Developing a Dark Web Collection and Infrastructure for Computational and Social Sciences

Yulei Zhang[1], Shuo Zeng[1], Chun-Neng Huang[1], Li Fan[1], Ximing Yu[1], Yan Dang[1], Catherine A. Larson[1], Dorothy Denning[2], Nancy Roberts[2], and Hsinchun Chen[1]
[1]Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721, USA
[2]Department of Defense Analysis, Naval Postgraduate School, Monterey, CA 93943
{ylzhang, shuozeng, cnhuang, fanli, ximyu, ydang}@email.arizona.edu, {dedennin, nroberts}@nps.edu, {cal, hchen}@eller.arizona.edu

***Abstract*— In recent years, there have been numerous studies from a variety of perspectives analyzing the Internet presence of hate and extremist groups. Yet the websites and forums of extremist and terrorist groups have long remained an underutilized resource for terrorism researchers due to their ephemeral nature and access and analysis problems. The purpose of the Dark Web archive is to provide a research infrastructure for use by social scientists, computer and information scientists, policy and security analysts, and others studying a wide range of social and organizational phenomena and computational problems. The Dark Web Forum Portal provides web enabled access to critical international jihadist and other extremist web forums. The focus of this paper is on the significant extensions to previous work including: increasing the scope of data collection, adding an incremental spidering component for regular data updates; enhancing the searching and browsing functions; enhancing multilingual machine-translation for Arabic, French, German and Russian; and advanced Social Network Analysis. A case study on identifying active participants is shown at the end.**

Keywords- Dark Web archive, incremental forum spidering, multilingual translation, social network visualization

I. INTRODUCTION

In recent years, there have been numerous studies from a variety of perspectives analyzing the Internet presence of hate and extremist groups. The use of the Internet by such groups has provoked interest in terrorism researchers in various social sciences including psychology, sociology, criminology, and political science; computational scientists studying web mining and information extraction; and security analysts and others concerned with homeland and national policies and security.

Yet the websites and forums of extremist and terrorist groups have long remained an underutilized resource due to their ephemeral nature and access and analysis problems. They emerge quickly, often just as quickly disappearing or, in many cases, seeming to disappear by changing their uniform resource locators (URLs) but retaining much of the same content [1]. Furthermore, some are hacked or closed by the ISPs. Thus, many researchers, students, analysts, and others face difficulties in identifying, collecting, and analyzing this content. Since terrorist and extremist groups are increasingly using the Internet to promulgate their agendas, it has become imperative that persistent access as well as user-friendly searching be provided to this content. Given the sheer volume of sites, their dynamic and fugitive nature, different languages, and noise, it has become clear that systematic and automated procedures for identifying, collecting, and searching these sites must be provided.

Therefore, as reported in our previous paper, the purpose of the Dark Web archive is to provide a research infrastructure for use by social scientists, computer and information scientists, policy and security analysts, and others [2]. The archive is currently comprised of 13 million postings from 29 international jihadist web forums. These forums collectively host 340 thousand members whose discussions cover a wide range of socio-political, cultural, ideological, and religious topics. The forums collected for this project are in Arabic, English, French, German, and Russian, and have been carefully selected with significant input from terrorism researchers, security and military educators, and other experts. The Arabic-language forums selected include major jihadist websites, some of which have English-language sections. The English-language forums represent both extremist and more moderate groups in order to facilitate study of radicalization processes over time. Three French forums, and the single forums in German and Russian, provide representative content for extremist groups producing content in these languages. The content is updated regularly in order to remain fresh and relevant, and the infrastructure, described later in this paper, includes tools for searching, browsing, translation, analysis, and visualization.

In the following sections, this paper will describe the extensions to the previous work, which include greatly increasing the scope of data collection; adding an incremental spidering component for regular data updates; enhancing the searching and browsing functions; enhancing multilingual machine-translation for Arabic, French, German and Russian; and adding advanced Social Network Analysis.

