A Systematic Survey of Online Data Mining Technology Intended for Law Enforcement

MATTHEW EDWARDS, AWAIS RASHID and PAUL RAYSON, Lancaster University

As more and more crime takes on a digital aspect, law enforcement bodies must tackle an online environment generating huge volumes of data. Manual inspections becoming increasingly infeasible, law enforcement bodies are optimising online investigations through data-mining technologies. Such technologies must be well-designed and rigorously grounded, yet no survey of the online data-mining literature exists which examines their techniques, applications and rigour. This article remedies this gap through a systematic mapping study describing online data mining literature which visibly targets law enforcement applications, using evidence-based practices in survey-making to produce a replicable analysis which can be methodologically examined for deficiencies.

Categories and Subject Descriptors: [**Applied computing**]: Surveillance mechanisms; [**Social and professional topics**]: Government surveillance; [**General and reference**]: Surveys and overviews

General Terms: Documentation, Security, Measurement

Additional Key Words and Phrases: Systematic survey, literature review, online data mining, OSINT, open-source intelligence, law enforcement, cybercrime

ACM Reference Format:
Edwards, M., Rashid, A., and Rayson P., 2015. A Systematic Survey of Online Data Mining Technology Intended for Law Enforcement *ACM Comput. Surv.* 48, 1, Article 1 (July 2015), 56 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

The increasing fusion of digital and physical life presents two key challenges to law enforcement agencies: the population's online presence means law enforcement must learn to adapt to crimes taking place only online, and yet this increasingly digital interaction also provides a valuable and increasingly necessary evidential resource for officers investigating both physical and online crimes. With Internet accessibility widening and more and more crime taking on a digital aspect, online investigation is becoming a critical tool for law enforcement organisations, and scientific examination of such processes becomes ever more a key issue.

With manual inspection of online information being labour-intensive and the unprecedented scale of information online, law enforcement agencies seek to optimise their surveillance or investigation of online data sources through the use of various data mining technologies. This behaviour has diverse implications, including raising social and ethical questions about privacy and the role of state surveillance as well as posing unique technical challenges for the data mining technologies being employed.

The field of computer science, particularly data mining research, has a key role to play in shaping the future of these investigations. To date there is no comprehensive

Authors' address: Security Lancaster, School of Computing and Communications, InfoLab21, Lancaster University, UK

DOI: http://dx.doi.org/10.1145/0000000.0000000

survey which draws together those problems already being addressed in the computer science literature around this field and highlights which areas need to be given more attention. Our aim in this paper is to remedy this gap by identifying open research problems and a research agenda for the community at large.

While the broader aim of our study is to survey the literature for gaps, guiding questions were developed to help target the extraction of information:

(1) What are the problems (crimes, investigative requirements) are being addressed in the literature?
(2) Which online data sources are being used?
(3) What are the methods (data mining techniques) which are being employed to provide solutions?
(4) Are studies making use of multiple data sources?
(5) Are studies validating their contribution's utility to law enforcement practitioners?

Drawing on a base of computer science literature, we carry out a comprehensive search for and evaluation of peer-reviewed studies concerning the mining of digital data sources for law enforcement purposes. Our study is aimed at examining the visible state-of-the-art with regards to both techniques and the criminal activities being addressed. This review will not only help identify gaps to be addressed in the state of the art but also inform ongoing public debate about the privacy implications of such technologies.

Taking inspiration from a recent trend towards evidence-based practice in software engineering, we perform a systematic mapping study (SMS), with the intent of producing a survey which not only covers the available literature and has replicable results, but also can be methodologically examined for deficiencies.

It is an important part of a review's design to make clear not only the scope of the survey, but the intended purpose. Our primary concern is that the results of the survey identify gaps in the published research regarding data mining of online sources for crime detection or investigation purposes. The results of the study can then be used to inform ongoing work in this area.

Two terms should be considered as key here. Firstly, the specification of *online* data sources, meaning data which can be gathered from examination of Internet-based sources. This separates our study from other areas of research such as work which makes use of restricted criminal records or other police databases, as well as distancing the study from many areas of digital forensics which focus on the investigation of hard disks or active machine memory. Secondly, the specification of data mining with application in *crime* detection. While many data mining methods have plausible application in this domain, we only consider publications which make an explicit reference to such employment. We should also make it clear that, for our purposes, we are excluding from this definition data mining which is performed for purely Information Security reasons, thus leaving aside a mature literature on Intrusion Detection Systems and related work which has already been surveyed [Axelsson 2000].

In Section 2, we outline the automated search that we carried out, and the subsequent systematic screening, quality analysis and composition of results. In Section 3, we present an overview of the literature covered in the study, identifying the data-mining methods being used and the applications to which they are being put. In Section 4, we describe broad patterns and publication trends visible across our corpus and in Section 5 we conclude with an analysis of the state-of-the-art we have uncovered, while noting limitations to the scope of our review.

Table I. Search queries were constructed by the combination of quoted forms of the following term-sets

First Term-set
Crime
Police
Law Enforcement

Second Term-set
Artificial Intelligence
Data Fusion
Data Mining
Information Fusion
Natural Language Processing
Machine Learning
Social Network Analysis
Text Mining

2. METHOD

A systematic literature review (SLR) attempts to provide answers to a specific research question through a transparent and objective approach to the collection and synthesis of existing scientific literature on the topic. This method can be contrasted with non-systematic literature reviews, whose contents may be unrepresentative of a field of research due to, for example, narrative-driven distortion, where reviewers include only papers whose findings support their line of argument; or narrowness of study, where reviewers are unaware of a large number of relevant publications because they were never personally exposed to them.

In designing our survey, we drew heavily on the work of [Okoli and Schabram 2010], which recommends an explicit eight-step system to SLRs. Certain key deviations from this procedure adapted the process to the specific form of systematic literature review we engaged in, namely, a systematic mapping study (SMS). A description of the main features of SMSs is provided by [Budgen et al. 2008], but the key distinction between the two types can be summarised as an SMS being an SLR which aims to more broadly survey the available literature rather than answer specific research questions.

2.1. Search Procedure

The search process, carried out between December 2012 and January 2013, was designed as an automated search, targeting four key computer science publication databases: IEEExplore [IEEE 2013], The ACM Digital Library[ACM 2013], SpringerLink[Springer 2013] and ScienceDirect[Elsevier 2013]. In each database, 24 queries were carried out[1], as defined in Table I and the resulting papers' metadata collected. In total, 13,246 unique results were collected[2].

The title and abstract of each result were then examined by a reviewer and classified as either relevant or irrelevant to the study according to the following criteria. If the answer to either of these questions was no, then the study was not included.

(1) Does the study appear to address or make use of online data — that is, types of data which may be discovered online (either on the Web or otherwise)? Specifically excluded are data such as disk images from a crime scene and restricted databases.
(2) Does the study have a stated or heavily implied application in law enforcement, crime detection, monitoring or investigation? For the purposes of this study, studies dealing primarily with attacks against computer infrastructure (intrusion detection systems) are specifically excluded.

[1]These were identified through a series of pilot searches to determine the relevance and effectiveness of particular search strings.
[2]In the ScienceDirect search engine, queries matching over 1000 results will only return the first 1000 items, and a similar effect at 2000 results is observed in the IEEExplore search engine. None of our search queries reached such limits.

(3) Does the study appear to have a methodology which involves either fully automated or machine-assisted processing of data?

Following the screening process, we gathered all references from each of the 116 accepted papers, along with all papers identified by Google Scholar as having cited the accepted items. These were also put through the screening process above.

2.2. Analysis Procedure

Following the search and screening stages, we obtained the full text of each accepted paper, along with a full citation. Items whose full text could not be located, those which turned out to be in a non-English language or which on review of the full text did not meet the screening criteria, were discarded. This resulted in a final included list of 206 accepted papers. Each paper was examined to answer specific questions regarding its quality. Each paper was given a value of 0, 0.5 or 1 for each of the following points, with 0 being a negative response, 0.5 being a partial positive response and 1 being a positive response. The overall quality rating for each item is the sum of its individual scores on these responses.

(1) Does the paper outline its method in a replicable manner?
(2) Does the paper make its evaluation replicable?
(3) Where evaluation is qualitative, does the evaluation make use of domain experts?
(4) Where evaluation is quantitative, is an appropriate statistical assessment of results carried out?

Alongside the quality analysis, questions related to this review's main aims were answered for each paper.

(1) What problem[s] is (/are) being addressed?
(2) Which data source[s] is (/are) being used?
(3) What method[s] is (/are) being employed?
(4) Does it make use of multiple data sources?
(5) Does it validate the contribution's utility to practitioners?

All examination was carried out by the same examiner, making use of a predefined data extraction form made up of these questions. As well as grouping results based on paper authorship, we examined groupings according to whether papers were addressing a similar problem and whether they were using the same methods or data sources.

2.3. Limitations

There are a number of limitations inherent in the methodology of this review which should be considered when examining its results. The most prominent is that the judgement of whether any candidate article met the inclusion criteria was performed by only one person. This increases the risk of reviewer bias and human error affecting results, especially with regards to the classification of borderline cases. An alternate methodology using multiple reviewers could mitigate this risk, but raises new problems of inter-reviewer consistency, not to mention practical considerations of funding and training. A sample of candidate articles were independently reviewed by a second reviewer following the same protocol, with a disagreement rate of 15% between the two reviewers over the acceptance of articles.

The range of our query terms was necessarily limited in order to produce a manageable volume of candidate articles to be considered for inclusion. While efforts were made to select the most relevant query terms, a side effect of this necessarily limited

range is that some relevant works may not have been considered for inclusion in the review.

The use of search terms which explicitly reference crime and law enforcement meant that only papers which explicitly identified their methods as being relevant to such an application area would be included. As such, papers describing methods which are or may be useful to law enforcement in online data mining tasks may have escaped inclusion. Additionally, we must consider the possibility that research behind tools developed specifically for law enforcement application in this domain may have been withheld from publication, either due to commercial confidentiality or for considerations of the public good. Indeed, since the collection of data for this review was completed, mainstream attention has been drawn to the revealed reality of this possibility [Greenwald et al. 2013].

The initial sources considered for analysis were four large computer science publication databases. While other sources were therefore initially excluded, there was no source restriction (beyond attainability) placed on items from papers' citation lists or papers identified by Google Scholar as having cited included titles. In total, 93 conferences and 59 journals provided articles. The most prominent venues in the accepted paper corpus were the IEEE Intelligence and Security Informatics conference, and the Digital Investigation journal.

3. RESULTS

Within the 206 papers reviewed, we identified 8 broad problem topics that the papers sought to address; represented as columns in Table II:

(1) Financial crime, which relates to fraud or crimes like copyright infringement whose principle damage is economic. Financial criminal activity does not always leave visible traces in online data sources, as much financial information is kept private. However, a number of specific areas are visible. Primarily there is the example of copyright infringement, one of the more widespread criminal activities visible online, can be examined via a number of public interfaces, not least the P2P filesharing mechanisms often used to commit it. Another online lens into financial crime comes via online auction sites, whose transactions are to some extent available to the public for scrutiny. Finally, when an allegation of fraud is being investigated, the email records of suspects can indicate collusion and/or implicate co-conspirators.
(2) Cybercrime, which is intended to cover crimes focused on information systems. While a majority of cybercrime will take place online, and leave traces in data sources such as firewall and server logs, much of this category of crime is excluded from our study, as it involves a vast body of work in intrusion detection and similar fields. The works which we did consider in the scope of this study which dealt with cybercrime mostly focused on the social and economic background to cybercriminal activity, often mined from online fora where criminals share or sell information.
(3) Criminal threats or harassment. The type of threat dealt with in this category ranges from the identification of serious bomb and murder threats in messages, to filtering instances of 'trolling', where the aim is merely to provoke shock. Identifying such messages has proven more difficult than the identification of spam mail, due to the varied possible representations of threats, and some form of sentiment analysis may prove critical to any solution to this problem.
(4) Police intelligence — the creation of tools to support government or law-enforcement in the general detection of crime – is one of broader problem categories when it comes to criminal acts addressed. Generally speaking, the interest is in either the investigation of criminal organisations or some spatially-restricted

prediction of crime, but other minor crimes are also addressed. A large body of this work aims to augment police investigations by filtering knowledge from web-based news articles, the intent being to provide situational awareness and keep investigators abreast with public information.

(5) Crimes against children, including grooming and child trafficking. Online grooming of children has become very high-profile, and a number of publications focus on means of detecting it – either by identifying the age of a conversational partner or by directly modelling predatory behaviour in instant messaging conversation. Other online data is also examined in relation to these crimes, most significantly filesharing networks which are often used to distribute images or videos of child abuse.

(6) Criminal or otherwise links to extremism and terrorism. The specific nature of the problem addressed is entirely focused on either white supremacists from the United States or else Islamic fundamentalists. In both cases, the primary online lenses into the groups are the online fora they use to discuss matters pertaining to their ideologies, and much of the research effort is in examining their social networks and analysing the persuasive techniques they have employed.

(7) Identification of online individuals in criminal contexts. The Internet being a theoretically anonymous medium, a critical issue for many criminal matters is identifying a person. Given the highly-textual nature of much online activity, means for identifying a person from their writing dominate this problem, but data sources can be diverse, including email, online posts, instant messaging logs and even images.

Table II. Number of studies for each Method and Problem category. Row totals indicate the number of papers per method, but rows will not add up to these totals due to multiple-problem studies. Similarly, the totals will not sum to 206 because of multiple method categories being assigned to several papers and the column totals indicate papers per problem, not counting duplicates across rows.

	Finance	Cybercrime	Harassment	Unclear	Intelligence	Children	Extremism	Identification	Total
CV	3	0	2	2	0	6	2	6	21
SNA	2	2	0	0	8	0	9	1	22
IE	2	4	0	2	14	6	14	0	39
ML	3	1	7	4	3	4	10	11	43
NLP:AA	0	1	0	0	0	0	1	28	30
NLP:AP	0	0	1	1	0	14	2	1	19
NLP:SA	0	0	5	0	0	0	8	0	13
NLP:TC	2	1	4	1	3	7	4	1	23
NLP:O	0	0	0	7	0	1	0	0	8
ETC	2	6	0	2	3	5	4	9	29
Total	12	14	14	15	24	37	47	48	206

Additionally, some papers made reference to criminality in a broad sense, but appeared not to address specific crimes or categories of crime. These were labelled as 'Unclear'. Some papers were labelled as addressing multiple problem topics. The most prominent problem topic was online identification — broadly, the problem of identifying individuals based on only online data, a problem which was particularly related to the analysis of malicious emails and language-based classifiers. This topic was closely matched numerically by those papers addressing extremism or terrorism. Generally, the crimes most often focused on were terrorism or extremism related, or else linked to crimes against children.

Also identified were five broad classes of method common to several papers, including different types of online data being gathered and analysed. These are represented as rows in Table II. Of the method categories, the largest was natural language processing (NLP), with machine learning (ML), information extraction (IE), social net-

work analysis (SNA) and computer vision (CV) falling far behind. Some papers did not fit into these five categories neatly, so a miscellaneous ('ETC') category houses them. The NLP subsection, due to being much greater in size, is broken down into Authorship Attribution (AA), Author Profiling (AP), Sentiment Analysis (SA), Text Classification (TC) and Other Methods (O). Data types observed included web page and forum contents, including data from social networks, email data, instant messaging data and network traces.

We next provide a short analysis of each paper collected. These reviews are collected by topic area in order to facilitate access for researchers from different fields. Each of the following subsections discusses papers grouped by the broad class of the method they follow, with the nested subsections further dividing the topic by the broad class of the problem addressed.

3.1. Computer Vision

3.1.1. Identification with Computer Vision. Identification tasks in computer vision mostly rest on visual identity, a troublesome concept in an environment where a new face – or no face at all – is so simple to obtain. The majority of the uncovered literature looks at a particular subset of visual identity – recognising the visual representation of a person in a particular online environment, such as the game Second Life. The challenges here have significant overlap with the development of facial recognition systems in general, including automating adequate pre-processing to find comparable facial images and minimizing the runtime of any face-matching system. Additional problems are raised by the possibility of 'different worlds' where the same person may use a different visual identity, something not generally possible in the real world. Some authors posit that avatars made by the same person may be consistent across different services, or in some way connected to their actual visual appearance, but this appears unproven.

[Klare et al. 2011] discuss the application of computer vision to the facial recognition of online avatars, justifying the research topic with reference to criminals and especially terrorist groups using virtual environments — particularly the online game Second Life — as training simulators, and studying two key applications. The first of these involved inter-reality avatar-to-photograph matching, where avatar faces generated from photographs were matched against other photographs of the same subject. Off-the-shelf face recognition technology sufficed here, given that the avatar was generated automatically from an actual image of the target. While a useful result, this merely suggests that their automatic avatar generation system preserves key information for facial recognition, and not that users will do so when crafting their own avatars. Addressing this, the second application used a collection of actual Second Life avatars and attempted to match different images of these avatars. The authors discovered acceptable classification accuracy, although they reported a performance bottleneck in face and eye detection, with significant improvement in accuracy coming from manual eye location. A remaining question left unanswered is whether avatars can be recognised across different digital platforms.

[Baili et al. 2011] introduce a new method for avatar facial recognition, employing wavelet transforms alongside a hierarchical multi-scale local binary pattern (HMLBP). The authors build on other developments in this area, including earlier propositions of the use of wavelet transforms and local binary patterns. The study focuses on re-detecting avatars from Second Life and Entropia — another online world — from pictures in different poses. While the results show improvements over earlier work, the method still relies on correctly-cropped input, achieved in the dataset through manual effort. This hurdle must be addressed for a completely automated system for detection and recognition of avatar faces.

[Mohamed and Yampolskiy 2012a; Mohamed and Yampolskiy 2012b] both continue with the use of wavelet-based local binary patterns, for re-detecting avatar faces, but with new variations, one making use of Eigenfaces and the other making use of directional statistical features. Both experiments re-used the Second Life and Entropia datasets presented in the previous publications. In the first of these two publications, the authors mention the design of a fully automated system for addressing the cropping problem as a target for ongoing work. In the second, the authors include a comparison of the classification time for a number of leading techniques, an important consideration for any near-realtime detection and recognition system.

In summary, this niche area of facial recognition has shown significant development with regards to the core task of re-identification of avatars from within an online environment like Second Life or Entropia. Still awaiting research is meaningful deployment of these classification systems, with work evaluating automated cropping and exploring usable interfaces to the online environment appearing on the horizon. In crossing realities, some initial work evaluating the traceability of the results of automatic avatar generation has been undertaken (e.g. [Klare et al. 2011]), but it remains to be seen if links between a user's visual appearance and virtual avatar can be determined, or indeed if avatars are consistent across online environments.

Other computer vision work on identification includes that of [Zhang et al. 2009] and [Wei et al. 2008]. [Zhang et al. 2009] focus specifically on image spam — spam emails which make use of text presented as images in order to avoid text-based filtering techniques. They cluster spam images by visual features, and report a high success rate with respect to a manually-identified ground truth. Their approach analyses images for evidence of various types of template being reused by spammers, as divined by layout, colour and texture. What they do not report in this paper is whether typical spam filtering, or indeed linguistic analysis, can be applied to text extracted via optical character recognition. [Wei et al. 2008] describe a more general spam-origin toolkit which makes use of website image comparison (from following links in spam emails) as one tool in its arsenal, alongside WHOIS and IP lookup information and more typical email attributes such as subject lines. While their analysis of a researcher-gathered dataset appears to reveal interesting clusters of spam, and the utility of the website image comparison in particular is demonstrated by all but one cluster centring on one website image, an evaluation against manually-identified ground truth would be stronger justification of the method's validity.

3.1.2. Computer Vision and Crimes against Children. Computer vision's role in preventing crimes against children is mostly connected to the recognition of child abuse imagery in an online population of images. This can be either searching for known child abuse imagery in order to filter it or identify distributors, or else identifying new examples. A common problem in the second domain is distinguishing between ordinary adult pornographic material and images of children in pornographic context, which is highly visually similar.

[Haggerty et al. 2008] present the FORWEB system, which focuses on forensic applications of existing signature analysis and web-crawling systems, the key motivation of the authors being to automate the search and discovery process involving networked servers. They clearly distinguish their approach from established storage media analysis tools like *EnCase* in much the same way as this review separates these areas of study. Their file-fingerprinting scheme aims to identify images based on properties which are much less likely to be affected by the simple alterations which throw off hash-based file comparisons, and combined with the spidering bot this becomes a useful tool for detecting known malicious images. Like all such tools, this relies on the existence of an up-to-date database of known suspect files (a resource which may in

itself bring significant performance overhead) and does not address the aim of identifying unknown media of a suspect nature.

[Ibrahim 2009] takes a quite different approach, detailing a method whereby image files are identified on the network, reconstructed and then classified as either child abuse media or not by both a machine learning system and an image matching system similar to FORWEB's fingerprinting scheme — the intent being that such a system would be installed on network boundaries to filter child abuse material. Aside from concerns about network performance, a major weakness of their trial of the system is that for legal reasons their system only attempted to distinguish between nude and non-nude images, which is clearly a far easier task than distinguishing child features from adult ones.

This also applies to [Uke and Thool 2012], which opens with a motivation of preventing child abuse, but in a sudden switch focuses on identifying pornographic video scenes as a proxy. The paper neither provides an implementation nor an evaluation of the system, merely outlining methods to be explored.

