Elsevier Editorial System(tm) for Social Science Research
Manuscript Draft

Manuscript Number: SSR-D-15-00403R1

Title: Liking and Hyperlinking: Community Detection in Online Child Sexual Exploitation Networks

Article Type: SI: Big data in SS

Keywords: Internet mediated research; child sexual exploitation; child pornography; cybercrime; social networks; online communities

Corresponding Author: Dr. Bryce G Westlake,

Corresponding Author's Institution: San Jose State University

First Author: Bryce G Westlake

Order of Authors: Bryce G Westlake; Martin Bouchard

Abstract: The online sexual exploitation of children is facilitated by websites that form virtual communities, via hyperlinks, to distribute images, videos, and other material. However, how these communities form, are structured, and evolve over time is unknown. Collected using a custom-designed webcrawler, we begin from known child sexual exploitation (CE) seed websites and follow hyperlinks to connected, related, websites. Using a repeated measure design we analyze 10 networks of 300+ websites each - over 4.8 million unique webpages in total, over a period of 60 weeks. Community detection techniques reveal that CE-related networks were dominated by two large communities hosting varied material -not necessarily matching the seed website. Community stability, over 60 weeks, varied across networks. Reciprocity in hyperlinking between community members was substantially higher than within the full network, however, websites were not more likely to connect to homogeneous-content websites.

Liking and Hyperlinking: Community Detection in Online Child Sexual Exploitation Networks

Bryce G Westlake[*a], Martin Bouchard[a]

[a]School of Criminology, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6

ABSTRACT

The online sexual exploitation of children is facilitated by websites that form virtual communities, via hyperlinks, to distribute images, videos, and other material. However, how these communities form, are structured, and evolve over time is unknown. Collected using a custom-designed webcrawler, we begin from known child sexual exploitation (CE) seed websites and follow hyperlinks to connected, related, websites. Using a repeated measure design we analyze 10 networks of 300+ websites each – over 4.8 million unique webpages in total, over a period of 60 weeks. Community detection techniques reveal that CE-related networks were dominated by two large communities hosting varied material –not necessarily matching the seed website. Community stability, over 60 weeks, varied across networks. Reciprocity in hyperlinking between community members was substantially higher than within the full network, however, websites were not more likely to connect to homogeneous-content websites.

KEYWORDS

Internet mediated research; child sexual exploitation; child pornography; cybercrime; social networks; online communities

*Corresponding author.
Email Address: Bryce.Westlake@sjsu.edu
Permanent Address: Department of Justice Studies, San Jose State University
One Washington Square, San Jose, California, 95192-0050

HIGHLIGHTS

- An automated data collection tool allowed us to analyze over 4.8 million webpages
- We map the hyperlink networks surrounding child sexual exploitation (CE) websites
- We investigated the website characteristics that comprise CE-related communities
- Networks were structured around 2 large core communities and 3-5 smaller ones
- Hyperlinks were not formed around content similarities between websites

*Corresponding author.
Email Address: Bryce.Westlake@sjsu.edu
Permanent Address: Department of Justice Studies, San Jose State University
One Washington Square, San Jose, California, 95192

Liking and Hyperlinking: Community Detection in Online Child Sexual Exploitation Networks

Research on the sociology of the Internet has shown that online communities matter; perhaps just as much as offline communities and social interactions. Previously seen as small and bound by a neighbourhood or village, the minimization of communication barriers afforded by the Internet (Rheingold, 2006) have facilitated the transition of communities into large, global, social networks (Wellman, Boase, and Chen, 2002). This transition significantly impacted how people with similar interests connected. For those conducting illicit activities, the Internet opened up a new avenue for conducting business and targeting victims. For distributors and consumers of child pornography, the Internet transitioned a traditionally solitary crime into a globally communal crime, where material could be accessed instantaneously (Hillman, Hooper, and Choo, 2014). Previously isolated and facing constant challenges to find new material, the Internet provided consumers with the opportunity to connect across long distances with like-minded individuals, who provided moral validation, social support, and access to a constant stream of new material, often free of charge (Beech, et al., 2008; Estes, 2001; Quayle and Taylor, 2011; Taylor and Quayle, 2003; Tremblay, 2006). Central to the opportunities and support provided online are the networks that form between individuals *and* entities (e.g., websites), and the communities that develop from these networks. As a result, in this paper we take a unique approach to understanding communities by focusing on the larger, macro-level, networks created between website entities, via hyperlinks. Specifically, we examine the structure and long-term functionality of child exploitation (CE) related website communities.

