

US Domestic Extremist Groups on the Web: Link and Content Analysis

Yilu Zhou, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai, *University of Arizona*

Although US domestic extremist and hate groups might not be as well-known as some international groups, they nevertheless pose a significant threat to homeland security. Increasingly, these groups are using the Internet as a tool for facilitating recruitment, linking with other extremist groups, reaching global audiences, and spreading

hate materials that encourage violence and terrorism. According to the Southern Poverty Law Center (www.splcenter.org), 708 active extremist and hate groups operated in the US in 2002.¹ The groups had 443 Web sites in 2002, which increased by 12 percent in 2003 to 497 sites.

Researchers and watchdog organizations such as SPLC, the Simon Wiesenthal Center, and SurfControl are finding it ever more time consuming to track existing and new Web sites and to explore their content and usage patterns.² As such Internet content proliferates, researchers need better tools to monitor, analyze, and predict changes in extremist and hate groups' Web use and influence.

As part of the University of Arizona's Dark Web (ai.eller.arizona.edu/research/terror/index.htm), a project dedicated to the study of terrorist and extremist groups' Internet activities, we are studying automatic and semiautomated procedures and systematic methodologies for capturing extremist groups' Web site data for subsequent analysis. By analyzing the site content and visualizing the hyperlinks at the collection level, our methodology formalizes the knowledge discovery process. At the same time, we sought to better understand how domestic extremist groups use 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 difficulties for researchers seeking to understand their structure and dynamics.³ However, many groups are active on the Internet, so analysis of their Web-based activities should prove valuable for supplementing other research activities.

Background

Research on how social movements use the Internet is in its early stages, and little work exists on extremist and hate groups. Table 1 summarizes findings from four studies that used systematic methodologies to identify domestic extremist groups and determine how they exploited Internet technology. For example, Gustavson and Sherkat used an *egocentric* network approach for data collection and research on white supremacy movement Web sites.⁴ This approach analyzes networks from the perspective of a person or group at the center of their community—an ego—and counts the number and diversity of relations.⁵ Gustavson and Sherkat selected the Aryan Nations site as the ego because the group has ties to major factions of the white supremacy movement and a core goal of supporting coalitions among them.

In general, these four studies showed that US extremist and hate groups are constantly seeking ways to make their communications and information operations more effective^{3,4} and to facilitate collective identity, solidarity, and leaderless resistance.^{2,4} Such groups have continuously exploited technology to enhance their operations. For example, they were among the early adopters of computer bulletin boards.² Stormfront.org, a neo-Nazi Web site set up in 1995, is considered to be the first major domestic "hate site," and it remains a central node within the white supremacist community.³

In addition to Web sites, extremists use the Internet to access private message boards, email, research, and listservs and to sell merchandise. For example, the e-commerce site for Resistance Records, a music production company affiliated with the National

A study of semiautomated methodologies to capture and organize domestic extremist Web site data revealed interorganizational structures and cluster affinities that coincided with both domain expert knowledge and earlier manual research.

Table 1. Summary of research on US extremist and hate groups' Internet usage.

Methodology	Finding
Observation	Michael Rhine ⁶ traced early Internet use by extremists and identified usage patterns centered on racial computer games, Usenets, bulletin boards, and Web sites.
Content analysis (157 Web sites)	Phyllis Gerstenfeld, Diana Grant, and Chau-Pu Chiang ² found that most sites contained external links to other extremist sites; half included multimedia content, and half contained racist symbols. The groups used Web sites to expand their reach to international audiences, link to diverse extremists groups, and give the groups maximum image control.
Network and content analysis (80 Web sites)	Val Burris, Emery Smith, and Ann Strahm ³ reported that Internet hyperlinks appeared to provide a reasonably accurate representation of a group's interorganizational structure. They also found that Internet use helped create an international virtual extremist community.
Egocentric network and content analysis (226 Web sites)	Aleta Gustavson and Darren Sherkat ⁴ found white supremacist factions engaged in coalition building. They used an egocentric network-sampling technique that began with the Aryan Nations Web site, rather than a seed URL set from politically motivated watchdog organizations. The method generated a large sample and showed the Internet playing a major role in ideological resource sharing.

Alliance, a white supremacist group, is estimated to have had about \$1 million in sales revenue in 2001.²

Groups like the National Alliance have a significant Internet presence with several hundred sites, which range from single Web pages to large sites containing extensive documents, discussion groups, and music collections.⁵ The research literature depicts the white supremacist movement as a fragmented, decentralized, and often sectarian network of organizations that fall into one of three categories: Ku Klux Klan, neo-Nazis, and skinheads.³ Christian Identity theology, which teaches that white people are the only true children of God, is an important unifying aspect of the movement.³

Leftist environmental and animal liberation groups also use the Web as a tool for propaganda and violent leaderless resistance.²

Most of the studies in table 1 used manual processes to gather Web sites, classify them, code their content, and visualize the Web traffic patterns. Few research tools are available to explore Web site content and usage patterns.⁷ However, Web search engines and watchdog organization sites offer some capabilities for integrating the Dark Web resources and supporting information fusion.

