ChatCoder: Toward the Tracking and Categorization of Internet Predators

April Kontostathis, Department of Mathematics and Computer Science
Lynne Edwards, Department of Media and Communication Studies
Amanda Leatherman, Department of Media and Communication Studies

Ursinus College, Collegeville PA 19426
akontostathis,ledwards,amleatherman@ursinus.edu

Abstract—We describe the preliminary results from a new research project which studies the communicative strategies of online sexual predators. These pedophiles approach children via Internet technologies, such as instant messaging or chat rooms. This article presents the software we used to facilitate analysis of chat log transcripts and the development of a communicative theory of online predation. This software is used to label and analyze chat transcripts. Our preliminary experimental results show that we can distinquish between predator and victim communication, but not as reliably as we would like to. We can, however, confidently distinquish between predatory and non-predatory discussion. In a second set of experiments, we used k-means clustering to discover that there are four types of online predation communication.

1 Introduction

According to the most recent 2008 online victimization research by the National Center for Missing and Exploited Children, approximately 1 in 7 youth (ages 10- to 17-years-old) experience a sexual approach or solicitation by means of the Internet [15]. The NCMEC has set up a CyberTipLine for reporting cases of child sexual exploitation, and the magnitude of the calls is staggering. From March 1998, when the CyberTipLine began operations, until October 27th, 2008, there were 40,353 reports of "Online Enticement of Children for Sexual Acts", one of the reporting categories. There were 122 in the week of October 27th, 2008 alone [15].

This paper describes early progress in a multi-disciplinary study that seeks to analyze, index, and categorize the communicative strategies employed by cyber-predators to lure children into sexually abusive relationships. We have developed automated tools to assist in the analysis of chat log transcripts by communications researchers, and analyzed transcripts with this software.

We also describe the use of text mining techniques to distinguish between the predator and victim communication, as well as methods for distinguishing between predatory and non-predatory Internet chat. Additional experiments using clustering techniques for identifying multiple predator types were performed and are described. The preliminary results of these clustering experiments show that there are potentially four different strategies of Internet predation communication.

In the next section we describe related work in both communication and computer science literature. Section 3 describes the data we use in our experiments. Sections 4 and 5 describe our approach for the operationalization of the communicative theory, as well as the ChatCoder software that is used to facilitate the testing of this theory in the online environment. Our experimental approach and results for both the categorization and the clustering experiments are described in Section 6. Section 7 summarizes our conclusions.

2 Related Work

This project integrates communication and computer science theories and methodologies to develop tools to protect children from, and empower them against, cyber-predators.

2.1 Communications

The theory of luring communication provides a model of the communication processes that child sexual predators use in the real world to entrap their victims [16]. This model consists of 3 major stages:

1) gaining access to the victim,
2) entrapping the victim in a deceptive relationship,
3) initiating and maintaining a sexually abusive relationship.

During the gaining access phase, the predator maneuvers himself into professional and social positions where he can interact with the child in a seemingly natural way, while still maintaining a position of authority over the child. For example, gaining employment at an amusement park or volunteering with a community youth sports team. The next phase, entrapping the victim in a deceptive relationship, is a communicative cycle that consists of grooming, isolation and approach. Grooming involves subtle communication strategies that desensitize victims to sexual terminology and reframe sexual acts in child-like terms of play or practice. In this stage, offenders also isolate their victims from family and friend support networks before approaching the victim for the third phase: sexual contact and long-term abuse.

In previous work, we expanded and modified the luring theory to accommodate the difference between online luring and real world luring [11]. For example, the concept “gaining access” was revised to include the initial entrance into the online environment and initial greeting exchange by offenders and victims, which is different from meeting kids at the amusement park or through a youth sports league. Communicative desensitization was modified to include the use of slang, abbreviations, net speak, and emoticons in online conversations. The core concept underpinning entrapment is the ongoing deceptive trust that develops between victims and offenders. In online luring communications, this concept is defined as perpetrator and victim sharing personal information, information about activities, relationship details, and compliments.

