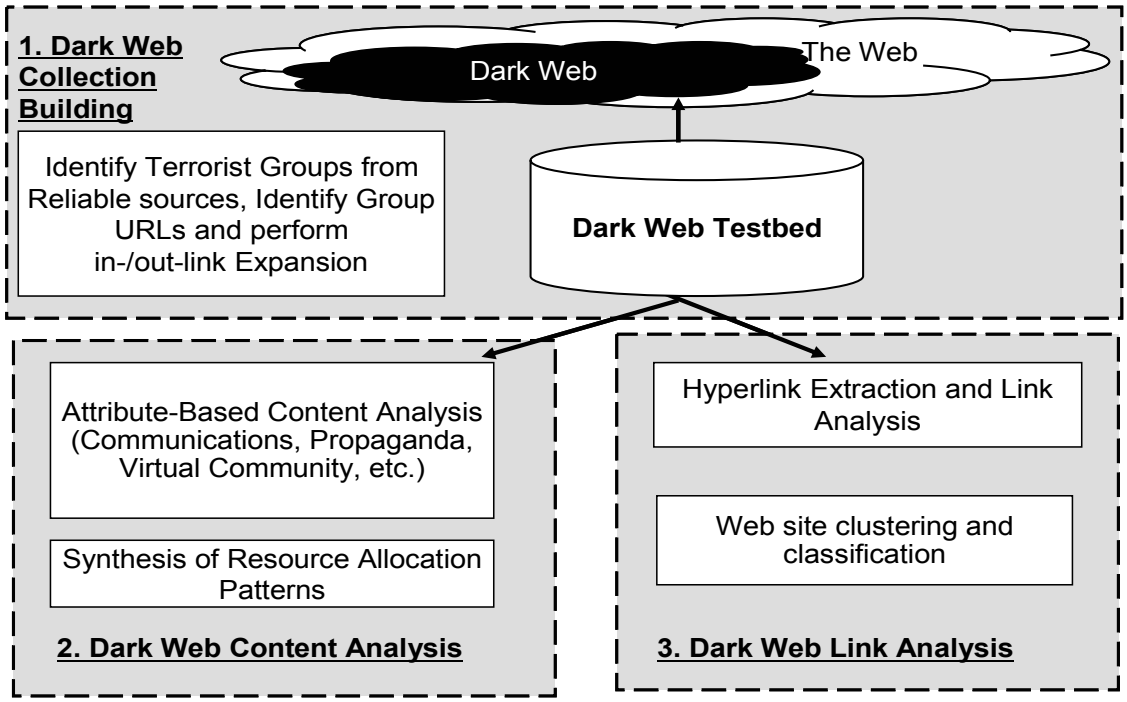

Figure 1: Architecture of Proposed Approach

To accomplish this, we first employ a semi-automatic procedure for harvesting and constructing a high quality domestic extremist Web site collection. We then perform link analysis and run a node clustering algorithm on the collection for the study of

hyperlinked terrorism Web communities. In the last step, we conduct an attribute-based systematic content analysis of our collection to study various facets of the domestic extremists' Web usage.

3.1 Collection Building

Our first goal is to construct a high-quality collection of terrorism Web sites. "High-quality" refers to the comprehensiveness and relevancy of the collected Web sites. It is desired to have a collection representing the majority of known U.S. domestic extremist groups with a presence on the Web, while keeping the collection free of unrelated Web sites. Because terrorism Web sites are often hidden and dynamic, we propose to use a recursive collection building procedure which combines both manual selection and automatic Web harvesting methods. We employ the following four steps:

1) Identify seed URLs of organizations: The first task is to find an initial set of domestic terrorism Web sites. We mainly search for URLs listed on the Web sites of major watchdog organizations such as SPLC and the Anti-defamation League (ADL) which continuously update their lists of domestic terrorism Web sites. We obtained the lists of URLs in December 2003 which served as seeds for Step 2.

2) Conduct out-link and in-link expansion: After identifying the seed URLs, the out-links and in-links of the seed URLs were automatically extracted using link-analysis programs. The out-links are extracted from the HTML contents of "favorite link" pages under the seed Web sites. The in-links are extracted from Google in-link search service through Google API. Automatic out-link and in-link expansion is an effective way to expand the scope of our collection.