Web harvesting

The first step in studying the terrorism Web infrastructure is to *harvest* extremist Web sites. This means gathering unstructured information from Web pages and data and organizing it on a local repository for further analysis. Previous studies of extremist and hate groups' Web sites used a manual approach for this task.^{2-4,6} For example, Burris, Smith, and Strahm³ built their collection by manually downloading documents from seed URLs for a two-week period in 1997.

They identified seeds from seven watchdog organizations that monitor extremist and hate groups. Hatewatch is one such organization, operating under SPLC (www.splcenter.org/intel/hatewatch/hatewatch.jsp). Manual approaches are, of course, time-consuming and inefficient.

In other relevant domains, such as e-government, researchers have used automatic Web harvesting methods. The Paradigma project aims to archive Norwegian legal documents on the Web.⁸ It employed a focused Web crawler to automatically discover and download relevant Web sites by following the HTML links from a starting set of pages. The project then extracted metadata and used it to rank the Web sites in terms of relevance. The automatic approach is more efficient than the manual approach, but current techniques for focused crawling often introduce noise (off-topic Web pages) into the harvest results.

The US Center for Research Libraries employed a semiautomatic approach to harvesting domain-specific Web sites.⁹ The CRL project aims to develop a methodology for constructing an archive of broad-spectrum Web-based political communication. Domain experts provided seed URLs as well as typologies for constructing metadata to use in the crawling process.

We adopted a semiautomatic approach for harvesting terrorism Web sites, because it combines the high accuracy of manual approaches with the high efficiency of automatic approaches.

Web link and content analysis

Once the Web sites are harvested, two types of analysis helped us study how a group is using the Web.

The first, *Web link analysis*, is based on

hyperlink structure. Researchers have used it to discover hidden relationships among communities.^{10,11} One study of scholarly communications defined two classes of Web link analysis studies: relational and evaluative.¹² *Relational analysis* gives insight into the strength of relations between Web entities in particular Web sites, while *evaluative analysis* reveals a Web entity's popularity or quality level. Relational analysis works well for terrorism research because it illuminates the relations between extremist Web sites and organizations. Researchers have applied relational link analysis in various domains outside terrorism. For example, David Gibson, Jon Kleinberg, and Prabhakar Raghavan¹⁰ used the Hyperlink-Induced Topic Search in an automated methodology to discern Web communities. The HITS tool searches for authoritative hypermedia on a given broad topic. Edna Reid also used hyperlink-based topologies to uncover companies' noncustomer online communities.¹¹ However, she categorized the communities manually. The vast amount of information on the Web makes this qualitative methodology difficult to apply to large-scale studies.

The second analysis, *Web content analysis*, is the systematic study of site content. Chris Demchak, Christian Friis, and Todd La Porte¹³ provide a well-defined methodology for analyzing communicative content in government Web sites. Their work focuses on measuring openness in these sites. They developed a Web site attribute system tool to support their work, which basically consists of a set of high-level attributes, such as transparency and interactivity. Each high-level attribute is associated with a second layer of more refined low-level attributes.

We adopted Demchak's approach to guide the content analysis reported here.