2.2 Computer Science

Much of the social networking research in computer science has focused on chat room data [13] [9]. Some of this work has centered on identifying discussion thread sub-groups within a chat forum [1] [3], while other researchers have focused on the technical difficulties encountered when trying to parse chat log data [5] [21].

When we looked for domain specific research, we found several interesting projects that analyze chat data in order to detect terrorist organizations [10] [4]. However, surprisingly few researchers have attempted to deal with the creation of specific applications for analysis and management of Internet predation. In fact, we found only two researchers who are working on such applications. In the first, Pendar has had some success when analyzing chat log transcripts in order to differentiate between the aggressor and the intended victim [17]. The second is focused on child pornography; Hughes, et al. have worked on identifying child pornography distribution via peer-to-peer networks [8].

Technologies that analyze chat, identify threatening situations, and help parents and victims defend themselves are essential for both preventing the exploitation of minors and for identifying predators. The few commercial products that profess to provide this service are woefully lacking. eBlaster^{TM}records everything that occurs on a monitored computer and forwards the information to a designated recipient, but does not provide a mechanism for filtering or analyzing all the data it collects [6]. Net Nanny^{TM}can record everything, and offers multiple levels of protection for different users [14]. The latest version of Net Nanny^{TM}claims to offer alerts for potential predators and cyber-bullies, but the alerts appear to be based on simple keyword matching [12], not based on communication theory, nor does Net Nanny^{TM}suggest appropriate responses to the user. Young victims would particularly benefit from empowering technologies that simultaneously provide them with effective communicative responses, designed to register their non-compliance, and halt the abuse. Our project goals include development of open source tools, with solid theoretical foundation in both communication and computer science, that not only detect predatory action while it is occuring, but also provide prompts intended to teach youth how to respond to, and protect themselves from, predators.

3 Data

There is little reliable labeled data pertaining to predator communications; much of the work that has appeared in both computer science and communication studies forums is focused on anecdotal evidence and chat log transcripts from Perverted Justice (PJ) [18]. Perverted-justice.com began as a grass-roots effort to identify cyber-predators. PJ volunteers pose as teens and tweens in chat rooms and respond when approached by an adult seeking to begin a sexual relationship with a child. When these activities result in an arrest and conviction, the chat log transcripts are posted online.

We downloaded 288 chat logs that were available from the PJ website as of August 2008. These transcripts are labeled with personal information about the predators. The predators who participated in the chats have been convicted based, at least in part, on the content of the chat logs, which provides the PJ data with a measure of credibility. It is important to note, however, that the raw data is not available. The PJ organization claims that the weblogs are unaltered and complete, but there is no way to verify this statement. Repeated requests for more information and access to the raw data have been ignored by Perverted-justice.com. In fact, the logs have comments (clearly identified as such), from the PJ volunteers, dispersed throughout. Many of these cases take months or years to develop, and multiple chat sessions are often included on a single web page. Parsing the chat logs is an interesting problem that we continue to struggle with.

We have also obtained chat room data from Dr. Susan Gauch, University of Arkansas, who collected the data during a chat room topic detection project [2]. Dr. Gauch's project included the development of a crawler that downloaded chat logs (ChatTrack). Unfortunately, the software is no longer available. This chat data, although somewhat dated, was used as a baseline for some of our preliminary studies in Section 6.

4 Modeling Predation

Communications researchers define two primary goals for content analysis [20]:

1) To describe the communication
2) To draw inferences about its meaning

In order to perform a content analysis for Internet predation, we developed a codebook and dictionary to distinguish among the various constructs defined in luring communication theoretical model. The coding process occurs in several stages. First, a dictionary of luring terms, words, icons, phrases, and netspeak for each of the three luring communication stages was developed. Second, a coding manual was created. This manual has explicit rules and instructions for assigning terms and phrases to their appropriate categories. Finally, software that mimics the manual coding process was developed (this software is described in detail in Section 5). Twelve transcripts from the Perverted Justice website were carefully analyzed for the development of the dictionary. These 12 online conversations ranged from 349 to 1500 lines of text. The perpetrators span from 23 to 58 years of age, were all male, and were all convicted of sexual solicitation of minors over the Internet.