II. LITERATURE REVIEW

A. *Incremental Forum Spidering*

Spiders [3] are defined as "software programs that traverse the World Wide Web information space by following hypertext links and retrieving web documents by standard HTTP protocol." There are six important characteristics (i.e., accessibility, collection type, content richness, URL ordering features, URL ordering techniques, and collection update

procedure) of spider programs [4]. A functional spider program should therefore be able to handle the registration requirement of targeted forums (accessibility), extract desired information from various types of collected data (collection type), filter out file types which are not of interest (content richness), sort queued URLs based on given heuristics (URL ordering features and techniques), and keep the collection up-to-date (collection update procedure).

The goal is to provide comprehensive Dark Web forum data from various sources in a timely manner. Therefore, accessibility and the collection update procedure are the two most critical issues. Accessibility problems can be resolved by a human-assisted approach [5]. For the collection update procedures, two well adopted approaches are periodic and incremental spidering [6]. The periodic approach re-spiders the entire collection of targeted forums in a fixed time interval. Due to the collection size of the Dark Web forum portal, spidering the entire collection would demand a great deal of time and make it difficult to keep the collection up-to-date. In addition, the spidering process generally starts from a list of given URLs and then attempts to download all the contents which can be linked from the starting URLs. Malicious links could potentially trap spiders in a loop.

Adopting the incremental spidering approach can minimize the problems mentioned above and is a feasible solution for keeping the collection up-to-date [4]. Incremental spidering focuses on downloading only new content found in the targeted forums, so that the spidering process can be completed within a short time and even before the forums become aware of the spider. In addition, because incremental spidering targets only new content, the structure of the targeted pages is very clear and problems with malicious links can be avoided.

B. Multilingual Translation

The multilingual issue is critical in Dark Web research because much of Jihadist content is written in various languages such as Arabic, German, French, etc. In order to process multilingual content, different methods have been explored to execute translation tasks, such as the machine translation-based approach, corpus-based approach and dictionary-based approach [7]. The machine translation-based approach uses existing machine translation techniques to provide automatic translation. The corpus-based approach analyzes large document collections to construct a statistical translation model. In the dictionary-based approach, a bilingual dictionary is first constructed, and then a translation result is obtained by looking up a given term in the dictionary. Google Translation is one of the most popular machine translation tools (http://code.google.com/apis/ajaxlanguage/documentation/#Translation). The Google Translation Service provides translation functions for more than 80 languages. The language detection and language translation APIs it provides can be integrated into Web pages using Javascript. With this service, sentences in languages other than English can be translated to English automatically.

C. Social Network Analysis on Web Forums

Social Network Analysis (SNA) is a graph-based method to analyze the network structure of a group or population and its impact on social interactions [8]. SNA has been widely used to study various real-world networks [9]. The social networks formed in illegal organizations are referred to as "Dark Networks" [10]. The Dark Networks can be either in the real world, such as the criminal networks studied by Hu et al. [11], or in the virtual world, such as the social networks in web forums that are used by terrorists [12] to spread radical religious opinions, to organize terrorist activities or to share knowledge about making weapons. Thus terrorist web forums can be an important information resource for counter-terrorism tasks [13].

In web forums, a thread is a collection of posts, displayed by time in ascending order. The first post is the thread starter, and the other threads are all replies to the thread starter. The social network in web forums is based on the thread structure and is usually referred to as a "reply network" [14, 15]. The reply network is a bipartite directed graph that consists of nodes and links. Every node represents a user in web forum, and every link represents a reply-to relationship between two users, pointing from the author of the thread starter to the author of the reply.

