Seeing the Forest Through the Trees: Identifying Key Players in the Online Distribution of Child Sexual Exploitation Material

Chapter · December 2016

CITATION
1

READS
220

2 authors:

Bryce Garreth Westlake
San Jose State University
13 PUBLICATIONS **207** CITATIONS
SEE PROFILE

Richard Frank
Simon Fraser University
76 PUBLICATIONS **911** CITATIONS
SEE PROFILE

Seeing the Forest Through the Trees: Identifying Key Players in the Online Distribution of Child Sexual Exploitation Material

Bryce G. Westlake[a] and Richard Frank[b]

Abstract:
The ever-increasing prevalence of child sexual exploitation material (CEM) in cyberspace requires that an interdisciplinary approach be adopted to improve combat efforts. Central to this is the incorporation of technologies that can reduce the physical, mental, and resource strain experienced by law enforcement, including intelligently automating some of the detection process, minimizing visual contact with CEM, and prioritizing targets. To maximize the impact of law enforcement activities against online CEM distribution, combat strategies need to be identified that allow social control agencies to see the 'forest through the trees' to target key players within the massive distribution chain. This paper focuses on identifying key players (i.e., public websites) within online CEM distribution networks, through the adaptation of a composite measure known as Network Capital (NC). We use a custom-designed webcrawler tool to automatically scan and collect information on websites with known CEM. We then incorporate quantity and quality of CEM material being distributed, network connectivity, geographical location and website operator information to create a formula to identify targets, sensitive to jurisdictional constraints. We also show how NC is malleable to the requirements of the researcher or social control agencies to emphasize specific combat priorities.

Keywords: child sexual exploitation, child pornography; webcrawler, social network analysis, network capital

[a] San Jose State University, CA, USA
[b] Simon Fraser University, Burnaby, British Columbia, Canada
Corresponding Author:
Bryce Westlake, San Jose State University,
One Washington Square, San Jose, CA, USA, 95192-0050
Email: Bryce.Westlake@sjsu.edu

Seeing the Forest Through the Trees: Identifying Key Players in the Online Distribution of Child Sexual Exploitation Material

The online distribution of child sexual exploitation material (CEM), commonly referred to as child pornography, is conducted through a variety of mediums. Growth in the availability of cellular phones, cameras, webcams, and other image and video recording technology, coupled with the efficient global reach provided by cyberspace, has turned a traditionally isolated crime into an international community (Hillman, Hooper, & Choo, 2014). In turn, this has put considerable physical, mental, and resource strain on those combating the crime (Bourke & Craun, 2014; Burns, Morley, Bradshaw, & Domene, 2008; Craun, Bourke, & Coulson, 2015; Krause, 2009; Perez, Jones, Englert, & Sachau, 2010; Powell, Cassematis, Benson, Smallbone, & Wortley, 2015). The ever-increasing prevalence and complexity of online CEM distribution requires that an interdisciplinary approach be adopted to improve combat efforts. Central to this interdisciplinary approach is the incorporation of technologies that can reduce the strain experienced by child sexual exploitation investigators, including intelligently automating some of the detection processes, to increase efficiency and minimize visual contact with CEM and to prioritize targets.

To address some of the challenges associated with detecting and combatting CEM, several private organizations have partnered with law enforcement agencies to develop specialized tools. In conjunction with Toronto Police Services, Microsoft© developed a law enforcement and intelligence agencies repository of information and material related to child exploitation that can be shared internationally to improve investigation coordination (Microsoft, 2005). This *Child Exploitation Tracking System* is now used by multiple law enforcement agencies in Canada, Australia, United States, United Kingdom, and elsewhere (Microsoft, 2012).

In partnership with the National Center for Missing and Exploited Children, Microsoft© has also created *PhotoDNA*, which is able to efficiently analyze large quantities of images and detect modified versions of known CEM (Microsoft, 2009), while Google© has adapted pattern recognition software, used to identify copyrighted material on YouTube©, to detect CEM (Shiels, 2008). More recently, Facebook©, Google©, Microsoft©, Twitter©, and Yahoo© have incorporated the Internet Watch Foundation's database of known child sexual exploitation images into their products to delete/block CEM as soon as it is identified (O'Neil, 2015).

Beyond the simple detection of CEM is the prioritization of offenders, or offending entities (e.g., websites). The detection and subsequent removal of CEM is typically short-lived as it is easy for offenders to replace the deleted material or move it to another location in cyberspace. Difficulties with international jurisdiction can also hamper efficiency (Gillespie, 2011). Compounding the problem are the legal challenges researchers and social control agencies face, seeking to investigate the issue and aid law enforcement. While section 163.1(6) of the Canadian Criminal Code (1985) states that "the courts shall find the accused not guilty if the representation…that is alleged to constitute child pornography has…an educational, scientific, or medical purpose", this type of exception is not uniform across other countries, especially the United States. Although the research presented in this chapter, and conducting in Canada, did not require the 'possession' of child pornography, and thus did not have to rely on the aforementioned law, global research efforts to identify and test methods for detecting previously unknown material are hindered by legal constraints. Even in countries where the research can be conducted, there are complexities that need to be addressed and safeguards that need to be put in place. For our purpose, this meant discussions with lawyers, making the information technology staff at our university aware of our research topic, and implementing

multiple security measures within our research center and on the computers where the data collection was taking place. Together, international jurisdiction issues and research challenges mean that law enforcement resources are strained, with minimal impact on detection of new material, and overall distribution. To maximize impact, strategies need to be identified that allow social control agencies to see the 'forest through the trees' and target key players within the distribution chain. Through creative methods and interdisciplinary partnerships, such as those presented here, these types of goals can be achieved with minimal to no violations of criminal law.

This chapter combines the technology of automated CEM detection and identification with a social network analysis measure known as Network Capital (Schwartz & Rouselle, 2009), to identify key players (i.e., public websites) within the online networks of CEM distribution. In total, we analyze a network of 83 public websites that disseminate child sexual exploitation images, identified through MD5 hash values. We then use the combination of each website's distribution of videos and images with the frequency of CE-related keywords and connectivity to 'new' websites, to identify key targets that should be the focus of investigations. To account for jurisdictional issues in targeting strategies, we incorporate information about the geographical location of the website host and owner. We begin with an overview of the conceptualization of a key player and the approaches that have been used thus far to identify key players. We follow that with a discussion of the value in taking an interdisciplinary approach to combatting online child sexual exploitation and how that can be accomplished.