3) Filter the collection: Because bogus or unrelated sites can make their way into our collection, we have developed a robust filtering process based on evidence and clues from the Web sites. Aside from sites which explicitly identify themselves as the official sites of a extremist organization, a Web site that contains even minor praise of or adopts ideologies espoused by a extremist group is included in our collection. All other Web sites are excluded, for example, Web sites with pure religious content with no elements of violence or hate.

4) Perform automatic collection and processing of extremist Web sites: Once the extremist Web sites are identified, a spider program is used to automatically download all the contents of identified Web sites. Unlike the tools used in most previous studies, in order to enable deep and comprehensive studies on the extremist Web contents, our program was designed to download not only the textual files (e.g., HTML, TXT, PDF, etc.) but also multimedia files (e.g., images, video, audio, etc.) and dynamically generated Web files (e.g., PHP, ASP, JSP, etc.). Moreover, because extremist organizations set up forums within their Web sites whose contents are of special value to research communities, our program also can automatically log into the forums and download the dynamic forum contents.

3.2 Link Analysis

Our goal here is to shed light on the infrastructure of extremist and hate Web sites and to perform a sophisticated content analysis. We believe the exploration of hidden communities over the Web can give insight into the nature of relationships between Web sites from the same group, as well as relationships between Web sites of different extremist groups. In addition, hyperlinks between Web sites constitute an important cue for estimating the content similarity of any pair of Web sites in our collection. Hence, we employ this cue to confirm our initial manual classification of the Web sites under the categories shown in the Appendix.

Uncovering hidden Web communities involves calculating a similarity measure between all pairs of Web sites in our collection. We define similarity to be a real-valued multivariable function of the number of hyperlinks in Web site "A" pointing to Web site "B", and the number of hyperlinks in Web site "B" pointing to Web site "A". In addition, a hyperlink is weighted proportionally to how deep it appears in the Web site hierarchy. For instance, a hyperlink appearing at the homepage of a Web site is given a higher weight than hyperlinks appearing at a deeper level. Thus, the similarity between Web site "A" and "B" is calculated as follows:

$$Similarity(A,B) = \sum_{\substack{\text{All links L} \\ \text{b/w A and B}}} \frac{1}{1+\text{lv(L)}}$$

where lv(L) is the level of link L in the Web site hierarchy, with homepage as level 0 and the level increased by 1 with each level down in the hierarchy.

The similarity matrix is then fed to a multidimensional scaling (MDS) algorithm which generates a two dimensional graph of the Web sites. Multidimensional scaling is a data analysis technique which provides visual representation of proximities (dissimilarities) among objects so that objects that are more similar to each other are closer on the display and objects that are less similar to each other are farther apart [16]. This technique is often used in Social Network Analysis (SNA). When applied to Web site link analysis, the proximity of nodes (Web sites) in the graph reflects the level of similarity between Web sites. Gustavson and Sherkat [10] highlight that unreciprocal ties (direction of an edge in a directed graph) can clarify the exact nature of relationships for pairs of Web sites such as friendship, resource sharing, and coordination. These considerations will, however, be tackled in future extensions of this work.

3.3 Content Analysis

To better understand the goals and ways domestic extremists use the Web, we developed an attribute-based coding scheme for methodically capturing the content. The coding scheme consists of eight high level attributes: communications fundraising, sharing ideology, propaganda (inside), propaganda (outside), virtual community, command and control, and recruitment and training. These attributes are of interest to terrorism researchers and were identified by a terrorism research expert, who has 13 years of experience serving as a terrorism intelligence analyst in the CIA. Each high level attribute is composed of multiple fine grained low level attributes. For example, level of communication is measured by existence of email contact, telephone contact, multimedia files, outline feedback form, and documentation. These low level attributes were described in detail in the coding scheme and they do not require any specific terrorism domain knowledge to be identified from the Web sites. This attribute-based approach is similar to that employed in Demchak and Friis' [7] study on government Web site interactivity analysis. The Appendix shows the high level and associated low level attributes that were used in our study. This coding scheme tool enables the detection of the particular resource allocation patterns (e.g., fundraising, propaganda) that the domestic extremists use in the Web. Moreover, it allows for measuring the levels of

usage for particular purposes by assigning a weight to the low-level attributes. Gerstenfeld, Grant, and Chiang [8] pointed out that further research should be conducted to clarify the precise nature of the messages promoted on the Web sites.