Social network metrics are the various measurements that reflect the importance or connectivity of the node. Several types of metrics are commonly used in SNA for various purposes, including centrality, cohesion and reachability [16]. Centrality is a set of metrics that describes the importance of nodes based on their power of connecting the network, including betweeness, degree, closeness and so on. Cohesion of a social network describes how well a subset of nodes connects to each other in terms of forming a clique. Cohesion can be measured using a node's clustering coefficient, and a higher value indicates that if the node and the first neighbors of this node form a cluster, it is more likely that any two nodes in this cluster have a direct connection. Reachability describes the connectivity of two nodes, including metrics such as distance, which is the number of edges in the path that connects two nodes. In addition to these classic metrics, there are also other algorithms and metrics developed especially for estimating the importance in a web page network, such as PageRank scores and HITS scores.

III. MOTIVATION & RESEARCH QUESTIONS

Since there exists a large number of heterogeneous and widely distributed Dark Web forums, data integration and retrieval are still obstacles for researchers who wish to monitor Dark Web content [16]. The websites and forums of extremist and terrorist groups are underutilized as an information resource due to the issues described earlier in this paper. A systematic and integrated approach for searching, browsing, and analyzing these multilingual forums is important and in demand. Few studies have incorporated social network analysis into a real-time, online Dark Web forum analysis system.

Based on the research gaps discussed, we present the following research questions:

- Q1: How can regular data updates best be conducted through incremental spidering?
- Q2: How can different topics related to or associated with terrorist activities be identified using the topic search

component?

- Q3: How can the most influential forum members be identified using the SNA component?
- Q4: How can the integrated system be used as an infrastructure to support the research of scientists from various disciplines (e.g., social science and computer science) who are interested in this content?

IV. System Design

As shown in Figure 1, the Dark Web Forum Portal system contains three components: Data Acquisition, Data Preparation, and System Functionality. The overall system design is similar to our previous paper [2]. But we have added an incremental spidering component to regularly update the collection. We detail each component in the following sections.

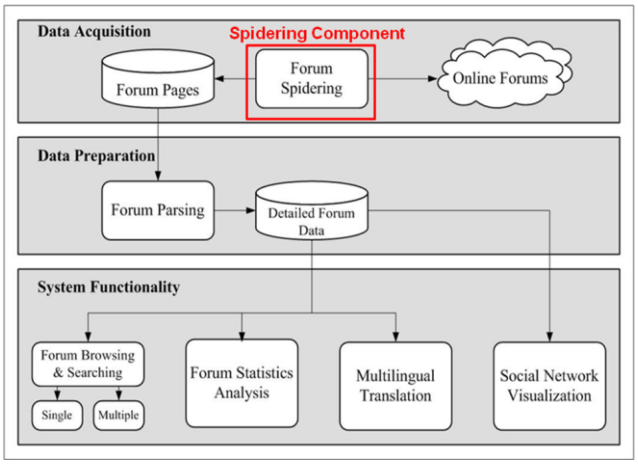


Figure 1. System Design of the Dark Web Forums Portal

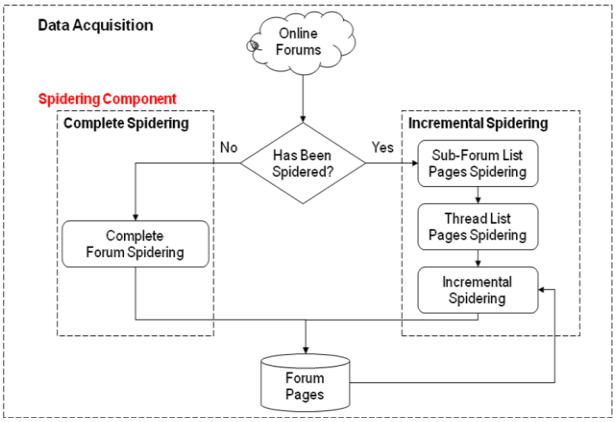


Figure 2. Framework of the spidering component

A. Data Acquisition

In this component, spidering programs are developed to collect the Web pages from online forums that contain Jihadist related content identified by domain experts. The spidering component is composed of complete spidering and incremental spidering (Figure 2). Complete Spidering is applied to forums the first time they are added to our collection, while incremental spidering is adopted if the forums already exist in the collection. When a forum is first added to our collection, the complete spidering is applied to collect all available postings. Incremental spiders are designed to identify and collect postings posted after the last updating time of the forum, so that only a small portion of forum data is collected and therefore makes the spidering process much more efficient. To achieve this goal, an incremental spider was developed for each forum in the collection.