This more difficult child-recognition task is tackled by [Islam et al. 2011]. They focus on detecting child exploitation material on social networks, but contribute an algorithm which could equally apply to P2P networks. Their skin detection technique is specifically tuned for the detection of child skin tones, and they also suggest techniques which help detect pornographic context. While these proposed methods are indeed critical research areas for computer vision in child protection, the authors do not provide the results of an evaluation or even a completed system.

Also attempting this task, [Shupo et al. 2006] aim at detecting child abuse material on the network level, as an alternative to manually searching suspicious venues or application-layer networks. Their classification system, consisting of a stochastic learning weak estimator combined with a linear classifier, was trained and trialled on a sanitised dataset provided by Candian law enforcement, a rare example of child abuse imagery being available for training. Notably, the classifier was tested on partial as well as whole images, taking into account likely fragmentation of images over a network link. While valuable for this alone, the classifiers being trialled still produced less-than-ideal rates of false positives for a tool to be deployed at the network level. Estimations of the base rates for child abuse material versus adult pornography suggest that alerts generated may be mostly incorrect – though this does not invalidate the utility of the classifier as a tool for network monitoring, given appropriate human supervision.

3.1.3. Computer Vision and Threats or Harassment. There is a specific use-case for computer vision in detecting visual (as opposed to verbal or written) forms of harassment in video communication. As these systems are intended to be deployed on large video-streaming populations, performance is critical to creating a deployable solution. The particular misbehaviour discussed in these papers has some particular visual challenges regarding lighting and the potential detection of faces.

[Xing et al. 2011] outline an unusual problem specific to the anonymous video-pairing site Chatroulette – users exposing themselves to strangers. They stress that a considerable proportion of Chatroulette's userbase would be classed as minors, and that site policy on age restriction and obscenity is difficult to enforce due to the anonymous nature of the service. The authors note that de-anonymising the service could solve this issue, but would damage one of the site's key features in the process, and so turn to video-analytic approaches for detecting offensive users. Their key observations include that misbehaving users usually hide their faces, and that misbehaving users' images differ from pornographic images in that they often stay partially clothed and only expose their genitals. Their system therefore focuses on detecting user faces as a

key feature in making a decision about the probability of misbehaviour, along with a novel skin detection system which takes into account the abnormal context of webcam images. While they evaluate their classification accuracy, they do not report on performance speed, an issue which would appear critical for their problem domain, as extra delay in connection would impair the appeal of the Chatroulette service.

[Cheng et al. 2012] refine this first approach into a fine-grained cascaded classification solution which filters out easily disambiguated images earlier in the process for the sake of efficiency. They also integrate new work on gathering contextual information from webcam images and a new fusion system for combining probabilities of misbehaviour. The improved system is evaluated against their older system and other contenders, showing significant improvement, particularly in regard to the previously unaddressed matter of classification latency.

3.1.4. Computer Vision and Terrorism/Extremism. A limited deployment of computer vision techniques in counter-terrorism is seen in the context of the analysis of propaganda videos released by jihadists. The problems they address are of coding the content of the videos in a pseudo-automated fashion, where correct identification can be an aide to intelligence work.

[Salem et al. 2006] present an exploratory study of jihadi videos which attempts to highlight the research and intelligence need for automatic exploration of jihadi video content, and produce a tool to support manual coding of videos for this purpose. The results are demonstrative of the effectiveness of their analysis on a set of terrorist videos and not that of the performance of their coding toolkit.

[Salem et al. 2008] provide an extended version of the same research, again with more focus on the content analysis than on the support tool. In both cases, while the authors' work is presented as a stage towards automated video content analysis, the requirements for progression from manual intervention are not fully detailed.

3.1.5. Computer Vision and Financial Crime. Computer vision has seen deployment in anti-piracy efforts. The systems in this section are attempting to detect copies of restricted material from being distributed online by comparing content to a stored visual fingerprint of pirated material – techniques also deployed in detecting known child abuse media. The problems addressed in these publications are primarily infrastructural, attempting to resolve detection efforts with minimal impact on legitimate traffic.

[Yin et al. 2009a; Hui et al. 2009] describe a system for large-scale online monitoring at the Content Distribution Network level, wherein videos are fingerprinted based on certain visual cues and compared to a blacklist of pirated material. Their system is particularly notable due to the fact that it was actually deployed on a large CDN, although the evaluation presented seems to be from laboratory results rather than real-world performance. Nonetheless, it would appear that their system is resilient to minor tampering such as is common with pirated material. Notably for a system to be deployed at a large scale, the performance overhead is quite significant, with fingerprinting and search time together incurring a 40 second delay, raising questions about usability.

[Hui et al. 2012] address problems linked to the computational and networking overhead of this large-scale video processing by deploying server clusters closer to the user in the content distribution network and distributing tasks between nodes based on proximity and computational load. This results in reduced processing time as compared to existing approaches, but still requires well over a minute to perform detection on movie-length items. It would appear that, despite ongoing work to address this issue, there is scope for improvement in the efficiency and scalability of video copy detection.

3.1.6. Other Computer Vision Applications. [Hu et al. 2007] target the detection of pornography, but do so with reference to illegal or offensive activity — whether the authors suspect that pornography is illegal, or target illegal pornography particularly but work with proxy data, is unclear. Their method addresses not only image recognition, but also the text processing of suspected pornographic web content, combining this information in their classifier. Their contour-based detection method appears to perform better than region-based skin detection, specifically with regard to false positive rates including bikini or face-focused images.

[Wang et al. 2012] describe existing general-purpose information filtering systems which they suggest could be used to defend users against various types of information, insult or crime. A range of methods and systems for information filtering are outlined, but neither methods nor systems are subject to a great deal of scrutiny. How information filtering technologies such as those presented can be linked to the prevention of crime is also not clearly outlined.

3.2. Social Network Analysis

3.2.1. SNA and Terrorism/Extremism. As a set of tools for analysing communities and graphs, social network analysis has seen particular deployment in counter-terrorism context, where the analysis of groups can be useful in identifying key nodes and group behaviour. Particularly, it is applied to graphs which are mined from online forums and blogs, where relationships between individuals can be determined structurally from links.

[Chau and Xu 2007; Xu and Chau 2006], focus on mining and analysing online communities in blogs, specifically communities of blogs frequented by hate groups. These two studies both make use of the same 28 anti-black blogrings from the Xanga blogging platform. While the studies include semi-automatic detection of hate groups as a key aim, the selection process presented relies on manual filtering of search results. A more automated means of selecting hate groups could aid in making their approach generalisable.

[Ríos and Muñoz 2012], is the first of a number of studies making use of the Dark Web Forum Portal collection. The authors focus on detecting overlapping communities by using latent dirichlet allocation to detect topics, with a positive evaluation on an English-language forum from the Dark Web Portal. The treatment of networks as allowing members to be part of more than one community is perhaps a useful model, but whether topics of conversations reflect actual networks rather than simply ideological leanings is not clarified.

[Lenselink 2011] focuses on the process behind online radicalisation. This work includes a well-written motivating example, and a review of current theory related to online radicalisation, but most importantly for this review it also includes a social network analysis using forum data from two Dark Web fora, one from the middle-east and one from Europe. Interestingly, the author reports technical issues with a module of the Dark Web Portal. The analysis suggests that radicalisation is happening between the most involved members of the community, as identified by several measures of centrality.

[Yang and Ng 2008] gathered discussion data from MySpace, using the DBSCAN algorithm to cluster topics as points for a social network visualisation tool. While the level of detail in the description of the algorithm is adequate, the authors' choice of example in their demonstration of the tool is the only link specifically to terrorism. Further detail on what may constitute interesting patterns within the network resulting from their clustered topics would make the tool's utility to terrorism investigators clearer.

[Patil et al. 2012] describe the Dark Web Attribute System which applies content and link attributes to items from the Dark Web collection, calculating measures of technical sophistication for various linked terrorist websites. The evaluation lacks rigour, however, and doesn't effectively demonstrate what might well be useful annotation work.

[Chaurasia et al. 2012] describe the application of SNA techniques as part of a system for identifying and monitoring terrorists at the ISP level, also advocating their system's use for targeted disruption of terrorist networks through identifying key nodes. The paper describes only a theoretical system and provides no evaluation. Most pressingly for a paper advocating large-scale surveillance, they include no discussion of the likely rate of false positives. Their baseline is also likely to be misleading, as they base their threshold of typical terrorist behaviour on only terrorist content, ignoring the possibility that terrorist individuals may access other sites. A more behaviourally sound model of terrorist web usage would be of use in improving such a system.

[Negnevitsky et al. 2005] describe a method utilising social network analysis for detecting changes in a group's behavioural patterns, as observed via email communications. They particularly highlight homeland security and intelligence applications of this method. They do not provide an evaluation in this paper, but discuss their ongoing development of a simulated email dataset for that purpose. As they discuss, their current model does not handle dynamic social networks such as those they expect in real data, an area which needs redressing. A key limitation of any such simulation would be its validity as a predictor of performance on a real email network – it would seem more advisable to work with real email datasets in developing the analysis methods outlined, even where this means working with proxy data rather than actual terrorist network data.

[Sureka et al. 2010] analyse YouTube's social graph to discover extremist videos and communities. Their system works from a seed list of videos to discover YouTube videos which are hate speech and users advocating acts of aggression. The authors discuss the network properties of the connections they found – including the different types of YouTube network – alongside brief topic analysis of user comments. The main contribution here is the development of search support tools for an intelligence analyst, adding structure and ranking content, but there is limited comment on the scope of the approach.

In summary, social network analysis has been applied to a number of terrorism-related datasets with some success, but current studies tend to present either toolsets which, due to the nature of terrorist content, often cannot be evaluated easily, or else exploratory analyses of a particular network which demonstrate some value but do not generalise. A theme common to a small number of papers has been using topic analysis of text to better subdivide communities of interest, but it would appear that this approach has yet to be validated in a meaningful manner.

3.2.2. SNA and Police Intelligence. As with terrorist organisations, social network analysis has been applied to online information about criminal organisations, often mined from news reports or other unstructured text documents. This provision provides for opportunities – additional information on time or space of interactions may be available – but also additional challenges in that relationships are not necessarily correctly represented in such secondary sources.

[Peng and Wang 2008] provide a case study where link analysis – with links in the form of webpage co-occurrence – is used to trace a notorious violent criminal, producing link charts for known members of his gang and related individuals. The method presented relies on Google search results to identify relevant web pages, which may

lead to narrowed results due to personalisation if countermeasures are not taken. A comparison with other methods for identifying web sources could prove useful.

[Hosseinkhani et al. 2012] provide a review of web mining for input into criminal network analysis, and propose a framework which integrates the identification of crime hot spots and criminal communities into the workflow of a web crawling agent. Detail on how the more relevant tagging modules will be implemented is omitted.

[Tseng et al. 2012] focus on term networks, presenting a novel algorithm for key term extraction, and presenting a case study similar to that of [Peng and Wang 2008] where news related to a particular gangster was gathered and mined to describe relationships between gangsters. The term model presented appears more powerful than simple entity collocation, but the study presented does not make a convincing case for the utility of this method, demonstrating only simple relationships as could be found through more traditional means.

From this sample, the area of web mining criminal networks, like terrorism network analysis, appears to suffer from a lack of rigorous evaluation. Identification of a means of better evaluating the performance of information-gathering agents such as these could help focus research efforts. A standard marked dataset suitable for evaluation could be considered an initial step.

[Lauw et al. 2005] describe attempts to discover the social networks of criminals by mining spatio-temporal events such as web usage. A detailed explanation of the problem and algorithmic approach are given, and the theory is validated against a data set collected from a university campus' wireless network. While their system appears technologically sound and is well-presented, the intended deployment scenario is not clear.

[Karran and Llewellyn-Jones 2009] discuss integrating SNA concepts into common digital forensics practice for investigation of email. The validating case study involves transforming the Enron email dataset into a form suitable for social network analysis and highlighting key actors from within that dataset. As the thesis itself acknowledges, social network information is not 'hard' evidence which can be considered directly in court, being instead useful in guiding further investigation. The analysis of the Enron dataset (which we discuss further in Section 4.1.3) presented does show some utility, but it is worth noting that the analyst's interpretation of results seems likely to be informed by previous knowledge of the dataset's context. A blinded study would mitigate such issues.

[Dudas 2013] makes use of Twitter data and geolocation for building a social network based on ongoing terrorist events, and then provides a modifiable visualisation to aid interpretation. Several areas for ongoing development are highlighted, including incorporation of temporal and sentiment dimensions into the visualisation tool.

[Barbian 2011] theoretically demonstrates a means of detecting hidden friendships — relationships in a network which are not formal connections. While a potentially valuable intelligence tool, the paper does not provide an evaluation of this method's efficacy.

[Stolfo et al. 2006a] use SNA as part of a range of tools for investigating email data for various crime-related purposes. The social network analysis component is only one part of the tool, which is described only very briefly and not evaluated.

[Al-Zaidy et al. 2012] describe the process of mining and analysing criminal networks from collections of unstructured text documents, in an approach which relies on the recognition of named entities and the detection of prominent communities of connected names. Their approach was validated in a case study from a real cybercrime investigation, with an instant messaging database provided by law enforcement and their investigation being compared to an expert's manual analysis of the chat logs. It is notable that the analysis was guided by the researchers' own identification of sus-

picious information – while fully automated analysis is not necessarily desirable, for purposes of evaluation it is necessary to distinguish the performance of the support tool from the performance of the authors. A blinded study with a number of analysis engine users compared to a number of manual analysis users would provide more objective assessment of their network-mining engine's utility.

3.2.3. SNA and Cybercrime. [Ma et al. 2011] focus on the construction of social networks from email and blog data linked specifically to cybercriminal activity. The paper refers most often to cybercrime as its motivation, but also to terrorists who 'upload obscene pictures'. The degree to which authorship identification techniques were applied is unclear.

[Nirkhi et al. 2012] apply SNA techniques — as part of a toolkit with other subsystems — to help identify cybercriminals from email data. They appear to have implemented their system and even gathered a dataset (Enron) to trial it on, but provide no evaluation in this paper.

3.2.4. SNA and Financial Crime. [Gray and Debreceny 2007] address financial crime through application of social network analysis in mining corporate emails to prevent fraudulent transactions. Taking the approach that data outside the accounting information system should help protect against fraud involving senior management figures, they strive to mine both the textual content and social networks of email data. They provide a competent review of relevant work and use the Enron email dataset as an illustrative example.

[Pandit et al. 2007] also use SNA in detecting fraud, but as applied to transaction data from online auctions. Their method, working as a third-party service, applies Markov Random Fields to model the networks and belief propogation to detect fraud within the network. Their positive evaluation includes both a synthetic dataset and a transaction dataset scraped from the popular auction network eBay. A third-party approach such as this would appear to allow their system to adapt to a number of auction platforms, but with a risk of being rendered ineffectual by changes to site templates or APIs.

3.2.5. SNA and Identification. [Hadjidj et al. 2009] turn SNA methods to forensic (i.e. identification) analysis of temporal email data. The SNA component of this mostly text-mining tool is employed to provide behavioural, temporal and geographic modelling information. A partial evaluation of a different module of the toolset is provided using the Enron email database, but the SNA component is presented merely as a useful analytics and visualisation workbench.

3.2.6. SNA and Crimes Against Children. [Frank et al. 2010] examine the structure of online child exploitation networks, building networks of websites based on their links and a set of predefined bad keywords, with the ultimate goal of identifying the major nodes whose removal would most disrupt online exploitation. They demonstrate their deployment on four networks crawled from websites identified through search results, identifying the key nodes through the top 10 values for in-degree and for severity of content as identified through keywords. They also find that centrality does not correlate with severity of content, but severe websites were highly linked to each other, suggesting scope for targeting subnetworks of the most extreme material where law enforcement resources are scarce.

3.3. Information Extraction

3.3.1. Information Extraction in Terrorism and Extremism. A variety of information extraction techniques can be applied in analysis of terrorists and extremism, including topic mining and summarization. Websites and forums frequented by these groups are par-

ticularly rich source of information. The main body of research tends to focus on either white supremacist groups in the US or Islamic fundamentalists.

[Marcus 1998] describes a system called ProfileMiner for combating cyberterrorism. This system amounts to an interface or series of interfaces to a database of online information, the compilation of which is left unspecified but appears to be tied to a commercial product called MAVIS. No evaluation is provided, nor is it clear whether the interface was actually constructed rather than simply designed.

[Zhou et al. 2005a] briefly outline the motivation for and design of the Dark Web Portal, a resource used in several papers addressing information extraction from terrorist and extremist sites. [Zhou et al. 2005b] describe a semi-automated system for collecting and analyzing information on 'Dark Web' sites, and apply this to a selection of United States extremist web sites. Their results and methodology are subjected to an expert evaluation with a positive outcome. While their automated collection stage (outlined in more detail itself in [Zhou et al. 2007]) appears effective, their approach to filtering the results of said searches involves manual filtering of hundreds of URLs followed by a second stage of search to manually bulk out the results. If value over that of a typical search engine is to be added in a semi-automated collection and filtering system, it must be to reduce such loads on the analyst. The work by [Chen et al. 2008] appears to be linked, in which a case study is carried out to collect and analyse examples of Arabic fora. The same levels of expert evaluation and manual workload are evident, suggesting that the only key difference in the two works is the community being analysed.

[Qin et al. 2007; Chen et al. 2008] take a similar approach in what may be a continuation of the same line of research. A coding system referred to as the Dark Web Attribute System is developed to look specifically for signs of technical sophistication and content richness in the design of the websites of extremist groups. The first paper uses this framework to compare terrorist sites to those of US government agencies while the second compares the internet presence of extremist organisations drawn from three geographical regions. The latter's detailed analysis of these technical indicators highlights how relatively innocuous details can be of interest when examined at scale. A combination of this attribute system with an automatic collection system could prove useful in identifying new groups that show above-average sophistication, perhaps thus better helping identify key emerging threats.

[Gerstenfeld et al. 2003] report on the observed typical behaviours of those holding supremacist or separatist beliefs, as determined through an examination of 157 purposefully-selected extremist sites. Their findings include interesting results such as disavowal of racism and a low rate of direct incitement to violence. The authors also comment on the utility of the internet to widely-scattered extremist groups. Though their motivation is given in terms of extremism generally, their sample appears focused particularly on a certain group of white supremacist sites, with Islamic extremists appearing only in an 'Other' category.

[Yang et al. 2009] describe how web crawling technology is integrated with NLP techniques to extract common topics from websites hosted by extremists or terrorists. Though their evaluation does include an attempt to assess the compactness of topics, a notable problem with their LDA results is the generation of several reasonably similar topics. A means of better combining (or representing the distinction between) such groups could be considered an area for study in the field of topic extraction in general.

[Yang and Ng 2009] relate how a clustering opinion-extraction method targeted specifically at opinions expressed in online discussion is trialled on a corpus drawn from Myspace, including discussions about terrorism. Their clustering method attempts to overcome some limitations with the DBSCAN clustering mechanism, focusing on a distance-base clustering method. More detail on the TFIDF mechanism by

which web opinions can be represented as a vector of core concepts could help generalise their method to non-clustering applications.

[Jayanthi and Sasikala 2011] describe a process of mining hyperlinks from terrorist web pages as part of link analysis. However, only a simulation of output from a toolkit is provided.

[Inyaem et al. 2009] focus on gathering open-source information about terrorism via summarisation of web-based news articles. Though some details appear obscured by poor translation, the authors seem to find support for an ontological approach to detection of terrorist events, comparing this approach to a gazetteer and some form of grammatical parser in an evaluation on Thai news articles.

Of additional interest to this problem topic and method is an information extraction tool which is designed for general intelligence use [Skillicorn and Vats 2007]. It makes use of terrorist subjects as an illustrative case study. The network study of blogging sites [Xu and Chau 2006], which focuses on extremist hate groups, is also relevant.

3.3.2. Information Extraction and Police Intelligence. Information extraction can be applied to online sources for intelligence on organised criminal activity. That the core of this is a collection of web-mining toolkits suggests common research interest around synthesis of web-based news articles for intelligence purposes, heavily reliant on named entity recognition techniques, with some recurring problems including the identification of relevant articles for processing and the reliance on a domain lexicon for identifying key information.

[Ge et al. 2010] focus on extracting 'story' patterns from web-based news articles as part of open-source intelligence efforts, with pattern-matching begun by the detection of trigger words indicating certain events. Their system is demonstrated on a collection of Chinese news articles on the 2008 Mumbai attacks, but what their published results show is unclear. Their system's reliance on trigger words seems to suggest that individual implementation will either require existing domain knowledge, reducing utility for emerging events, or else be general terms which may not fully capture specific narratives. The question of how appropriate news articles are gathered for processing is also critical for deploying such a system.

[Ku et al. 2008] focus on extracting crime information – in the form of key phrases – from narrative reports, which is applied primarily to police and witness reporting, but is noted for potential application to web news. Similarly to the previous study, their approach struggles with a scalable system for managing a crime lexicon, which they resolve with manually-created lists supplemented by dictionary resources.

[Atkinson et al. 2010] also describe efforts to gather structured knowledge from web-based news articles for EU security purposes, clustering articles based on textual similarity and geographical location, then applying other event extraction tools. There is a lack of detail on the operation of these tools.