The online distribution of CE material, depicting the sexual abuse of children, is conducted using a variety of public and private platforms. Central to the online distribution of CE material are the producers and distributors. However, within the World Wide Web (WWW), the practice of distribution is heavily influenced by the websites that host the material. Although initially uploaded by individuals, consumers do not directly access the supplier. Rather, they rely on the intermediary website to acquire material, and use said website to gain access to other websites with similar (or different) material. Therefore, the process of CE distribution, within the WWW, cannot be fully understood without considering the role websites play, as they provide the initial means for accessing distributors and like-minded offenders. In fact, websites may play an even greater than suspected role in distribution, as they can dictate, to some degree, who and what individual consumers have access to, via their connections (hyperlinks) to other suppliers (i.e., websites). Therefore, the analysis of websites is central to understanding CE distribution within the WWW and cyberspace. More specifically, the analysis of the structure of the communities websites form, how these communities function, and how they evolve over time can have implications for the development of social policies, the methods used by researchers to study online child sexual exploitation, the strategies used by law enforcement agencies, and the role of public and private companies in detection and removal.

In this study, we examine the communities that form around public websites involved in the distribution of CE material. Although inclusion criteria are used, it is worth nothing that not all websites we examine are actively distributing CE images. Instead, we study the communities *surrounding* 10 known child sexual exploitation websites and thus characterize the networks as related to them. Connected via hyperlinks, we explore the social structure of website communities, how these communities are formed, and how they evolve and adapt over time. A supplementary objective of this study is to introduce the Child Exploitation Network Extractor (CENE), an automated data collection tool that can be useful in increasing the efficiency of data collection processes for online research on social problems, and studying graphic and/or traumatizing topics (e.g., child sexual abuse). In this study, CENE allowed us to process automatically the content and hyperlink information of 4,831,050 webpages.

Online Communities and Illegal Websites

Website communities are formed through hyperlinks connecting websites directly to one another. Hyperlink network analysis (Park & Thelwall, 2003) has shown that hyperlinking practices are purposive, following underlying communicative rationale and providing opportunities to create and foster alliances (Foot et al., 2003; Park, Kim, and Barnett, 2004; Park, Thelwall, and Kluver, 2005). Hyperlinked websites often share ideological similarities and are rarely connected directly to dissimilar websites (Ackland and Shorish, 2009; Adamic and Glance, 2005; Hargittai, Gallo, and Kane, 2008). As a result, hyperlinking can be viewed as a tool used to create smaller communities, within a larger social network, and provide website visitors access to like-minded others and targeted content. In social network terms, the action of hyperlinking is a recognition of the existence of another, important (or strategic) enough to warrant a public display whereby visitors of the first website can easily transfer to the second.

A central component of the World Wide Web, websites have been appropriated by offenders in three important ways. First, social networking websites have been exploited by sexual offenders for grooming victims (Whittle, Hamilton-Giachritsis, and Beech, 2014a; 2014b; Whittle, et al., 2013), while e-commerce websites are used to acquire financial information (Holt and Turner, 2012; Pratt, Holtfreter, and Reisig, 2010). Second, criminal websites have been created to provide offenders with *convergent settings* (Felson, 2003) where they can ply their trade (Decary-Hetu and Dupont, 2012) or acquire social support (Tremblay, 2006; Wortley and Smallbone, 2012). Third, websites are developed for the primary purpose of distributing illegal goods and services (Dolliver, 2015; Steinmetz and Tunnell, 2013; van Hout and Bingham, 2013) and promoting illicit activities (Milrod and Weitzer, 2012; Chow-White, 2006).