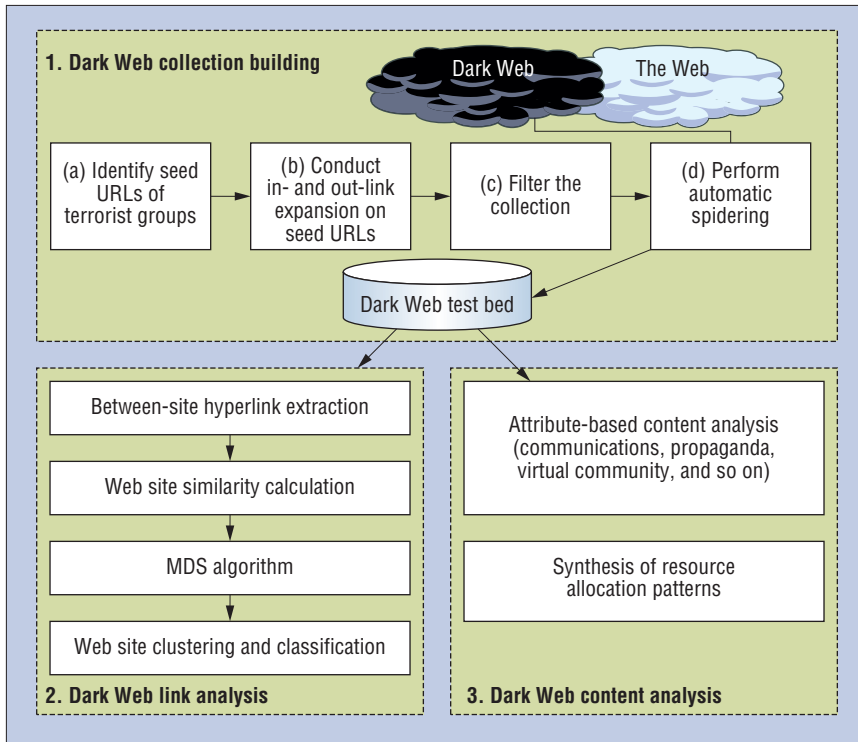


Figure 1. Architecture of proposed study approach. The Dark Web project provides a framework for building a collection of US extremist and hate group Web sites and for analyzing their hyperlinks and content.

Proposed approach

Figure 1 presents an overview of our proposed approach, which is part of the University of Arizona’s Dark Web project. Our project goals are to understand how US domestic extremist and hate groups are using the Internet, to identify appropriate techniques for collecting high-quality Web pages from the groups, and to automate systematic procedures for analyzing and visualizing individual site content. Our approach consists of three components: collection building, link analysis, and content analysis.

Collection building

Our first goal is to construct a high-quality collection of terrorism Web sites—that is, a collection that is comprehensive and relevant. The collection should represent most known US domestic extremist groups that have a Web presence while keeping free of unrelated sites. Because terrorism sites are dynamic and have short shelf lives—they emerge overnight, disappear by changing their URLs, and are often shut down by ISPs or hacked by hackers. We therefore propose a recursive procedure that builds the collection by combining both manual selection and automatic Web harvesting methods. The process has four steps.

Identify seed URLs. First, we identified an initial set of domestic terrorism Web sites. We mainly searched URLs listed on the Web sites of major watchdog organizations, such as the SPLC and the Anti-Defamation League (www.adl.org). The SPLC and ADL continuously update their lists of domestic terrorism Web sites. We obtained the URL lists for December 2003.

Conduct out-link and in-link expansions. After identifying the seed URLs, we used link-analysis programs to automatically extract the URL out-links and in-links. The out-links are from the HTML contents of “favorite link” pages under the seed Web sites; the in-links are from the Google in-link search service through the Google API.

Filter the collection. Because bogus or unrelated sites can make their way into the collection, we have developed a robust filtering process based on evidence and clues from the expanded Web site collection. As part of this process, we manually check each identified site’s content and add URLs for sites that contain even minor praise of ideologies espoused by an extremist group. We exclude all other sites from our collection—for exam-

ple, sites with purely religious content and no elements of violence or hate.

Automatically collect and process extremist Web sites. Once the extremist Web sites are identified, a spider program automatically downloads their contents. We designed the program to download not only text files (HTML, PDF, plain text, and so on) but also multimedia files and Web files generated dynamically from scripting languages such as PHP and JSP. Moreover, our program can automatically log into and download the content from the Web forums that extremist organizations often set up within their sites.

Link analysis

Analyzing hyperlink structures sheds light on extremist and hate site infrastructures. We believe it also reveals hidden communities in the relationships among different Web sites for the same group and the interactions with other extremist group sites. In addition, hyperlinks between sites constitute an important cue for estimating the content similarity of any Web site pair in a collection. We employed this cue to confirm our initial manual site classification according to the SPLC categories.

Uncovering hidden Web communities involves calculating a similarity measure between all Web site pairs 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 site *B* pointing to site *A*. In addition, we weight a hyperlink proportionally to how deep it appears in the Web site hierarchy. For instance, a hyperlink appearing at a site’s homepage has a higher weight than hyperlinks appearing at a deeper level. Thus, we calculate the similarity between Web site *A* and *B* as

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

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

Next we feed the similarity matrix to a *multidimensional scaling* algorithm, which generates a two-dimensional graph of the Web sites. MDS is a data analysis technique that visually represents proximities (dissimilarities) among objects. It displays objects

Table 2. Attributes used in the study.