[Wenhua and Na 2010] discuss an extension of the Encase forensics toolset to allow analysis of web pages regarding some form of illegal gambling activity. They attempt to mine not only entities, but also the relationships between entities, in an unsupervised manner. However, their approach to information extraction appears to be overly tailored to their chosen problem domain for it to generalise to other scenarios. No evaluation is provided.

[Skillicorn and Vats 2007] focus on the problem of gathering *novel* information about a topic, relative to a specified set of existing knowledge. They make use of web search engine results regarding the known topic and use these web pages to form new queries based on prominent nouns, clustering the results based on descriptive nouns. Their ATHENS approach is trialled on terrorism topics, but could equally apply to other law-enforcement or defense uses. Their method for selecting descriptive nouns compares

the frequency in the web pages under review to the frequency in the British National Corpus, a standard English corpus. While this approach is domain independent, strict comparisons are likely to lead to spurious noun-phrases being identified, so it would be better to search only for nouns whose frequency is significantly greater than in the reference corpus in order to prevent common variations diluting search terms. In the same vein, the BNC relies on texts now well over a decade old, and is not likely to include a number of now-common proper nouns. A different reference corpus, perhaps drawn specifically from web sources, might make a more suitable baseline.

[Wang et al. 2012c] use Twitter as a source of general crime prediction, drawing on automatic semantic analysis, event extraction and geographical information systems to map crime hotspots. In an evaluation on actual hit-and-run crime data, their system outperforms a baseline uniform model. While there may be scope for improvement in the predictive technique, more interesting developments are likely to be found in modification of the model for deployment on a streaming Twitter feed. [Wang et al. 2012a] do so, using Twitter data to model criminal incidents geographically. They apply a spatio-temporal generalised additive model to a combination of geographical and demographic features of an area and textual features extracted from the Twitter feed of a news agency, evaluating their performance against actual crime incidence rates. Their analysis shows that the textual features provided by the Twitter data improve prediction accuracy as compared to a previous model only using geographic and demographic information.

[Giacobe et al. 2010] aim at gathering information in extreme events, describing an approach to open-source intelligence which was applied in an artificial competition environment (searching for red balloons), and how experiences in the challenge may relate generally to intelligence-gathering, particularly with regard to false-reporting. Their overview is high-level and rather specific to their challenge, but includes reference to a number of techniques and technologies not otherwise captured by this review.

[Dudas 2013] focuses on the detection and analysis of 'dark networks', with specific focus on visualisation tools for handling networks parsed from Twitter and placed by geolocation. No formal evaluation is provided, but the paper discusses trial usage on real networks of interest.

[Johnson et al. 2012] look at finding relationships between unstructured law enforcement texts (emails) and using said relationships to help augment information of interest, analysing the semantic relatedness of documents and linking identified entities. A demonstrative application is presented, acting on a sanitized corpus of real law enforcement emails. Given appropriate consideration of scalability, this information linking tool would appear to be an impressive resource for augmentation of police intelligence.

3.3.3. Information Extraction and Crimes against Children. Information extraction techniques are sparsely deployed in child protection, mostly aimed at child sexual trafficking. Their aims include identifying children known to be missing by monitoring trafficking networks and chatrooms for mentions of their names or other identifying details.

[Wang et al. 2012b] present an approach to combating the sexual trafficking of children through examination of open sources such as classified advertisement sites and bulletin boards. They examine such resources for evidence of trafficking networks and introduce techniques to search for victims under aliases and misspelt names. Though the authors do not present an evaluation, they discuss ongoing trial deployment, highlighting challenges specifically related to anonymisation of the toolkit's interactions with sites to prevent counter-intelligence, and with scaling their approach to wider monitoring.

[Romaniuk 2000] approaches the same problem from a different angle, applying intelligent agents to identify missing children on the internet by connecting information in open databases of missing children with web crawling and IRC chat monitoring. The approach was partially implemented as the SADIE system at the time of publication. The proposed ecosystem of agents dealing with specific data sources appears flexible, but the exact means of calculating results' similarity to a short query – a very key detail for any of the agents – is left unspecified.

A common theme to both these sexual trafficking technologies is the integration of information from multiple sources, but both publications appear to focus on different sources for their information. This may, in part, be due to the large time gap between the two papers. The SADIE system outlines a high-level approach to multiple data source integration, but leaves many implementation matters unresolved.

[Marjuni et al. 2009] attempt to extract crime information (Who, Where, When, How, What, Why) from chat logs, drawing on published examples of sexual abuse from adult dating and scam interactions as their data source. Tokenisation and part-of-speech tagging of the data is discussed. Classification accuracy results for their crime information categories are also presented, though how these results were derived is unclear. While mining instrumental crime information as would fit the given categories could well prove useful to investigators, the paper does not present a coherent solution for the purpose.

3.3.4. Information Extraction and Cybercrime. Cybercriminal activity has also been mined from online data sources, the primary source being fora where they sell or exchange information.

[Spencer 2008] uses an XML framework to mark up relationships extracted from hacker fora, essentially mapping the social network of said fora for usage by police. The paper focuses heavily on the choice of technology and representation for the task – XPath queries on tidied HTML source of the fora – with no real evaluation of the value to law enforcement. Furthermore, the approach relies on manual exploration of the XPath query space for page sources, a process which could, at the least, have been guided through generated templates.

[Zhuge et al. 2009] present a study of the cybercriminal economy on the Chinese web, attempting to model the extent of this black market and the amount of malicious code involved in its constituent websites. In addition to these contributions, the authors offer detailed description of a cybercriminal infrastructure. Their estimation of the value of cybercriminal assets, and particularly their attempt at breaking down these totals by classification allows law enforcement and cyber-security vendors to focus their efforts where greatest impact can be effected.

Additionally, some papers discussed earlier fall into this category.[Ma et al. 2011] construct networks of cybercriminal activity from email and blog data while [Marcus 1998] describes a system designed for fighting cyberterrorism through handling of collected intelligence sources.

3.3.5. Information Extraction and Finance. [Bernard et al. 2011] address a financial criminal matter (money laundering) by helping track financial services through web mining. Their tool crawls the web, identifying online financial trading sites through a generalised linear model applied to textual features. While the presented accuracy appears impressive, the results were obtained via an artificially balanced dataset wherein roughly a quarter of all websites were actually OFT (Online Financial Trading) sites – a situation which is unlikely to be the case when the system is deployed on the web generally.

3.3.6. Information Extraction and Online Identification. [Zhu 2007; Aggarwal et al. 2007; Bali 2007] all cover various aspects of an email de-anonymisation workbench built on a set of mature UNIX tools. This UnMask toolkit focuses specifically on detecting and countering spoofing attempts within email messages, examining email bodies and headers particularly for examples of spoofed links, forms and headers, and storing evidence in a manner suitable for law enforcement use. Their aim of achieving this through combining a variety of pre-existing tools is laudable for its software reuse, but there is a lack of rigorous analysis of the performance of the anti-spoofing components. The papers, however, do include a case study demonstrating potential uses of the toolkit during an investigation.

3.3.7. Other Information Extraction Applications. [Broadway et al. 2008] describe a study aimed at improving the analysis of forensic network traces from investigations, presenting a high-level packet analysis tool which has been developed and compared to existing tools, but not formally evaluated.

[alias Balamurugan et al. 2007] cover the detection of 'suspicious' or 'deceptive' emails. They do not provide a clear definition of what that may mean, but an ominous reference to national security. They apply a series of classifiers to an insufficiently explained dataset, and report high classification accuracies, particularly for the IBk decision tree.

3.4. Machine Learning

3.4.1. Machine Learning and Online Identification. Machine learning techniques have been applied in online identification tasks, often working with email data, attempting in particular to identify scammers and phishers from their campaign output.

[Airoldi and Malin 2004] present the ScamSlam project. It focuses on identifying the common origins of scams, particularly advance fee fraud, through the use of unsupervised hierarchical clustering on scam emails detected with a Poisson filter. Their method appears to detect a small number of scammers (20) sending most of the advance fee fraud messages in a corpus of 534 such scams, but they are unable to verify this result. It is not clear how broadly their system may be applied, as the advance fee fraud they focused on has a fairly large text body, which may be atypical of scam email.

[Stallings et al. 2012] cover the same use-case, but make use of email headers to build clusters of scam originators using WHOIS data. Using this approach, they identified 12 email addresses which were key in registering spam-origin domains. Such an approach holds benefits in that it may be applied to many scam or spam emails without requiring specific additional feature in the body of the scam, but also risks vulnerability to email spoofing.

[Yearwood et al. 2010] present an approach looking at profiling rather than simply detecting phishing attacks. Their study makes use of hyperlinks from the body of an email as well as structural features and WHOIS information in a pair of classifiers. They profile phishing emails by having the classifiers apply multiple labels to each email regarding the presence of scripts, images, etc. and the apparent legitimacy of linked sites from WHOIS information. The strength of the paper lies in its clear formulation of the problem of profiling phishers rather than merely detecting phishing.

[Dazeley et al. 2010] present a combined approach for profiling large volumes of phishing email. The results of several independent unsupervised clustering algorithms working on a random subset of a large dataset are combined with a variety of consensus algorithms and, in turn, used to train a number of fast classification algorithms for use on the whole dataset. This approach of using unsupervised clustering to prime supervised clustering would appear to work well for classification of emails into clus-

ters detectable in the training set, but may suffer in a deployment where new clusters of phishing email begin to appear.

The other identification studies using machine learning more generally address the identification of criminals from email data. The approaches discussed below are somewhat unusual as compared to the more standard classifiers discussed in the later NLP section, but contain some overlap.

[Iqbal et al. 2010] make use of an existing speaker recognition framework from the field of speech processing in an attempt at authorship analysis, using several classifiers. The framework is evaluated against the Enron email dataset with results indicating a competitive approach, although the requirement of 200 training emails per author is not insignificant.

[Schmid 2012] explores the application of associative classification to authorship attribution of text, in an approach which requires the extraction and amalgamation of rules. However, the system performs poorly on a multi-author dataset, with a best classification accuracy of 50% on only 10 possible authors.

[Stamatatos 2006] covers authorship attribution in a trial dataset of an online newsletter. The approach uses classifier ensembles, demonstrating how a range of diverse classifiers can be constructed through exhaustive disjoint subspacing, and showing that the approach outperforms a simple SVM model using word frequencies. The author goes on to enhance the model with a cross-validated committees technique.

3.4.2. Machine Learning and Terrorism and Extremism. Machine learning is deployed both in detecting terrorism-related activities and in identifying terrorists from their online footprint.

[Cheong and Lee 2011] explore microblogging within the terrorism informatics domain. They perform an observational analysis of the Twitter network's response to two real-life terrorist events, and use this as inspiration for the design of an information-gathering framework. They later apply the framework to a synthetic dataset of events which share some properties with terrorism events. They also apply a variety of common machine learning analyses to their dataset in an exploratory manner.

[Sahito et al. 2011] link streams of Twitter data to other resources through Open Data mechanisms. They apply named entity recognition to the content of Tweets. They mention the terrorism domain, their aim being to allow for structural links to entities to be be imposed on unstructured Twitter data to better allow law enforcement to parse and respond to events detected via Twitter. However, the implementation of this is relegated to future work.

[Nizamani et al. 2013] evaluate a number of machine learning methods (the ID3 decision tree algorithm, logistic regression, Naive Bayes and SVM) for the purpose of detecting suspicious emails. As well as developing a terrorism-related email dataset for the purposes of this comparison (including real messages gathered from newsgroups), they develop a feature selection system that provides consistent improvement to the results of all of the tested classifiers. They report that for their application, logistic regression and ID3 outperformed the Naive Bayes and SVM classifiers.

[Shen and Boongoen 2012] use a qualitative formalism as the basis for a fuzzy analysis, applying this to link analysis and the determination of aliases. They evaluate their system against unspecialised unsupervised learning systems on a constructed terrorism dataset gathered from web articles, an author publication dataset (DBLP) and an email dataset. Their system appears to outperform a number of similar link-based algorithms.

[Elovici et al. 2004; Elovici et al. 2010] employ web usage data to identify terror-related activities, training a classifier on the web usage of ordinary users and a collection of known terrorist web sites. The aim is to deploy a system which monitors

the web access of users (at an ISP or organisational access provider level) and raises alerts whenever a user accesses abnormal content. The civil liberty implications of such a mass-monitoring system could rightly be challenged, but more practical issues may prevent adoption. Their detection system reached an AUC of 91% on their experimental dataset, rising to 99.7% with additional components. Given the large number of normal users and relatively tiny number of real, detectable terrorist usages of actual networks, even such a classifier would produce an unreasonable volume of false alerts for every true event it captured. This issue is not unique to the work of these authors, but applies to all systems of this kind. Nonetheless, these two papers consist of a coherent description of the development of a high-performance classifier for web usage data.

[Tinguriya and Kumar 2010] suggest a self-organising map approach to classifying web users from usage data. They provide no evaluation or appraisal of their proposed system, and indeed minimal description of its proposed operation.

[Endy et al. 2010] have developed what they term an intelligent search procedure for webmining cyber-terrorism information, feeding a vector representation of 600 articles, half related to cyberterrorism, into a self-organising map, the results of which they then briefly dissect. Their presentation of the SOM as a heat-coloured grid seems ill-suited for law enforcement analysts.

[Yang et al. 2012] focus on identifying extremist content in social media sites, drawing their design inspiration from biological immune systems. They build a mathematical representation of lymphocytes which incorporates lexical, sentiment and syntactic features of text as a precursor to a semi-supervised classification system. In an evaluation of this system on violent messages scraped from a white supremacist web forum, their system outperformed two benchmark labelling systems.

3.4.3. Machine Learning and Harassment. Machine learning can be deployed to detect threatening textual communications, the aim in most cases being to produce a classifier which separates threatening messages from normal communication.

Appavu and Rajaram [Appavu alias Balamurugan and Rajaram 2008] compare a decision tree classifier with SVM and Naive Bayes classifiers, using two email corpora and two different feature selection mechanisms (information gain and term frequency variance). They find decision trees to outperform SVM and Naive Bayes in detecting examples of threatening email. A follow-on paper [Appavu et al. 2009] repeats this analysis, but includes the Ad Infinitum algorithm, which outperforms the other methods. [Banday et al. 2011] later revisit this work, looking also at the detection of threatening emails. The authors compare the data of Appavu and Rajaram to their own Naive Bayes approach, which makes use of different features (single and multiple keywords as well as weighted keywords with context matching). Measuring the accuracy of results with the F1-score rather than simple percentage accuracy, they find that their weighted multiple keyword system with context matching performs in a manner competitive with the better methods from Appavu and Rajaram's analysis. They do not make a direct comparison due to the different datasets underlying results, but a review of F1 scores indicates that some of the methods presented by Appavu and Rajaram may be better classifiers.

[Xu et al. 2012b] identify emotions common in cyber-bullying, and develop a training procedure to help recognise these emotions from text without reference to a labelled training set. They evaluate their zero-label-trained SVM system on a labelled Wikipedia corpus, finding that it has lower cross-validation error than three baseline methods. They also apply it to Twitter traces involving bullying, finding that only a relatively small proportion of said traces showed emotion, and that where emotion was detected it did not necessarily reflect severity or sincerity. [Yin et al. 2009b] in-

stead focus on a supervised learning approach to detecting cyber-bullying, using term frequency as a primary measure, and supplementing it with sentiment and contextual features. Their model performs fairly poorly on their web datasets, with the best F1-measure accuracy being less than 50%.

[Wang et al. 2006] take inspiration from biological immune systems in much the same manner as [Yang et al. 2012], also integrating term frequency into their mathematical adaption of it. Though they claim 'good results', there is no evidence of any evaluation.

[Shekar and Imambi 2008] turn to a more conventional Naive Bayes classifier, testing it on a small corpus and a bag-of-words feature set which appears to be extended with some user-level attributes. The presentation is somewhat ambiguous, describing classification rules for detecting a 'threat' class of message, but presenting classification results for 'movie' 'food' and 'travel' topic classes, none of which are alluded to a-priori.

3.4.4. Machine Learning and Crimes Against Children. A small number of rule-based systems have been generated to help with detecting predators in textual exchanges.

[Hidalgo and Díaz 2012] present a knowledge-based system for detecting sexual predators, with a Naive Bayes subsystem with reasonable classification accuracy. Interestingly, their hand-coded rules for predator characterization were originally written in and for Spanish, but were automatically translated to apply to English, and appear to still be effective in identifying the main predation phases.

[McGhee et al. 2011] compare previously-developed rule-based classifiers to decision tree and a k-nearest neighbours classifiers. They find that the machine learning systems improve classification of predation when working with specific transcripts, but fail to reject the null hypothesis in a more general case comparison against their rule-based system. The average accuracy of their rule-based classifier is 68%.

[Peersman et al. 2012] move away from rule-based systems, applying and combining two separately-trained SVM classifiers in a weighted manner. They achieve an F1-score of 0.9 for the task of classifying authors as predators, but much lower accuracy for detecting specific grooming posts.

3.4.5. Machine Learning and Financial Crime. [Modupe et al. 2011] use SVMs and Random Forests to detect advanced fee fraud scams in an email dataset. They report high classification accuracy on a synthetic dataset where roughly a third of all mail was advanced fee fraud messages, and find that their SVM classifier outperformed the Random Forests classifier. An evaluation more comparable to real deployment base rates would be preferred.

[Walgampaya et al. 2010] focus on click fraud prevention using web usage data. They detail a multi-level data fusion mechanism which takes input from a click map module, an outlier detection module and a knowledge-based rule module, and stores levels of suspicion regarding specific IP addresses, referrers and countries. They provide a detailed analysis of the results of their system as applied to publicly-available click-through data.

[Bernard et al. 2011] track online financial services through web mining, gathering textual features to reach conclusions about the probability of a site being an online financial transaction site. Their evaluation against human subjects shows demonstrable benefits in terms of speed, and generally high precision.

3.4.6. Machine Learning and Police Intelligence. [Chung 2012] studies appropriate machine learning systems for categorising temporal events collected from web data. Using a case study involving web articles related to an incident of domestic terrorism, the performance of Naive Bayes, SVM and neural network methods at applying temporal

group labels across a range of feature set sizes is demonstrated. The results show that while all three systems performed in a satisfactory manner, SVM and Naive Bayes increased in accuracy as the number of features increased, while the neural network peaked at 70 features.

[Stolfo et al. 2006b] describe the application of a general-purpose email-mining toolkit to behavioural analysis, with a case study in detecting viral emails in an archive of the emails provided by 15 users. The system performs well when introduced to sudden and abnormal flows, but struggles to detect slow campaigns for the delivery of email. The degree to which virality can sensibly be detected from such a small user corpus is debatable. The paper also provides a lengthy demonstration of the overall capabilities of the email-mining toolset.

3.4.7. Other Machine Learning Applications. [Do et al. 2004] approach the problem of pornographic web page identification with two classifier components. One component classifies web pages into various predefined categories, which can then be used to filter these web pages from access. The other component analyses the behaviour of users with respect to the category of sites accessed. They test their system – with a variety of classifiers – against commercially-available web filters, and find that their best classifier outperforms them.

[Tian et al. 2010] focus on the effect of pornography on young people as their motivation. They note failures of strict rule-based and keyword-based systems for filtering undesirable information, and propose a system which gathers a broader range of features from a page to assist in classification. They do not evaluate the performance of this proposed system.

[Lim et al. 2007] attempt to detect abnormal patterns of email traffic using a hierarchical fuzzy system. They develop three different system architectures, and trial these systems on a selection of threads from the Enron email dataset, finding that all three agree with each other in the ranking of abnormality of communication links. Whether such a test holds external validity is hard to determine.

[Benjamin and Chen 2012] apply machine learning to recognise the traits of key actors in hacker communities. Their regression analysis of the social structure of hacker fora from the United States and China, determines that involvement in a number of threads, total message volume and number of attachments uploaded are the major factors which explain the reputation score of members of the community. While the paper focuses on cybercrime as a domain, its results could be said to apply more to certain online forum communities, of a criminal or otherwise nature.

3.5. Natural Language Processing

Due to the size of this category, subcategories of papers using similar NLP methods have been defined.

— *Authorship Analysis*: The identification of texts as belonging to a particular author.
— *Author Profiling*: The identification of qualities of a text's author.
— *Sentiment Analysis*: The analysis of opinion or emotion markers in text.
— *Text Classification*: Classification of a text into one of a range of categories.
— *Other*: NLP-based methods which do not fall under the other categories.

Authorship Attribution.

3.5.1. Authorship Attribution and Online Identification. Authorship attribution is naturally tied closely to the problem domain of online identification, and a wide range of techniques have been applied on a number of datasets. Typical problems for such studies include deciding upon the most appropriate feature set to use in classification, and finding appropriate methods for different data sources. Other for or highly related to

this field include authorship verification, authorship similarity detection and stylometric comparison, with different clusters focusing on either the conflation of author identities or the assignment of specific texts to an author, but the technical challenges of both tasks are expressible within the same framework.

[Ma et al. 2009b] apply authorship attribution specifically to phishing emails, aiming to cluster messages based on othographic features using an adapted form of the K-means algorithm. They reason that the semantics of phishing emails are often too similar to be useful for disambiguation. They provide an evaluation on a collection of 2048 known phishing emails, with several differing initial parameters for their clustering algorithm and gradually refined feature sets. While their method appears to produce reliable clusters, a validated dataset would be useful for verification purposes.