The transcripts downloaded from Perverted-justice.com were in netspeak, a writing style used by individuals over the Internet, for both positive and negative reasons. Therefore, we also developed a guide to translate these netspeak abbreviations into standard English. As part of this translation, we convert emoticons (emotional symbols and faces) into text. These alterations in the transcripts did not distort the meaning of words or sentences written by either victim or predator. After this transformation, we captured key terms and phrases that were frequently used by online sexual predators, and identified their appropriate category labels within the luring model: deceptive trust development, grooming, isolation and approach [16] [11]. The dictionary included terms and phrases common to net culture in general, and luring language in particular. Some examples appear in Table 1.

The current version of coding dictionary contains 475 unique phrases. A breakdown of the phrase count by category appears in Table 2.

TABLE 2
Dictionary Summary - Phrase Count by Category

Category	Phrase Count
Activities	11
Approach	56
Communicative Desensitization	220
Compliment	35
Isolation	43
Personal Information	29
Reframing	57
Relationship	24

5 ChatCoder

The prototype computer application we developed, ChatCoder, proved particularly helpful for creating a sound codebook for analysis of the chat log text, and thus for the operationalizatoin of the communicative theory for online predation. Figure 1 shows the single transcript view within ChatCoder. We used this view during dictionary development, because the analyst was able to quickly add new phrases and recode the transcript to see the results. Transcripts could also be saved in HTML format to retain the color highlighting.

The analysts noticed that there were occasionally large gaps in the colors – when those gaps were reread, subtleties like requests for AIM info and cell phone numbers were noticed. Thus, the software was used to easily identify important phrases that had been missed in the manual coding process.

In addition to providing a color for each category, ChatCoder allows the user to define a nominal value for each category (see Figure 2). In the Perverted Justice project the colors and nominal values were used to identify relative aggression on the part of the predator. The software also supports output of just the nominal values in a transcript.

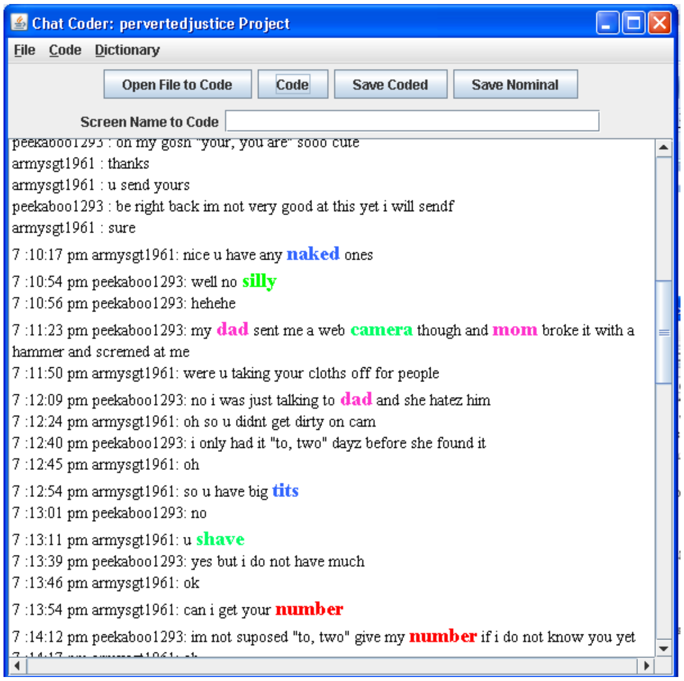
ChatCoder also includes a Batch Code feature which codes an entire directory of transcripts in one pass. The software stores the coded transcript, the nominal data, and a summary of the categories that were seen in the batch of transcripts. The summary data can then be imported into Excel. The chart in Figure 3 demonstrates the percentage of transcripts containing each coding category. Communicative desensitization appears in almost every transcript, but relationship appears in less than 20 percent.

Thus far ChatCoder has been used primarily for the online predation project, but the software is capable of managing codebooks for multiple projects. It is our intention to distribute the software for use in communications research.