High-level attribute	Low-level attribute	Description	Weight
Communications	Email	Email address	1
	Telephone	Telephone number	3
	Multimedia	Video clips of bombings, games, animated pictures, and so on	3
	Online feedback form	Allow users to give feedback to or ask questions of the site owner or maintainer	1
	Documentation	Reports, books, letters, memos, and so on	1
Fundraising	External aid mentioned	Other groups, individuals, associations, or governments supporting the organization	1
	Fund transfer	Fund transfer methods, bank accounts, and so on	1
	Donation	Donations such as direct bank deposits	1
	Charity	Donations to religious or public health and welfare organizations	1
	Support group	Suborganizational structure charged with fund raising	1
Sharing ideology	Mission	Organization goals (for example, destruction of an enemy state)	1
	Doctrine	Group beliefs, such as religious or political convictions	1
	Justification for use of violence	Condoning the use of violence to accomplish goals (for example, suicide bombing)	1
	Pin-pointing enemy	Classifying others as either enemy or friend (for example, the US is an enemy)	1
Propaganda (insider)	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 leader's name	1
	Banner and seal	Banner depicting representative figure, graphical symbol, or seal	1
	Narrative about operation	Narrative of group's operations	1
Propaganda (outsider)	Reference to Western media coverage	Western media coverage of event	1
	News reporting	Group's interpretation of event	1
Virtual community	Listserv	Automatic mailing list server that broadcasts to everyone on the list*	1
	Text chat room	Virtual room for chat sessions (for example, ICQ)	3
	Message board	Electronic message center	1
	E-conferencing	Electronic conference	3
	Web ring	Series of Web sites linked together in a ring; clicking through all sites in the ring eventually brings the visitor back to originating site*	2
Command and control	Tactics	Pointer to communication or operational pattern regarding an operation	1
	Organization structure	Organizational hierarchy, such as list of the leader and lieutenants	1
	Multimedia from group's senior member	Multimedia of leadership meeting and other activities (for example, 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 headquarters or operation location	1
	Explicit invitation	Invitation to join or attend meeting, etc.	1

*Description from www.webopedia.com

that are more similar to each other closer together and objects that are less similar to each other farther apart.¹⁴ MDS is a common tool in social network analysis. 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 Sherkat⁴ show how directed

graph edges can clarify relationships such as friendship, resource sharing, and coordination between Web site pairs. We expect to tackle such considerations in future extensions of our work.

Content analysis

To better understand domestic extremists'

uses and goals for the Web, we developed an attribute-based coding scheme for methodically capturing the content. The coding scheme consists of eight high-level attributes, listed in table 2: communications, fundraising, sharing ideology, propaganda (inside), propaganda (outside), virtual community, command and control, and recruit-

Table 3. Reliability test results for four US extremist Web sites.

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

ment and training. We selected these high-level attributes with help from a terrorism research expert, who has 13 years' experience as a CIA terrorism intelligence analyst.

Each high-level attribute comprises multiple fine-grained low-level attributes. For example, as table 2 shows, the communication attribute is a function of email contact, telephone contact, multimedia files, outline feedback form, and documentation. We developed a coding scheme that describes these low-level attributes in detail, and identifying them on Web sites doesn't require any specific terrorism domain knowledge.

This attribute-based approach is similar to that employed in a study analyzing government Web site interactivity.¹³ The coding scheme tool helps detect particular resource allocation patterns, such as fundraising or propaganda, reflected in how groups are using the Web. Moreover, assigning a weight to the low-level attributes lets us measure the usage levels for particular purposes. Gerstenfeld, Grant, and Chiang² noted the need for further research to clarify the precise nature of the messages promoted on the Web sites.

To ensure the coding scheme's reliability, we asked four student coders to use it for content analysis on four randomly selected US domestic extremist Web sites. We compared the content analysis results for each site and calculated a Cronbach's alpha reliability score, a popular tool for assessing the reliability of scales or measures like our coding scheme. A score of 0.7 indicates acceptable

reliability.¹⁵ The high average Cronbach's alpha of 0.807 shown in table 3 indicates high reliability in the coding scheme.

Test bed: Collection of extremist group Web sites

Using our proposed approach, we created a Web site collection for US extremist and hate groups. We manually extracted a set of URLs from relevant literature, identifying a total of 266 seed URLs from the SPLC and ADL Web sites as well as the Google directory. We performed a link expansion of this initial set, which increased the count to 386 URLs. We validated this set by filtering out irrelevant URLs. We finally deemed 97 URLs to be relevant.

Then, using an automatic Web crawling toolkit called SpidersRUs (<http://ai.bpa.arizona.edu/research/spider/index.htm>), we downloaded all the Web documents within these sites. As a result, our final collection contains about 400,000 documents.

We based our link analysis on all 97 Web sites crawled. However, time constraints kept us from performing content analysis on all 97 Web sites. Instead, we categorized the sites according to an SPLC scheme, shown in table 4, and selected the largest Web sites from each category to form a subset of 44 sites for content analysis. The 44 sites are representative of the domestic extremist groups maintaining a presence on the Web.

We manually coded the attributes for each Web site.

Table 4. Summary of the collection with SPLC categories.