Identifying Key Players

Generally speaking, the identification of a key player within a criminal network is grounded in the desire to maximize reduction in criminality. This can be accomplished through

intelligence gathering, network disruption, and/or removal of material (e.g., drugs, weapons, images, videos, etc.). However, identifying key players within any criminal organization can be a complex task as the definition of a key player can differ depending on the goals of the identification. Traditionally, key players have been characterized as those most 'central' (Bonacich, 1972; 1987; Freeman, 1979; Katz, 1953; Wasserman & Faust, 1994) to the network. To some extent, this centrality-based definition remains prevalent in the offline study of illegal drug distribution (Morselli, 2010), co-offending (Tayebi, Bakker, Glasser, & Dabbaghian, 2011) and adolescent delinquency (Liu, Patacchini, Zenou, & Lee, 2012).

The conceptual link between centrality and key players is also prevalent online. For example, PageRank formulas, like those used by Google (Sobek, 2003, Page, Sergey, Rajeev, & Terry, 1999) and social network researchers (Heidemann, Klier, & Probst, 2010), are based on eigenvector centrality (see Bonacich, 2007; Hanneman & Riddle, 2005). Social science research of online activist groups (Nouh & Nurse, 2015) and child sexual exploitation distribution networks (Joffres, Bouchard, Frank, & Westlake, 2011) have also incorporated centrality measures in key player analyses. However, one of the limitations of many centrality measures is that they conflate visibility with being synonymous with key player. In many criminal networks, the key players are not necessarily those who are the most visible (Malm & Bichler, 2011; Medina & Hepner, 2008; Morselli, 2009; Natarajan, 2006). Moreover, criminal organizations have moved towards a more decentralized network framework (Bouchard & Nash, 2015; Decary-Hetu, Morselli, & Leman-Langlois, 2012; van Dijk, Spapens, Reichel, & Albanese, 2014) effectively disconnecting the notion that the most visible are the key players. While social network measures of centrality are useful in gathering intelligence or fragmenting criminal

networks, their omission of resource and connectivity variability between network nodes is problematic.

To address the problem of centrality measures defining key players, Borgatti (2006) proposed that fragmentation (see Borgatti, 2003) and inter-set cohesion (reach) –"a direct measure of the amount of connection between a set [of nodes] and the rest of the [network]" (p.28) – be used to identify the most optimal *set* of nodes. Comparing fragmentation, hub, and bridge network disruption strategies, Joffres et al, (2011) found that the effectiveness of each was dependent on the goals/objectives of law enforcement in reducing density, clustering, reachability, or cohesion. Although Borgatti's (2006) proposal did address the connectivity goals of law enforcement disruption strategies, it did not take into consideration the resources (e.g., material or content) provided, or available, to each node; an important consideration for law enforcement targeting strategies.

To address Borgatti's (206) omission of resource-sharing, Schwartz and Rouselle (2009) incorporating weighted measures of each node's attributes and connections. Offline, it may be difficult to ascertain an accurate measure of a node's (e.g., offender or organization) resources; online this process becomes easier. In studying piracy, hacking, fraud, or other cybercrimes the resources available to each node can be accurately calculated (e.g., Décary-Hétu & Dupont, 2012; Décary-Hétu & Morselli, 2011; Holt, Strumsky, Smirnova, & Kilger, 2012. For example, resources for websites involved in the distribution of child sexual exploitation material may include the amount or type of content being distributed, the strength or frequency of connections, and/or the physical location or speed of the website server and those operating the website. Using methods like those described by Schwartz and Rouselle (2009), each of these resources can be quantified and given a weighted value. Of course, the ability to measure available resources is

contingent on the completeness, thoroughness, and accuracy of the methods used to collect the data.

An Interdisciplinary Approach to Cybercrime Research

The use of Internet-mediated research (IMR) methods continue to grow for both primary and secondary research purposes (Hewson, Vogel, & Laurent, 2015). While gaining in usage, one roadblock for adoption by some social science researchers is that many are not adept at developing data collection techniques and methods for use in cyberspace. As a result, an opportunity has arisen to foster interdisciplinary research partnerships, especially in the study of cybercrime. For example, interdisciplinary partnerships between social and computer scientists have led to the creation of fictionally vulnerable computer systems (honeypots) to observe the motives and techniques used by hackers (Almutairi, Parish, & Phan, 2012; Marin, Naranjo, & Casado, 2015; Provos & Holz, 2007; Spitzner, 2003), as well as the creation of automated data collection tools to examine child sexual exploitation on public websites (Frank, Westlake, & Bouchard; Westlake, Bouchard, & Frank, 2011) and peer-to-peer networks such as Gnutella (Steel, 2009), eDonkey (Fournier, et al., 2014), and BitTorrent (Rutgaizer, Shavitt, Vertman, & Zilberman, 2012).

Among the primary reasons for the increase in IMR is the amount of data that is readily available to researchers (Hewson, Vogel, & Laurent, 2015). However, it is this very advantage that also provides, potentially, the greatest challenge. The abundance of data available requires a modification to the ways that social science researchers traditionally gather, organize, and analyze data. While some studies have concluded that the quality and representativeness of Internet data is comparable to offline survey data (e.g., Change & Krosnick, 2009; Dillman, 2007), even proponents of IMR have raised concerns that adapted techniques and methodologies

will not undergo rigorous validation (Schonlau, van Soest, Kapteyn, & Couper, 2009; Shropshire, Hawdon, & Witte, 2009).

There are also ethical concerns about how IMR data is collected (Ess, 2013). For newly developing domains (e.g., Internet), the appropriate ethical practices are still being determined. Among cyber-researchers, ethical arguments often center on the general topic of consent. For example, Hewson (2003) posits whether it is ever ethically justified to use publicly available data, if the data has not been voluntarily, and *deliberately*, made available. Included within that discussion is where is the line between public and private Internet data? Within cybercrime research, Holt (2010) highlights the issue of observation and whether identifying oneself as a researcher, in a deviant online group, runs the risk of data contamination and to their cyber and/or personal safety. As a result, university ethics review boards may not be clear on the appropriate protocol for protecting researchers from legal issues that may arise from the data collection (e.g., child pornography) necessary for studying online illegal activities.

Specific to automated data collection techniques, the tool and method of identification need to be shown to be reliable and valid. In studying the distribution of child sexual exploitation material, this means that data collection is optimized with multiple criteria (see Westlake, Bouchard, & Frank, 2012; Westlake, Bouchard, & Girodat, in press) and is able to distinguish between relevant and irrelevant data (see Westlake, Bouchard, & Frank, 2015). That is, distinguish between child exploitation and non-child exploitation data (i.e., websites).

Current Study

The ever-increasing prevalence of child sexual exploitation material (CEM) coupled with the lack of manpower, resources, and cooperation between government bodies and countries have been cited as barriers to successful CEM investigations (Jewkes & Andrews, 2007; Wortley

& Smallbone, 2012). Together, these issues point to the need for a combat strategy of target prioritization rather than 'blindly swinging at websites'. Adapting Schwartz and Rouselle's (2009) concept of network capital to the distribution of CEM on public websites, Westlake, Bouchard, and Frank (2011) began to take a target prioritization approach, formulating a measure that considered the content (severity) and connections (connectivity) of each website within a larger network.