TABLE 1
Sample excerpt from Codebook for Internet Predation

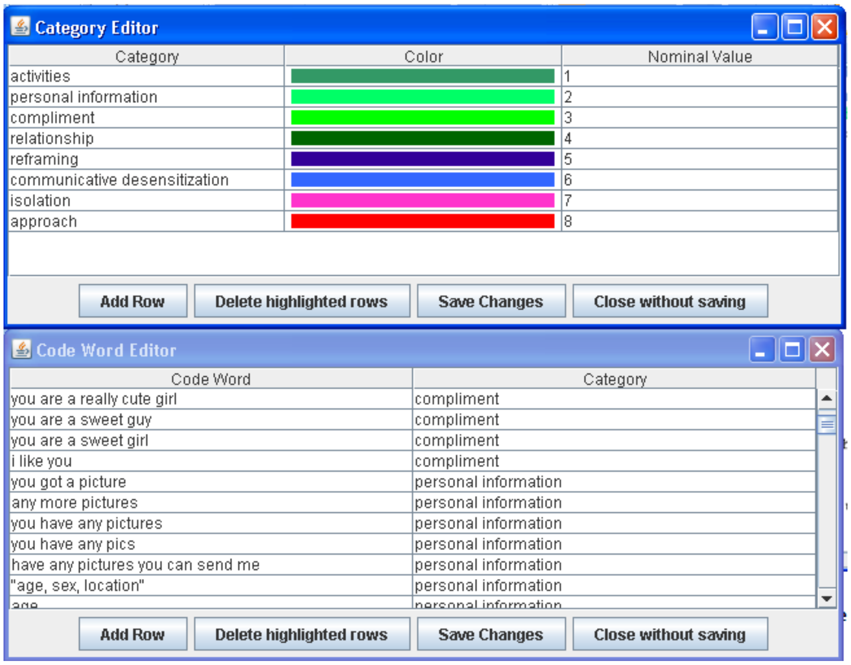
Phrase	**Coding Category**
are you safe to meet	Approach
i just want to meet	Approach
i just want to meet and mess around	Approach
how cum	Communicative Desensitization
if i don't cum right back	Communicative Desensitization
i want to cum down there	Communicative Desensitization
i just want to gobble you up	Communicative Desensitization
you are a really cute girl	Compliment
you are a sweet girl	Compliment
are you alone	Isolation
do you have many friends	Isolation
let's have fun together	Reframing
let's play a make believe game	Reframing
there is nothing wrong with doing that	Reframing

Fig. 1. A coded transcript for the Perverted Justice project



peekaboo1293 : oh my gosh "your, you are" sooo cute
armysgt1961 : thanks
armysgt1961 : u send yours
peekaboo1293 : be right back im not very good at this yet i will sendf
armysgt1961 : sure
7 :10:17 pm armysgt1961: nice u have any **naked** ones
7 :10:54 pm peekaboo1293: well no **silly**
7 :10:56 pm peekaboo1293: hehehe
7 :11:23 pm peekaboo1293: my **dad** sent me a web **camera** though and **mom** broke it with a hammer and scremed at me
7 :11:50 pm armysgt1961: were u taking your cloths off for people
7 :12:09 pm peekaboo1293: no i was just talking to **dad** and she hatez him
7 :12:24 pm armysgt1961: oh so u didnt get dirty on cam
7 :12:40 pm peekaboo1293: i only had it "to, two" dayz before she found it
7 :12:45 pm armysgt1961: oh
7 :12:54 pm armysgt1961: so u have big **tits**
7 :13:01 pm peekaboo1293: no
7 :13:11 pm armysgt1961: u **shave**
7 :13:39 pm peekaboo1293: yes but i do not have much
7 :13:46 pm armysgt1961: ok
7 :13:54 pm armysgt1961: can i get your **number**
7 :14:12 pm peekaboo1293: im not suposed "to, two" give my **number** if i do not know you yet