To ensure that the coding scheme is reliable, we asked four individual student coders to perform content analysis on four randomly selected U.S. domestic extremist Web sites using the coding scheme. For each of the four Web sites, the corresponding sets of content analysis results were compared and a reliability score (Cronbach's Alpha) was calculated. The results of the experiment are shown in Table 2. The high average Cronbach's Alpha of 0.807 shows that the coding scheme has high reliability.

Table 2: Reliability Test Results

Web site	United Nuwaubian Nation Of Moors Web site	Kingdom Identity Ministries Web site	Texas League of the South Web site	Knights of Ku Klux Klan Web site	**Average**
Cronbach's Alpha	0.825	0.794	0.863	0.746	**0.807**

4. Testbed: Collection of Domestic Extremists' Web sites

Following the proposed approach described in Section 3.1, we created a domestic extremists' Web sites collection. We manually extracted a set of URLs from relevant literature. A total of 266 seed URLs were identified from SPLC and ADL Web sites as well as in the Google directory. This procedure is similar to Gerstenfeld, Grant and Chiang's study [8] where they used several non-profit watchdog organizations and the Yahoo's category of White Pride and Racialism.

A link expansion of this initial set was performed and the count increased to 386 URLs. The resulting set of URLs is validated through filtering the irrelevant URLs introduced by the Google search and out-/in-link expansion. A total of 97 URLs were deemed relevant. We then used an automatic Web crawling toolkit called SpidersRUs (ai.bpa.arizona.edu/research/spider/index.htm) to download all the Web documents within the identified Web sites. As a result, our final collection contains around 400,000 documents.

Our link analysis is based on all 97 Web sites crawled. However, because of the time constraint, we could not perform content analysis on all 97 Web sites. We selected the largest Web sites from each category to form a subset of 44 Web sites for performing the content analysis. The 44 Web sites are representative of the domestic extremist groups maintaining a presence on the Web. Table 3 provides the summary and a categorization (based on SPLC) of the Web sites. We manually coded the attributes in each Web site.

Table 3: Summary of the Collection with Categories

Category	**Initial Count Before Selection**	**Final Count of URLs**	**Example Group**
Black Separatist	2	2	Nation of Islam
Christian Identity	17	13	Kinsman Redeemer Ministries
Militia	15	8	Michigan Militia
Neo Confederate	17	4	Texas League of the South
White Supremacy	29	7	Ku Klux Klan
neo-Nazi	15	9	American Nazi Party
Eco-terrorism /Animal Rights	2	1	Earth Liberation Front
Total	97	44	

5. Analysis Results

5.1 Link Analysis Results

Our link analysis aim to visualize and analyze hidden domestic terrorism hyperlinked communities and inter-community relationships. Following the analysis approach proposed in Section 3.2, five communities are identified by a terrorism domain expert in the network shown in Figure 2. On the top left side of the network resides the Southern Separatists' cluster. This cluster mainly consists of the Web sites of New Confederate organization in the southern U.S. They espouse a separatist ideology, promoting the establishment of an independent state in the south. In addition, they share white-supremacy ideas with other non-neo-confederate racist organizations such as the Ku Klux Klan (KKK), the most prominent hate group in history. A cluster of neo-Nazi and White Supremacy Web sites inhibits the top right corner of the network such as the Stormfront and the White Aryan Resistance (www.resist.com). In the bottom right corner we identified a cluster that is primarily grouping Christian Identity Web sites. A clear separation between Christian Identity, neo-Nazi, and White Supremacy groups is, in general, hard to make. This observation agrees with previous social movements studies

[4,10]. Thus, within the Christian Identity cluster appears neo-Nazi Web sites (www.aryannationsknight.com, www.aryan-nations.org), and a White Supremacy Web site of the Knights of the Ku Klux Klan (www.kkkk.net).