[De Vel et al. 2001] make use of both structural and linguistic features and an SVM classifier. They validate their approach on a collection of emails to particular newsgroups, finding high accuracy in most cases. They additionally investigate the use of word collocation and the dimensionality of function words in a bid to improve classification accuracy. However, this does not improve performance.

[Zheng et al. 2003] compare decision trees, neural networks and support vector machines on a corpus drawn from English email messages and both English and Chinese BBS postings. The best results are for SVM classification of the English newsgroup postings, with neural network performance lagging slightly behind. They note a drop in performance in their Chinese dataset, which they ascribe to fewer style features for that language. A follow-up paper [Zheng et al. 2005] makes use of an extended set of features and the same set of classifiers, again finding that the SVM classifier outperforms the C4.5 decision trees and the neural network.

[Corney 2003] also covers the reduction of authorship attribution to a pattern of certain writing features, applying this approach with an SVM classifier to public-domain books, theses and the author's own email collection, with good results in each case. The results show that function words appear to make the best features.

[Teng et al. 2004] suggest the use of an SVM classifier for authorship attribution on emails, explaining the operation of the classifier and listing some structural features of email which might be useful, but providing no evaluation. Given the prior existence of work such as [De Vel et al. 2001], this would appear to be of at best explanatory value.

[Stamatatos 2006] explores authorship attribution via an ensemble of SVM classifiers and a feature set subspacing approach. Exhaustive disjoint subspacing is compared with the k-random classifiers method of ensemble construction, finding that the former outperforms the latter and also outperforms an SVM classifier when small subset sizes are chosen.

[Stamatatos 2008] covers the class imbalance problem in authorship attribution, where the volume of available training text for some candidate authors is extremely low. A new method for handling imbalanced datasets through variable-length sampling of training data is presented. The method is compared against a re-sampling variant to the existing under- and over-sampling methods, making use of both English and Arabic datasets. The results show that the method resulting in the best net improvement to the accuracy of an SVM classifier trained on the resulting training set is the random re-sampling of text from the available training data.

[Abbasi and Chen 2008] provide a useful review of the state-of-the-art and go on to demonstrate a classification method which makes use of individual author-level feature subsets from a large feature space. They compare this method to an SVM classifier with a feature set drawn from previous literature and to an ensemble of SVM classifiers with an extended feature set, using a range of online text forms (the Enron email dataset, eBay comments, posts from an online forum and chat logs). Their system out-

perform both competitive methods on the email, comment and chat datasets, but not on the forum messages, where the ensemble of SVM classifiers performed best. Alongside the identification experiment, they also distinguish the task of detecting similarity, and perform a similar evaluation for that purpose, finding their method outperforms the competitive baseline methods.

[Dardick et al. 2007] focus particularly on blogs, covering the ethical debate over why bloggers may legitimately seek anonymity, and why law enforcement may wish to circumvent this barrier of anonymity. The paper covers technical approaches to stylometry only briefly and at a high level. [Dreier 2009] explores the same topic with more technical detail, creating a baseline model of authors based on frequency of characters and words, and using individual deviation from this baseline as the features for classification. Both Naive Bayes and SVM classifiers are evaluated, finding low average accuracy across all authors, but that certain authors were extremely well-predicted.

[Ma et al. 2008; Ma et al. 2009a] focus on applying authorship attribution to Chinese online texts. The first paper focuses on authorship attribution in email, covering issues such as the lack of explicit word boundaries in Chinese text and the selection of sequential patterns from texts, passing said patterns to an SVM classifier. In their evaluation they provided 30 training examples for three authors, and had 20 further emails classified as belonging to one of these three authors, with a classification rate of 90%. In the second paper, the authors also apply their classifier to blog and BBS messages, drawing a comparison between three classifiers, one of which uses linguistic features, another which uses structural features, and one which uses both. They find the classifier using the combined feature set outperformed the others, though they all performed at above 65% accuracy. They also examine the effect of varying the number of authors to classify texts, finding that larger numbers of authors caused accuracy to drop.

[Iqbal et al. 2008; Iqbal et al. 2010; Iqbal et al. 2013; Iqbal 2011] all address authorship attribution through frequent pattern mining. The first of the publications focuses on the notion of frequent patterns as a means of ensuring the forensic worth of authorship attribution techniques, objecting to the lack of intuitive explanation in an SVM classifier. They use a combination of lexical, syntactic, structural and content-specific features in their method, detecting frequent writing patterns in an author's text and filtering out frequent patterns which are common to a large number of authors. They validate the viability of their method in an evaluation on the Enron email dataset. In the second publication, the authors use standard clustering algorithms to group texts together as a prerequisite to mining frequent patterns for author identification. They examine the accuracy of the resulting output as a means of evaluating which clustering mechanism is best-suited to the task, again using the Enron email dataset as a source. The third publication presents frequent-pattern writeprints as a 'unified' solution to authorship analysis. The authors describe use cases involving small and large training samples, and also extend their system to discovering characteristics of an author. The evaluations on the Enron email dataset are repeated, and alongside these results a trial of the characterisation application is carried out. The results show that for gender prediction the approach performs slightly better than random assignment, and for location prediction, with three classes, it again performs with accuracy slightly above that which one would expect for random assignment. Finally, [Iqbal 2011] combines this research into one volume, providing greater detail on the difference in approach between two versions of the classifier, with extensions covering somewhat separate problems of extracting cliques and topics from chat logs.

[Orebaugh 2006; Orebaugh and Allnutt 2009; Orebaugh and Allnutt 2010] cover authorship analysis on instant messaging communications. The first publication focuses on examining character frequency as a stylometric feature, examining the frequencies

of characters in a small four-author dataset and testing for whether frequency of characters is distinct. The results show that uppercase characters, numbers and special characters are distinguishing and may be used as a form of intrusion detection system. In the second publication, the authors analyse what appears to be the same dataset, but with an extended range of features, including sentence structure and pre-defined sets of special characters. They apply three classifiers – the J48 decision tree, the IBk nearest neighbour classifier and a Naive Bayesian classifier – to these features, and find a high accuracy in each case, though given the sample size this would not be unexpected. They analyse the distinguishing features and find that abbreviations are the best discriminators, followed by the use of special characters. In the final publication, the authors expand their evaluation to include two larger datasets, examining the useful features for accurate classification in each system, and using different classifiers. They find high accuracy with an SVM classifier trained on a range of 356 features, including lexical and syntactic features as well as the previously-used structural and frequency attributes. On a dataset of 105 authors, they achieve 84.44% accuracy.

[Layton et al. 2010] cover a particularly constrained form of authorship attribution which is particular to online discourse – attribution of Tweets to their authors. They detail the structural properties of Tweets and present a preliminary analysis of the viability of attribution using Tweets. They find classification accuracy of approximately 60% for 20 training examples, and highlight that adding training Tweets increased accuracy up to 120 examples, after which increases appear not to be significant.

[Chen et al. 2011] apply a frequent-pattern mining approach on the Enron email dataset. Writeprints – consisting of numeric representations of the relative frequency of stylistic features extracted from an author's text – are constructed and then compared in order to determine whether authors are similar enough to be the same. They compare SVM, PCA, K-NN, DT and K-means approaches, finding SVM to have superior classification accuracy.

[Khan 2012] uses a bag-of-words model of email bodies and applies a Naive Bayes ensemble method to attribute emails drawn from the Enron email corpus. The method chieves a respectable classification accuracy, outperforming previous work on the same dataset, but it does perform best when given messages over 100 words, which slightly limits application to online texts.

[Pearl and Steyvers 2012] cover the detection of 'authorship deception', which includes both a normal attribution use case and an imitation attack whereby authors attempt to imitate the writing style of a victim. Their method involves building a writeprint of stylometric and content features, and applying logistic regression as a classifier. Evaluation on a blog dataset shows good performance in the classic attribution case, and a small evaluation of the imitation case shows highly positive results.

[Liu et al. 2012] attempt to address issues with the difficulty of writeprint comparison through a novel semi-random subspace method, which also aims to overcome redundancy in feature sets. A detailed description and theoretical analysis of the methodis provided, followed by an empirical evaluation on a subset of a large English corpus, displaying accuracy results with regard to both the number of authors and the number of texts available per author. In all cases they compare their method to other well-performing classifier ensembles, with a positive result.

Given that the aims and methodologies of many authorship attribution papers targeting identification are comparable, results and methods from various studies may be contrasted with each other. Table III gives an overview of some of the best results from each paper, noting the dataset and number of classes being attempted. Important additional information including length of texts used, texts per author, and features used in classification are all left to the original texts. Items are sorted chronologically.

Table III. Summary Comparison of Authorship Attribution Approaches

Source	Classifier	Dataset (#Messages)	Authors	Accuracy (%)
[De Vel et al. 2001]	SVM	Newsgroup (1,259)	4	-
[Corney 2003]	SVM	Email (253)	4	85.2
[Zheng et al. 2003]	SVM	Newsgroup (153)	9	96.08
[Zheng et al. 2003]	SVM	Email (70)	3	91.43
[Zheng et al. 2003]	SVM	BBS (70)	3	82.58
[Zheng et al. 2005]	SVM	Newsgroup (c.960)	20	97.69
[Zheng et al. 2005]	SVM	BBS (532)	20	88.33
[Stamatatos 2006]	EDS Ensemble	Web news (200)	10	99
[Iqbal et al. 2008]	AuthorMiner	Enron Email (120)	6	90
[Iqbal et al. 2008]	AuthorMiner	Enron Email (100)	10	90
[Abbasi and Chen 2008]	Writeprint	Enron Email	25	92
[Abbasi and Chen 2008]	Writeprint	Enron Email	50	90.4
[Abbasi and Chen 2008]	Writeprint	Enron Email	100	83.1
[Abbasi and Chen 2008]	Ensemble	eBay comments	25	96
[Abbasi and Chen 2008]	Writeprint	eBay comments	25	96
[Abbasi and Chen 2008]	Writeprint	eBay comments	50	95.2
[Abbasi and Chen 2008]	Writeprint	eBay comments	100	91.3
[Abbasi and Chen 2008]	SVM	Java forum	25	94
[Abbasi and Chen 2008]	SVM	Java forum	50	86.6
[Abbasi and Chen 2008]	Ensemble	Java forum	100	53.5
[Abbasi and Chen 2008]	Writeprint	CyberWatch Chat	25	50.4
[Abbasi and Chen 2008]	Writeprint	CyberWatch Chat	50	42.6
[Abbasi and Chen 2008]	Writeprint	CyberWatch Chat	100	31.7
[Ma et al. 2008]	SVM	Email (150)	3	90
[Ma et al. 2009a]	SVM	Blog (1,379)	7	89.49
[Ma et al. 2009a]	SVM	BBS (410)	5	73.97
[Ma et al. 2009a]	SVM	Email (95)	5	80.61
[Dreier 2009]	SVM	Blog (c44,000)	100	54.53
[Orebaugh and Allnutt 2009]	NB	IM (?)	4	99.29
[Orebaugh and Allnutt 2010]	SVM	IM (950)	19	88.42
[Orebaugh and Allnutt 2010]	SVM	CyberWatch Chat (1250)	25	84.44
[Layton et al. 2010]	SCAP	Tweets (100,000)	50	72.9
[Iqbal et al. 2010]	EM	Enron Email (200)	5	80
[Iqbal et al. 2010]	K-m	Enron Email (200)	5	88
[Iqbal et al. 2010]	Bisecting K-m	Enron Email (200)	5	83
[Iqbal et al. 2013]	AuthorMiner2	Enron Email (160)	4	$x \approx 90$
[Iqbal et al. 2013]	AuthorMiner2	Enron Email (800)	20	$x \approx 70$
[Chen et al. 2011]	SVM	Enron Email (750)	25	88.31
[Khan 2012]	NB	Enron Email (6,109)	10	86.92
[Khan 2012]	NB	Enron Email (5,799)	9	87.05
[Liu et al. 2012]	PSemi-RS	Text corpus (2500)	50	77.04

3.5.2. Authorship Attribution and Cybercrime. [Gray et al. 1997] cover the application of authorship analysis techniques to software source code, demonstrating how the means of expression can vary even when programmers are solving the same problem. Their motivation is the attribution of malicious code, just as in natural language analyses the application is attribution of malicious or incriminating messages. They identify a number of features which could be useful for authorship attribution, and present two short case studies of events where malicious code has been examined for attribution to its owner.

A number of other papers focusing on online identification of authors cited cybercrime in a general sense, but made no specific link to the domain as defined in this study, and hence have not been included in our analysis.

3.5.3. Authorship Attribution and Terrorism and Extremism. [Abbasi and Chen 2005] discuss authorship attribution with particular application to the forum postings of extremist

organisations, with a focus on selecting an appropriate feature set for classifying Arabic text. They describe a study using SVM and C4.5 classifiers, applied to both English text from a Klu Klux Klan forum and Arabic text from posts associated with the Palestinian Al-Aqsa Martyrs group. They find slightly better performance at classifying English text authors than Arabic authors, and note that SVM significantly outperformed C4.5, going on to dissect the important features in classification for both languages.

Author Profiling.

3.5.4. Author Profiling and Crimes Against Children. Author profiling is widely deployed in the detection of sexual predators from chat transcripts. The typical classification is between text written by a child and text written by an adult. Some problems arise from attempting to parse net-speak with traditional linguistic tools, although there are indications that use of such language can itself be a useful age-determining feature.

[Pendar 2007] applies SVM and k-nearest neighbour classifiers to binary classification of predator and victim in chat logs from a vigilante website. They make use of word n-grams, with minimal pre-processing, as their input to a feature extraction function. They find their best classification rate (94.3%) comes from the k-NN classifier with a k of 30 and 10,000 features, both inputs being the largest of various levels tried.

[Tam 2009] detail a method for distinguishing between teen and adult conversations, with application to detecting sexual predators. Using a chat corpus, they attempt to distinguish between teens and chat users of different age brackets, using word and character n-grams in a Naive Bayes classifier and then an SVM classifier, finding that the SVM classifier outperformed Naive Bayes. Unsurprisingly, the most difficult-to-distinguish age group were the authors in their 20's.

[Kontostathis et al. 2009] compare a rule-based approach to log classification to a human analysis. They describe how a new iteration in a rule-based analysis of chat logs differs from a previous version of the tool in more appropriately identifying combinations of keywords in chat lines. The inter-coder reliability between their new tool and human analysis is reported as much improved. A follow-up work by the same authors [Kontostathis et al. 2010] surveys the literature regarding both sexual predation and cyberbullying, and as such covers some work on profiling sexual predators.

[Michalopoulos and Mavridis 2011] focus on detecting a grooming author by classifying messages into one of three attack categories and then combining classification probabilities. They perform a comparative evaluation with a number of different classification algorithms, finding SVM to perform poorly next to k-NN, Naive Bayes, Maximum Entropy and Expectation Maximisation. They consider their Naive Bayes approach the most suitable.

[Peersman et al. 2011] aims at predicting both age and gender of chat authors, with application to checking the truthfulness of reported profiles on social media sites. Working with a corpus drawn from a Belgian social network, they discuss several issues particular to online chat corpora, including shortness of texts and the variability of Dutch net-speak. They avoid issues of stemming and more involved linguistic analysis by utilising word and character n-grams. They find that word unigrams are the features best used for distinguishing between age and sex categories, achieving good accuracy in both cases.

[Inches and Crestani 2011], is notable in covering both author characterisation and topic detection, addressing general text-based surveillance. They describe a short characterisation experiment wherein Twitter users who are informative for a particular topic are identified, using a broad feature set and the expectation maximisation clustering algorithm.

[Bogdanova et al. 2012a; Bogdanova et al. 2012b] focus on two different subproblems in the identification of sexual predators. The first draws on the concept of

Table IV. Summary Comparison of Author Profiling Approaches applied to Crimes Against Children

Source	Classifier	Dataset (#Documents)	Task	Accuracy (%)
[Pendar 2007]	SVM	PervertedJustice (1,402)	Victim/Predator	90.8
[Pendar 2007]	k-NN	PervertedJustice (1,402)	Victim/Predator	94.3
[Tam 2009]	SVM	Lin2006 (2,161)	Author Age	78.6
[Tam 2009]	NB	Lin2006 (2,161)	Author Age	69.8
[Peersman et al. 2011]	SVM	Netlog (1,537,283)	Author Age	66.3
[Villatoro-Tello et al. 2012]	SVM+NN	PervertedJustice	Victim/Predator	93.5

fixated discourse – that predators will return to the subject of sex throughout grooming conversations. The authors apply a sentiment similarity measure to lexical chains identified from text. They hypothesise that long lexical chains related to sex are indicative of authors being sexual predators. They find some evidence for this in a comparison of the length of sex-fixated lexical chains in both a sexual predator corpus and a cyber-sex corpus. The second publication turns to the use of sentiment and emotion features in conversations involving sexual predators. The authors identify from related work that several sentiment features are linked to sexual predation, and construct a feature set based on sentiment markers. This feature set is compared to a number of simple character and word-based feature sets in a Naive Bayes classifier running over a corpus of chat logs from a vigilante website combined with ordinary cyber-sex logs.

[Inches and Crestani 2012] cover the 2012 International Sexual Predator Identification Competition, detailing a common evaluation framework against which 16 methods for identification of sexual predators could be evaluated in a comparable manner. The competitors were provided with a sample of 30% of a synthetic dataset constructed from a vigilante site and publicly-available IRC logs, and evaluated based on the F-score of their method's performance on the remainder of the set. The paper provides an overview of participants' approaches as well as their results. A more detailed account of the method used by the winning competitor [Villatoro-Tello et al. 2012] is also included in our review, as are the methods used by two other competitors [Morris and Hirst 2012; Peersman et al. 2012] who were placed 4th and 6th respectively. A master's thesis [Morris 2013] by the author of the 4th-ranked paper provides additional detail on their method's unsuccessful behavioural analysis addition to an SVM classifier using unigram and bigram features.

Some of the approaches taken in author profiling in this domain can be broadly compared to each other, so we provide a summary comparison table in Table IV. The two types of task attempted are distinguishing predators from their textual contributions and determining the age of authors as part of such a system. Note that many important details on the precise features and processes used in classification are best explained in the original publications, and note also that the figures given for author age classification are figures focused on general child-versus adult classification, and the results for more specific age groups vary from this figure within publications – the general effect being that older adults are easier to distinguish from teens and children. Additional comparable approaches can be seen in the results reported by [Inches and Crestani 2012], our figure for [Villatoro-Tello et al. 2012] being their (the topmost-ranked) performance on that evaluation.

3.5.5. Author Profiling in Terrorism and Extremism. [Gawron et al. 2012] use a vocabulary of group membership markers to rank documents by the degree of militancy of the author, with the aim of building more efficient search tools for such material. Working with a corpus of white extremist websites, they find that these hand-selected features, when weighted by TF-IDF, correlate more closely with human rankings of militancy than full feature-sets or feature-sets selected from those words with the highest mutual

information. They also outperform a variant using weights based on a cosine similarity measure. They also find that an SVM using TF-IDF and the full vocabulary performs best at classifying texts as militant.

3.5.6. Author Profiling in Threats and Harassment. [Chen et al. 2012] aim to profile users who are likely to send out abusive messages. They do this through a combination of lexical and syntactic features and independent sentence offensiveness measures which draw upon a form of sentiment analysis. They distinguish the degree to which profanity determines offensiveness and produce some tailored rules for identifying name-calling. Their evaluation against manual markup of 249 Youtube users' comments shows high abusiveness classification accuracy.

3.5.7. Other Author Profiling Applications. [Pham et al. 2009] do not make specific reference to a particular type of crime – beyond a generic formulation of 'cybercrime' – but undertake a range of author profiling, ambitiously attempting to derive not only age and gender information, but also the occupation of the author. They gather a corpus of well-used Vietnamese blogs, and run a large number of classifiers in a comparative evaluation for each classification task. They find, for the most part, that an IBk decision tree algorithm is best-performing, with the exception of the occupation classification, which is best served by a random forests classifier.

Sentiment Analysis.

3.5.8. Sentiment Analysis and Terrorism and Extremism. [Abbasi 2007; Abbasi et al. 2008] explore sentiment analysis on US and Middle Eastern web forum postings. In the first paper, the authors focus on the detection of emotions or affects in web-based discourse. The authors manually construct a lexicon mapping terms to a score of intensity in a particular category of sentiment. They proceed to a case study comparing the US and Middle Eastern extremist groups based on the intensity of the hate and violence intensity of their postings, finding a linear relationship in both cases and a strong one in the case of the Middle Eastern groups. In the second paper, the authors move towards automated sentiment classification of English and Arabic content. They use a range of stylistic and syntactic features and make use of a genetic algorithm to aid in feature selection. SVM classification using this system performed well at classification on a benchmark movie review dataset and on manually tagged English and Arabic forum postings.

[Yang et al. 2011] study detection of radical opinions in web forum postings. They particularly focus on detecting the features which are most relevant to such a specific form of classification task, working with a full range of lexical, structural, syntactic and content features. They validate their method on two U.S. based hate group fora, and find that a choice of lexicon for context is highly important. Their experimentation with a number of classifiers found SVM to outperform Naive Bayes and Adaboost.

[Rohn and Erez 2012] provide a very high-level description of mining of online information about agro-terrorism, providing no detailed implementation steps nor evaluation.