Fig. 2. Categories and Sample Code Words for the Perverted Justice Project



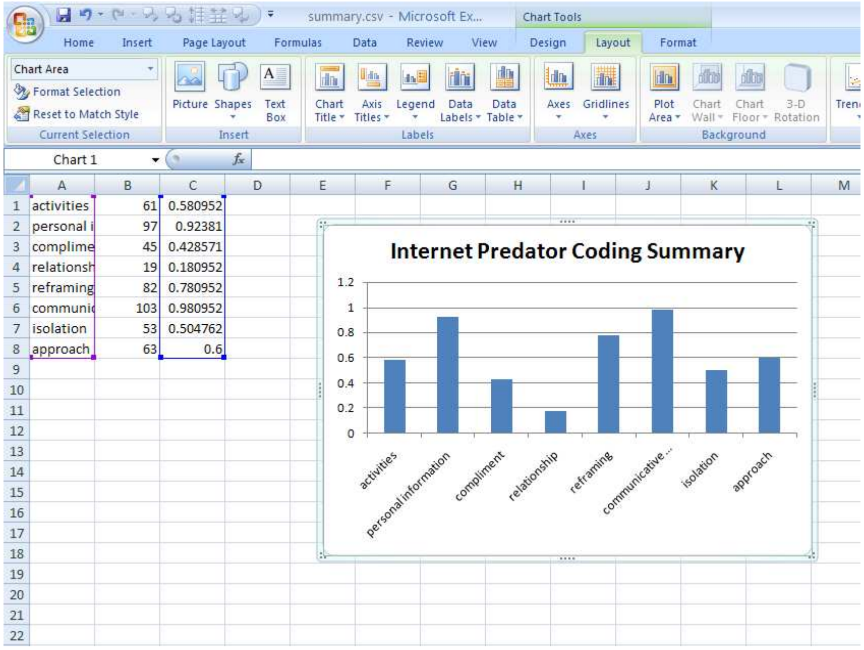
Category Editor

Category	Color	Nominal Value
activities	🟩	1
personal information	🟩	2
compliment	🟩	3
relationship	🟩	4
reframing	🟪	5
communicative desensitization	🟦	6
isolation	🟪	7
approach	🟥	8

Code Word Editor

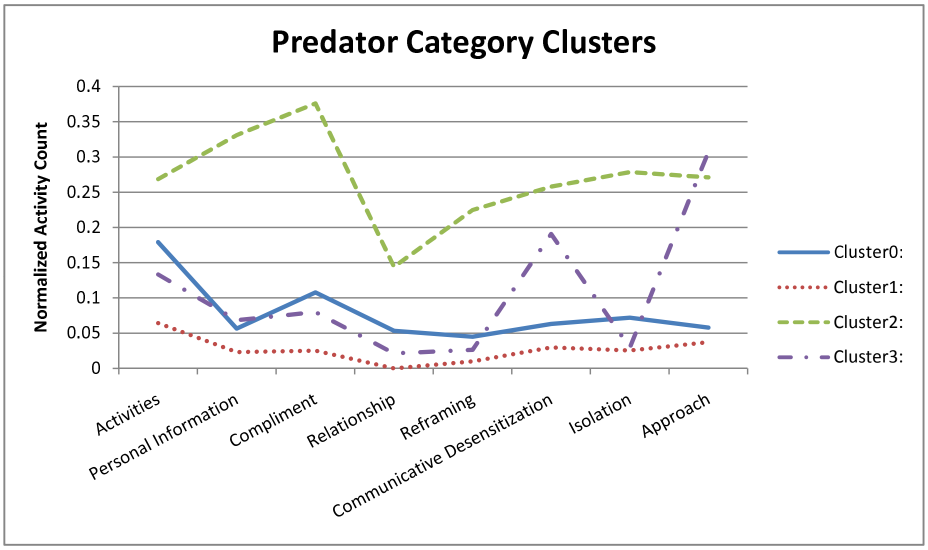
Code Word	Category
you are a really cute girl	compliment
you are a sweet guy	compliment
you are a sweet girl	compliment
i like you	compliment
you got a picture	personal information
any more pictures	personal information
you have any pictures	personal information
you have any pics	personal information
have any pictures you can send me	personal information
"age, sex, location"	personal information
age	personal information

Fig. 3. Summary Statistics from a Batch of Transcripts



	A	B	C
1	activities	61	0.580952
2	personal i	97	0.92381
3	complime	45	0.428571
4	relationsh	19	0.180952
5	reframing	82	0.780952
6	communic	103	0.980952
7	isolation	53	0.504762
8	approach	63	0.6

Fig. 4. Initial clustering of predator type



6 Experimental Methodologies and Results

We have performed two sets of text mining experiments using this data. The first attempts to categorize communication strategies, trying to distinguish between predator and victim, or predatory and normal chat. The second uses clustering to determine whether or not different communicative strategies are used for luring children.