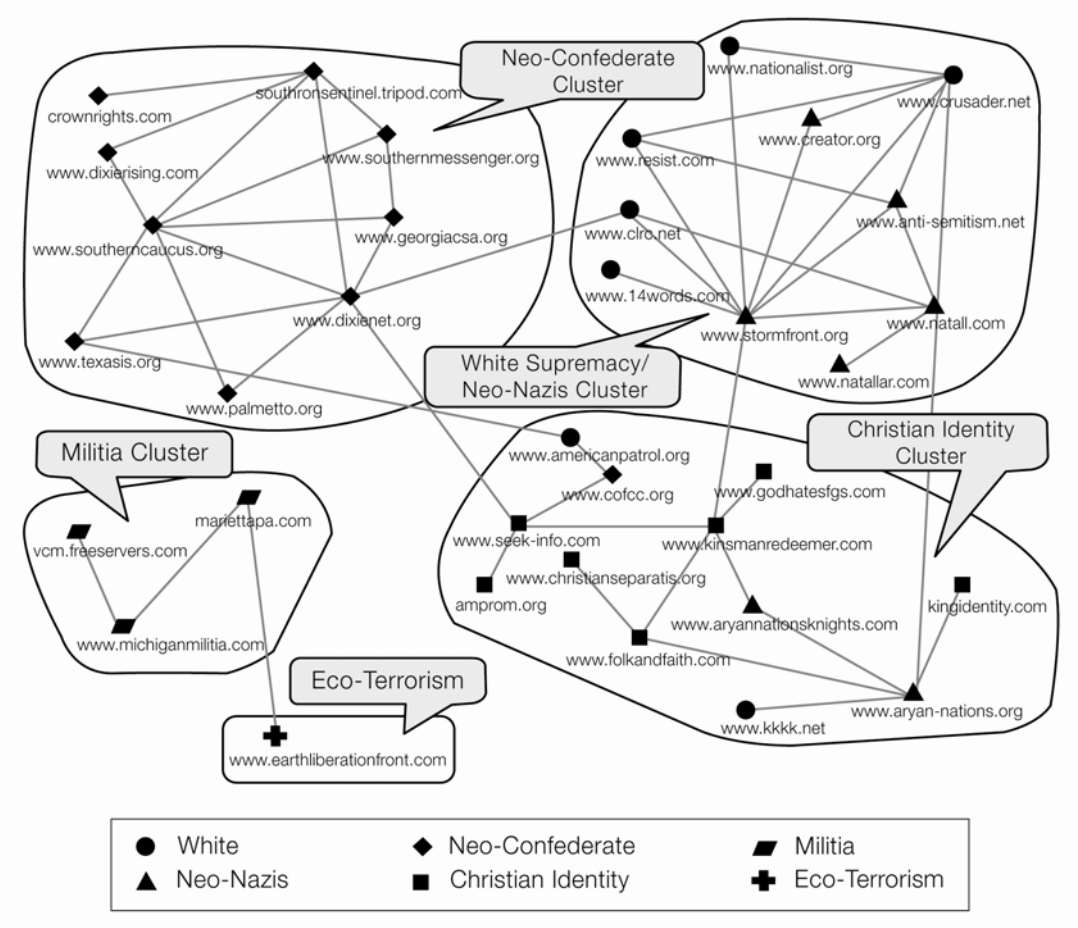

Figure 2: Web Community Visualization of Domestic Extremist and Hate Groups

Links between communities do not necessarily represent cooperation between them. An example is the few links between the neo-Confederates and Christian Identity/neo-Nazi/White-Supremacy clusters. When investigating such links, we found that Web site owners share common interests in some issues. For instance, the link between www.texasls.org (neo-Confederate) and www.americanpatrol.org (White Supremacist) reflects the common interest in "protecting" the southern border and bitterness felt towards Hispanic illegal immigrants. The numerous links between the neo-Nazi/White supremacists and Christian Identity are, on the other hand, more likely to represent good relations between the communities. Both communities have a similar ideology and researchers sometimes group them together. Two isolated communities can be seen on

the bottom left corner of the network: the Militia and Eco-Terrorism clusters. These communities have different interests and ideologies. This agrees with results of Burris, Smith, and Strahm's research [4] which concluded that bridges between the White Supremacy movement and other extremists such as the Militia are virtually non-existent.

A frequently recurring question in social network analysis is that of the existence of central or prominent nodes. We identified two such nodes in our network of U.S. domestic extremist Web sites. The first node and by far the most famous among terrorism researchers is www.stormfront.org. This Web site has many in-links indicating its popularity among White Supremacists which is in agreement with results from earlier research [4]. The second Web site is that of National Alliance (www.natall.com), a neo-Nazi Web site which also has a very high number of in-links testifying to its prominence. Owners of White Supremacy Web sites tend to cite and acknowledge other Supremacists' literature which may be residing on other Web sites. They mainly intend to gain more credibility by referring to other Supremacists with whom they share the same ideology.