Category	Initial URL count before selection	Final URL count	Example group
Black separatist	2	2	Nation of Islam
Christian identity	17	13	Kinsman Redeemer Ministries
Militia	15	8	Michigan Militia
Neoconfederate	17	4	Texas League of the South
White supremacist	29	7	Ku Klux Klan
Neo-Nazi	15	9	American Nazi Party
Ecoterrorist/animal rightist	2	1	Earth Liberation Front
Total	97	44	

Link analysis results

Our link analysis aims to visualize and analyze hidden domestic terrorism communities and intercommunity relationships among all 97 Web sites in our collection. Figure 2 shows the networked clusters for five communities identified by a terrorism domain expert. We generated the visualization using NetDraw open source network visualization program called (www.analytictech.com/netdraw.htm) based on our link analysis results.

The neoconfederate cluster in the top left corner consists mainly of sites espousing a separatist ideology to establish an independent state in the southern US. The neoconfederates share white-supremacist ideas with other racist organizations such as the Ku Klux Klan—the most prominent US hate group. The White Supremacy/neo-Nazi cluster in the network's top right corner includes the Stormfront site and the White Aryan Resistance site. The bottom right corner identifies a cluster of primarily Christian Identity Web sites. As other researchers have noted,^{3,5} it's generally difficult to make clear separations between Christian Identity, neo-Nazi, and white supremacist groups. Thus, the Christian Identity cluster includes some neo-Nazi Web sites and a white supremacist site.

Links between communities don't necessarily represent cooperation between them. Figure 2 shows only a few links between the neoconfederate cluster and the Christian identity and white supremacy/neo-Nazi clusters. In investigating these links, we found that Web site owners shared only a few common interests. For instance, the link between www.texasls.org (neoconfederate) and www.americanpatrol.org (white supremacist) reflects a common interest in "protecting" the southern border and a common bitterness toward Hispanic illegal immigrants. The numerous links between the white supremacy/neo-Nazi and Christian identity clusters, on the other hand, are more likely to represent strong affinities between the communities. Both have a similar ideology and researchers sometimes group them together.

Figure 2 shows two isolated communities in the network's bottom left corner: the mili-