However, Westlake, Bouchard, and Frank's (2011) research was limited in its scope as it did not consider two underlying attributes central to combat. First, some websites are operated by the same individuals or groups of people. As such we would expect to see higher rates of connectivity and overlap in content on these websites. Moreover, this overlap in ownership impacts target prioritization. If five websites are operated by the same ownership, than all five need to be shutdown at once as the removal of only one will result in minimal impact on network distribution. Second, disrupting CEM distribution is complicated by jurisdictional boundaries. As such, target prioritization strategies need to consider the physical location of the material and the offender.

The research presented in this chapter improves on and extends the work of Westlake, Bouchard, and Frank (2011) through the creation of a custom-written webcrawler tool to map networks of public websites distributing CEM on the Internet. The Location Extraction of Child Exploitation Networks (LECEN) webcrawler improves on the webcrawler designed by Westlake et al., by refining the website inclusion criteria and collecting information on the geolocation of the domain hosting the website, the image hosting service, and the Whois registrant information for the domain. This additional location information allows social control agencies to quickly identify the appropriate jurisdiction. We also extend the concept of network capital (NC),

incorporating the improved inclusion criteria and geolocation data, and highlight the flexibility of NC to be adapted to the needs of researchers and social control agencies.

METHODS

Webcrawler

Data were collected using the Location Extraction of Child Exploitation Networks (LECEN) custom-written webcrawler, which functions similarly to those used by search engines to automatically navigate the Internet and collect information about the content found on websites and the webpages comprising each website. Like other webcrawlers, LECEN scans a website's underlying code and catalogues the type and location of text and media. Unlike many commercial webcrawlers, ours has been developed with built-in analysis functions that allow researchers to interpret the criminal aspect of the Internet and the corresponding social networks.

LECEN updates and refines the image hash value[1] and keyword inclusion criteria of the Child Exploitation Network Extractor webcrawler (see Westlake, Bouchard, & Frank, 2012; 2015). Further, two additional functionalities were added in LECEN. First, as each webpage is retrieved, it is geo-located allowing the researchers to geographically identify the location of the website server and the server hosting the child sexual exploitation material (CEM). Second, each new domain encountered is queried against the public Whois information database to retrieve and store the contact information of the organization who registered the website domain. Through these improvements, we can revise our formula for identifying key players within CEM distribution networks, to focus on specific website characteristics, and prioritize targets (offenders, websites, domains, servers) within certain jurisdictions.

[1] A hash value is the result of a mathematical procedure whereby data is broken into a 32-hexidecimal code unlikely to be shared between files (Rivest, 1991; Tretyakov, et al., 2013). When any file, CEM image or otherwise, is modified, a new hash value is created, as the file is different than the original. In essence, a hash value acts like a file's fingerprint.

LECEN required *seed* websites to begin data collection. For this study seed websites from Westlake, Bouchard, and Frank (2011), which were found to contain CEM, were chosen. These seeds were analyzed by LECEN and hyperlinks to adjoining websites were followed, with the scanning process repeating. For each webpage encountered, the source hyperlink text markup language (HTML) was retrieved and examined to determine if it met our inclusion criteria. For a webpage to be included in our data it was required to contain at least one child exploitation image, from our law enforcement supplied hash value database, or at least seven of our 82 keywords related to CEM. If the webpage met the criteria the rest of the website was scanned and the hyperlinks from said website were followed recursively. If criteria requirements were not met, the webpage was dropped from the queue and no further analysis was performed on it, or any of its' linked webpages.

Inclusion Criteria

LECEN integrated an MD5 hash value database, provided by the Royal Canadian Mounted Police (RCMP), to verify that a website contained CEM. For each webpage analyzed by LECEN, images were loaded into the computer's memory[2], hashed and checked against the RCMP database, and then discarded. . Last updated June 1st, 2012, the RCMP database contained more than 52 million hash values, classified into three distinct categories (Table 1). Category 1 contained 702,997 hash values that met the Canadian Criminal Code's definition of child pornography, under section 163.1(1). Category 2 contained 2,109,813 hash values, often referred to as 'gray-area' images, depicting an individual engaged in explicit sexual activity where the age of the individual was uncertain. Category 3 contained 49,419,190 hash values of images that

[2] By loading the image only into the computer's memory and not writing it to the hard drive, we were able to check the hash value associated with the image without being in possession (i.e., downloading) of the child exploitation image.

found alongside CEM but not meeting the criminal definition. For example, a picture of a child prior to undress would be included in this Category. For the purpose of our research, only the presence of Category 1 and 2 images were used as a criterion.

Research into CEM distribution has focused on the presence of specific keywords to identify content (e.g., LeGrand, Guillaume, Latapy, & Magnien, 2009; Vehovar, Ziberna, Kovacic, & Dousak, 2009). We used a selection of 82 keywords from Westlake, Bouchard, and Frank (2011) that were found to be prevalent on CEM websites. An analysis of thresholds to minimize the potential for false-positives and false-negatives found that the presence of seven keywords was an appropriate balance (Frank, Westlake, & Bouchard, 2010). Keywords were comprised of three different types (see Table 1). Category 1 (Code) had 45 keywords used by producers and distributors to uniquely identify CEM. These included terms such as 'Babyj', 'pthc', and 'qwerty'. Category 2 (Sexual Abuse) had 22 keywords that described sexual activities (e.g., 'anal' and 'pussy') or violent sexuality (e.g., 'abuse', 'cries', and 'torture'). Category 3 (General) contained 14 keywords commonly associated with CEM but indirectly (e.g., 'boy', 'little', 'young').

[TABLE 1]

Data Collection

The dataset used for this research is a subset of the network of data collected by LECEN, totaling 2,242 websites. The subset focused on in this chapter are those websites that had either Category 1 or Category 2 images. This resulted in a final sample of 83 websites. For each of these 83 websites, hyperlinks to websites not included in the final sample were removed as were self/internal hyperlinks.

For each website, a Whois service query for the domain registrant and a latitude/longitude geolocation for the domain's Internet Protocol (IP) address was conducted. The Whois service query, originally referred to as Nicname, was a text-based query-response protocol that allowed us to find out the registrant information for a website domain (Sullivan & Kucherawy, 2012). It provided the registrant (legal owner), administrative (primary contact outside owner), and technical contact (contact that maintains the website's operation and functionality) information for the website domain. This allowed us to trace the IP address beyond the basic connection to the hosted website and provide us with details regarding who owned the account associated with the website. Although the name associated with each of these three are often the same, it is possible for the registrant to be a third-party company hired to provide services, improving anonymity. Nevertheless, it is important to collect information on all three as differing contact information can impact jurisdictional considerations by providing multiple locations where law enforcement can intervene. As Whois registration information was derived from the Internet Corporation for Assigned Names and Numbers (ICANN) it can be considered to be up-to-date and accurate.