6.1 Categorizing Predator vs. Victim

In order to provide a baseline for the usefulness of the code book for detection of online predation, we ran two small categorization experiments. In the first experiment, we coded 16 transcripts in two ways: first we coded the predator dialog (so only phrases used by the predator were recorded), and then we coded for the victim. Thus, we had 32 instances, and each instance had a count of the phrases in each of the coding categories (eight attributes). Our class attribute was binary (predator or victim).

We used the J48 classifier within the Weka suite of data mining tools [22] to build a decision tree to predict whether the coded dialog was predator or victim. The J48 classifier builds a C4.5 decision tree with reduced-error pruning [19]. This experiment is similar to [17], but Pendar used a bag-of-words approach and an instance-based learner. The classifier correctly predicted the class 60 percent of the time, a slight improvement over the 50 percent baseline. This is remarkable when we consider the fact that we were coding individuals who were in conversation with each other, and therefore the terminology used was similar. Stratified three-fold cross validation, as implemented within Weka, was used to evaluate the results.

In a second experiment we built a C4.5 decision tree to distinguish between PJ and ChatTrack transcripts. We coded 15 PJ transcripts (both victim and predator dialog) and 14 transcripts from the ChatTrack data set [2]. The classifier that was built was able to distinguish the PJ transcripts 93 percent of the time. We also used stratified 3-fold cross validation for evaluation in these experiments.

These are simple features and algorithms, and the small sample sizes are clearly not sufficient for evaluation of the categorization model. Thus, in future work we plan to develop better training and test data sets, as well as build better classifiers using larger samples.

6.2 Clustering Predator Communication

As we analyzed the PJ transcripts, we noticed recurring patterns within the dialog used by the suspects and began to wonder if we could cluster different types of predators via their language pattern usage.

We chose the k-means [7] clustering algorithm because it is known to be both simple and effective. The k-means algorithm partitions a set of objects into k sub-classes. It attempts to find the centers of natural clusters in the data by assuming that the object attributes form a vector space, and minimizing the intra-cluster variance. Thus, k-means generally forms tight, circular clusters around a centroid, and the algorithm outputs this centroid. k-means is particularly applicable to numeric attributes, and all of our attributes are numeric.

In our experiments, we counted the number of phrases in each of the eight coding categories for all 288 transcripts (predator only), and created an 8-dimensional vector for each instance. Thus, we used the same attributes that were used in the categorization experiments, but we were able to use all of the PJ transcripts. The vectors were column normalized by dividing by the maximum value in each column (i.e., all *activities* values were divided by the maximum value for *activities*). These vectors were then input to the k-means algorithm, and a set of clusters was determined.

The user must provide a value of k to the k-means clustering tool, and we were unsure about the number of categories of suspects that we might find, so we tried various values for k. We found that $k = 4$ produced the best result (the minimum intra-cluster variance), suggesting the hypothesis that there are four different types of Internet predators. More work is needed to determine labels for these categories of suspects. The centroid for each cluster appears in Figure 4. This figure clearly shows that some suspects spend more time overall with the victim (lines that are higher on the graph) and also that suspects in different clusters used different strategies during their conversations (as determined by line shape). For example, cluster 2 has a higher ratio of compliments vs. communicative desensitization as compared to cluster 3.

7 Conclusions

We have described an interesting and relevant project on cyber-violence. We have also described the first steps toward the development of technology that will detect cyber-violence as it occurs. Our research is well-grounded in both communication theory and text mining practice. Eventually we hope to develop open-source software that will notify parents and authorities when a child is being targeted by a sexual predator over the Internet.