3.5.9. Sentiment Analysis and Threats and Harassment. Sentiment analysis has a particular role to play in detecting threats and harassment in text, due to its ability to detect the tone of conversation. It has been applied with some success to posts in both online fora and social media.

[Sobkowicz and Sobkowicz 2010] study the dynamics of political discussions on Polish internet fora, drawing on them as a source of strongly bipolar exchanges. They perform a topological assessment of the discussion network, and undertake a detailed analysis of the nature of user interactions and thread popularity based on political

affiliation of participants. They note a connection to analyses of hate groups, and contradict existing understanding of contrasting views leading to averaging of opinion.

[Warner and Hirschberg 2012] focus on detecting hate speech on the web, discussing issues with clearly defining hate speech — such as distinguishing reclamation or discussion of racial slurs from their offensive deployment. They perform a manual coding of hate speech related to Jews, and compare an SVM classifier using a number of feature sets to this ground truth, finding acceptable classification accuracy on a unigram feature set.

[Ptaszynski et al. 2010] cover cyber-bullying, detailing the design of a tool to help assist parents and school personnel in spotting malicious online posts. Drawing on a dataset of manually-gathered cyber-bullying instances, they perform a comparative affect analysis to distinguish the degree of emotion associated with cyber-bullying texts, drawing on an existing affect analysis framework with emoticon support. They found that there were not notably more emotive items in the positively labelled set, but that there were significantly more vulgarities. Interestingly, they also found evidence of sarcasm in their bullying dataset, with the category of 'fondness' ranking unexpectedly high. Based on this analysis, they build a machine learning system to be integrated into a web crawler for classifying malicious posts.

[Xu et al. 2012a] present social media as a valuable resource for facilitating academic study of bullying, and highlight a number of key challenges for the NLP community to overcome, using Twitter as a source for example data and a number of exploratory analyses. Their detailed exploration of the topic is a broad starting-point for researchers to expand on.

Text Classification.

3.5.10. Text Classification in Crimes against Children. A number of publications focus on estimating the volume of child abuse media in filesharing networks, identifying files which may contain child abuse based on their filenames. A critical component behind many of these approaches is the collection of appropriate keywords to identify in filenames, these terms being drawn from a specialised vocabulary used by sharers of this media.

[Steel 2009] focuses on evaluating the volume of child abuse material on the Gnutella network based on a keyword-based evaluation of filenames and search queries, also investigating a number of common claims regarding characteristics of such material on peer-to-peer networks. They find that just under 1% of queries and 1.45% of files were related to child abuse material.

[Prichard et al. 2011] specifically attempt to identify the path to the use of child abuse material, presenting results drawn from a three-month study of the isoHunt filesharing network's top 300 search terms, where 3 of 162 terms were linked to child abuse material.

[Panchenko et al. 2013] aim at building an automatic classifier for child abuse material based on filenames. As an early step, their SVM-based and logistic regression-based classifiers are trained and evaluated on pornographic filenames as a proxy, with promising initial results. Other work by the same authors [Panchenko et al. 2012] presents more detail on the implementation of the filename normalisation and classification procedure, but no new results evaluating the viability of their classifier in distinguishing child abuse material filenames from adult pornography.

[Fournier et al. rint] perform a comparison of paedophile activity in KAD and eDonkey, two distinct filesharing networks. Using an existing classification tool to label queries, they find that eDonkey contained more child abuse-related queries (0.25%) than KAD (0.09%).

This collection of studies, though working on different networks, tend to arrive at similar results regarding the extent of sharing of child abuse material, with a small but significant percentage of a number of filesharing platforms appearing to contain child abuse material.

[Penna et al. 2010] focus on detecting predation – rather than predators – in chat logs taken from online games such as World of Warcraft. Their method uses a keyword-lookup system whereby suspicious messages are those which reveal personal information. A small trial evaluation found that their system highlighted two synthetic suspicious messages inserted into ordinary chat logs, though it would also appear that a large false-positive rate is inherent to their approach.

3.5.11. Text Classification in Terrorism and Extremism. [Skillicorn 2004] focuses on detecting related messages, using a form of term frequency analysis to correlate and cluster messages using certain unusual words. Their focus is on detecting groups such as terrorists, that are aware of being monitored by keyword systems and are thus unnaturally altering their word usage. They demonstrate their approach on a synthetic dataset. Stronger demonstration that the word usage behaviour expected exists in communication traces would help validate the approach. [Fong et al. 2008] similarly aim to detect word substitutions in messages, a measure which might be adopted by those seeking to avoid keyword-based surveillance. They draw upon a range of weak sentence oddity indicators which, combined in a decision tree classifier, achieve good classification accuracy for sentences drawn from the Brown and Enron corpora where a noun has been replaced with another noun with a similar frequency.

[Skillicorn 2010] applies a number of word-usage models to posts on an English-language forum, drawing on measures of radicalisation and deception to rank forum posts and providing some analysis of the distribution of posts. They find that highly-radical posts are ranked low for deception, signalling sincerity.

3.5.12. Text Classification in Police Intelligence. [Sun and Ng 2011] focus on the prevention of drug abuse through monitoring social media. Their framework is designed to identify the popularity of posts within specified topics. Specifically, they focus on the prediction of comment arrival as a proxy for popularity, finding good results in an evaluation on Twitter and the Hong Kong Discussion forum.

3.5.13. Text Classification in Threats and Harassment. [Appavu alias Balamurugan and Rajaram 2008; Appavu et al. 2009], and [Wang et al. 2006; Banday et al. 2011; Shekar and Imambi 2008] cover building classifiers to detect threatening emails, all having been covered together under machine learning applications to harassment above.

3.5.14. Text Classification in Cybercrime. [Chang et al. 2012] appear to address cybercrime – though not clearly in the sense we use the term in this review – applying Naive Bayes, C4.5 and SVM classifiers to the somewhat ambiguous question of deciding whether or not texts are useful to cybercrime investigations. A trial on manually-coded case descriptions from a United States Department of Justice website suggests that all three classifiers have acceptable performance, with Naive Bayes the best-performing.

3.5.15. Text Classification in Finance. [Watters et al. 2011] focus on copyright infringement, sampling the BitTorrent network to gather information on the number of shared files, assigning files individual categories, and then checking a random sample of filenames manually to determine how many files appeared to contain copyright-infringing material. They find that the vast majority of shared files contain infringing content.

Other NLP methods for general application. [Wong and Xia 2008] deal with the normalisation of Chinese net-speak, a task which falls somewhere between automated

spellchecking and automated translation. They delve into the issue of chat normalisation in some depth, and construct a normalisation system based on a source channel model and phonetic mapping. Their system performs well compared to the ordinary source channel model.

[Kramer 2010] presents an anomaly detection framework for time-dependant datasets, applying this to a collection of forum posts from the Dark Web Portal as a trial application. The approach detects significant shifts in forum posting trends, drawing out some of its potential applications.

[Appavu et al. 2007] turn to an association rule mining approach for detecting deception, which uses a keyword-based feature selection trained on existing data to learn rules for classification. A synthetic, unspecified, dataset is used to generate a single rule, which is reported, but its utility at classification is not reported.

[Keila and Skillicorn 2005] cover deception detection from linguistic cues in the Enron email dataset. There is a detailed explanation of the application of singular value decomposition but no in-depth analysis of the approach is provided.

[Shukur et al. 2011] focus on NLP means of detecting deception in online chat software, speculating that an ontology-driven approach may be best, though no implementation is carried out. The paper consists mostly of a discussion of a possible approaches and an outline of an ontology.

3.6. Other methods

Additional publications using a variety of hard-to-categorise methods are described in Supplementary Material C.

4. ANALYSIS

In this section, we summarise key information which can be inferred from the corpus of publications included in our study.

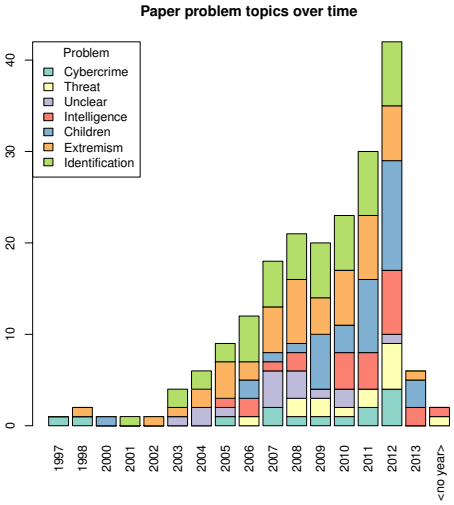
4.1. Research Questions

The first questions which may be answered are the guiding research questions outlined in Section 2.

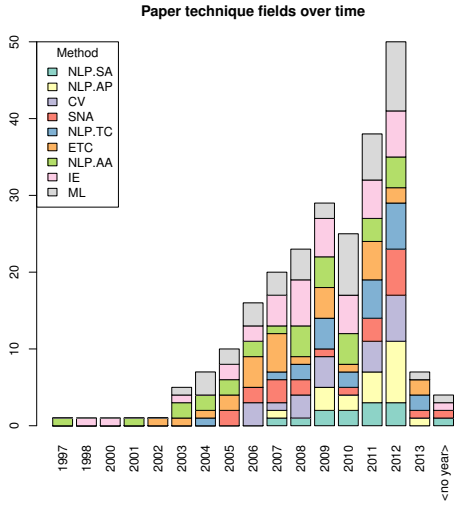
4.1.1. What are the problems (crimes, investigative requirements) are being addressed in the literature?. As can be seen in Figure 1(a), a number of high-impact crimes such as terrorism and the sexual predation of children are prominent topics, alongside more broadly applicable aims such as identification of offenders using online data.

In addition, Figure 1(a) shows the most common problem topics over time. It can be seen that topics such as the identification of internet users and the investigation of terrorism or extremism remain relatively stable (as a percentage of research output) over time, while the attention to crimes against children appears to have increased since around 2009.

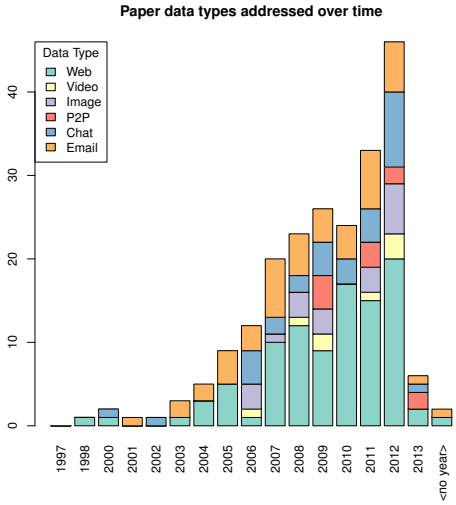
The problem of online identification was most often associated with NLP approaches to uncovering the author of a given written text, an aim relevant to legal debates about incriminating texts such as emails or blog posts. The authorship attribution literature connected to this aim appears to be rich and mature, with a number of comparable studies. This topic is extended somewhat in combined NLP and ML work – sometimes including computer vision techniques – which aims to cluster spam or phishing email campaigns to identify common origins, and similar aims motivate more traditional IP-lookup approaches. A very much distinct body of research by A.A. Mohamed and R.V. Yampolskiy also addresses identification, their focus being on approaches to identifying people via online avatars in virtual games.



(a) The most common problem topics over publication years



(b) The most common techniques over publication years



(c) Data type usage over publication years

Fig. 1. Problems, techniques and data types over publication years

Papers addressing extremist or terrorist problems almost uniformly apply themselves to investigating and monitoring online communities as part of information-gathering efforts. Several studies look at the links between different sites and communities, while some others look at means of identifying the most radical members of groups where discussion is visible. The two key demographics targeted are groups linked to Jihadist terrorism and far-right extremist groups in the United States, suggesting a U.S.-centered publication bias. The predominant trend in the Cybercrime

publications was also investigating and monitoring online communities, which suggests that there may be a binding theme of investigating online criminal communities.

There are two main categories of crimes against children visible in the reviewed publications. The first, and most common, is the detection of sexual predators engaged in online conversation with children, the aim being to detect attempts at grooming children for contact, a problem which by its nature draws heavily on NLP approaches. The second is the detection of child abuse material, which includes both CV attempts to discern such content from images and videos and filename-based attempts to quantify the volume of such content on a number of P2P filesharing networks.

Financial crime publications either address copyright infringement on P2P networks, or else the detection of fraud, usually from auction sites. Intelligence tools are most often concerned with either mining criminal social networks from open sources, or providing alerts about potential criminal activity, often with respect to certain geographic or temporal limits.

Those papers whose focus was least obviously a criminal matter often made reference to pornography, which may be indicative of different legal frameworks and cultural backgrounds. Such results in this review might be considered to address parental control systems rather than strictly handle criminal content.

It is worth noting that 'terrorism' and 'cybercrime' were both often used as general motivations, not necessarily specific to the paper's focus, with 74 papers containing a reference to 'terrorism' and 50 papers referencing 'cybercrime' compared to 47 and 12 papers actually labelled as addressing these topics.

4.1.2. What are the methods which are being employed to provide solutions?. Natural Language Processing (NLP) is highly dominant in our results, with around half of all collected papers making some use of NLP techniques in some way. The presentation in Figure 1(b) breaks down this category along closer lines. The heavily textual nature of most electronic communications makes this a somewhat unsurprising result. Machine Learning techniques are also well-represented, with common classifiers like SVMs and Naive Bayes being applied to a variety of problems.

There are 21 papers in the review (10.2% of the corpus) which make some use of computer vision or image processing techniques. The low proportion of such papers may be linked to the choice of search terms in the discovery phase of the review — there was no CV-linked term included, but there were NLP and SNA terms. Of these 21, 16 papers made use of only CV techniques. As is to be expected, most of the data sources used in this area were forms of image and video, with only a couple of exceptions where web and email data was processed visually.

22 papers in the review (10.7%) made use of some form of social network analysis (SNA). This number appears relatively low given a search term specifically selected for these methods, perhaps indicating a research area which requires further exploration. Of these 22 papers, 13 were labelled as only using SNA methods, the others overlapping with techniques from the domains of information extraction and natural language processing. The data used in these papers were primarily web data, including blogs and fora, but email data also formed a sizeable proportion of the study.

39 papers in the review (18.9%) focused on helping combat crime by mining information from public resources. Of these, 24 were labelled as solely oriented towards information extraction, while the remainder also used methods involving natural language processing and social network analysis. The vast majority of information extraction studies made use of web-based data, including online fora and social networking services like Twitter.

There are 43 papers (20.9%) which make use of machine learning techniques, only 20 of which make exclusive use of such techniques. Of the other 23, 18 use some form of

NLP technique, indicating a significant overlap between those papers labelled as using machine learning and those labelled as using natural language processing. Such a relationship is retrospectively unsurprising given the close relationship between these fields. The 43 papers were fairly evenly divided with respect to the data types studied, with email and web data each featuring in nearly half of all studies. Four studies handled chat data and two studies — one overlapping with CV techniques — made use of image data.

92 studies from the review (44.7% of the total) used some form of natural language processing, making this by far the largest category of methods. Of these 92, 65 used only NLP techniques, making this also the category with least overlap with other methods (closely followed by the much smaller group of computer vision). The large number of NLP-related studies collected may be linked to the inclusion of two terms in the search procedure which link to NLP.

There were 27 papers which did not fit within any of the broader technique categories. Chat data and network trace data were more prominent amongst these papers than in the main categories. Often these papers described frameworks or abstract processes for combating a threat.

4.1.3. Which online data sources are being used?. A breakdown of the different broad categories of data is provided in Figure 1(c). Most commonly examined was web data, with nearly half of all publications making some use of textual or semistructured web data. Within this category, simple web pages are most favoured, with social media – particularly Twitter – second and online fora third most popular. Behind Web data comes the other significant data source, email, which has the advantage of being both long-established and well-used. Chat data from instant messaging applications forms the third key data type under analysis, with comparatively few papers making use of images or videos.

With regard to specific data sources, a few common elements were observed across papers.

— **The Enron email dataset** is a dataset of roughly half a million email messages from roughly 150 users. It was originally made public as a result of Federal Energy Regulation Commission's investigation into the Enron corporation. A full explanation of the dataset is provided by [Klimt and Yang 2004]. As a labelled dataset of authors and a large volume of messages produced by them, this corpus was often a standard reference for studies attempting authorship attribution, but also used in some studies based on social network analysis. A total of 22 papers make reference to the Enron dataset.
— **PervertedJustice**[3] is a vigilante website where volunteers run sting operations by posing as minors and luring paedophiles into volunteering identifying and incriminating information. They publish a large and growing corpus of chat logs involving attempted grooming, which have been used frequently in studies attempting to identify sexual predators or analyse stages of grooming attacks. The common use of this resource points to a common issue for researchers working on such topics – the lack of actual case data to work with means that researchers must work with proxy data. The degree to which predator-volunteer conversations accurately reflect predator-child conversations is unclear, in part also due to this lack of real case data to verify results. 15 papers make reference to this resource.
— **The Dark Web Forum Portal** is a search and summarisation interface to a collection of 28 fora which are linked to extremist or terrorist material. As a standard

[3]perverted-justice.com

collection of online forums, it is of particular interest to researchers studying the organisation of terrorism online. 11 papers make reference to this resource.

4.1.4. How many studies are making use of multiple data sources?. Relatively few papers (21 or 10.2% of the total) combine different types of data or present methods which would apply to different types of data. Among those which did, the approach was typically either general monitoring of all network traffic such as in ECHELON and similar wiretapping, or else the use of NLP methods which could apply to online texts of many kinds. Slightly more papers (24 or 11.7%) were marked as 'partial' responses for this question, due to using a variety of data sources of the same type – for instance, using both English and Arabic text corpora to evaluate a hypothesis.

While in many cases a paper's contribution will be limited to one particular subject area, and thus would not be expected to apply to distinct data types, this result hints that more effort at synthesis of otherwise distinct forms of online data could well prove a fruitful area of research.

4.1.5. How many studies validate their contribution's utility to law enforcement practitioners?. Very few (10 or 4.9%) papers reported a positive evaluation of their tool's utility by a law enforcement practitioner or similar expert authority. It was possible to infer from the means of evaluation or similar references that a further (14 or 6.8%) of papers were written with co-operation of law enforcement, implicitly crediting the work with some level of practitioner support.

These figures do not necessarily reflect the true rate of interaction between researchers and practitioners, and it is possible that trials with law enforcement are only conducted after successful publication, or that law enforcement use of tools is not widely publicised in the name of reducing criminal awareness. Even with such qualifying scenarios in mind it seems problematic that papers specifically reporting themselves as supporting law enforcement and intelligence applications so rarely report on evaluations by the relevant professionals.

5. DISCUSSION

This paper has presented the results of a large and systematic survey of the computer science literature which relates to data mining methods and tools used in a law enforcement scenario.

Online criminality is often linked to the relative anonymity of electronic interaction, and in response to this the reviewed computer science literature, and particularly natural language processing, reveals a mature field of authorship analysis for online texts, with many rigorously evaluated methods for determining the author of a given text building on and referencing each other, with feature sets and reference corpora being shared between papers. Consideration has been given to the standards of evidence required for legal use. Other approaches to identification of online individuals for criminal matters show similar levels of evaluation.

The detection of sexual predators in online chat transcripts shows similar levels of interest, with multiple studies applying a range of methods to the same goal, and even a number of publications recording a specific competition, with methods using the exact same dataset so as best to be compared. It is notable, however, that the most common data source for these publications was a form of proxy data – the Perverted Justice website's transcripts between people outside of law enforcement agencies and sexual predators. This seems to indicate that a willing research community is having to work around legal or other restrictions on gaining access to actual criminal chat data.

Similar legal obstacles seem to be faced by researchers attempting to develop means to automatically detect child abuse media – many forced to use less-helpful forms of

proxy data such as adult pornography – and even studies merely attempting to quantify the presence of child abuse media, where filename-only approaches are dominant.

Publications investigating terrorism or extremism also have access to a common data source in the form of the Dark Web Forum Portal, though it appears to be less uniformly drawn upon. With this topic, widely mentioned even in publications not directly addressing it, scarcity of ground truth information about real-world threats appears to have diverted many efforts in the open literature into exploration of the networking and rhetorical properties of self-identified online extremist communities.

As is revealed in the breakdown of the quality analysis in Appendix A, a significant proportion of publications reviewed had deficiencies in evaluation and indeed a quarter of publications had no evaluation. While in a minority of cases this may be because the paper proceeds via theoretical proof, or because the format of the publication does not allow sufficient space, in others poor adherence to scientific standards are evident. Especially when designing methods for use in law enforcement or intelligence deployments, where lives may directly be ruined by underperforming analysis tools, researchers must be focused on the best way to identify objective truth regarding their methods.

In some cases, with papers relying on social network analysis or information extraction methods, and particularly where the method designed was semi-automated or involved visualisation, evaluation sections presented only demonstrative case studies as support for their tool or method's utility, which is sufficient only for exploratory presentations. If the contribution is increased performance of the analyst interfacing with the software, sufficiently defensible user trials must be presented, and the same might be said for tools which aim to help an analyst seek resources on the Web.

In other cases, laboratory evaluations of classifiers were presented, but insufficiently comparable to the real-world deployment scenarios. With the exception of copyright infringement, most fields of study in this review hold an inherent class imbalance problem – there are far fewer traces of criminals in the online world than there are traces of innocent netizens, and classifiers operating on a general population must thus overcome the likelihood of high rates of false alerts. Synthetic but unrealistic datasets may demonstrate a classifier's theoretical ability, but evaluations should always be linked to actual deployment scenarios.