Geolocation refers to the process of identifying the location of a device connected to the internet and involves mapping the IP address to a real-world geographic location for the host (Mueller & Chango, 2008). The end result is an address in the form of city/state/country and/or a longitude/latitude pair. As IP addresses are typically reserved for specific service providers and not end-users, the IP address may only provide a rough estimate of the end-user's actual location. Therefore, we focused on the state and country levels of analysis. We integrated MaxMind's GeoLite (2014) database into LECEN as it is reported to have a 99.8 per cent accuracy at the country level and 90 per cent at the state level (Poese, Uhlig, Kaafar, Donnet, & Gueye, 2011).

Network Capital

Network capital (NC) is a composite measure that takes into account the resources made available by each website, the cohesiveness of the network, and the relationship between the websites. The more a website contributes to NC, the more they are considered a 'key player' within the overall network. A previous application of NC to the online distribution of CEM took into consideration a website's keywords, images, and videos (Westlake, Bouchard, & Frank, 2011). The current chapter extends this application by considering the impact of physical location and website ownership in determining key players, to address the issue of international jurisdiction and multiple websites being operated by the same individual/group/company.

NC is comprised of two components, *node_resources* and *node_connectivity*, which are summed together and compared to the total potential NC. Although we weighted *node_resources* and *node_connectivity* equally, coefficient weights can be added into the NC formula, to emphasize the priorities of social control agencies, or weights can be added at the individual resource-level to emphasize certain types of resources (e.g., known image hash values).

$$network\_capital = \frac{node\_resources + node\_connectivity}{N + N(N-1)}$$

Where:

N	Denotes the total number of websites (nodes) within the network.

Node_resources refers to the resources (e.g., CEM) that a website has available to share with the rest of the network. The more a website shares, the greater its impact and value to the overall network. We include five resource, measured per webpage: 1) category 1 & 2 images, 2) videos, 3) category 1 keywords, 4) category 2 keywords, and 5) proportion of hyperlinks to websites with different domain registration information. The flexibility of NC allows for each

resource to be weighted, should one be deemed more important than the other; however, for the purpose of the demonstration within this chapter, all resources were weighted equally. A website's value on each of the four content-based resources (images, videos, and keywords) was standardized against the highest scoring website within the network. For example, the website with the most videos per webpage received a score of 1.0 while all other websites' scores were represented as a proportion of this, based on their comparative videos per webpage count. A website's overall *node_resources* score was the average of the five resources, represented by the following formula:

$$node\_resources = \sum_{n=1}^{NAW_i} \frac{RW_{ni}}{NRW_i}$$

Where:

i	Denotes website (node).
RW_{ni}	Denotes the resource(s) weights attributed to node *i*, each ranging from 0.0 to 1.0.
NRW_i	Denotes the number of resources weighted for node *i*.

Node_connectivity refers to the ability of a website to contribute its resources to the overall network, based on the connections it has with other websites within the network. Although Schwartz and Rouselle (2009) included indirect connections, the size of the network analyzed here and the high direct connectivity between public websites means that we only included direct connections between websites. Mathematically, calculating *node_connectivity* from website *i* and *j* is done by multiplying website *i*'s *node_resources* by the proportion of resources they share (in our demonstration this is 1.0) and by any 'hyperlink weights' specific to the connection between *i* and *j*. The formula for this *node_connectivity* is as follows:

$$node\_connectivity_{ij} = \left(\left[\frac{\sum_{n=1}^{NRW_i} RW_{ni}}{NRW_i}\right] * RSL\right)\left(\frac{\sum_{m=1}^{NHW_{ij}} HW_{mij}}{NHW_{ij}}\right)$$

Where

i and j	Denotes websites (nodes)
RW_{ni}	Denotes the resource(s) weights attributed to node *i*, each ranging from 0.0 to 1.0
NRW_i	Denotes the number of resources weighted for node *i*.
RSL_i	Denotes the proportional resource-sharing level for node *i*.
HW_i	Denotes the hyperlink weights between 0.0 and 1.0 for node *i*[3].
NHW_{ij}	Denotes the number (2) of hyperlink weights applied to the *i* and *j* connection.

As each website was public, we set the resource sharing level to 1.0. We also included two hyperlink weights. The first was applied to all outgoing connections: the proportion of hyperlinks that connected to websites with different registration and domain contact information. The reason for including this hyperlink weight was that we felt a website's ability to connect with websites operated by other people is an important measure of being a key player. That is, hyperlinking to 'new' websites provided the best opportunity for diversifying the co-offending network. The second hyperlink weight included was dependent on the two websites connecting and was the proportion of the originating website's total outgoing hyperlinks that were destined for the connected website. We believed that the proportion of a website's hyperlinks that were

[3] As hyperlinks are unidirectional, only the hyperlink weights of the originating website are considered.

directed towards another website was indicative of the relationship strength between the two websites and the opportunities to distribute resources.

RESULTS

One of the primary difficulties in combating cybercrime is the ease with which offenders can disguise, or anonymize, their activities. Likewise, it is possible for the operators of a website to hide their identity through the use of private website registration services companies. Of the 83 websites analyzed in this study, 44 (53 per cent) used a Whois-masking registration company. However, as we utilized a geolocation search we were still able to obtain latitude and longitude coordinates for each domain. Table 2 summarizes the descriptive characteristics of each website located in the United States and outside the United States, separated by use of a private registration service while Figure 1 displays the network categorized by US/Non-US and Private/Public registration. Despite these separations, there were very few significant differences. Publicly registered websites located in the United States had a greater percentage of outgoing hyperlinks going to websites with the same Whois registration and geolocation (26 per cent to seven per cent) and more keywords per webpage than privately registered websites located outside the United States. These findings suggest that private registration services or geographic locations are not necessarily evidence of increased illegal activity. However, they may still impact the selection process for target prioritization.

[Figure 1]

[Table 2]

Network Capital

Network capital is a composite measure of resources and connectivity. In this study, resources included videos, child sexual exploitation hash values (images), keywords, and unique

connections. Figure 2 visualizes the network of 83 websites and highlights the key players based on different criteria. The size of each node represents their total contribution to overall network capital, with the ten most connected nodes represented by a circle, the ten most resource-rich by a square, and those in the top ten for both connectivity and resources represented by a triangle. Of the top ten contributors to network capital, eight were located in the United States while the other two were located in the Netherlands. Four used private Whois registration services while the other six did not.

[Figure 2]

Table 3 summarizes network capital, broken down by resources and connectivity. We show the percentage reduction in network capital by the strategic removal of the top three resource contributors, top three connectivity contributors, or top three overall contributors. We also compare these reductions to a non-strategic prioritization technique. Using a random number generator, we removed three nodes from the network, repeating this technique five times. The removal of the top three contributors to resources, connections, or overall resulted in an average reduction of seven per cent whereas random removal resulted in four per cent.