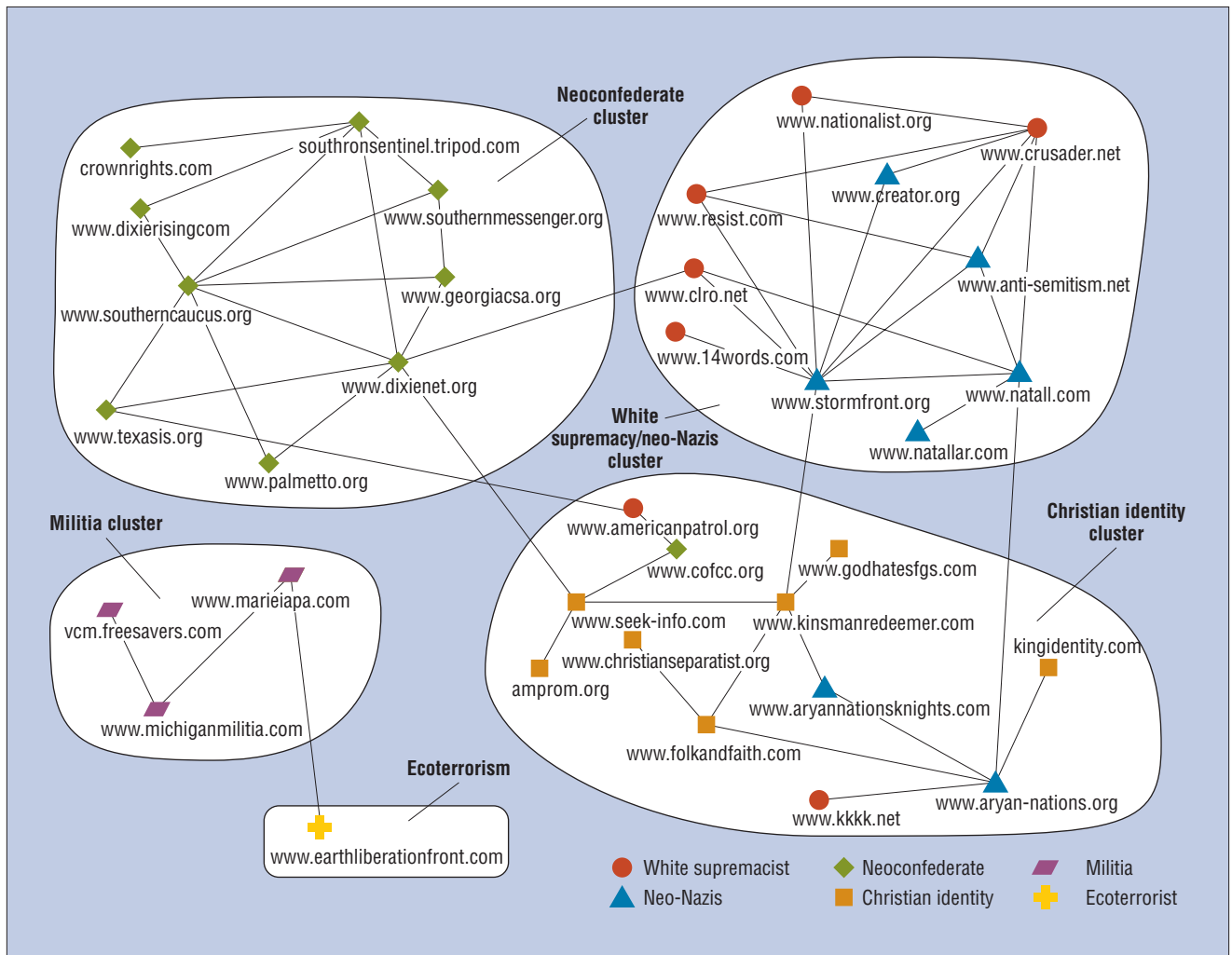


Figure 2. Web community visualization of domestic extremist and hate groups according to link analysis.

tia and ecoterrorist clusters. These communities have different interests and ideologies. This result agrees with earlier research,³ which found no bridges between the White Supremacy movement and other extremists such as the Militia.

Social network analysis looks for central or prominent nodes. Our network showed two. The first and by far foremost is www.stormfront.org. It has many in-links indicating its popularity among white supremacists. This result also agrees with earlier research results.³ The second most prominent site belongs to the National Alliance (www.natall.com). It is a neo-Nazi site that also has many in-links, testifying to its popularity. White supremacy Web site owners tend to cite and acknowledge supremacist literature on other Web sites, which accounts for some of the traffic. Their purpose seems mainly to gain more credi-

bility by referring to other documents that express the same ideology.

The link analysis results also showed many relatively isolated Web sites within a single cluster. (Because our focus was on link structure between sites, we ignored the isolated sites in this study.) Linking to other Web sites can increase a site owner's credibility or enforce a sense of solidarity within a usually geographically dispersed extremist community. However, Burris, Smith, and Strahm's study concludes that this doesn't always hold true.³ Some Web sites might not want to compete over a potential population of future members or consumers for goods being sold on the Web sites. For instance, we found that 14 Words, a white supremacy group that publishes and sells white supremacist literature, doesn't have a single out-link to other White Supremacist sites. Burris, Smith, and Strahm posit that it's because the

site owner doesn't want to encourage users to visit its competitors.³

Content analysis results

Two graduate students recruited from the University of Arizona's business school coded each of the 44 Web sites selected for the content analysis collection. They used the attribute coding scheme and recorded the presence of low-level attributes based on this scheme. For instance, www.stormfront.org contains a forum and a bulletin board, which contribute to the virtual community attribute.

After completing the coding scheme for the collection, we compared the content of each extremist community described in figure 2. We aggregated data from all Web sites belonging to a cluster and calculated the normalized content levels into six dimensions. Each dimension represents a normalized activity scale between 0 and 1, showing the degree of