The incremental spidering process consists of three main steps: Sub-Forum List Page Spidering, Thread List Page Spidering and Incremental Spidering. *Sub-Forum List Page Spidering:* Forums generally contain one or more sub-forums representing different discussion themes. In this step, incremental spiders first spider and parse sub-forum list pages of a forum and identify the URLs of sub-forums. *Thread List Page Spidering*: Thread list pages contain the metadata of discussion threads (such as title, date of the last update, and author name) which are sorted in descending order by the date of the last update. For each sub-forum, the incremental spider starts from downloading the first thread list page of the sub-forum and dates of the last update of discussion thread are then extracted. Threads updated later than the date of the latest posting in the database are considered to be new threads and their URLs are collected. If every thread listed in the first thread list page is a new thread, the spidering will move to the next thread page. Otherwise, the spidering of this sub-forum is complete. *Incremental Spidering*: After collecting all the URLs of new threads, the incremental spider begins to download all of the postings within the new threads.

We conducted an experiment of incremental spidering using Hanin Net forum (http://www.hanein.info/vb/), which is the most active forum identified by our domain experts. Because this forum is the most active in our forum list, the experiment result can be considered as the upper bound of the time required for incremental spidering. In the experiment, we collected postings from 11/1/2009 through 12/10/2009 (about 6 weeks). The experiment results showed that during the time span, 3,504 threads and 29,016 postings were collected. The entire incremental spidering process was completed in 39 minutes. The results indicate that incremental spidering is a promising solution for keeping our forum data up-to-date.

B. Data Preparation

In this component, forum parsing programs are developed to extract the detailed forum data from the raw HTML Web pages and store it in a local database. For each forum, the structured, detailed forum data extracted include thread names, main message bodies, member names, and post dates.

C. System Functionality

Different functions are developed and incorporated into the system as real-time services, including single and multiple forums, browsing and searching, forum statistics analysis, multilingual translation, and social network visualization. The Dark Web Forum portal is implemented using Apache Tomcat and the database is implemented using Microsoft SQL Server 2008. For forum statistics analysis, Java applet-based charts are created to show the trends based on the numbers of messages produced over time. The multilingual translation function is implemented using Google Translation Service, which can

automatically detect non-English texts and translate them into English. The social network visualization function provides dynamic, user-interactive networks implemented using JUNG (http://jung.sourceforge.net/) to visualize the interactions among forum members.

V. DATA SET

Table 1 lists the forums incorporated into our system. Currently, the portal contains 29 forums, among which 17 are Arabic forums, 7 are English forums, 3 are French forums and the other 2 are in German and Russian, respectively.