Another observation is the occurrence of relatively isolated Web sites within a single cluster. Linking to other Web sites can be of benefit to the Web site owners, such that they gain credibility or they enforce the sense of solidarity within a usually geographically dispersed extremist community. However, Burris, Smith, and Strahm's study points that this does not always hold true [4]. In particular, some Web sites may be competing over a potential population of future members and/or consumers of goods that are being sold on the Web sites. For instance, we found that www.14words.com, which publishes and sells White Supremacist literature does not have a single out-link to other White Supremacist Web sites. This finding is similar to that from Burris, Smith, and Strahm's study. They posit that the site, being an e-commerce one, may not want to recommend competitors or encourage users to go to other sites.

5.2 Content Analysis Results

Two graduate students recruited from the business school at the University of Arizona coded each Web site based on the coding schema in our collection and recorded the presence of low-level attributes based on our coding scheme. For instance, the neo-Nazi

Web site www.stormfront.org shown in Figure 3 contains a forum and a bulletin board. The presence of these attributes contributes to the richness of the virtual community attribute.

Figure 3: Web site of www.stormfront.org (1st White Supremacist Web site)

After completing the coding scheme for the 44 Web sites in the collection, we compared the content of each of the extremist communities described in Figure 2. We aggregated data from all Web sites belonging to a cluster and calculated the normalized content levels in the six dimensions. Each of these six dimensions represents a normalized activity scale between 0 and 1, showing the degree of activity on the dimensions. The activity scale of cluster *c* on dimension *d* was calculated by the following formula:

$$ActivityScale(c,d) = \frac{\sum_{i}^{n}\sum_{j}^{m} w_{i,j}}{m \times n}$$

$$\text{where } w_{i,j} = \begin{cases} 1, & \text{Attribute i occurs in site j} \\ 0, & \text{Otherwise} \end{cases}$$

n = total number of attributes in dimension *d*;

m = total number of Web sites belonging to cluster *c*.

Figure 4 shows the content levels for the six categories of extremist groups: Black Separatists, Christian Identity, Militia, neo-confederates, neo-Nazi, and Eco-terrorism.

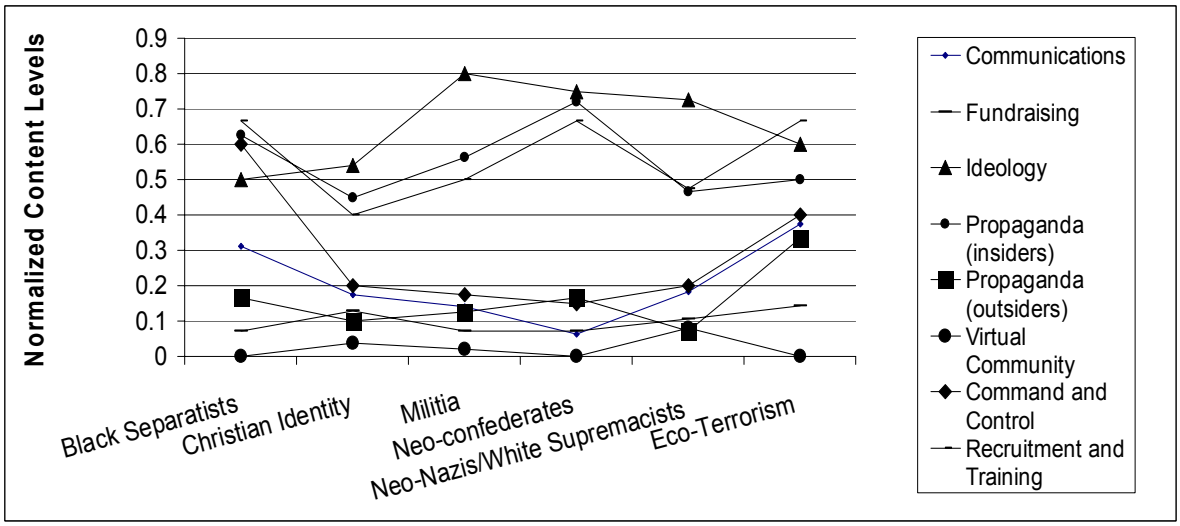

Figure 4: Content Analysis of Web Communities

As Figure 4 shows sharing ideology is the attribute with the highest frequency of occurrence in domestic extremist group Web sites. Basically, the sharing ideology attribute encapsulates all communication media devoted to portraying the goals of the extremist group, defining its general policies, and presenting the foundational ideology. A major goal of extremist and hate groups is to expose their own definitions of the movements.