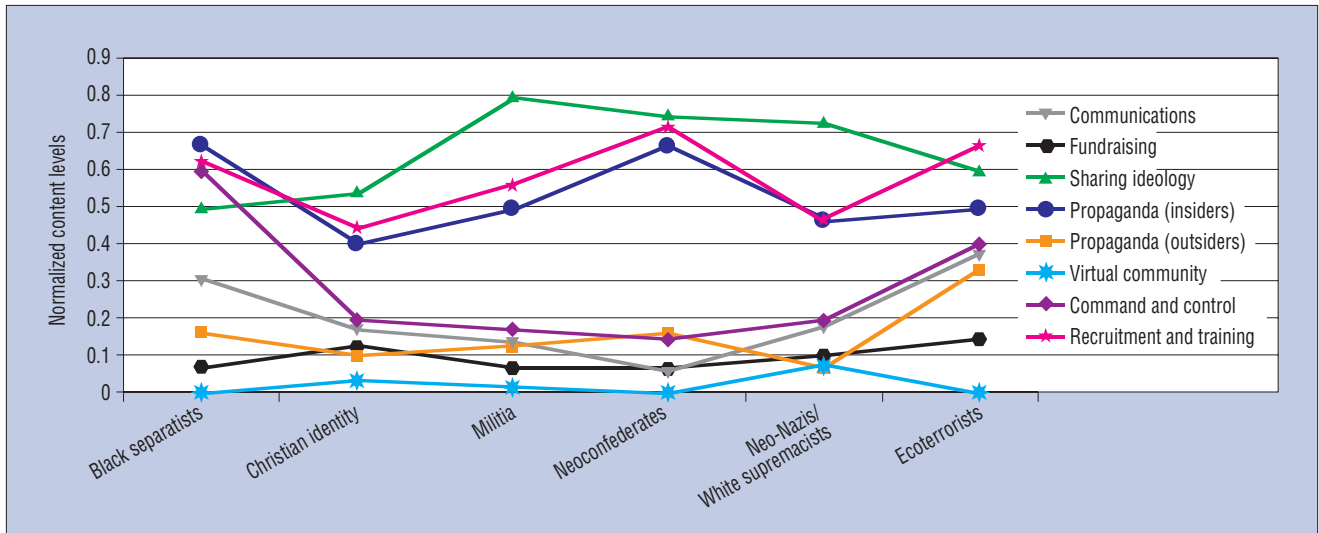


Figure 3. Content analysis of extremist Web communities. The communications attribute rates highest for all groups.

activity on the dimension. We calculated the activity scale of cluster c on dimension d by the following formula:

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

where

$$w_{i,j} = \begin{cases} 1, & \text{attribute } i \text{ occurs in site } j \\ 0, & \text{otherwise} \end{cases}$$

and n is the total number of attributes in dimension d , while m is the total number of Web sites belonging to cluster c .

Figure 3 shows the content levels for six extremist group categories: black separatist, Christian identity, militia, neoconfederate, neo-Nazi/white supremacist, and eco-terrorist.

As Figure 3 shows, the attribute of highest frequency in this collection is sharing ideology. Basically, this attribute encapsulates all communication media devoted to portraying the group’s goals, defining its general policies, and presenting its ideology. A major goal of extremist and hate groups is to expose their own definitions of their purposes.

Domestic terrorist Web sites have little content reflecting propaganda directed toward outsiders. The one exception is eco-terrorist sites, which perhaps expect a broader audience for their message than racist groups do. The virtual community content for all groups was limited. The freedom of movement and speech within the US keeps domestic terrorism groups from depending

heavily on virtual communities for resources, in contrast with international extremist groups like al Qaeda.

We also observed much higher levels of communications and command and control attributes in the case of eco-terrorist/animal rights and black separatist groups. The communications attribute tells how much the Web site owners and users rely on resources such as email and chat. In general, most Web masters provide an email contact for feedback purposes. Moreover, Web sites—for example, those maintained by the Nation of Islam—reach a higher level of sophistication by posting recordings and videos of group leaders. These multimedia resources also contribute to the communications attribute because they constitute an effective method for transmitting ideas and policies through the organization’s hierarchy to lieutenants and group members.

Our validation of a systematic methodology for the study of domestic extremist Web site content showed a topological infrastructure for US domestic extremist and hate group Web sites that seems to closely match domain experts’ knowledge. Because this study involved a sample of 97 Web sites, future work should enlarge the sample and further verify the results.

We have planned several future research directions. We want to automate the content analysis process by applying data mining tech-

niques, also expect to apply more sophisticated link analysis algorithms and Web community mining algorithms and to experiment with other network visualization techniques. Finally, we’d like to extend our work to international extremist groups and compare their Web usage with that of US domestic groups. ■

Acknowledgments

This research has been supported in part by the following grants: US Dept. of Homeland Security/Corp. for National Research Initiatives, “BorderSafe Initiative,” Oct. 2003–Mar. 2005; National Science Foundation/Information Technology Research, “COPLINK Center for Intelligence and Security Informatics—A Crime Data Mining Approach to Developing Border Safe Research,” EIA-0326348, Sept. 2003–Aug. 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.

References

1. Southern Poverty Law Center, “Hate Groups, Militias on Rise as Extremists Stage Comeback,” 2004; www.splcenter.org/center/splreport/article.jsp?aid=71.
2. P.B. Gerstenfeld, D.R. Grant, and C. Chiang, “Hate Online: A Content Analysis of Extremist Internet Sites,” *Analysis of Social Issues and Public Policy*, vol. 3, no. 1, 2003, pp. 29–44.
3. V. Burriss, E. Smith, and A. Strahm, “White Supremacist Networks on the Internet,” *Sociological Focus*, vol. 33, no. 2, May 2000, pp. 215–235.