Criminal-based websites have a tendency to operate in isolation from mainstream and dissimilar illegal websites, preferring to form small communities with like-minded others (Burris, Smith, and Strahm, 2000; Chau and Xu, 2008; Zhou, et al., 2005). Research into online criminal communities has found that even for solitary crimes, such as hacking, there exist strong ties whereby tools and information are actively shared across intricate network connections (Holt, 2009). Within these website communities offenders acquire criminal capital (prestige, notoriety, and status) through the sharing of tools and information with other offenders (Decary-Hétu, Morselli, and Leman-Langlois, 2012; Dupont, 2013; Holt, 2007). In acquiring criminal capital, offenders are more able to form friendships with others that may later translate into co-offending opportunities (Holt, Blevins, and Burkert, 2010; McCarthy and Hagan, 1995). Subsequently, websites may play an important role in the co-offending selection process given they function to facilitate offender transactions by creating points of congregation. As a result, websites can been characterized as also acquiring criminal capital, through being seen as a place to acquire tools and information, and to connect with other offenders. By hyperlinking to homogeneous others, the websites create the boundaries of the community, facilitating and controlling some criminal opportunities afforded to offenders.

Among offline offender, those who subsequently co-offend often share homogeneous personal attributes, such as age, sex, ethnicity, residency, or criminal experience (Carrington, 2014; Schaefer, 2012; van Mastrigt and Farrington, 2011; Weerman, 2003). However, research has been divided as to whether this proclivity towards homogeneity is the result of preference (Reiss and Farrington, 1991) or structural opportunities (Carrington, 2002). Recently, van Mastrigt and Carrington (2014) found that homogeneous tendencies are not strictly the result of opportunity structures (e.g., more young men involved in crime) but rather that offenders also make conscious choices to select co-offenders similar to themselves.

It is unclear whether proclivities towards homogeneity translate to cyberspace. This is because the structure of the Internet means that co-offenders rarely interact in-person, instead relying on more secure digital communication methods (Decary-Hetu and Dupont, 2012). As a result, personal characteristics such as age, sex, ethnicity, residency, or experience are not readily available –and may not actually matter. Instead, online partnerships may be formed from broad interests in specific types of crimes, such as racism (e.g., white supremacists) and extremism, or, like between some offline gang members, facilitated by joint memberships to the same website.

Given that criminal-based websites hyperlink to similar websites (e.g., Burris, Smith, and Strahm, 2000) and offline criminal enterprises co-offend more along functional ties than ethnic ties (Malm, Bichler, and Nash, 2011), it follows that the first step in studying CE-related website communities is to identify the attributes that form the foundation for hyperlinking. As websites are innate, they do not have any personal characteristics that can be used to promote connectivity. However, CE-related websites can be characterized by their homogeneity in sex of victim (boy/girl), type of content (explicit/non-explicit), or medium of distribution (images/videos/text). Although, it is also possible that CE-related communities are solely based on broad interests and that connectivity between websites is more akin to a pure availability model.

There is support for the hypothesis that illicit website communities may function closer to a pure availability model. Within large piracy communities, websites, such as The Pirate Bay, appear indiscriminate with regards to the type of content they provide or the websites they associate. Even those who provide specific types of content, such as television shows (e.g., EZTV), do not appear to function outside the larger piracy network, or specifically among a community of television-based websites. However, within smaller Warez-release communities (Basamanowicz and Bouchard, 2011) and CE communities (Wortley and Smallbone, 2006), the connectivity selection process appears to be more discriminate. In the distribution of CE-related material, public websites may try to maximize embeddedness within the larger network by creating communities with any website, irrespective of content. A third possibility is that they congregate with dissimilar websites to avoid direct competition and to maximize general appeal, increasing overall consumer traffic.

CURRENT STUDY

The distribution of CE-related material in cyberspace is conducted using a variety of methods operating strictly through online means or through a combination of offline and online tactics (van Wijk, Nieuwenhuis, and Smeltink, 2009). While a large percentage of CE-related material is distributed through Internet Chat Relay, the Deep Web, and other private networks, Carr (2004) identified the World Wide Web –which includes blogs, discussion forums, live webcam feeds, and photo galleries- as the second most prominent source for obtaining CE images. Supporting Carr’s assertion, recent research suggests that a substantial amount is still distributed through public, or semi-public, methods via peer-to-peer networks (Latapy, Magnien, and Fournier, 2013; Steel, 2009; Wolak, Liberatore, and Levine, 2014), BitTorrent networks (Rutgaizer, et al., 2012) and public websites (Westlake, Bouchard, and Frank, 2011). .

While parallels can be drawn between virtual and non-virtual communities, we propose that virtual communities are distinct enough to warrant separate conceptual and empirical treatment. One key reason for this distinction is that virtual communities do not require direct contact between members nor do they even directly require people for functionality. Instead, websites connect, through hyperlinks, with one another to form virtual ‘entity-based’ communities. Through these hyperlinks, public CE-related websites form communities, through hyperlinks, that facilitate the transmission of content, information, and offender connections, criminal opportunities, belonging, and validation.