[Table 3]

While many websites are operated by different people, some are operated by the same individual(s) and/or serviced by the same hosting company. Simply removing the websites contributing the most to network capital may not be the best strategy as it may be more effective to target the individual(s) registered to a cluster of websites and/or the hosting company servicing multiple websites. This is because the removal of a website may only temporarily inhibit the distribution chain, as the registrant can create a new website and resubmit all of the material from the removed website. However, by targeting the registrant or the domain, the

impact on the overall distribution network can be greater as multiple websites can be eliminated using the same amount of resources as shutting down one website. Using this approach, we summarized the contribution of each registrant and domain to network capital to identify which should be prioritized and by whom (i.e., jurisdiction).

There were three American-based registrants that we will focus on here. The first was in Utah, the second in Kansas, and the third in Arizona. The Utah registrant operated four websites, who if removed would result 5.95 per cent reduction in network capital. The Kansas registrant operated six websites, whose removal resulted in a 5.17 per cent reduction in network capital. Finally, removing the Arizona registrants three websites corresponded to a 3.99 per cent reduction in network capital. However, it is worth noting that the registrant of the Arizona-hosted websites was also the registrant on two websites hosted in California. Therefore, focusing on the registrant rather than the server location, in this situation, would result in a 5.86 per cent reduction in network capital. Comparing these findings to pursuing the top three website contributors to resources, connectivity, or network capital, presented in Table 3, targeting a specific registrant and/or domain had a similar impact on NC reduction, and a greater impact on NC reduction than randomly selecting targets.

Taking our analysis to an international level, we found that sharing information about specific registrants and domains can also be useful in disrupting the distribution chain. The top registrant/domain contributor to NC was located in the Netherlands (two websites), whose removal resulted in a 5.92 per cent reduction while the removal of a Canadian registrant (two websites), located in Quebec, was associated with a 2.83 per cent reduction. It is worth noting that the Netherlands and Canadian domains were registered to private companies, making the process slightly more complicated. However, there was a public registrant in England operating

two websites, whose removal reduced NC by 1.86 per cent. Combining this with the American-based registrants, a joint operation (i.e., sharing of data) between the United States, Netherlands, Canada, and England would result in the removal of 21 websites and a reduction to NC of 27.58 per cent. In this scenario, while the United States would be responsible for three targets, every other country would have one target and thus a lower impact on resources with a substantial impact on CEM distribution on the Internet.

DISCUSSION

The quantity and complexity of child sexual exploitation material (CEM) distribution in cyberspace necessitates an interdisciplinary approach focused on maximizing identification and network disruption and minimizing direct visual contact with material by investigators. To address these, we proposed the use of automated data collection tools that follow user-specified inclusion criteria and a flexible algorithm that improves target prioritization strategies. The algorithm, and subsequent measure Network Capital (NC), builds on existing strategies of key player identification by accounting for the variety of resources each player (i.e., website) shares with the network and the connections they have with other network players. We also addressed the complications of jurisdictional boundaries by incorporating the geolocation of websites into target prioritization strategies. Our demonstration of automated data collection and key website identification, through NC, highlight the opportunities for collaboration between social and computer scientists in understanding cybercrime. We conclude with a discussion of the importance and challenges of a) automated data collection; b) flexibility in the criteria for identifying key players; c) flexibility in what constitutes a key player; and d) continued research on child sexual exploitation.

The fast-paced nature of the Internet coupled with the abundance of data available means that finding ways to automate data collection processes should be at the forefront of priorities for researchers investigating cybercrime. However, these efforts need to be approached with caution as the reliability and validity of automated data collection tools need to be ensured. One potential way to address these issues is through interdisciplinary partnerships that maximize the expertise of researchers in different fields. For example, when it comes to studying various phenomena using data collected from the Internet, manual efforts to do so can only yield a very limited dataset.

Computer programs, in the form of webcrawlers, can automate much (if not all) of the data collection process. However, how much researchers should rely on the collected data depends largely on the study's purpose. If the study is focused on data collected from a single website, than reliability can be very high since all data collected is on 'target'. If the study is focusing more on the content of many websites, or trying to sample the Internet, than reliability is not ensured.

Context matters. Keyword-based inclusion can fail if the same keywords are used in different places, in different ways, within different contexts. For example, the keyword 'bomb' can be used by websites supporting terrorists in the context of making bombs, while government websites would be expected to discuss bombs in the context of disarmament (see Westlake, Bouchard, & Frank, 2015 for CEM-specific examples). Although image hash values provide fool-proof evidence of a webpage containing previously known CEM, keywords still matter as not all CEM images are known (and thus could be missing from the image hash value database). Additionally, the CEM website could opt to post content that while still illegal contains no CEM images. We attempted to minimize this type of context-related confusion through the use of a

seven keyword threshold, based on comparative analyses conducted by Westlake, Bouchard, and Frank (2012). Even with such a high thresholds, false positives[4] can occur, and hence some manual verification is necessary.

The concept of targeted law enforcement strategies is not new. However, traditional efforts to target key players have omitted the resource component of what constitutes a key player, focusing solely on the centrality within a network. While a network-linkage based key player selection would remove the most connected websites (i.e., hubs), not taking into account the actual content present on those websites would be naïve, as those hubs do not necessarily contain the most CEM. Similarly, going after the website with the most CEM might not align with the priorities of the social control agency targeting the CEM network. Taking into consideration the varying definitions of what constitutes a key player, we proposed an algorithm that is flexible to the needs of the researcher, social control agency, and even type of cybercrime. Network Capital considers the resources and the connections while allowing the flexibility to choose which to select and how each should be weighted relative to one another. That is, if social control agencies desire to focus on content-rich websites, they can increase the weight of *node_resources* compared to *node_connectivity*, or they can increase the weight of individual resources. As a result, the proposed network capital algorithm takes into account the traditional centrality concept of key players while accounting for resource-specific priorities and adjustments as required.

Finally, the importance of investigating online CEM distribution is evident; however, the challenge is developing methods for identifying non-image media, new material (i.e., new victims), and website/user patterns. Interdisciplinary partnerships provide an excellent

[4] Pages deemed to be CEM when in fact are not.

opportunity to address some of these challenges through the development of audio and video-based criteria that can look for patterns in CEM across multiple websites. Adding these criteria into automated data collection tools could aid in finding new material, based on audio between known and unknown content and commonalities in the background of videos. To increase the confidence of the data collection method, various online communities could be studied in order to determine 'patterns' that could be used to describe them and differentiate them from other communities. These patterns could be based on keywords, and perhaps including sentiment through the use of natural language processing and sentiment analysis (see Choi & Cardie, 2008; Wilson, Wiebe, & Hoffmann, 2009). However, given that these websites are embedded in a network, this information should also be used in the creation of these patterns.