A small number of long-term projects and toolsets were referenced in multiple papers gathered by this review. In some cases, these publications report on significant incremental improvements on developed approaches, with fresh evaluations (e.g., [Bali 2007; Mohamed and Yampolskiy 2012a; Mohamed and Yampolskiy 2012b]). In more modular systems, such as the Email Mining Toolkit, the discussion is limited to brief description of individual components of a larger toolset (e.g., [Stolfo et al. 2006a; Stolfo and Hershkop 2005]). The lack of detail makes it difficult to ascertain the strengths and weaknesses of such extensions with respect to each other as well as other comparable approaches. Further work and more detailed evaluations are needed to fully understand the effectiveness of such extensions.

Impacting and underlying this study's results is the rapid pace of development in online mediums. While certain technologies such as email have remained fundamentally consistent over the years, the same cannot be said for all online activity which draws law enforcement attention. Individual games and social networking platforms can become popular, draw law enforcement attention, and then become unpopular even as researchers devise the appropriate tools to analyse this content — some of the data sources in reviewed papers are from what might now be thought of as essentially dead communities. Certain methods, such as analysis of written text, can be generalised across a number of platforms and data sources, and are as such especially valuable.

The current volume of papers making use of multiple data sources is low, with information extraction studies being the most likely to attempt this. The community should put greater focus on tools which generalise to different applications. Many methods may already be transferable, but studies attempting to replicate the performance of a method on a new type of data are very rare. The cross-examination of different data sets might also help standards of evaluation for researchers working in areas where accurate ground truth is not readily available.

A number of papers on textual analysis and information extraction subjects demonstrate that their methods work with multiple languages, the most common being English and Arabic. English being dominant globally, online and in science makes it a clear target for analysis, whereas Arabic is clearly targeted by law enforcement and intelligence interest in counter-terrorism applications. The linguistic challenges behind textual analysis should not be forgotten or assumed solved when dealing with less-analysed tongues, but law enforcement should be aware that such technology exists outside of what is collected in this review, even if it does not advertise itself as applicable to law enforcement.

Finally, the extremely low overall level of engagement with actual law enforcement bodies or domain experts appears problematic for a corpus of papers specifically selected for referencing their intended deployment with law enforcement. This is not necessarily a problem which may be overcome by the research community alone, but attempts should be made to involve relevant professionals in the evaluation of tools being designed for their use.

5.1. Further Work

Replication of this work could prove useful to the community, firstly in the ordinary sense of validating this study's results and conclusions and helping shape the method, and secondly in identifying ongoing trends in publication which follow from the end of this corpus. The area of crime informatics appears to be expanding rapidly, and a significant volume of new contributions should be expected over the coming years, which will no doubt alter the picture presented in this paper's results. For instance, studies making use of real case data from law enforcement and live evaluations in law enforcement settings, e.g. [Hurley et al. 2013; Rashid et al. 2013; Peersman et al. 2014] and use of data from forums exclusively used by criminals, e.g. [Afroz et al. 2014] have started to appear. It would be interesting to study if this trend continues and whether the research community can build shared real-world data sources for both evaluating the tools and techniques and for pursuing a research agenda on multi-source data synthesis for law enforcement applications.

Additionally, this study could be used as a basis for deeper systematic exploration of the literature regarding a particular subgroup of topics or techniques identified in our results. Researchers with domain knowledge in a particular area may wish to expand our results through refined keyword selection or deeper following of references in already-identified work.

5.2. Bibliography and Supplementary Material

Some additional analysis of this dataset – including the results of a quality analysis, reviews for literature which was not easily classified, a guidance table for reading the results by problem topic, and a description of some publication patterns – were excluded from this document due to space considerations. They are made available as supplementary electronic material.

A. QUALITY ANALYSIS

The quality analysis outlined in Section 2.2 indicated high variation in the quality of papers, with an average overall score of 2.15, with a large standard deviation of 0.95. This agrees with reviewer impressions of the highly differing quality of reviewed publications.

While most studies (85%) outlined their proposed method with appropriate detail (those which did not nearly always being short position papers) and nearly three-quarters of all included papers described some form of evaluation, 47% of the studies included appeared to use neither an appropriate statistical evaluation nor domain experts in their evaluation. Though it should be acknowledged that this figure includes many studies where neither was appropriate, having such a large proportion of studies lack rigorous evaluation is problematic.

Table V. Quality Analysis results by problem topic

Topic	Average Quality Score	SD
Harassment	2.54	0.75
Identification	2.35	0.84
Children	2.24	1.02
Finance	2.21	0.94
Extremism	2.1	0.94
Intelligence	1.91	0.92
Cybercrime	1.5	1.04

As shown in Table V, individual problem topics differed from the mean, though few in a manner which can be shown to be statistically significant, due in part to the high variance in scores. We can see that within this review, studies promoting general intelligence toolkits, tackling cybercrime and addressing extremism were on average of lower quality, and that studies exploring harassment and cyberbullying or attempting to identify individuals were on average of higher quality, but it would be dangerous to infer much from such summary measures given the continually high deviation.

Table VI. Quality Analysis results by field of technique

Topic	Average Quality Score	SD
NLP:AP	2.74	0.84
NLP:AA	2.6	0.75
ML	2.52	0.74
NLP:SA	2.46	0.9
NLP	2.4	0.82
NLP:TC	2.37	0.63
CV	2.05	0.8
IE	1.88	0.94
NLP:O	1.69	1.07
SNA	1.55	0.71

Table VI shows a similar breakdown for the different fields of technique. It appears that NLP and Machine Learning studies have generally higher than average quality scores. In the case of NLP subfields like Author Profiling and Authorship Analysis, there are enough studies that these deviations can be be considered significant at the 5% level. The fields of Social Networking Analysis and Information Extraction appear to suffer from a lack of means for concrete evaluation of the performance of their methods on criminal datasets, appearing with below-average scores for paper quality.

B. PUBLICATION PATTERNS

B.0.1. Temporal. Analysis of the publications over time shows a clear bias for the past decade, the oldest paper with a known publication date being from 1997. Our queries of academic databases were only limited in date to include publications from 1970 onwards, so this would appear to be representative of publication trends. It would also appear that this area of research has been expanding rapidly over the past decade.

B.0.2. Venues. The venue which produced the most papers included in this review is the Intelligence and Security Informatics (ISI) conference, with 15 publications, closely followed by the Journal of Digital Investigation with 8 publications. Following these two were three venues with four publications apiece: the Journal of the American Society for Information Science and Technology, the proceedings of the International Conference on Digital Government Research and the ACM SIGKDD Workshop on Intelligence and Security Informatics.

B.0.3. Authors. Of the 473 individual authors, 103 contributed to more than one paper, 35 contributed to more than two papers and 19 contributed to more than three papers included in our review — these authors are noted in Table VII. The author appearing on the most papers was Hsinchun Chen, of the University of Arizona, with contributions to 16 papers, most of which dealt with web and web forum mining for the monitoring of violent extremists. His co-author Edna Reid also makes the list as the second most represented author in our review.

Table VII. Authors with contributions to more than 3 papers included in the review

Author	Number of Papers
Chen, Hsinchun	16
Reid, Edna	7
Iqbal, Farkhund	6
Qin, Jialun	6
Zhou, Yilu	6
Debbabi, Mourad	5
Fung, Benjamin C. M.	5
Skillicorn, D. B.	5
Watters, Paul A.	5
Yampolskiy, Roman V.	5
Abbasi, Ahmed	4
Appavu, S.	4
Edwards, Lynne	4
Kontostathis, April	4
Lai, Guanpi	4
Ma, Jianbin	4
Rajaram, R.	4
Teng, Guifa	4
Yearwood, John L.	4

C. OTHER METHODS

C.1. Other Identification Methods

[Dickson 2006b; Dickson 2006b; Dickson 2006a] examine the operation of common online chat protocols and programs to provide aid to online investigations. The focus of these papers leans towards forensics, providing specific details of default directories and operations where evidence such as logs and 'buddy lists' may be found. Some details could also be relevant to off-the-wire capture and analysis. [Van Dongen 2007] is similar in nature, appearing to primarily act as an update in that it examined Windows Live Messenger 8.0 rather than MSN Messenger 7.5.

[Prusty et al. 2011] describe a vulnerability in the OneSwarm P2P filesharing network which allows law enforcement to identify the source of shared content, with reference to the sharing of child exploitation material on that network, but also with reference to private investigation for copyright infringement. Their attacks require substantial investment from law enforcement, with a requirement to compromise 25% of the network.

[Stolfo and Hershkop 2005] outline the general framework of the Email Mining Toolkit. While the EMT has several intelligent modules detailed in other work within this review, this particular paper includes only the highest-level overview of such functionality, and includes no evaluation in itself.

[Lalla 2011] proposes a forensic methodology for handling email data, including consideration of data mining and classification stages while preserving the legal admissibility of evidence.

C.2. Other Studies on Cybercrime

[Morris et al. 2011] cover the prediction of cyber threats for businesses, describing a framework which would draw information from a dynamic collection of blogs, social media and web news articles using an ontology-driven model. No evaluation of the benefits of this approach is provided, with performance tests noted as future work.

[Roussinov and Robles-Flores 2007] delve into the field of detecting malevolent content as an application of information retrieval, comparing keyword searches with question-answering technology. They argue that answers to specific questions are more likely to be of an illegal nature than the results of a keyword-based search. They base their argument in favour of question-answering technology on both a comparison of standard algorithms and a well-described blinded empirical study with a small number of volunteers.

[Thorat and Manore 2010] describe a mass-monitoring system to be deployed on an organisational network, explicitly storing a large amount of user information. This includes a scanned copy of the users' identification documents alongside web browsing information, along with an indicator of the sensitivity of their activities, a measure whose derivation is not specified. The evaluation seemingly draws upon real-world deployments, but this only graphs the rate of alerts at different reporting thresholds.

[James and Kalutarage 2012] focus on the analysis of online conflicts by studying network traffic with a Bayesian reasoning approach to combining information from multiple sources. They provide an experimental evaluation with simulated attacks, and explore a number of methods for detecting the attacks, but their classifier has low precision.

[Chung and Wang 2007] summarise intended work on profiling the relationships between cyber-criminals, as well as visualising this information to help provide investigative leads, but provide no further detail.

[Tompsett et al. 2005] focus, in a broad sense, on the motivation for a framework for profiling the activities of cybercriminals, including geographical profiling, while listing

a few potential data sources. The paper only identifies possible areas of exploration, and does not in itself describe a system.

C.3. Other Studies on Crimes Against Children

[Chopra et al. 2006] propose to combat child abuse material at the network level by maintaining a time-relative obscenity score for every source IP address. They aim to derive obscenity measures from feature extraction on images reconstructed from IP packets. No evaluation of the approach is provided.

[Garcia-Ruiz et al. 2009] focus on describing the distribution of child abuse material on the online game Second Life, reviewing various ethical issues with monitoring the game as well as comparing Second Life's tools with measures which are common in other social networks.

[Latapy et al. 2013] quantify child abuse material on the eDonkey filesharing network. They base their analysis on two separate network captures of queries sent to two of the main eDonkey servers, with an expert analysis to aid with their keyword-based quantification of paedophile queries. They use the combination of IP address and port to work back from keywords and estimate the number users of eDonkey issuing such queries. They report that roughly 0.2% of eDonkey users search for paedophilic material. While the use of experts in validation strengthens confidence in the results as presented, this method would appear difficult to replicate.

[Gupta et al. 2012] aims to provide a deep understanding of paedophile grooming conversational strategy, working from a manual analysis of online paedophile chat logs to identify the different stages of grooming in chat. As with many studies in this area, their data is drawn from the Perverted Justice vigilante site. As well as a set of six categories of grooming conversational phase, they define conditional probabilities for transition between different stages.

C.4. Other Studies on Terrorism and Extremism

[Glaser et al. 2002] focus on the advocation of racial violence in online fora, approaching the issue through semi-structured interviews with members of racist internet chat rooms. They capitalise on the candour in opinions expressed online to gather close perspective on precise conditions leading to the advocation of violence. They provide a narrative analysis of their results.

[Freiburger and Crane 2008] contribute a systemic model of terrorist use of the internet, framing evidence from previous studies in terms of social learning theory, including comment on the power of group membership to the isolated, and the use of the uncensored internet for promotion of terrorist aims. They lay out some angles of approach for counter-terrorism online, including the infiltration of planning stages conducted in public.

[Cao 2008] proposes a field of behavior informatics and analysis, including terrorism and monitoring online communities among its motivations. The proposal motivates a range of data mining around behaviour, but is a broad appeal rather than a description of specific technology.

[Oh et al. 2011] present a case study covering the Mumbai terrorist attack based on Twitter data. They provide a deep analysis of the events of the attack as they relate to the social network, including the use of Twitter for situational awareness by the attackers' remote handlers and by mainstream media organisations. While they thus demonstrate the utility of Twitter as an information source for online data mining within the intelligence and law enforcement domain, they only go so far as to provide a conceptual framework for information control.

[Prentice et al. 2011] cover linguistic analysis of a corpus of violence-inciting online texts related to a specific event. The authors use a semi-automated coding system

which identifies persuasive content, distinguishing between different argument categories. They then use a semi-automated system for exploring the key concepts captured from the texts. Among other details, they find a strong tendency for extremist texts to use moral proof arguments to justify violence.

[Zolfaghar et al. 2009] aim somewhat broadly at establishing a framework for international counter-terrorism online through the use of honeypots and web mining. Most of their suggestions pertain to legislation, but they also suggest rather broad monitoring of web usage and deployment of honeypots.

C.5. Other Studies on Financial Crime

[Bauer et al. 2009] focus on monitoring Bittorrent traffic from the perspective of copyright holders attempting to enforce take-down notices. They note that existing policies of polling trackers for lists of peers and then distributing take-down notices to said IP addresses can be misled by malicious trackers or by malicious users registering false IP addresses with the tracker. In response, they propose an active-probing system which attempts to verify that a particular address is indeed sharing components of a file. They validate this approach in a comparative experiment.

[Ieong et al. 2009] also cover the investigation of peer-to-peer networks, but discuss the range of issues facing investigations in a survey manner rather than addressing a specific research question.

C.6. Other Studies on Police Intelligence

[Suzumura and Oiki 2011] describe a general framework for linking streams of open data in order to gain real-time information on a variety of topics, listing a number of application areas, with crime intelligence among them. As a case study, they trial their system on the contents of the Twitter Streaming API, displaying Twitter messages on a geographical map of their location.

[Monroy-Hernández et al. 2013] analyse the response of Twitter to calamitous events. They focus on a more prolonged event, the Mexican Drug War, highlighting the use of Twitter to warn of violence, and discuss the rises and falls in tweets with relation to events in the conflict. They also discuss their observations of certain users arising as curators of information.

C.7. Other Studies applied to ill-defined crimes

[Nabbali and Perry 2003; Nabbali and Perry 2004] are both two parts of a single narrative regarding the Carnivore electronic surveillance system and the ECHELON global interception run by the United States' Federal Bureau of Investigations. These publications cover both details of implementation and deployment and the legal rationale underlying the programmes' wide-ranging surveillance capabilities.

D. GUIDANCE TABLE

Section 3 and the related Appendix C orders the literature by the methods used. Readers interested in particular problems rather than methods may wish to use the following table to find discussion of relevant studies.

Problem	Sections
Financial Crime	3.1.5, 3.2.4, 3.3.5, 3.4.5, 3.5.15, C.5
Cybercrime	3.2.3, 3.3.4, 3.5.2, 3.5.14, C.2
Threats and Harassment	3.1.3, 3.4.3, 3.5.6, 3.5.9, 3.5.13
Intelligence	3.2.2, 3.3.2, 3.4.6, 3.5.12, C.6
Crimes against Children	3.1.2, 3.2.6, 3.3.3, 3.4.4, 3.5.4, 3.5.10, C.3
Terrorism and Extremism	3.1.4, 3.2.1, 3.3.1, 3.4.2, 3.5.3, 3.5.5, 3.5.8, 3.5.11, C.4
Online Identification	3.1.1, 3.2.5, 3.3.6, 3.4.1, 3.5.1, C.1
Unclear Problems	3.1.6, 3.3.7, 3.4.7, 3.5.7, 3.5.15, C.7

REFERENCES

A. Abbasi. 2007. Affect Intensity Analysis of Dark Web Forums. In *Intelligence and Security Informatics, 2007 IEEE*. IEEE Conference Publications, 282–288. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4258712

A. Abbasi and H. Chen. 2005. Applying authorship analysis to extremist-group Web forum messages. In *Intelligent Systems, IEEE*. IEEE Journals & Magazines, 67–75. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1512002

Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems* 26, 2 (2008), 7.

Ahmed Abbasi, Hsinchun Chen, and Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)* 26, 3 (2008), 12.

ACM. 2013. ACM Digital Library. (2013). http://dl.acm.org

Sadia Afroz, Aylin Caliskan-Islam, Ariel Stolerman, Rachel Greenstadt, and Damon McCoy. 2014. Doppelgänger Finder: Taking Stylometry To The Underground. In *IEEE Symposium on Security and Privacy*. IEEE, 212–226.

S. Aggarwal, J. Bali, Zhenhai Duan, and L. Kermes. 2007. The Design and Development of an Undercover Multipurpose Anti-spoofing Kit (UnMask). In *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*. IEEE Conference Publications, 141–150. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4412984

Edoardo Airoldi and Bradley Malin. 2004. *ScamSlam: An Architecture for Learning the Criminal Relations Behind Scam Spam*. Carnegie Mellon University, School of Computer Science,[Institute for Software Research International].

Rabeah Al-Zaidy, Benjamin Fung, Amr M Youssef, and Francis Fortin. 2012. Mining criminal networks from unstructured text documents. *Digital Investigation* 8, 3 (2012), 147–160.

S Appavu alias Balamurugan, R Rajaram, G Athiappan, and M Muthupandian. 2007. Data mining techniques for suspicious email detection: a comparative study. In *Proceedings of the European Conference on Data mining*. Portugal, 213–217.

S Appavu, Muthu Pandian, and R Rajaram. 2007. Association rule mining for suspicious email detection: a data mining approach. In *Intelligence and Security Informatics, 2007 IEEE*. IEEE, IEEE Conference Publications, 316–323.

S Appavu, R Rajaram, M Muthupandian, G Athiappan, and KS Kashmeera. 2009. Data mining based intelligent analysis of threatening e-mail. *Knowledge-Based Systems* 22, 5 (2009), 392–393.

S. Appavu alias Balamurugan and R. Rajaram. 2008. Learning to Classify Threaten E-mail. In *Modeling & Simulation, 2008. AICMS 08. Second Asia International Conference on*. IEEE Conference Publications, 522–527. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4530530

Martin Atkinson, Jenya Belayeva, Vanni Zavarella, Jakub Piskorski, Silja Huttunen, Arto Vihavainen, and Roman Yangarber. 2010. News mining for border security Intelligence. In *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on*. IEEE Conference Publications, 173–173. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5484744

Stefan Axelsson. 2000. *Intrusion detection systems: A survey and taxonomy*. Technical Report. Technical report.

N. Baili, D. D'Souza, A.A. Mohamed, and R.V. Yampolskiy. 2011. Avatar Face Recognition Using Wavelet Transform and Hierarchical Multi-scale LBP. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*. IEEE Conference Publications, 194–199. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6146968

Jasbinder Singh Bali. 2007. *Automation of Email Analysis Using a Database*. Master's thesis. Florida State University.

M Tariq Banday, Jameel A Qadri, Tariq Jan, Nisar Shah, and others. 2011. Detecting Threat E-mails using Bayesian Approach. *International Journal of Secure Digital Information Age* 1, 2 (2011), 10.

G. Barbian. 2011. Detecting Hidden Friendship in Online Social Network. In *Intelligence and Security Informatics Conference (EISIC), 2011 European*. IEEE Conference Publications, 269–272. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6061247

K. Bauer, D. McCoy, D. Grunwald, and D. Sicker. 2009. BitStalker: Accurately and efficiently monitoring bittorrent traffic. In *Information Forensics and Security, 2009. WIFS 2009. First IEEE International Workshop on*. IEEE Conference Publications, 181–185. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5386457

V. Benjamin and Hsinchun Chen. 2012. Securing cyberspace: Identifying key actors in hacker communities. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*. IEEE Conference Publications, 24–29. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6283296

K. Bernard, A. Cassidy, M. Clark, K. Liu, K. Lobaton, D. McNeill, and D. Brown. 2011. Identifying and tracking online financial services through web mining and latent semantic indexing. In *Systems and Information Engineering Design Symposium (SIEDS), 2011 IEEE*. IEEE Conference Publications, 158–163. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5876870

Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. 2012a. Modelling fixated discourse in chats with cyberpedophiles. In *Proceedings of the Workshop on Computational Approaches to Deception Detection (EACL 2012)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 86–90. http://dl.acm.org/citation.cfm?id=2388616.2388629

Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. 2012b. On the impact of sentiment and emotion based features in detecting online sexual predators. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis (WASSA '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 110–118. http://dl.acm.org/citation.cfm?id=2392963.2392986

J. Broadway, B. Turnbull, and J. Slay. 2008. Improving the Analysis of Lawfully Intercepted Network Packet Data Captured for Forensic Analysis. In *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*. IEEE Conference Publications, 1361–1368. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4529503

David Budgen, Mark Turner, Pearl Brereton, and Barbara Kitchenham. 2008. Using mapping studies in software engineering. In *Proceedings of PPIG*, Vol. 8. Lancaster University, 195–204.