4. A.T. Gustavson and D.E. Sherkat, "Elucidating the Web of Hate: The Ideological Structuring of Network Ties among White Supremacist Groups on the Internet," paper presented at Ann. Meeting Am. Sociological Assoc., 2004.
5. L. Garton, C. Haythornthwaite, and B. Wellman, "Studying Online Social Networks," *Doing Internet Research: Critical Issues and Methods for Examining the Net*, Steven Jones, ed., Sage, 1999, pp. 75–105.
6. M. Whine, "Far Right on the Internet," *Governance of Cyberspace*, B. Loader, ed., Routledge, 1997, pp. 209–227.
7. J. Qin et al., "The Dark Web Portal Project: Collecting and Analyzing the Presence of Terrorist Groups on the Web," *Intelligence and Security Informatics: IEEE Conf. Intelligence and Security Informatics (ISI 2005)*, LNCS 3495, Springer, 2005, pp. 623–624.
8. K. Albertsen, "The Paradigma Web Harvesting Environment," *Proc. 3rd European Conf. Research and Advanced Tech. for Digital Libraries (ECDL) Workshop on Web Archives*, 2003; <http://bibnum.bnf.fr/ecdl/2003>.
9. B. Reilly et al., "Political Communications Web Archiving: Addressing Typology and Timing for Selection, Preservation and Access," *Proc. 3rd ECDL Workshop on Web Archives*, 2003; <http://bibnum.bnf.fr/ecdl/2003>.
10. D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring Web Communities from Link Topology," *Proc. 9th ACM Conference on Hypertext and Hypermedia*, ACM Press, 1998, pp. 225–234.
11. E.O.F. Reid, "Identifying a Company's Non-customer Online Communities: A Prototypology," *Proc. 36th Hawaii Int'l Conf. on System Sciences*, IEEE CS Press, 2003; <http://e-business.fhbb.ch/eb/publications.nsf/id/214>.
12. C.L. Borgman and J. Furner, "Scholarly Communication and Bibliometrics," *Ann. Rev. Information Science and Technology*, vol. 36, 2002; books.infotoday.com/asist/arist36/sample.pdf.
13. C.C. Demchak, C. Friis, and T.M. La Porte, "Webbing Governance: National Differences in Constructing the Face of Public Organizations," *Handbook of Public Information Systems*, G. David Garson, ed., Marcel Dekker, 2000.
14. J. Xu and H. Chen, "CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery," *ACM Trans. Information Systems*, vol. 23, no. 2, Apr. 2005, pp. 201–226.
15. J. Nunnally, *Psychometric Theory*, McGraw-Hill, 1978.

The Authors



Yilu Zhou is a doctoral candidate in the University of Arizona's Department of Management Information Systems, where she is also a research associate of the Artificial Intelligence Lab. Her research interests include multilingual knowledge discovery, Web mining, and human computer interaction. She received her BS in computer science from Shanghai Jiaotong University. Contact her at the Artificial Intelligence Lab, Dept. of Management Information Systems, Univ. of Arizona, Tucson, AZ 85721; yiluz@eller.arizona.edu.



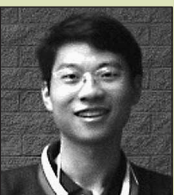
Edna Reid is a research scientist in the Artificial Intelligence Laboratory and the Hoffman E-Commerce Laboratory at the University of Arizona. Her research interests include competitive intelligence, Web mining, and e-learning systems. She has her doctorate in library science from the University of Southern California. She is a member of several professional associations and a founding member of Society of Competitive Intelligence Professionals, Singapore. Contact her at the Artificial Intelligence Lab, Dept. of Management Information Systems, Univ. of Arizona, Tucson, AZ 85721; ednareid@eller.arizona.edu.



Jialun Qin is a doctoral student at the University of Arizona's Department of Management Information Systems. His research interests include knowledge management, Web knowledge discovery and extraction, digital libraries, and social network analysis. He received his BS in computer science from Shanghai Jiaotong University in China. Contact him at the Artificial Intelligence Lab, Dept. of Management Information Systems, Univ. of Arizona, Tucson, AZ 85721; qin@eller.arizona.edu.



Hsinchun Chen is McClelland Professor of Management Information Systems at the University of Arizona. His research interests include intelligence analysis, data/text/web mining, digital library, knowledge management, medical informatics, and Web computing. He received his PhD in information systems from New York University. Contact him at the Artificial Intelligence Lab, Dept. of Management Information Systems, Univ. of Arizona, Tucson, AZ 85721; hchen@eller.arizona.edu.



Guanpi Lai is a doctoral student at the University of Arizona's Department of Systems and Industrial Engineering. His research interests include information retrieval, data mining, Petri nets, embedded systems, and intelligent transportation systems. He received his MS in industrial engineering from the University of Arizona. He is a member of the IEEE. Contact him at the Dept. of Systems and Industry Eng., Univ. of Arizona, Tucson, AZ 85721; guanpi@email.arizona.edu.

QUESTIONS?

COMMENTS?

IEEE Intelligent Systems wants to hear from you!

EMAIL

isystems@computer.org