Patterns, such as 'a website likely contains CEM if more than 50 per cent of its links are pointing to webpages containing known CEM' could be developed and used to increase the validity of data collection. While important for law enforcement purposes, organizations with transnational reach, such as the Virtual Global Taskforce (VGT) and Internet Watch Foundation, can use this information to aid in the coordination of international investigations. Through advancements in website profiling (i.e., pattern identification), these organizations could act as central hubs for identification and forward the data to the appropriate law enforcement agency. This would increase the efficiency of removing CEM and focus local investigations on arresting perpetrators rather than finding CEM online. For example, VGT could conduct an automated data collection and identify websites that may contain new content. After visual inspection, a list of key targets could be provided to the appropriate agency and/or country, based on a) location of the server/website and b) markers found in the CEM that identify a specific location (e.g., road sign in a specific language).

Interdisciplinary partnerships, specifically between social and computer scientists, play an important role in facilitating future research into the online distribution of child sexual exploitation material. Combined with the continued support of private organizations, such as Microsoft©, Google©, and National Center for Missing and Exploited Children, and government agencies, such as the victim identification program of Homeland Security, a better understanding of how these networks function can be determined and tools and techniques for identifying children currently being exploited can be developed.

REFERENCES

Almutairi, A., Parish, D., and Phan, R. (2012) 'Survey of high interaction honeypot tools: Merits and short-comings', Retrieved from: http://www.cms.livjm.ac.uk/pgnet2012/Proceedings/Papers/1569604821.pdf.

Bonacich, P. (1972) 'Factoring and weighting approaches to status scores and clique identification', *Journal of Mathematical Sociology*, 2: 113-120.

Bonacich P. (1987) 'Power and centrality: A family of measures', *American Journal of Sociology*, 92: 1170-1182.

Bonacich, P. (2007) 'Some unique properties of eigenvector centrality', *Social Networks*, 29: 555-564.

Borgatti, S. (2003) 'The key player problem', in R. Breiger, K. Carley, and P. Pattison (eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers,* Washington D.C.: National Academy of Science Press.

Borgatti, S. (2006) 'Identifying sets of key players in a social network', *Computational and Mathematic Organization Theory,* 12: 21-34.

Bouchard, M., and Nash, R. (2015) 'Researching terrorism and counter-terrorism through a network lens', in M. Bouchard (ed.) *Social Network, Terrorism and Counter-Terrorism: Radical and Connected,* New York: Routledge.

Bourke, M.L., and Craun, S.W. (2014) 'Secondary traumatic stress among Internet Crimes Against Children task force personnel', *Sexual Abuse: A Journal of Research and Treatment,* 26: 586-609.

Burns, C.M., Morley, J., Bradshaw, R., and Domene, J. (2008) 'The emotional impact on coping strategies employed by police teams investigating internet child exploitation', *Traumatology,* 14: 20-31.

Canadian Criminal Code, RSC. (1985). c C-46 s163.

Chang, L., and Krosnick, J.A. (2009) 'National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality', *Public Opinion Quarterly*, 73: 641-678.

Choi, Y. and Cardie, C. (2008), 'Learning with compositional semantics as structural inference for subsentential sentiment analysis', in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 793-801. Association for Computational Linguistics.

Craun, S.W., Bourke, M.L., and Coulson, F.N. (2015) 'The impact of Internet crimes against children work on relationships with families and friends: An exploratory study', *Journal of Family Violence,* 30: 393-402.

Décary-Hétu, D., and Dupont, B. (2012) 'The social network of hackers', *Global Crime,* 13: 160-175.

Décary-Hétu, D., and Morselli, C. (2011) 'Gang presence in social network sites', *International Journal of Cyber Criminology,* 5.2: 876-890.

Décary-Hétu, D., Morselli, C., and Leman-Langlois, S. (2012) 'Welcome to the scene: A study of social organization and recognition among warez hackers', *Journal of Research in Crime and Delinquency*, 49: 359-382.

Dillman, D.A. (2007) *Mail and Internet surveys: The tailored design method,* 2nd edn, New York: John Wiley and Sons.

Fournier, R., Cholez, T., Latapy, M., Chrisment, I., Magnien, C., Festor, O., and Daniloff, I. (2014) 'Comparing pedophile activity in different P2P systems', *Social Sciences,* 3: 314-325.

Frank, R., Westlake, B.G., and Bouchard, M. (2010) 'The structure and content of online child exploitation', in *Proceedings of the 16th ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD 2010)*, Article 3. Washington, DC: ACM. doi: 10.1145/1938606.1938609.

Freeman, L.C. (1979) 'Centrality in social networks: Conceptual clarifications', *Social Networks,* 1: 215-239.

Gillespie, A. A. (2011) *Child pornography: Law and policy,* New York: Routledge.

Hanneman, R.A. and Riddle, M. (2005), *Introduction to social network methods,* Riverside, CA: University of California, Riverside.

Heidemann, J., Klier, M., and Probst, F. 'Identifying key users in online social networks: A PageRank based approach', paper presented at the 31st International Conference on Information Systems, St. Louis, MO, December 2010.

Hewson, C., Vogel, C., and Laurent, D. (2015) *Internet research methods*. Thousand Oaks: Sage.

Hillman, H., Hooper, C., and Choo, K.R. (2014) 'Online child exploitation: Challenges and future research directions', *Computer Law & Security Review,* 30: 687-698.

Holt, T. J., Strumsky, D., Smirnova, O., and Kilger, M. (2012) 'Examining the social networks of malware writers and hackers', *International Journal of Cyber Criminology,* 6: 891-903.

Jewkes, Y., and Andrews, C. (2007) 'Internet child pornography: International responses', in Y. Jewkes (ed.) *Crime Online*, Portland: Willan Publishing.

Joffres, K., Bouchard, M., Frank, R., and Westlake, B.G. (2011) 'Strategies to disrupt online child pornography networks', in *Proceedings of the EISIC - European Intelligence and Security Informatics*, 163-170. Athens, Greece: IEEE.

Katz, L. (1953) 'A new index derived from sociometric data analysis', *Psychometrika*, 18: 39-43.

Krause, M. (2009) 'Identifying and managing stress in child pornography and child exploitation investigators', *Journal of Police and Criminal Psychology,* 24: 22-29.

LeGrand, B., Guillaume, J., Latapy, M., and Magnien, C. (2009) 'Dynamics of paedophile keywords in eDonkey queries'. Retrieved from: http://antipaedo.lib6.fr/.

Liu, X., Patacchini, E., Zenou, Y., and Lee, L.F. (2012) 'Criminal networks: Who is the key player?'. CEPR Discussion Paper No. 8772.

Malm, A., and Bichler, G. (2011) 'Networks of collaborating criminals: Assessing the structural vulnerability of drug markets', *Journal of Research in Crime and Delinquency*, 48: 271-297.