Longbing Cao. 2008. Behavior Informatics and Analytics: Let Behavior Talk. In *Data Mining Workshops, 2008. ICDMW '08. IEEE International Conference on*. IEEE Conference Publications, 87–96. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4733926

Weiping Chang, Yungchang Ku, Sinru Wu, and Chaochang Chiu. 2012. CybercrimeIR–A Technological Perspective to Fight Cybercrime. In *Intelligence and Security Informatics*, Michael Chau, G.Alan Wang, WeiThoo Yue, and Hsinchun Chen (Eds.). Lecture Notes in Computer Science, Vol. 7299. Springer Berlin Heidelberg, 36–44.

Michael Chau and Jennifer Xu. 2007. Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies* 65, 1 (2007), 57 – 70. http://www.sciencedirect.com/science/article/pii/S1071581906001248 Information security in the knowledge economy.

Nisha Chaurasia, Mradul Dhakar, Astha Chharia, Akhilesh Tiwari, and RK Gupta. 2012. Exploring the Current Trends and Future Prospects in Terrorist Network Mining. In *Proceedings of The Second International Conference on Computer Science, Engineering and Applications (CCSEA 2012), Delhi, India*, Vol. 2. CSCP, 7.

Hsinchun Chen, Wingyan Chung, Jialun Qin, Edna Reid, Marc Sageman, and Gabriel Weimann. 2008. Uncovering the dark Web: A case study of Jihad on the Web. *Journal of the American Society for Information Science and Technology* 59, 8 (2008), 1347–1359.

Hsinchun Chen, Jialun Qin, Edna Reid, and Yilu Zhou. 2008. Studying Global Extremist Organizations' Internet Presence Using the DarkWeb Attribute System. *Terrorism Informatics* 18 (2008), 237–266.

Xiaoling Chen, Peng Hao, R Chandramouli, and K Subbalakshmi. 2011. Authorship Similarity Detection from Email Messages. In *Machine Learning and Data Mining in Pattern Recognition*, Petra Perner (Ed.). Lecture Notes in Computer Science, Vol. 6871. Springer Berlin Heidelberg, 375–386.

Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. 2012. Detecting offensive language in social media to protect adolescent online safety. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 71–80.

Hanqiang Cheng, Yu-Li Liang, Xinyu Xing, Xue Liu, Richard Han, Qin Lv, and Shivakant Mishra. 2012. Efficient misbehaving user detection in online video chat services. In *Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12)*. ACM, New York, NY, USA, 23–32.

Marc Cheong and Vincent CS Lee. 2011. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers* 13, 1 (2011), 45–59.

Munish Chopra, Miguel Vargas Martin, Luis Rueda, and Patrick CK Hung. 2006. A source address reputation system to combating child pornography at the network level. In *IADIS Internl. Conf. on Applied Computing*. 6.

Wingyan Chung. 2012. Categorizing temporal events: A case study of domestic terrorism. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*. IEEE Conference Publications, 159–161. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6284279

Wingyan Chung and G.A. Wang. 2007. Profiling and Visualizing Cyber-criminal Activities: A General Framework. In *Intelligence and Security Informatics, 2007 IEEE*. IEEE Conference Publications, 376–376. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4258736

Malcolm Walter Corney. 2003. *Analysing e-mail text authorship for forensic purposes*. Ph.D. Dissertation. Queensland University of Technology.

Glenn S Dardick, Claire R La Roche, and Mary A Flanigan. 2007. Blogs: Anti-forensics and counter anti-forensics. In *Proceedings of the 5th Australian Digital Forensics Conference*. School of Computer and Information Science, Edith Cowan University, Perth, Western Australia.

Richard Dazeley, John Yearwood, Byeong Kang, and Andrei Kelarev. 2010. Consensus clustering and supervised classification for profiling phishing emails in internet commerce security. Springer, 235–246.

Olivier De Vel, Alison Anderson, Malcolm Corney, and George Mohay. 2001. Multi-topic e-mail authorship attribution forensics. In *Proceedings of ACM Conference on Computer Security-Workshop on Data Mining for Security Applications*. ACM.

Mike Dickson. 2006a. An examination into AOL Instant Messenger 5.5 contact identification. *digital investigation* 3, 4 (2006), 227–237.

Mike Dickson. 2006b. An examination into Yahoo Messenger 7.0 contact identification. *digital investigation* 3, 3 (2006), 159–165.

T.D. Do, K. Chang, and S.C. Hui. 2004. Web mining for cyber monitoring and filtering. In *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*. IEEE Conference Publications, 399–404 vol.1. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1460448

David J Dreier. 2009. *Blog fingerprinting identifying anonymous posts written by an author of interest using word and character frequency analysis*. Master's thesis. Monterey, California; Naval Postgraduate School.

Patrick M Dudas. 2013. Cooperative, Dynamic Twitter Parsing and Visualization for Dark Network Analysis. In *Network Science Workshop*. IEEE, 172–176.

Y Elovici, A Kandel, M Last, B Shapira, and O Zaafrany. 2004. Using data mining techniques for detecting terror-related activities on the web. *Journal of Information Warfare* 3, 1 (2004), 17–29.

Yuval Elovici, Bracha Shapira, Mark Last, Omer Zaafrany, Menahem Friedman, Moti Schneider, and Abraham Kandel. 2010. Detection of access to terror-related Web sites using an Advanced Terror Detection System (ATDS). *Journal of the American society for information science and technology* 61, 2 (2010), 405–418.

Elsevier. 2013. ScienceDirect. (2013). http://sciencedirect.com

E. Endy, C. Lim, K.I. Eng, and A.S. Nugroho. 2010. Implementation of Intelligent Searching Using Self-Organizing Map for Webmining Used in Document Containing Information in Relation to Cyber Terrorism. In *Advances in Computing, Control and Telecommunication Technologies (ACT), 2010 Second International Conference on*. IEEE Conference Publications, 195–197. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5675810

SzeWang Fong, D. Roussinov, and D.B. Skillicorn. 2008. Detecting Word Substitutions in Text. In *Knowledge and Data Engineering, IEEE Transactions on*. IEEE Journals & Magazines, 1067–1076. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4538219

Raphaël Fournier, Thibault Cholez, Matthieu Latapy, Clémence Magnien, Isabelle Chrisment, Ivan Daniloff, and Olivier Festor. 2014 (2012 preprint). Comparing paedophile activity in different P2P systems. *Social Sciences (arXiv preprint)* 3, 3 (2014 (2012 preprint)), 314–325.

Richard Frank, Bryce Westlake, and Martin Bouchard. 2010. The structure and content of online child exploitation networks. In *ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD '10)*. ACM, New York, NY, USA, Article 3, 9 pages.

Tina Freiburger and Jeffrey S Crane. 2008. A systematic examination of terrorist use of the internet. *International Journal of Cyber Criminology* 2, 1 (2008), 309–319.

M.A. Garcia-Ruiz, M.V. Martin, A. Ibrahim, A. Edwards, and R. Aquino-Santos. 2009. Combating Child Exploitation in Second Life. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*. IEEE Conference Publications, 761–766. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5444398

Jean Mark Gawron, Dipak Gupta, Kellen Stephens, Ming-Hsiang Tsou, Brian Spitzberg, and Li An. 2012. Using Group Membership Markers for Group Identification in Web Logs. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence, 467–470.

A. Ge, W. Mao, and D. Zeng. 2010. Story extraction from the Web: A case study in security informatics. In *Service Operations and Logistics and Informatics (SOLI), 2010 IEEE International Conference on*. IEEE Conference Publications, 306–310. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5551561

Phyllis B Gerstenfeld, Diana R Grant, and Chau-Pu Chiang. 2003. Hate online: A content analysis of extremist Internet sites. *Analyses of Social Issues and Public Policy* 3, 1 (2003), 29–44.

N.A. Giacobe, Hyun-Woo Kim, and A. Faraz. 2010. Mining social media in extreme events : Lessons learned from the DARPA network challenge. In *Technologies for Homeland Security (HST), 2010 IEEE International Conference on*. IEEE Conference Publications, 165–171. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5655067

Jack Glaser, Jay Dixit, and Donald P Green. 2002. Studying hate crime with the internet: what makes racists advocate racial violence? *Journal of Social Issues* 58, 1 (2002), 177–193.

Andrew Gray, Philip Sallis, and Stephen MacDonell. 1997. Software forensics: Extending authorship analysis techniques to computer programs. University of Otago.

Glen L Gray and Roger Debreceny. 2007. Data Mining Of Emails To Support Periodic And Continuous Assurance. *College of Business and Economics, California State University at Northridge, Working Paper* (2007).

Glenn Greenwald, Ewen MacAskill, and Laura Poitras. 2013. Edward Snowden: the whistleblower behind the NSA surveillance revelations. *The Guardian* 9, 6 (2013).

Aditi Gupta, Ponnurangam Kumaraguru, and Ashish Sureka. 2012. Characterizing Pedophile Conversations on the Internet using Online Grooming. *arXiv preprint arXiv:1208.4324* (2012).

Rachid Hadjidj, Mourad Debbabi, Hakim Lounis, Farkhund Iqbal, Adam Szporer, and Djamel Benredjem. 2009. Towards an integrated e-mail forensic analysis framework. *digital investigation* 5, 3 (2009), 124–137.

John Haggerty, David Llewellyn-Jones, and Mark Taylor. 2008. FORWEB: file fingerprinting for automated network forensics investigations. In *Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop (e-Forensics '08)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, Article 29, 6 pages. http://dl.acm.org/citation.cfm?id=1363217.1363256

José María Gómez Hidalgo and Andrés Alfonso Caurcel Díaz. 2012. Combining Predation Heuristics and Chat-Like Features in Sexual Predator Identification. In *CLEF (Online Working Notes/Labs/Workshop)*.

Javad Hosseinkhani, Suriayati Chaprut, and Hamed Taherdoost. 2012. Criminal Network Mining by Web Structure and Content Mining. In *11th WSEAS International Conference on Information Security and Privacy*. WSEAS, 24–26.

Weiming Hu, Ou Wu, Zhouyao Chen, Zhouyu Fu, and Steve Maybank. 2007. Recognition of pornographic web pages by classifying texts and images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29, 6 (2007), 1019–1034.

Wen Hui, Hao Yin, and Chuang Lin. 2009. Design and deployment of a digital forensics service platform for online videos. In *Proceedings of the First ACM workshop on Multimedia in forensics (MiFor '09)*. ACM, New York, NY, USA, 31–36.

Wen Hui, Haiying Zhao, Chuang Lin, and Yang Yang. 2012. ViDeCloud: E cient Support for Large-scale Video Copy Detection. *Journal of Computational Information Systems* 8, 3 (2012), 1055–1062.

Ryan Hurley, Swagatika Prusty, Hamed Soroush, Robert J Walls, Jeannie Albrecht, Emmanuel Cecchet, Brian Neil Levine, Marc Liberatore, Brian Lynn, and Janis Wolak. 2013. Measurement and analysis of child pornography trafficking on p2p networks. In *Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 631–642.

Amin Abdurahman Ibrahim. 2009. *Detecting and preventing the electronic transmission of illicit images*. Master's thesis. University of Ontario Institute of Technology.

IEEE. 2013. IEEExplore. (2013). http://ieeexplore.ieee.org

Ricci Ieong, Pierre KY Lai, KP Chow, Michael Kwan, Frank Law, H Tse, and K Tse. 2009. Forensic Investigation of Peer-to-Peer Networks. *Handbook of Research on Computational Forensics, Digital Crime and Investigation: Methods and Solution* (2009), 355.

Giacomo Inches and Fabio Crestani. 2011. Online conversation mining for author characterization and topic identification. In *Proceedings of the 4th workshop on Workshop for Ph.D. students in information & knowledge management (PIKM '11)*. ACM, New York, NY, USA, 19–26.

Giacomo Inches and Fabio Crestani. 2012. Overview of the international sexual predator identification competition at PAN-2012. In *CLEF 2012 Evaluation Labs and WorkshopWorking Notes Papers. Rome, Italy*.

U. Inyaem, Phayung Meesad, C. Haruechaiyasak, and Dat Tran. 2009. Ontology-Based Terrorism Event Extraction. In *Information Science and Engineering (ICISE), 2009 1st International Conference on*. IEEE Conference Publications, 912–915. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5454498

Farkhund Iqbal. 2011. *Messaging Forensic Framework for Cybercrime Investigation*. Ph.D. Dissertation. Concordia University.

Farkhund Iqbal, Hamad Binsalleeh, Benjamin Fung, and Mourad Debbabi. 2010. Mining writeprints from anonymous e-mails for forensic investigation. *digital investigation* 7, 1 (2010), 56–64.

Farkhund Iqbal, Hamad Binsalleeh, Benjamin CM Fung, and Mourad Debbabi. 2013. A unified data mining solution for authorship analysis in anonymous textual communications. *Information Sciences* 231 (2013), 98–112.

Farkhund Iqbal, Rachid Hadjidj, Benjamin Fung, and Mourad Debbabi. 2008. A novel approach of mining write-prints for authorship attribution in e-mail forensics. *digital investigation* 5 (2008), S42–S51.

Farkhund Iqbal, Liaquat A. Khan, Benjamin C. M. Fung, and Mourad Debbabi. 2010. e-mail authorship verification for forensic investigation. In *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10)*. ACM, New York, NY, USA, 1591–1598.

Mofakharul Islam, Paul A. Watters, and John Yearwood. 2011. Real-time detection of childrens skin on social networking sites using Markov random field modelling. *Information Security Technical Report* 16, 2 (2011), 51 – 58. http://www.sciencedirect.com/science/article/pii/S1363412711000550 Social Networking Threats.

A.E. James and S.A. Kalutarage, H.K. andQin Zhou andShaikh. 2012. Sensing for suspicion at scale: A Bayesian approach for cyber conflict attribution and reasoning. In *Cyber Conflict (CYCON), 2012 4th International Conference on*. IEEE Conference Publications, 1–19. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6243988

SK Jayanthi and Ms S Sasikala. 2011. XGraphticsCLUS: Web Mining Hyperlinks and Content of Terrorism websites for Homeland Security. *Int J. Advanced Networking and Applications* 2, 6 (2011).

J.R. Johnson, A. Miller, L. Khan, and B. Thuraisingham. 2012. Measuring Relatedness and Augmentation of Information of Interest within Free Text Law Enforcement Documents. In *Intelligence and Security Informatics Conference (EISIC), 2012 European*. IEEE Conference Publications, 148–155. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6298825

Alexander John Karran and David Llewellyn-Jones. 2009. *A Digital Forensics Analytical Process Model for the Investigation, Analysis and Visualisation of Social Networks Derived from E-mail*. Master's thesis. Liverpool John Moores University.

P. S. Keila and D.B. Skillicorn. 2005. Detecting unusual email communication. In *Proceedings of the 2005 conference of the Centre for Advanced Studies on Collaborative research (CASCON '05)*. IBM Press, 117–125. http://dl.acm.org/citation.cfm?id=1105634.1105643

AKM Mustafizur Rahman Khan. 2012. A simple but Powerful E-mail Authorship Attribution System. In *International Conference on Machine Learning and Computing*.

Brendan Klare, Roman V Yampolskiy, and Anil K Jain. 2011. Face recognition in the virtual world: recognizing avatar faces. *Michigan State University, East Lansing, MI2010* 1 (2011), 40–45.

Bryan Klimt and Yiming Yang. 2004. Introducing the Enron Corpus.. In *CEAS*.

April Kontostathis, Lynne Edwards, Jen Bayzick, Amanda Leatherman, and Kristina Moore. 2009. Comparison of rule-based to human analysis of chat logs. *communication theory* 8 (2009), 2.

April Kontostathis, Lynne Edwards, and Amanda Leatherman. 2010. Text mining and cybercrime. *Text Mining: Applications and Theory.* (2010), 149–164.

Steve Kramer. 2010. Anomaly detection in extremist web forums using a dynamical systems approach. In *ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD '10)*. ACM, New York, NY, USA, Article 8, 10 pages.

Chih Hao Ku, Alicia Iriberri, and Gondy Leroy. 2008. Natural language processing and e-Government: crime information extraction from heterogeneous data sources. In *Proceedings of the 2008 international conference on Digital government research (dg.o '08)*. Digital Government Society of North America, 162–170. http://dl.acm.org/citation.cfm?id=1367832.1367862

Himal Lalla. 2011. *E mail forensic authorship attribution*. Ph.D. Dissertation. University of Fort Hare.

Matthieu Latapy, Clmence Magnien, and Raphal Fournier. 2013. Quantifying paedophile activity in a large P2P system. *Information Processing & Management* 49, 1 (2013), 248 – 263. http://www.sciencedirect.com/science/article/pii/S0306457312000283

HadyW. Lauw, Ee-Peng Lim, HweeHwa Pang, and Teck-Tim Tan. 2005. Social Network Discovery by Mining Spatio-Temporal Events. *Computational & Mathematical Organization Theory* 11 (2005), 97–118. Issue 2.

R. Layton, P. Watters, and R. Dazeley. 2010. Authorship Attribution for Twitter in 140 Characters or Less. In *Cybercrime and Trustworthy Computing Workshop (CTC), 2010 Second*. IEEE Conference Publications, 1–8. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5615152

Lucas Lenselink. 2011. *Radicalization Online. Patterns of Social Interaction on the Al-Faloja and As-Ansar Forums*. Master's thesis. Utrecht University.

M.J.-H. Lim, M. Negnevitsky, and J. Hartnett. 2007. Detecting Abnormal Changes in E-mail Traffic Using Hierarchical Fuzzy Systems. In *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*. IEEE Conference Publications, 1–6. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4295556

Zhi Liu, Zongkai Yang, Sanya Liu, and Yinghui Shi. 2012. Semi-Random Subspace Method for Writeprint Identification. *Neurocomputing* 108 (2012), 93–102.

Jianbin Ma, Ying Li, Guifa Teng, Fang Wang, and Yang Zhao. 2008. Sequential Pattern Mining for Chinese E-mail Authorship Identification. In *Innovative Computing Information and Control, 2008. ICICIC '08. 3rd International Conference on*. IEEE Conference Publications, 73–73. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4603262

Jianbin Ma, Guifa Teng, Shuhui Chang, Xiaoru Zhang, and Ke Xiao. 2011. Social network analysis based on authorship identification for cybercrime investigation. *Intelligence and Security Informatics* (2011), 27–35.

Jianbin Ma, Guifa Teng, Yuxin Zhang, Yueli Li, and Ying Li. 2009a. A Cybercrime Forensic Method for Chinese Web Information Authorship Analysis. *Intelligence and Security Informatics* (2009), 14–24.

Liping Ma, J. Yearwood, and P. Watters. 2009b. Establishing phishing provenance using orthographic features. In *eCrime Researchers Summit, 2009. eCRIME '09*. IEEE Conference Publications, 1–10. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5342604

S. Marcus. 1998. Dynamic data mining for information exploitation. In *Information Technology Conference, 1998. IEEE*. IEEE Conference Publications, 79–82. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=713386

S.H. Marjuni, R. Mahmod, A. Ghani, A. Bin Mohd Zain, and A. Mustapha. 2009. Lexical criminal identification for chatting corpus. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*. IEEE Conference Publications, 360–364. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5234700

India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski. 2011. Learning to identify Internet sexual predation. *International Journal of Electronic Commerce* 15, 3 (2011), 103–122.

D. Michalopoulos and I. Mavridis. 2011. Utilizing document classification for grooming attack recognition. In *Computers and Communications (ISCC), 2011 IEEE Symposium on*. IEEE Conference Publications, 864–869. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5983950

A. Modupe, O.O. Olugbara, and S.O. Ojo. 2011. Exploring Support Vector Machines and Random Forests to Detect Advanced Fee Fraud Activities on Internet. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE Conference Publications, 331–335. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6137398

A.A. Mohamed and R.V. Yampolskiy. 2012a. Using discrete wavelet transform and eigenfaces for recognizing avatars faces. In *Computer Games (CGAMES), 2012 17th International Conference on*. IEEE Conference Publications, 143–147. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6314566

Abdallah A Mohamed and Roman V Yampolskiy. 2012b. Wavelet Based Statistical Adapted Local Binary Patterns for Recognizing Avatar Faces. In *Advanced Machine Learning Technologies and Applications*. Springer, 92–101.

Andrés Monroy-Hernández, Emre Kiciman, Munmun De Choudhury, Scott Counts, and others. 2013. The new war correspondents: The rise of civic media curation in urban warfare. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1443–1452.

Colin Morris. 2013. *Identifying Online Sexual Predators by SVM Classification with Lexical and Behavioral Features*. Master's thesis. Department of Computer Science, University of Toronto.

Colin Morris and Graeme Hirst. 2012. Identifying Sexual Predators by SVM Classification with Lexical and Behavioral Features. In *CLEF (Online Working Notes/Labs/Workshop)'12*. 29.