With the exception of Eco-terrorism, an interesting phenomenon in domestic terrorism Web sites is the low level of content concerning the propaganda directed towards outsiders. This may be because Eco-terrorism groups have a much wider audience as compared to racist groups such as White Supremacist who only address very specific communities. For all groups, there was limited content in terms of the virtual community attribute. With their freedom of movement and speech within the U.S., domestic terrorism groups are not heavily dependent on virtual communities for resources unlike international extremist groups like al Qaeda that depend a lot on virtual communities.

Another interesting observation is the much higher levels of communications and command and control attributes in the case of Eco-terrorism/Animal Rights and Black Separatist groups. The communications attribute tells how much the owners and users of

the Web site are reliant on communication resources such as: email and chat. In general, most Web masters provide an email for feedback purposes. Moreover, some Web sites like those maintained by the Nation of Islam reach a higher level of sophistication through posting recordings and videos of group leaders. These multimedia resources also contribute to the communication attribute as they constitute an effective method of transmitting ideas and policies from the organization's hierarchy to lieutenants and members of the extremist/hate group.

6. Discussion and Future Work

In this study, we validated a systematic methodology for the study of domestic extremist Web site content. We employed focused collection building techniques, Web link analysis and attribute-based quantitative content analysis. Since there were several areas in which our findings support earlier social movement research, we are able to conclude that the topological infrastructure of the U.S. Domestic extremist and hate group Web sites seems to match domain experts' knowledge very well because the communities are formed by groups that are known to share similar ideologies or have close relationships with each other. Visualizing hyperlinked communities lead to an easier and more complete understanding of the underlying Web infrastructure of domestic extremist groups. In addition, it showed the existence and strength of the relationships between various hyperlinked communities and helped to identify likely relationships between extremist groups in the world.

The results of this research has also been useful for our work on the Dark Web Portal testbed, in that it provided systematic methodologies for capturing, classifying, and analyzing domestic extremist Web site data. Because this study involved a sample of 97 Web sites, future studies of this kind should endeavor to enlarge the sample and verify if similar outcomes can be achieved.

We have several future research directions. First, we plan to automate the content analysis process by applying data mining techniques. We also plan to apply more sophisticated link analysis algorithms and Web community mining algorithms on the extremist Web site analysis and experiment with other network visualization techniques.

Last, we would like to conduct similar study on other international extremist groups and compare their use of Web with that of U.S. domestic extremist groups.

7. Acknowledgement

This research has been supported in part by the following grants:

- DHS/CNRI, "BorderSafe Initiative," October 2003-March 2005.
- NSF/ITR, "COPLINK Center for Intelligence and Security Informatics – A Crime Data Mining Approach to Developing Border Safe Research," EIA-0326348, September 2003-August 2005.

We would like to thank all members of the Artificial Intelligence Lab at the University of Arizona who have contributed to the project, in particular Wei Xi, Feng Huang, Homa Atabakhsh, Cathy Larson, Chun-Ju Tseng, and Shing Ka Wu.

8. References

[1] K. Albertsen, "The Paradigma Web Harvesting Environment," *3rd ECDL Workshop on Web Archives*, Trondheim, Norway, August 2003.

[2] R. Blitzer, "Domestic Preemption," *Terrorism Threat and U.S. Government Response: Operational and Organizational Factors*. J.M. Smith and W.C. Thomas, eds. Colorado: U.S. Air Force Academy, INSS, 2001. http://www.usafa.af.mil/nss/terrorism.htm. Accessed January 15, 2005.

[3] C. L. Borgman, J. Furner, "Scholarly Communication and Bibliometrics," *Annual Review of Information Science and Technology.* B. Cronin, ed. Information Today, Inc, 2002.

[4] V. Burris, E. Smith, A. Strahm, "White Supremacist Networks on the Internet," *Sociological Focus*, vol. 33, 2:215-235.

[5] H. Chen, J. Qin, E. Reid, et al., "Dark Web Portal: Collecting and Analyzing the Presence of Domestic and International Terrorist Groups on the Web," *IEEE Intelligent Transportation System Conference*, Washington, D.C., 2004.

[6] CNN, U.S. Hate Groups Hard to Track, July 1999.

[7] C. C. Demchak, C. Friis, T. M. La Porte, "Webbing Governance: National Differences in Constructing the Face of Public Organizations," *Handbook of Public Information Systems*, G. David Garson, ed., New York: Marcel Dekker Publishers, 2000.