Marín, J.M.F., Naranjo, J.Á.M., and Casado, L.G. (2015) 'Honeypots and honeynets: analysis and case study', in M.M. Cuz-Cunha & R.M. Portela (eds.), *Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance,* Hershey, PA: IGI Global.

MaxMind GeoLite (2014) 'GeoIP2 Databases'. Retrieved from: https://www.maxmind.com/en/geoip2-databases.

Medina, R., and Hepner, G. (2008) 'Geospatial analysis of dynamic terrorist networks', in I. Karawan, W. McCormack, & S.E. Reynolds (eds.) *Values and Violence: Intangible Aspects of Terrorism,* Berlin, Germany: Springer.

Microsoft. (2005 April 7) 'Microsoft collaborates with global police to develop Child Exploitation Tracking System for law enforcement agencies'. Retrieved from http://news.microsoft.com/2005/04/07/microsoft-collaborates-with-global-police-to-develop-child-exploitation-tracking-system-for-law-enforcement-agencies/.

Microsoft. (2012 March 20) 'Microsoft PhotoDNA technology helping law enforcement fight child pornography'. Retrieved from: https://blogs.microsoft.com/cybertrust/2012/03/20/microsoft-photodna-technology-helping-law-enforcement-fight-child-pornography/.

Morselli, C. (2009) *Inside criminal networks,* New York: Springer.

Morselli, C. (2010) 'Assessing vulnerable and strategic positions in a criminal network', *Journal of Contemporary Criminal Justice*, 26: 382-392.

Mueller, M., and Chango, M. (2008) 'Disrupting global governance: The Internet Whois service, ICANN, and privacy', *Journal of Information Technology & Politics*, 5: 303-325.

Natarajan, M. (2006) 'Understanding the structure of a large heroin distribution network: Quantitative analysis of qualitative data', *Journal of Quantitative Criminology,* 22: 171-192.

Nouh, M., and Nurse, J.R. (2015) 'Identifying key-players in online activist groups on the Facebook social network', in *Proceedings of the 15th International Conference on Data Mining Workshops*, 969-978. Atlantic City, NJ: IEEE.

O'Neil, P.H. (2015 August 11) 'U.S. tech companies are now using a massive database to stop child pornography'. Retrieved from: http://www.dailydot.com/politics/facebook-twitter-google-child-pornography-iwf-hash-list/.

Page, L., Sergey, B., Rajeev, M., Terry, W. (1999), 'The PageRank citation ranking: Bringing order to the web', Technical Report. Stanford InfoLab.

Perez, L.M., Jones, J., Englert, D.R., and Sachau, D. (2010) 'Secondary traumatic stress and burnout among law enforcement investigators exposed to disturbing media images', *Journal of Police and Criminal Psychology*, 25: 113-124.

Poese, I., Uhlig, S., Kaafar, M. A., Donnet, B., and Gueye, B. (2011) 'IP geolocation databases: Unreliable?', *ACM SIGCOMM Computer Communication Review*, 41: 53-56.

Powell, M., Cassematis, P., Benson, M., Smallbone, S., and Wortley, R. (2015) 'Police officers' perceptions of their reactions to viewing Internet child exploitation material', *Journal of Police and Criminal Psychology,* 30: 103-111.

Provos, N., and Holz, T. (2007) *Virtual honeypots: from botnet tracking to intrusion detection,* Boston: Pearson Education.

Rivest, R.L. (1991), 'MD5 unofficial homepage', http://userpages.umbc.edu/mabzug1/cs/md5/md5.html.

Rutgaizer, M., Shavitt, Y., Vertman, O., and Zilberman, N. (2012) 'Detecting pedophile activity in BitTorrent networks', *Lecture Notes in Computer Science,* 7192: 106-115.

Schonlau, M., van Soest, A., Kapteyn, A., and Couper, M. (2009) 'Selection bias in web surveys and the use of propensity scores', *Sociological Methods & Research,* 37: 291-318.

Schwartz, D.M., and Rouselle, T. (2009) 'Using social network analysis to target criminal networks', *Trends in Organized Crime,* 12: 188-207.

Shiels, M. (2008 April 14) 'Google tackles child pornography'. Retrieved from: http://news.bbc.co.uk/2/hi/7347476.stm.

Shropshire, K.O., Hawdon, J.E., and Witte, J.C. (2009) 'Web survey design: Balancing measurement, response, and topical interest', *Sociological Methods & Research,* 37: 344-370.

Sobek, M. (2003) 'The implementation of PageRank in the Google Search Engine'. Retrieved from: http://pr.efactory.de/e-pagerank-implementation.shtml.

Spitzner, L. (2003) *Honeypots: tracking hackers,* Volume 1, Reading: Addison-Wesley.

Steel, C. M. S. (2009) 'Child pornography in peer-to-peer networks', *Child Abuse & Neglect,* 33: 560-568.

Sullivan, A., and Kucherawy, M.S. (2012) 'Revisiting Whois: Coming to REST', *IEEE Internet Computing*, 16: 65-69.

Tayebi, M. A., Bakker, L., Glässer, U., and Dabbaghian, V. (2011) 'Locating central actors in co-offending networks', *Proceedings of Advances in Social Networks Analysis and Mining 2011 International Conference',* 171-179. Kaohsiung, Taiwan: IEEE.

Tretyakov, K., Laur, S., Smant, G., Vilo, J., and Prins, P. (2013) 'Fast probabilistic file fingerprinting for big data', *BMC Genomics,* 14: S2-S8.

van Dijk, J., Spapens, A. C. M., Reichel, P., and Albanese, J. (2014) 'Transnational organized crime networks', in P. Reichel and J. Albanese (eds.), *Handbook of Transnational Crime and Justice* (2nd edn), Thousand Oaks, CA: Sage.

Vehovar, V., Ziberna, A., Kovacic, M., Mrvar, A., and Dousak, M. (2009) 'Empirical investigation of paedophile keywords in eDonkey P2P network'. Retrieved from: http://antipaedo.lib6.fr/.

Wasserman, S., and Faust, K. (1994) *Social network analysis: Methods and applications,* Cambridge, England: Cambridge University Press.

Westlake, B.G., Bouchard, M., and Frank, R. (2011) 'Finding the key players in online child exploitation networks', *Policy and Internet,* 3(2), Article 6. doi: 10.2202/1944-2866.1126.

Westlake, B.G., Bouchard, M., and Frank, R. (2012) 'Comparing methods for detecting child exploitation content online', in *Proceedings of the EISIC - European Intelligence and Security Informatics*, 156-163. Odense, Denmark: IEEE.

Westlake, B.G., Bouchard, M., and Girodat, A. (in press) 'How obvious is it: The content of child sexual exploitation websites', *Deviant Behavior.*

Wilson, T., Wiebe, J. and Hoffmann, P. (2009) 'Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis', *Computational linguistics*, 35: 399-433.