TABLE I. STATISTICS OF THE FORUMS

Name	Language	Time Span	No. of Members	No. of Threads	No. of Messages
Al-Boraq	Arabic	01/08/2006 - 01/02/2010	3,503	52,322	223,648
Al-Fallujah	Arabic	09/19/2006 - 01/02/2010	5,853	74,899	547,712
Al-Firdaws*	Arabic	01/02/2005 - 02/06/2007	2,187	9,359	39,715
Midad al-Suyuf	Arabic	03/18/2006 - 1/02/2010	1,597	11,232	38,382
Alokab	Arabic	04/08/2005 - 12/31/2009	1,547	8,096	55,947
Al-Qimmah	Arabic	11/23/2007 - 01/02/2010	287	12,097	23,709
Alsayra	Arabic	04/05/2001 - 12/31/2009	66,705	147,598	1,227,207
Ansar	Arabic	11/07/2008 - 01/02/2010	1,224	12,041	46,928
At-tahadi	Arabic	04/14/2008 - 01/02/2010	313	2,599	5,406
Hanin Net	Arabic	11/27/2006 - 01/12/2010	2,837	96,239	821,478
Hawaa World	Arabic	01/01/2001 - 01/02/2010	113,579	40,501	2,251,553
Hadramout	Arabic	11/25/2000 - 12/29/2009	29,491	151,694	1,552,227
Ma'arik	Arabic	07/29/2007 - 01/03/2010	1,880	15,288	57,047
Al-Mujahidin	Arabic	11/09/2007 - 01/02/2010	4,259	29,980	140,930
Montada	Arabic	09/25/2000 - 12/29/2009	40,291	120,181	1,412,028
Ana al-Muslim	Arabic	10/08/1985 - 11/26/2009	12,215	179,791	1,343,370
Shumukh	Arabic	03/21/2007 - 01/02/2010	3,938	46,666	289,201
Ansar	English	12/08/2008 - 01/02/2010	377	11,133	29,056
Gawaher	English	10/24/2004 - 01/01/2010	6,790	210,656	569,709
Islamic Awakening	English	04/28/2004 - 12/31/2009	2,361	25,112	116,009
Islamic Network*	English	06/09/2004 - 05/07/2008	1,573	11,974	87,314
Islamic Web-Community	English	11/14/2000 - 12/31/2009	745	6,262	24,850
Turn To Islam	English	06/02/2006 - 01/01/2010	9,926	38,702	308,970
Ummah	English	04/01/2002 - 12/31/2009	14,349	71,218	1,192,583
Al Minha Dj	French	06/01/2008 - 01/04/2010	313	2,007	6,421
Forums d'aslama	French	10/06/2004 - 01/03/2010	2,665	20,468	131,559
Al-Mourabitoune	French	05/05/2002 - 03/27/2009	3,198	7,905	72,140
Ansar	German	02/27/2009 - 01/02/2010	62	726	1,645
KavkazChat	Russian	03/21/2003 - 01/03/2010	5,634	6,144	558,042
Total			**339,699**	**1,422,890**	**13,174,786**

** These forums are no longer active.*

The forums have been carefully selected with significant input from terrorism researchers, security and military educators, and other experts. The Arabic-language forums selected include major jihadist websites, some of which include English language sections. The English-language forums represent both extremist and more moderate groups. The French, German, and Russian forums provide representative content for extremist groups communicating in these languages, and provide additional opportunity to evaluate multilingual translation. The total number of messages is about 13M; approximately 3M postings will be added annually through incremental spidering.

VI. SYSTEM FUNCTIONALITY

As previously described, the system has four types of functions: single and multiple forum browsing and searching, forum statistics analysis, multilingual translation, and social network visualization. In this section, we describe our enhanced browsing and searching as well as social network visualization. For the other two functions, see the previous paper [2].

A. Forum Browsing & Searching

1) Single Forum Browsing & Searching

The search function allows users to search message titles or bodies using multiple keywords. Users can choose the Boolean operations of the keywords to be either "AND" or "OR." Users can also express their search terms in English even when the forum is, for example, mainly Arabic. In that case, the search will return matches for both the English terms and the Arabic translations of those terms.