In the current study, we investigate the communities that form within 10 networks of 300+ websites each, 4,831,050 webpages total, beginning with a known child sexual exploitation website. Using a custom designed web-crawler, we monitor these networks for 60 weeks, allowing us to study how the communities formed change over time. Drawing on community detection techniques, we a) describe the website characteristics that comprise and differentiate communities within the overall network; b) identify the stability of virtual communities over time; and c) whether hyperlinking tendencies are driven by homogeneous characteristics.

METHODS

Data

The 10 networks and 4.8 million plus webpages analyzed for this study come from a longitudinal project with the objective to analyze the evolution of the websites involved in the dissemination of child sexual exploitation (CE) material. Using a repeated measures design, we analyzed the networks ten times at an interval of 42 days (six weeks). As the data collection process was network-bound, the computational time required was not limited by the computer but rather the time to download the data. In order to manage each network’s size and relevance, we implemented three conditions. First, we created an exclusion list of popular websites (e.g., Facebook©) and false positives (e.g., Disney©) from previous CENE searches, known not to be directly associated with CE material. Second, we limited the size of the networks to ~300 websites and ~500,000 webpages. Third, an image had to be larger than 150 pixels by 150 pixels; or two inches (four centimeters) by two inches. Once the network size limitation were met, the data was aggregated up to the server level. In other words, the data on www.website.com/webpage1 and www.website.com/webpage2 were summed and listed under www.website.com. A list of all websites and webpages scanned were stored and reused at each data collection interval to ensure that the same webpages, if they were still online, were analyzed at each interval.

Data were collected using a custom-designed web-crawler that followed a snowball sampling method via hyperlinks between websites (Burris, Smith, and Strahm, 2000; Chau and Xu, 2007; Frank, Westlake, and Bouchard, 2010; Fu, Abbasi, and Chen, 2010; Zhou, et al., 2005). Referred to as the Child Exploitation Network Extractor (CENE), the webcrawler designed for this study operated similarly to automated data collection tools used by various search engines, to index websites. CENE followed a method similar to that of a person browsing the Internet, looking for CE material. Beginning with a *seed* website, known to be related to CE material, CENE scanned the hypertext markup language (HTML) for our pre-determined CE-identifying criteria. If the webpage was determined relevant, CENE continued to scan the website, collecting structural and website characteristics data. Similar to a person viewing the website, CENE then followed hyperlinks found on the website, to other websites, repeating the criteria search process. If the hyperlinked webpage was deemed irrelevant (i.e., did not meet the criteria), the website was discarded and CENE moved on to the next. A website was deemed to be CE-related if the webpage contained at least one known CE image, from a database of images provided by the Royal Canadian Mounted Police (RCMP), or at least seven keywords, from a list of 82, relevant to child sexual abuse.

Seed Websites

Each network began with a seed website selected from one of two sources. The first source, accounting for four seeds, was a list of websites, provided by the RCMP, known to distribute CE-related material. The second source, was a list of websites identified in previous CENE searches to be engaged in CE-related dissemination. This included, but was not limited to, image or video distribution or acting as an access point to CE websites. To partially account for the nature of the seed biasing the characteristics of the sample derived, five of our ten seeds were *blogs* while the other five were *sites*. A blog was defined as a website with user-generated posts in a traditional web-log setup. A site was defined as a website with interlocking webpages, which did not meet the criteria of a blog, including discussion forums and photo galleries. Each network began with an average of 305.10 (s.d. =2.33) websites and was re-crawled every 42.14 (s.d. =4.45) days.

Website Inclusion Criteria

Keywords: The 82 keywords used by CENE were found to be the most prevalent in online CE dissemination networks (Hurley et al., 2013; Latapy, Magnien, and Fournier, 2013; LeGrand, et al., 2009; Steel, 2009; Vehovar, et al., 2009) and categorized into three groups. The first group were *code keywords* (27) commonly used by offenders to alert one another to content, such as pthc (pre-teen hardcore). The second group were *thematic keywords* (23) not directly linked to child sexual abuse but typically present on such websites (e.g., boy, girl, child). The third group were *sex-oriented keywords* (32) referencing sexual organs and acts (e.g., pussy, cock, oral).