T.I. Morris, L.M. Mayron, W.B. Smith, M.M. Knepper, R. Ita, and K.L. Fox. 2011. A perceptually-relevant model-based cyber threat prediction method for enterprise mission assurance. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on*. IEEE Conference Publications, 60–65. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5753755

Talitha Nabbali and Mark Perry. 2003. Going for the throat: Carnivore in an Echelon World Part I. *Computer Law & Security Review* 19, 6 (2003), 456 – 467. http://www.sciencedirect.com/science/article/pii/S0267364903006034

Talitha Nabbali and Mark Perry. 2004. Going for the throat: Carnivore in an ECHELON world - Part II. *Computer Law & Security Review* 20, 2 (2004), 84 – 97. http://www.sciencedirect.com/science/article/pii/S0267364904000184

M. Negnevitsky, M.J.-H. Lim, J. Hartnett, and L. Reznik. 2005. Email communications analysis: how to use computational intelligence methods and tools?. In *Computational Intelligence for Homeland Security and Personal Safety, 2005. CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE Conference Publications, 16–23. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1500603

S.M. Nirkhi, R.V. Dharaskar, and V.M. Thakre. 2012. Analysis of online messages for identity tracing in cybercrime investigation. In *Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on*. IEEE Conference Publications, 300–305. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6246131

Sarwat Nizamani, Nasrullah Memon, Uffe Kock Wiil, and Panagiotis Karampelas. 2013. Modeling Suspicious Email Detection Using Enhanced Feature Selection. *International Journal of Modeling and Optimization* 2, 4 (2013), 371–377.

Onook Oh, Manish Agrawal, and H Raghav Rao. 2011. Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. *Information Systems Frontiers* 13, 1 (2011), 33–43.

Chitu Okoli and Kira Schabram. 2010. A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems* (2010), 49.

Angela Orebaugh. 2006. An Instant Messaging Intrusion Detection System Framework: Using character frequency analysis for authorship identification and validation. In *Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International*. IEEE Conference Publications, 160–172. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4105332

Angela Orebaugh and Dr Jeremy Allnutt. 2010. Data Mining Instant Messaging Communications to Perform Author Identification for Cybercrime Investigations. *Digital Forensics and Cyber Crime* (2010), 99–110.

Angela Orebaugh and Jeremy Allnutt. 2009. Classification of instant messaging communications for forensics analysis. *The International Journal of Forensics Computer Science* (2009), 22–28.

Alexander Panchenko, Richard Beaufort, and Cedrick Fairon. 2012. Detection of Child Sexual Abuse Media on P2P Networks: Normalization and Classification of Associated Filenames. In *Proceedings of the LREC Workshop on Language Resources for Public Security Applications*.

Alexander Panchenko, Richard Beaufort, Hubert Naets, and Cédrick Fairon. 2013. Towards detection of child sexual abuse media: categorization of the associated filenames. In *Advances in Information Retrieval*. Springer, 776–779.

Shashank Pandit, Duen Horng Chau, Samuel Wang, and Christos Faloutsos. 2007. Netprobe: a fast and scalable system for fraud detection in online auction networks. In *Proceedings of the 16th international conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 201–210.

GA Patil, KB Manwade, and Mr PS Landge. 2012. A Novel Approach for Social Network Analysis & Web Mining for Counter Terrorism. *International Journal* 4 (2012).

Lisa Pearl and Mark Steyvers. 2012. Detecting authorship deception: a supervised machine learning approach using author writeprints. *Literary and linguistic computing* 27, 2 (2012), 183–196.

Claudia Peersman, Walter Daelemans, and Leona Van Vaerenbergh. 2011. Predicting age and gender in online social networks. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents (SMUC '11)*. ACM, New York, NY, USA, 37–44.

Claudia Peersman, Christian Schulze, Awais Rashid, Margaret Brennan, and Carl Fischer. 2014. iCOP: Automatically Identifying New Child Abuse Media in P2P Networks. In *IEEE Symposium on Security and Privacy Workshops*. 124–131.

Claudia Peersman, Frederik Vaassen, Vincent Van Asch, and Walter Daelemans. 2012. Conversation level constraints on pedophile detection in chat rooms. *PAN* (2012).

Nick Pendar. 2007. Toward Spotting the Pedophile Telling victim from predator in text chats. In *Semantic Computing, 2007. ICSC 2007. International Conference on*. IEEE, 235–241.

Yi-Ting Peng and Jau-Hwang Wang. 2008. Link analysis based on webpage co-occurrence mining - a case study on a notorious gang leader in Taiwan. In *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on*. IEEE Conference Publications, 31–34. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4565025

L. Penna, A. Clark, and G. Mohay. 2010. A Framework for Improved Adolescent and Child Safety in MMOs. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*. IEEE Conference Publications, 33–40. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5562791

Dang Duc Pham, Giang Binh Tran, and Son Bao Pham. 2009. Author Profiling for Vietnamese Blogs. In *Asian Language Processing, 2009. IALP '09. International Conference on*. IEEE Conference Publications, 190–194. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5380763

Sheryl Prentice, Paul J Taylor, Paul Rayson, Andrew Hoskins, and Ben OLoughlin. 2011. Analyzing the semantic content and persuasive composition of extremist media: A case study of texts produced during the Gaza conflict. *Information Systems Frontiers* 13, 1 (2011), 61–73.

Jeremy Prichard, Paul A Watters, and Caroline Spiranovic. 2011. Internet subcultures and pathways to the use of child pornography. *Computer Law & Security Review* 27, 6 (2011), 585–600.

Swagatika Prusty, Brian Neil Levine, and Marc Liberatore. 2011. Forensic investigation of the OneSwarm anonymous filesharing system. In *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 201–214.

Michal Ptaszynski, Pawel Dybala, Tatsuaki Matsuba, Fumito Masui, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi. 2010. In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis. *International Journal of Computational Linguistics Research* 1, 3 (2010), 135–154.

Jialun Qin, Yilu Zhou, Edna Reid, Guanpi Lai, and Hsinchun Chen. 2007. Analyzing terror campaigns on the internet: Technical sophistication, content richness, and Web interactivity. *International Journal of Human-Computer Studies* 65, 1 (2007), 71–84.

Awais Rashid, Alistair Baron, Paul Rayson, Corinne May-Chahal, Phil Greenwood, and James Walkerdine. 2013. Who am I? Analysing Digital Personas in Cybercrime Investigations. *IEEE Computer* 46 (2013).

Sebastián A. Ríos and Ricardo Muñoz. 2012. Dark Web portal overlapping community detection based on topic models. In *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD '12)*. ACM, New York, NY, USA, Article 2, 7 pages.

Eli Rohn and Gil Erez. 2012. Fighting Agro-Terrorism in Cyberspace: A Framework for Intention Detection Using Overt Electronic Data Sources. In *International ISCRAM Conference*.

S.G. Romaniuk. 2000. Using intelligent agents to identify missing and exploited children. In *Intelligent Systems and their Applications, IEEE*. IEEE Journals & Magazines, 27–30. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=850824

Dmitri Roussinov and José A Robles-Flores. 2007. Applying question answering technology to locating malevolent online content. *Decision Support Systems* 43, 4 (2007), 1404–1418.

F. Sahito, A. Latif, and W. Slany. 2011. Weaving Twitter stream into Linked Data a proof of concept framework. In *Emerging Technologies (ICET), 2011 7th International Conference on*. IEEE Conference Publications, 1–6. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6048497

Arab Salem, Edna Reid, and Hsinchun Chen. 2006. Content analysis of jihadi extremist groups videos. *Intelligence and Security Informatics* (2006), 615–620.

Arab Salem, Edna Reid, and Hsinchun Chen. 2008. Multimedia content coding and analysis: Unraveling the content of Jihadi extremist groups' videos. *Studies in Conflict & Terrorism* 31, 7 (2008), 605–626.

Michael Schmid. 2012. *Computer-Aided Writeprint Modelling for Cybercrime Investigations*. Master's thesis. Concordia University.

DV Chandra Shekar and S Sagar Imambi. 2008. Classifying and Identifying of Threats in E-mails–Using Data Mining Techniques. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1. 5.

Qiang Shen and T. Boongoen. 2012. Fuzzy Orders-of-Magnitude-Based Link Analysis for Qualitative Alias Detection. In *Knowledge and Data Engineering, IEEE Transactions on*. IEEE Journals & Magazines, 649–664. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5677516

Z. Shukur, A.H. Nasution, and A.A. Wibowo. 2011. Approaches to develop oracle for detecting deception in online chatting software. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*. IEEE Conference Publications, 1–3. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6021557

Asaf Shupo, Miguel Vargas Martin, Luis Rueda, Anasuya Bulkan, Yongming Chen, and Patrick CK Hung. 2006. Toward efficient detection of child pornography in the network infrastructure. *IADIS International Journal on Computer Science and Information Systems* 1, 2 (2006), 15–31.

D.B. Skillicorn. 2004. Detecting related message traffic. In *Workshop on Link Analysis, Security and Counterterrorism, SIAM Data Mining Conference*. 39–48.

D.B. Skillicorn. 2010. Applying interestingness measures to Ansar forum texts. In *ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD '10)*. ACM, New York, NY, USA, Article 7, 9 pages.

D.B. Skillicorn and N. Vats. 2007. Novel information discovery for intelligence and counterterrorism. *Decision Support Systems* 43, 4 (2007), 1375 – 1382. http://www.sciencedirect.com/science/article/pii/S0167923606000637 Special Issue Clusters.

P Sobkowicz and A Sobkowicz. 2010. Dynamics of hate based Internet user networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 73, 4 (2010), 633–643.

Jonathan F. Spencer. 2008. Using XML to map relationships in hacker forums. In *Proceedings of the 46th Annual Southeast Regional Conference on XX (ACM-SE 46)*. ACM, New York, NY, USA, 487–489.

Springer. 2013. SpringerLink. (2013). http://link.springer.com

Tommy Stallings, Brad Wardman, Gary Warner, and Sagar Thapaliya. 2012. WHOIS Selling All The Pills. *International Journal of Forensic Computer Science* (2012).

Efstathios Stamatatos. 2006. Authorship attribution based on feature set subspacing ensembles. *International Journal on Artificial Intelligence Tools* 15, 05 (2006), 823–838.

Efstathios Stamatatos. 2008. Author identification: Using text sampling to handle the class imbalance problem. *Information Processing & Management* 44, 2 (2008), 790–799.

Chad Steel. 2009. Child pornography in peer-to-peer networks. *Child Abuse & Neglect* 33, 8 (2009), 560–568.

Salvatore J. Stolfo, Germán Creamer, and Shlomo Hershkop. 2006a. A temporal based forensic analysis of electronic communication. In *Proceedings of the 2006 international conference on Digital government research (dg.o '06)*. Digital Government Society of North America, 23–24.

Salvatore J. Stolfo and Shlomo Hershkop. 2005. Email mining toolkit supporting law enforcement forensic analyses. In *Proceedings of the 2005 national conference on Digital government research (dg.o '05)*. Digital Government Society of North America, 221–222. http://dl.acm.org/citation.cfm?id=1065226.1065291

Salvatore J. Stolfo, Shlomo Hershkop, Chia-Wei Hu, Wei-Jen Li, Olivier Nimeskern, and Ke Wang. 2006b. Behavior-based modeling and its application to Email analysis. *ACM Trans. Internet Technol.* 6, 2 (may 2006), 187–221.

Beiming Sun and V.T. Ng. 2011. Lifespan and popularity measurement of online content on social networks. In *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. IEEE Conference Publications, 379–383. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5984118

Ashish Sureka, Ponnurangam Kumaraguru, Atul Goyal, and Sidharth Chhabra. 2010. Mining YouTube to Discover Extremist Videos, Users and Hidden Communities. *Information Retrieval Technology* (2010), 13–24.

T. Suzumura and T. Oiki. 2011. StreamWeb: Real-Time Web Monitoring with Stream Computing. In *Web Services (ICWS), 2011 IEEE International Conference on*. IEEE Conference Publications, 620–627. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6009445

Jenny K Tam. 2009. *Detecting age in online chat*. Master's thesis. Monterey, California Naval Postgraduate School.

Gui-Fa Teng, Mao-Sheng Lai, Jian-Bin Ma, and Ying Li. 2004. E-mail authorship mining based on SVM for computer forensic. In *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*. IEEE Conference Publications, 1204–1207 vol.2. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1382374

Sandeep A Thorat and Samadhan R Manore. 2010. Internet Usage Monitoring for Crime Detection. *Information Processing and Management* 70 (2010), 420–423.

Xiao-Ping Tian, Guang-Gang Geng, and Hong-Tao Li. 2010. A framework for multi-features based Web harmful information identification. In *Computer Application and System Modeling (ICCASM), 2010*

International Conference on. IEEE Conference Publications, V11–614–V11–618. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5623130

Deepak Tinguriya and Binod Kumar. 2010. Detecting terror-related activities on the web using neural network. *Oriental Journal of Computer Science and Technology* 3, 2 (2010), 6.

B.C. Tompsett, A.M. Marshall, and N.C. Semmens. 2005. Cyberprofiling: offender profiling and geographic profiling of crime on the Internet. In *Security and Privacy for Emerging Areas in Communication Networks, 2005. Workshop of the 1st International Conference on*. IEEE Conference Publications, 21–24. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1588290

Yuen-Hsien Tseng, Zih-Ping Ho, Kai-Sheng Yang, and Chun-Cheng Chen. 2012. Mining term networks from text collections for crime investigation. *Expert Systems with Applications* 39, 11 (2012), 10082 – 10090. http://www.sciencedirect.com/science/article/pii/S0957417412002965

Nilesh J Uke and Ravindra C Thool. 2012. Detecting Pornography on Web to Prevent Child Abuse–A Computer Vision Approach. *International Journal of Scientific and Engineering Research* 3, 4 (2012), 1–3.

Wouter S Van Dongen. 2007. Forensic artefacts left by Windows Live Messenger 8.0. *Digital Investigation* 4, 2 (2007), 73–87.

Esaú Villatoro-Tello, Antonio Juárez-González, Hugo Jair Escalante, Manuel Montes y Gómez, and Luis Villasenor-Pineda. 2012. A two-step approach for effective detection of misbehaving users in chats. In *CLEF (Online Working Notes/Labs/Workshop)*.

Chamila Walgampaya, Mehmed Kantardzic, and Roman Yampolskiy. 2010. Real Time Click Fraud Prevention using multi-level Data Fusion. In *Proceedings of the World Congress on Engineering and Computer Science*, Vol. 1. Citeseer, 20–22.

Hao Wang, Congxing Cai, Andrew Philpot, Mark Latonero, Eduard H. Hovy, and Donald Metzler. 2012b. Data integration from open internet sources to combat sex trafficking of minors. In *Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o '12)*. ACM, New York, NY, USA, 246–252.

Ke-Jian Wang, Xian-Zhong Han, Xin-Sheng Sun, Shu-Hui Chang, and Hui-Fang Qi. 2006. Research on Forecasting the Dangerous Level to Illegal Email Based on Integrated Immune Evolution Algorithm. In *Machine Learning and Cybernetics, 2006 International Conference on*. IEEE Conference Publications, 2112–2116. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4028413

Ning Wang, Keyu Jiang, R. Meier, and Hongbiao Zeng. 2012. Information Filtering against Information Pollution and Crime. In *Computing, Measurement, Control and Sensor Network (CMCSN), 2012 International Conference on*. IEEE Conference Publications, 45–47. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6245812

Xiaofeng Wang, D.E. Brown, and M.S. Gerber. 2012a. Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*. IEEE Conference Publications, 36–41. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6284088

Xiaofeng Wang, Matthew Gerber, and Donald Brown. 2012c. Automatic crime prediction using events extracted from twitter posts. *Social Computing, Behavioral-Cultural Modeling and Prediction* 7227 (2012), 231–238.

William Warner and Julia Hirschberg. 2012. Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media (LSM '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 19–26. http://dl.acm.org/citation.cfm?id=2390374.2390377

Paul A. Watters, Robert Layton, and Richard Dazeley. 2011. How much material on BitTorrent is infringing content? A case study. *Information Security Technical Report* 16, 2 (2011), 79 – 87. http://www.sciencedirect.com/science/article/pii/S1363412711000616 Social Networking Threats.

Chun Wei, Alan Sprague, Gary Warner, and Anthony Skjellum. 2008. Mining spam email to identify common origins for forensic application. In *Proceedings of the 2008 ACM symposium on Applied computing (SAC '08)*. ACM, New York, NY, USA, 1433–1437.

Luo Wenhua and Li Na. 2010. Application of unstructured data processing and analyzing base on chinese in digital data evidence collecting. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*. IEEE Conference Publications, V7–780–V7–783. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5485757

Kam-Fai Wong and Yunqing Xia. 2008. Normalization of Chinese chat language. *Language Resources and Evaluation* 42 (2008), 219–242. Issue 2.

Xinyu Xing, Yu-Li Liang, Hanqiang Cheng, Jianxun Dang, Sui Huang, Richard Han, Xue Liu, Qin Lv, and Shivakant Mishra. 2011. SafeVchat: detecting obscene content and misbehaving users in online video chat services. In *Proceedings of the 20th international conference on World wide web (WWW '11)*. ACM, New York, NY, USA, 685–694.

Jennifer Xu and Michael Chau. 2006. Mining communities of bloggers: A case study on cyber-hate. In *International Conference on Information Systems, Milwaukee, WI (December 10-13)*. 11.

Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, and Amy Bellmore. 2012a. Learning from bullying traces in social media. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 656–666. http://dl.acm.org/citation.cfm?id=2382029.2382139

Jun-Ming Xu, Xiaojin Zhu, and Amy Bellmore. 2012b. Fast Learning for Sentiment Analysis on Bullying. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '12)*. ACM, New York, NY, USA, Article 10, 6 pages.

C.C. Yang and T.D. Ng. 2008. Analyzing content development and visualizing social interactions in Web forum. In *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on*. IEEE Conference Publications, 25–30. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4565024

C.C. Yang and T.D. Ng. 2009. Web opinions analysis with scalable distance-based clustering. In *Intelligence and Security Informatics, 2009. ISI '09. IEEE International Conference on*. IEEE Conference Publications, 65–70. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5137273

Li Yang, Feiqiong Liu, J.M. Kizza, and R.K. Ege. 2009. Discovering topics from dark websites. In *Computational Intelligence in Cyber Security, 2009. CICS '09. IEEE Symposium on*. IEEE Conference Publications, 175–179. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4925106

Ming Yang, Melody Kiang, Hsinchun Chen, and Yijun Li. 2012. Artificial immune system for illicit content identification in social media. *Journal of the American Society for Information Science and Technology* 63, 2 (2012), 256–269.

Ming Yang, Melody Kiang, Yungchang Ku, Chaochang Chiu, and Yijun Li. 2011. Social Media Analytics for Radical Opinion Mining in Hate Group Web Forums. *Journal of Homeland Security and Emergency Management* 8, 1 (2011).

J. Yearwood, M. Mammadov, and A. Banerjee. 2010. Profiling Phishing Emails Based on Hyperlink Information. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*. IEEE Conference Publications, 120–127. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5562782

Dawei Yin, Zhenzhen Xue, Liangjie Hong, Brian D Davison, April Kontostathis, and Lynne Edwards. 2009b. Detection of harassment on web 2.0. In *Proceedings of the Content Analysis in the WEB 2.0 (CAW2. 0) Workshop at WWW2009*. 20–24.

Hao Yin, Wen Hui, Quan Miao, Zheng Li, and Chuang Lin. 2009a. IVForensic: a digital forensics service platform for internet videos. In *Proceedings of the 17th ACM international conference on Multimedia (MM '09)*. ACM, New York, NY, USA, 1015–1016.

Chengcui Zhang, Wei-Bang Chen, Xin Chen, and Gary Warner. 2009. Revealing common sources of image spam by unsupervised clustering with visual features. In *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 891–892.

Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2005. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology* 57, 3 (2005), 378–393.

Rong Zheng, Yi Qin, Zan Huang, and Hsinchun Chen. 2003. Authorship analysis in cybercrime investigation. In *Proceedings of the 1st NSF/NIJ conference on Intelligence and security informatics (ISI'03)*. Springer-Verlag, Berlin, Heidelberg, 59–73. http://dl.acm.org/citation.cfm?id=1792094.1792100

Yilu Zhou, Jialun Qin, Guanpi Lai, and Hsinchun Chen. 2007. Collection of U.S. Extremist Online Forums: A Web Mining Approach. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*. IEEE Conference Publications, 70–70. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4076513

Yilu Zhou, Jialun Qin, Edna Reid, Guanpi Lai, and Hsinchun Chen. 2005a. Studying the presence of terrorism on the web: an knowledge portal approach. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries (JCDL '05)*. ACM, New York, NY, USA, 402–402.

Yilu Zhou, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai. 2005b. US domestic extremist groups on the Web: link and content analysis. *Intelligent Systems, IEEE* 20, 5 (2005), 44–51.

Zhenghui Zhu. 2007. *Deconstruction and Analysis of Email Messages*. Master's thesis. Florida State University.

Jianwei Zhuge, Thorsten Holz, Chengyu Song, Jinpeng Guo, Xinhui Han, and Wei Zou. 2009. Studying malicious websites and the underground economy on the Chinese web. *Managing Information Risk and the Economics of Security* (2009), 225–244.

K. Zolfaghar, A. Barfar, and S. Mohammadi. 2009. A framework for online counter terrorism. In *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*. IEEE Conference Publications, 1–5. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5402641