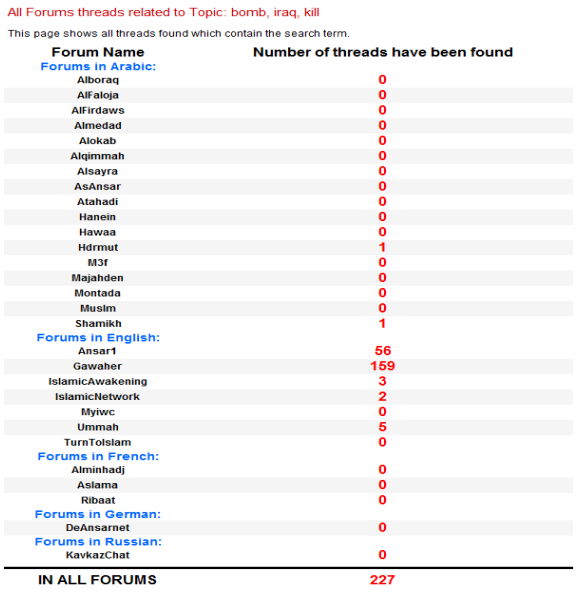
2) Multiple Forums Browsing & Searching

In addition to browsing and searching information in a particular forum, our portal also supports multiple forum searching across all forums in the portal. For example, a total of 227 threads (Figure 3) are retrieved across all forums that contain keywords "bomb," "Iraq," and "kill" (AND operation) in the thread titles or message bodies. Among them, 159 are from the forum "Gawaher," 56 are from forum "Ansar1," 5 are from forum "Ummah," etc. "Gawaher" has more discussions on this topic than any of the other forums. Detailed searching results for each forum on these keywords can be found by clicking the row corresponding to a particular forum. Figure4 shows the screenshot of the detailed result for forum "Gawaher" based on the cross forum search in Figure 3.

B. Social Network Visualization

The interface of the SNA function is shown in Figure 5. It consists of three parts: the *search panel* (top box), *analysis panel* (middle box), and *visualization panel* (bottom box).

The search panel allows the user to choose three search criteria: forum, keyword and time period. The threads that meet these search criteria are identified as "related threads" and are used to construct the social network. Any of the forums listed in the portal can be selected to perform SNA.


All Forums threads related to Topic: bomb, iraq, kill

This page shows all threads found which contain the search term.

Forum Name	Number of threads have been found
Forums in Arabic:	
Alboraq	0
AlFaloja	0
AlFirdaws	0
Almedad	0
Alokab	0
Alqimmah	0
Alsayra	0
AsAnsar	0
Atahadi	0
Hanein	0
Hawaa	0
Hdrmut	1
M3f	0
Majahden	0
Montada	0
Muslm	0
Shamikh	1
Forums in English:	
Ansar1	56
Gawaher	159
IslamicAwakening	3
IslamicNetwork	2
Myiwc	0
Ummah	5
TurnToIslam	0
Forums in French:	
Alminhadj	0
Aslama	0
Ribaat	0
Forums in German:	
DeAnsarnet	0
Forums in Russian:	
KavkazChat	0
IN ALL FORUMS	227

Figure 3. The screenshot of the cross forum search based on keywords "bomb", "iraq" and "kill" (AND operation)

Figure 4. The screenshot of the detailed result for forum "Gawaher" based on the cross forum search shown in Figure 3.

The keywords are selected by the user, in any language, separated with a space or comma. Thread names, user names, and postings are searched using these keywords, and a thread is identified as a related thread if the thread name, or at least one posting, or at least one user name, contain any of the keywords. The start date and the end date are used to constrain the postings in the search result. When related threads are returned, the social network will be constructed based on the structure of these threads.

The analysis panel allows the user to select different metrics for SNA, and to set the parameters for graph visualization.