8 Acknowledgements

This work was supported in part by the Ursinus College Summer Fellows program. The authors thank Dr. Susan Gauch and her students for providing the ChatTrack data, and Dr. Nick Pendar for his helpful advice on acquiring the Perverted-justice.com transcripts. Our thanks extend also to the many students and colleagues in both Mathematics and Computer Science and Media and Communication Studies departments at Ursinus College who have provided support and input to this project, as well as to the anonymous reviews who provided many helpful comments.

References

[1] E. Acar, S. Camtepe, M. Krishnamoorthy, and B. Yener. Modeling and Multiway Analysis of Chatroom Tensors. In *IEEE International Conference on Intelligence and Security Informatics*, 2005.

[2] J. Bengel, S. Gauch, E. Mittur, and R.Vijayaraghavan. ChatTrack: Chat room topic detection using classification. In *Second Symposium on Intelligence and Security Informatics*, 2004.

[3] S. Camtepe, M. Krishnamoorthy, and B. Yener. A tool for Internet chatroom surveillance. In *Second Symposium on Intelligence and Security Informatics*, 2004.

[4] R. Criado, J. Flores, M. Gonzlez-Vasco, and J. Pello. Choosing a leader on a complex network. *Journal of Computational and Applied Mathematics*, 1:10–17, 2007.

[5] N.W. Van Dyke, H. Lieberman, and P. Maes. Butterfly: a conversation-finding agent for Internet relay chat. In *Proceedings of the 4th international Conference on intelligent User interfaces*, 2009.

[6] eBlaster$^{\text{TM}}$, December 2008. http://www.eblaster.com/.

[7] J.A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[8] D. Hughes, P. Rayson, J. Walkerdine, K. Lee, P. Greenwood, A. Rashid, C. MayChahal, and M. Brennan. Supporting law enforcement in digital communities through natural language analysis. In *Proceedings of the 2nd International Workshop on Computational Forensics (IWCF'08)*, 2008.

[9] Q. Jones, M. Moldovan, D. Raban, and B. Butler. Empirical evidence of information overload constraining chat channel community interactions. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work*, 2008.

[10] V. Latora and M. Marchiori. How the science of complex networks can help developing strategies against terrorism. *Chaos, Solitons and Fractals*, pages 69–75, 2003.

[11] A. Leatherman. Luring language and virtual victims: Coding cyber-predators online communicative behavior. Technical report, Ursinus College, Collegeville, PA, USA, 2009.

[12] PC Mag, Dec 2008. http://www.pcmag.com/article2/0,2817,2335485,00.asp.

[13] M.J. Muller, M.E. Raven, S. Kogan, D.R. Millen, and K. Carey. Introducing chat into business organizations: toward an instant messaging maturity model. In *Proceedings of the 2003 international ACM SIGGROUP Conference on Supporting Group Work*, 2003.

[14] Net Nanny$^{\text{TM}}$, Dec 2008. http://www.netnanny.com/.

[15] NCMEC. National center for missing and exploited children, October 2008. http://www.missingkids.com/en_US/documents/CyberTiplineFactSheet.pdf.

[16] L.L. Olson, J. Daggs, B. Ellevold, and T. Rogers. Entrapping the innocent: Toward a theory of child sexual predators' luring communication. *Communication Theory*, 17(3):231–251, 2007.

[17] N. Pendar. Toward spotting the pedophile: Telling victim from predator in text chats. In *Proceedings of the First IEEE International Conference on Semantic Computing*, pages 235–241, 2007.

[18] Perverted-Justice.com. Perverted justice, August 2008. www.Perverted-justice.com.

[19] R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[20] D. Riffe, S. Lacy, and F.G. Fico. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Lawrence Erlbaum Associates, 1998.

[21] V.H. Tuulos and H. Tirri. Combining topic models and social networks for chat data mining. In *Proceedings of the 2004 IEEE/WIC/ACM international Conference on Web intelligence*, pages 235–241, 2004.

[22] I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, 2005.