[8] P. B. Gerstenfeld, D.R. Grant, C. Chiang, "Hate Online: a Content Analysis of Extremist Internet Sites," *Analysis of Social Issues and Public Policy*, vol. 3, 1:29-44.

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

[10] A. T. Gustavson, D.E. Sherkat, "Elucidating the Web of Hate: the Ideological Structuring of Network Ties Among Right Wing Hate Groups on the Internet," Annual Meetings of the American Sociological Association, 2004.

[11] R. Kay, "Web Harvesting," *Computer World*, June 21, 2004 http://www.computerworld.com. Accessed January, 15, 2005.

[12] E. O. F. Reid, "Identifying a Company's Non-Customer Online Communities: a Proto-typology," *Proceedings of the 36th Hawaii International Conference on System Sciences*, Springler, 2003. http://e-business.fhbb.ch/eb/publications.nsf/id/214. Accessed June 18, 2004.

[13] B. Reilly, G. Tuchel, J. Simon, C. Palaima, K. Norsworthy, Leslie Myrick, "Political Communications Web Archiving: Addressing Typology and Timing for Selection, Preservation and Access," *3rd ECDL Workshop on Web Archives*, Trondheim, Norway, August 2003.

[14] SPLC Report, "Hate Groups, Militias on Rise as Extremists Stage Comeback," 2004. http://www.splcenter.org/center/splcreport/article.jsp?aid=71. Accessed June 02, 2005.

[15] M. Whine, “Far Right on the Internet,” *Governance of Cyberspace.* B. Loader, ed., London: Routledge, 1997.

[16] J. Xu and H. Chen, “CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery,” *ACM Transactions on Information Systems*, Vol. 23, No. 2, April 2005, Pages 201–226.

9. Appendix

Attributes used in the study

High Level Attribute	Low Level Attribute	Description	Wgt.
Communications	Email	Email address	1
	Telephone	Telephone number	3
	Multimedia	Video clip of bombings, game, animated picture, etc.	3
	Online Feedback Form	Allow the user to give feedback or ask question to the site owner or maintainer	1
	Documentation	Report, book, letter, memo & other resources provided (e.g., PDF, Excel)	1
Fundraising	External Aid Mentioned	Other group, individual, association or government supporting the organization	1
	Fund Transfer	Fund transfer method, bank account, etc.	1
	Donation	Donation (e.g., direct bank deposit)	1
	Charity	Donation to religious welfare organization	1
	Support Group	Sub-organizational structure charged with the fund raising	1
Sharing Ideology	Mission	Goal of the organization (e.g., destruction of an enemy state)	1
	Doctrine	Belief of the group (e.g., religious, extreme right)	1
	Justification for use of violence	Ideology condones the use of violence to accomplish goals (e.g., suicide bombing)	1
	Pin-pointing enemy	Classify others as either enemy or friend (e.g., U.S. is an enemy)	1
Propaganda (insiders)	Slogan	Short phrase with religious or ideological connotation	1
	Date	Date in the history of the group	1
	Martyr	Name of member who died in related operation	1
	Leader	Group's leader name	1
	Banner and Seal	Banner depicting representative figure, graphical symbol or seal	1
	Narrative about operation	Narrative of operation & attack of the group	1
Propaganda (outsiders)	Reference to western media coverage	Western media coverage of event	1
	News reporting	Group's own interpretation of event	1
Virtual Community	Listserv	Automatic mailing list server that broadcasts to everyone on the list *	1
	Text chat room	Virtual room where a chat session takes place (e.g., ICQ)	3

High Level Attribute	Low Level Attribute	Description	Wgt.
	Message board	Electronic message center	1
	E-conferencing	Electronic conference	3
	Web ring	Series of Web sites linked together in a ring that by clicking through all of the sites in the ring the visitor will eventually come back to originating site *	2
Command and Control	Tactics	Pointer to communication or operational pattern regarding an operation	1
	Organization Structure	Hierarchy of organization (e.g., list of leader, lieutenant)	1
	Multimedia from group's senior member	Multimedia of leadership meeting and other activities (e.g., video of leader's message or instruction)	1
	Documentation of previous operation	Multimedia or text describing group's previous operation	1
Recruitment and Training	Operation's geographical area	Meeting, headquarter or operation location	1
	Explicit invitation	Invitation to join or attend meeting, etc.	1

* Description from www.webopedia.com