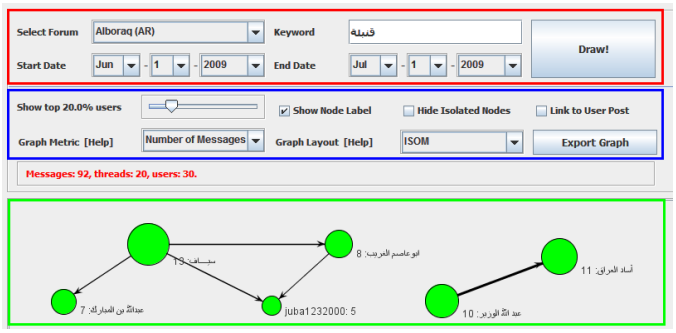


Figure 5. SNA Function Interface

Every node in the social network has a set of attributes, including the screen name, the number of postings, and various social network metrics. After the social network is constructed, all nodes are ranked in descending order based on the number of postings. Since the resulting social network usually contains a large number of message authors, which makes the graph too crowded for analysis, the slide bar can be used to display only a portion of the top authors based on the ranking in order to make the graph easier to read. The label as well as the value of the selected metrics can be displayed beside each node by checking the corresponding box. An isolated node is defined as a node that has no connections to any other node. Removing isolated nodes is a useful function when too many cause noise in the graph. Checking the "Link to User Post" box will change the color of every node from green to red, and a click on any node will pop up a new window that shows all postings by this user during the selected time period.

The visualization panel displays the graph based on the settings in the analysis panel, with the thickness of the link proportionate to the intensity of interactions between two nodes. Any node can be dragged to any position in the panel, and all connected nodes and corresponding links will be highlighted when holding the mouse button pressed during the move of the node. Different layouts are also provided for graph visualization. Four types of layout algorithms are integrated into the component, including static layout, circle layout, 3 force-based layouts (Fruchterman-Reingold, Kamada-Kawai, and Spring) and a self-organizing layout (ISOM). If users want to perform advanced analysis on the graph using other SNA tools such as UCINET, Pajek and so on, clicking the "Export Graph" button allows the graph to be exported to a ".net" format file, which is the Pajek graph file format recognized by most SNA tools.

VII. Case Study: Identifying Active Participants in Dark Web Forums using Social Network Analysis

In this case study, we demonstrated a scenario on how to use the SNA component to identify active users on a particular topic of interest (in this case, "religious beliefs"). We chose Islamic Awakening for this case study. We searched "Muslim, Islam, Sharia, Sunni and Shia" as keywords and generated the topic-based social network. The time period selected was

01/01/2009 -12/31/2009 (one year). As shown in Figure 6, "Bint ul Islam," "Iloveislam," and "Abuhannah" were the three most active participants on these topics based on the "Number of Messages." We can show the SNA based on other different graph metrics such as "betweeness," "PageRank," "in-degree," and "out-degree."

By checking the "Link to User Post" box, the user can see the detailed messages. Example messages on this religious topic include the following:

"Sister, since you are a new Muslim, Allah will test you as He Says in the Qur'an: 'Do men think that they will be left alone on saying, "We believe", and that they will not be tested?' We did test those before them, and Allah will certainly know those who are true from those who are false' [Al-Qur'an 29:2-3] You have to remain strong…" by "iloveislam."

"Between The Past And The Future Imam Ibn ul Qayyim al Jawziyyah al-Fawaa'id, pp 151-152 Al-Istiqaamah, No. 2 Your life in the present moment is in between the past and the future. So what has preceded can be rectified by tawbah (repentance), nadam (regret) and istighfar (seeking Allaah's forgiveness). ..." by "Bint ul Islam."

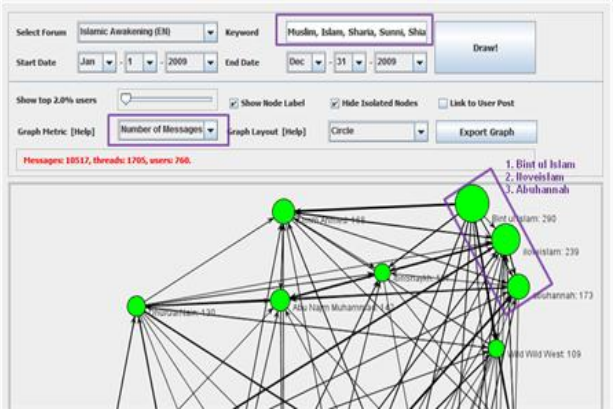


Figure 6. Active participants in the forum, "Islamic Awakening"

VIII. Conclusions and Future Directions

In this paper, we presented our work on developing a Dark Web collection and infrastructure for use by those in the computational and social sciences. Significant extensions of our previous work to date have included increasing the scope of our data collection; adding an incremental spidering component for regular data updates; enhancing the searching and browsing functions; enhancing multilingual machine-translation for Arabic, French, German and Russian; and advanced Social Network Analysis.