Wortley, R. K., and Smallbone, S. (2012) *Internet child pornography: Causes, investigation, and prevention,* Santa Barbara, CA: ABC-CLIO.

Table 1: Categories and corresponding quantity of keywords and image hash values used by LECEN during data collection.

	Keywords (Number)	**Image Hash Values (Number)**
Category 1	Child Exploiter-Code (45)	Child Exploitation (702,997)
Category 2	Thematic (22)	Child Nudity (2,109,813)
Category 3	Sex-Oriented (14)	Collateral (49,419190)

Table 2: Descriptive statistics for websites located in and outside the United States and using public or private Whois registration.

	United States (n=62)		Non United States (n=21)	
	Public Reg. (n=35)	**Private Reg. (n=27)**	**Public Reg. (n=4)**	**Private Reg. (n=17)**
Average Webpages	26.34	65.52	4.75	93.65
Cat 1 Images (Per Website)	23.29	4.74	4.25	35.82
Cat 2 Images (Per Website)	60.03	11.74	4.25	54.88
Cat 3 Images (Per Website)	675.56	173.56	31.75	2,406.06
Images (Per Webpage)	115.91	112.65	118.01	126.53
Videos (Per Webpage)	0.05	0.10	1.04	0.08
Cat 1 Keywords (Per Webpage)	223.19	225.87	35.50	1.83**
Cat 2 Keywords (Per Webpage)	560.78	435.24	83.43	88.48**
Cat 3 Keywords (Per Webpage)	2,535.44	1,606.93	73.17	62.65**
Outgoing to Own (%)	25.73	7.40*	1.79	13.63
Incoming from Own (%)	7.75	10.16	2.88	18.65
Hyperlinks to New (%)	82.89	90.03	97.48	79.47

*: Statistically significant ($p<0.05$) from publicly registered websites in the United States.
**: Statistically significant ($p<0.10$) from publicly registered websites in the United States.

Table 3: Change in Network Capital when removing three websites using targeted (top contributors) and non-targeted (random) strategies**.**

	Resources	Connectivity	Network Capital (NC)	Δ in NC*
Original	16.42	38.86	8.03	--
Top 3 Resources	14.67	37.04	7.51	6.93%
Top 3 Connections	15.04	36.67	7.51	6.91%
Top 3 Overall	14.76	36.83	7.49	7.16%
Random 1	15.78	37.27	7.70	4.22%
Random 2	15.78	37.15	7.68	4.47%
Random 3	15.76	37.19	7.69	4.43%
Random 4	16.02	37.75	7.81	2.82%
Random 5	15.57	37.04	7.64	5.09%

*Discrepancy due to rounding.

Figure 1: Publicly and non-publicly registered child sexual exploitation websites located inside or outside the United States.

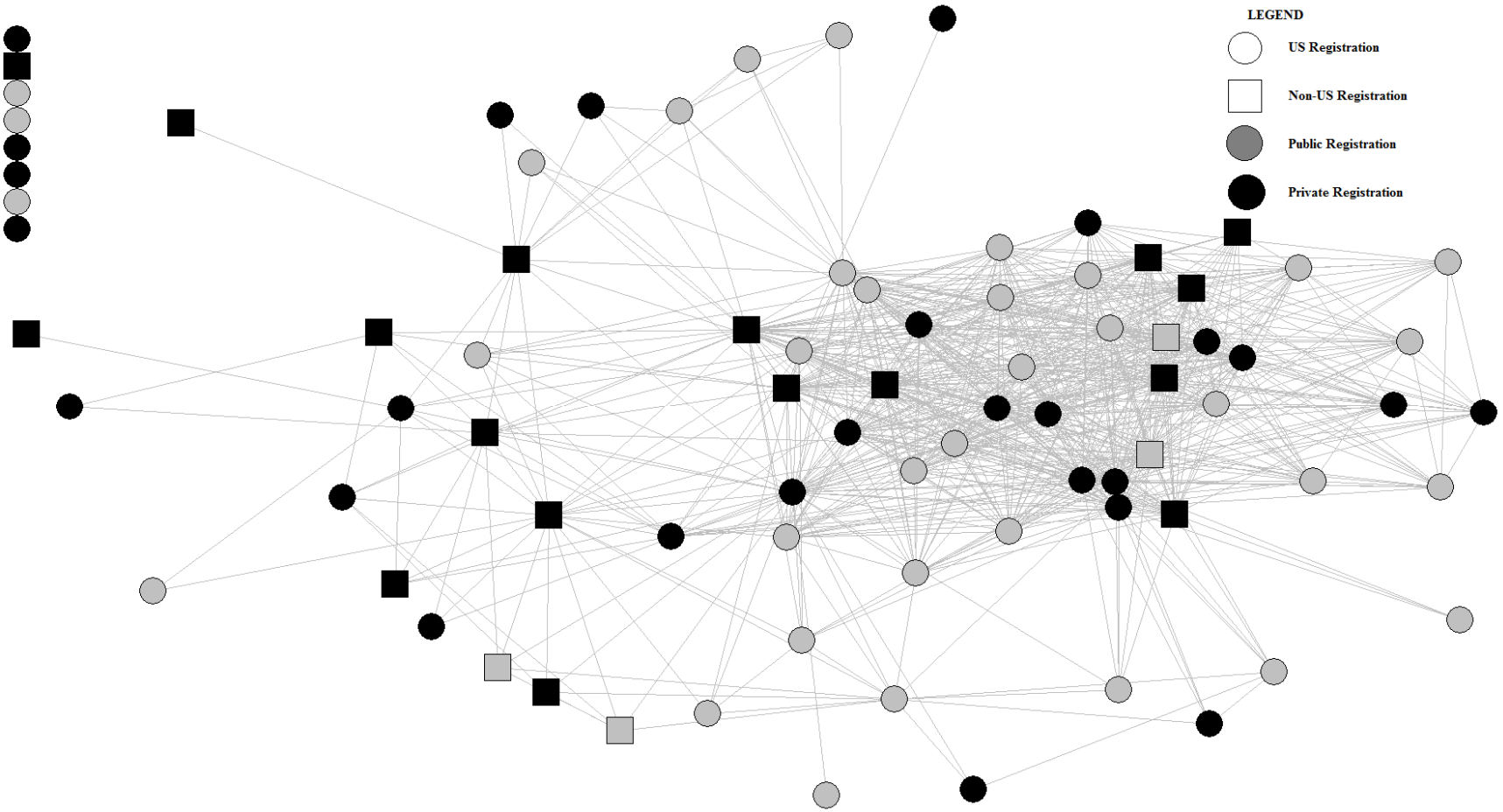

Figure 2: The top ten most resource-rich or connected child sexual exploitation websites and those that are top ten in both.