The Dark Web Forum Portal is an infrastructure which integrates heterogeneous forum data, and will serve as a strong complement to the current databases, news reports, and other sources available to the research community. Future work will include integrating a sentiment analysis engine, thus allowing for deeper text mining and analysis, and an improved user interface with automated translation functions.

Acknowledgments

This work is supported by the NSF Computer and Network Systems (CNS) Program, (CNS-0709338), September 2007 - August 2010 and HDTRA1-09-1-0058, July 2009 - July 2012. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or DOD.

References

[1] G. Weimann (2004). "www.terror.net: How Modern Terrorism Uses the Internet." Special Report, United States Institute of Peace. Retrieved October 31, 2006. http://www.usip.org/pubs/specialreports/sr116.pdf

[2] Y. Zhang, S. Zeng, L. Fan, Y. Dang, C. Larson, and H. Chen (2009). "Dark Web Forums Portal: Searching and Analyzing Jihadist Forums," in Proceedings of the IEEE International Intelligence and Security Informatics Conference (Dallas, Texas, June 8-11).

[3] F. C. Cheong (1996). Internet Agents: Spiders, Wanderers, Brokers, and Bots. Indianapolis, IN: New Riders Publishing.

[4] T.J. Fu, A. Abbasi, and H. Chen (forthcoming). "A Focused Crawler for Dark Web Forums," *Journal of the American Society for Information Science and Technology (JASIST)*.

[5] S. Raghavan and H. Garcia-Molina (2001). "Crawling the Hidden Web." In Proceedings of the 27th International Conference on Very Large Databases.

[6] J. Cho and H. Garcia-Molina (2000). "The Evolution of the Web and Implications for an Incremental Crawler." In Proceedings of the 26th International Conference on Very Large Databases.

[7] Y. Zhou, J. Qin, H. Chen et al. (2005). "Multilingual Web Retrieval: An Experiment on a Multilingual Business Intelligence Portal," in Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'2005).

[8] D. Liben-Nowel (2007). "The Link-prediction Problem for Social Networks," *JASIST*, 58(7), pp. 1019-1031.

[9] G. Kossinets, and D. J.Watts (2006). "Empirical Analysis of an Evolving Social Network," *Science*, vol. 311, pp. 88-90.

[10] J. Raab, and H. B. Milward (2003). "Dark Networks as Problems," *Journal of Public Administration Research and Theory*, Vol. 13, pp. 413-439.

[11] Hu, D., Kaza, S., and Chen, H. (2009). "Identifying Significant Facilitators of Dark Network Evolution." *JASIST*, 60(4), pp. 655-665.

[12] E. Reid, J. Qin, W. Chung et al. (2004). "Terrorism Knowledge Discovery Project: A Knowledge Discovery Approach to Addressing the Threats of Terrorism," in Proceedings of the 2nd Symposium on Intelligence and Security Informatics (Tucson, June 10-11), pp. 125-145.

[13] S. Coll and S. B. Glasser (2005). "Terrorists Turn to the Web as Base of Operations." *Washington Post*, August 7.

[14] Zhang, J., Ackerman, M. S., and Adamic, L. (2007). "Expertise Networks in Online Communities: Structure and Algorithms." In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12). WWW '07. ACM, New York, NY, pp. 221-230.

[15] Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. (2008). "Knowledge Sharing and Yahoo Answers: Everyone Knows Something." In Proceeding of the 17th international Conference on World Wide Web (Beijing, China, April 21 - 25). WWW '08. ACM, New York, NY, pp. 665-674.

[16] Albert, R. and Barabasi, A.-L. (2002), "Statistical Mechanics of Complex Networks," *Rev. Mod. Phys.* Vol. 74, pp. 47–97.