Child Exploitation Images: CE images were identified using a hash value database provided by the RCMP. A hash value is a 32-hexidecimal code that functions similar to a digital fingerprint. Each computer file is given a hash value based on its binary composition. When a file is edited, even minimally, a new hash value is created. Tretyakov, and colleagues (2013) state that the chances of two distinct files having the same hash value is ‘negligibly small’ ($1/2^{2048}$).

According to section 163.1 (1) of the Canadian Criminal Code (CCC, 1985), child pornography includes "… any written material, visual representation, or audio recording" of a person "under the age of eighteen years and is engaged in or is depicted as engaged in explicit sexual activity". Unlike the United States and Australia, the definition of child pornography is uniform across Canada and, like the United Kingdom and Australia, includes real or computer-generated visual representations (see Gillespie, 2012). Although section 163.1 (6) (a) of the CCC states that "No person shall be convicted of an offence… [if it] has a legitimate purpose related to the administration of justice or to science…", at no point did we store CE images. The data collection process consisted of retrieving the image into memory, analyzing it, and storing the summative image data (hash value and location).

Last updated on June 1st, 2012 (CENE was launched in July 2012), the database contained 2.25 million hash values classified into three groups (see Table 1). The first group (*Child Exploitation)* contained 618,632 images that, according to the CCC, met the definition of child pornography. The second group (*Child Nudity)* contained 652,223 images that would probably be considered child pornography but have not yet been brought before a judge. The third group (*Collateral)* contained 981,232 images that do not meet the definition but were important enough to be collected by offenders. Images in this category may have included initial photographs taken by an offender, of a child, prior to the removal of clothing.

Table 1. Description of the categories of keywords and hash values used by the web-crawler.

	Keywords (Number)	*Hash Values (Number)*
Category 1	Child Exploiter-Code (27)	Child Exploitation (618,632)
Category 2	Thematic (23)	Child Nudity (652,223)
Category 3	Sex-Oriented	Collateral

	(32)	(981,231)

Website Composite Variables

During data collection, CENE identified the total number of webpages, images, videos, keywords, and incoming and outgoing hyperlinks for each website. From these we created composite variables to describe the characteristics of websites and the communities they form.

Type of Focus

Sex (Boy/Girl): Using the relative frequency of specific keywords, we classified each website as being *boy* or *girl* oriented. The keywords used for these classifications were: boy, son, twink, penis, and cock, or girl, daughter, nimphets/nymphets, Lolita/lola/lolli/lolly, vagina, and pussy.

Content (Explicit/Non-Explicit): Using the relative frequency of specific keywords, we classified each website as being focused on *explicit* or *non-explicit* material. The composite *explicit* measure consisted of 21 keywords related to severe sexual abuse (e.g., cries, torture, and rape). The composite *non-explicit* measure consisted of 15 keywords related to personal characteristics (e.g., innocent, lover, smooth).

Medium (Image/Video/Story): Using the relative frequency of three types of media found on websites, we classified each website as primarily distributing images, videos, or stories. First, for *image* and *video*, we used the average number of instances of each medium found on a webpage. For *story*, we used the average number of our 82 keywords found on a webpage. Second, each website was given a standardized score (0.00 to 1.00) for each medium, relative to the other websites within the network. For example, the website with the most images per webpage was given a score of 1.00 while every other website was standardized against this website, on the same measure. For whichever medium a website received the highest score was its classification.

Connectivity

Incoming/Outgoing Hyperlinks: A website's connectivity was determined through the number of (unique) incoming and outgoing hyperlinks found. Whether Website A hyperlinked to Website B eight times or two times, the connections counted as one outgoing hyperlink for Website A and one incoming hyperlink for Website B. *Incoming* can be viewed as a measure of popularity while *Outgoing* a measure of a website's attempt to reach out to the community and integrate.

Community Detection

We considered a variety of community detection methods (e.g., Moody and White, 2003[1]) to identify the cohesive sub-groups that comprised a larger network of websites, and narrowed our selection to the faction analysis algorithm available in the UCINET software (Borgatti, Everett, and Freeman, 2002; also see de Amorim, Barthelemy, and Ribeiro, 1992; Glover and Laguna, 2013). Our preliminary analyses also included the Girvan-Newman method (Girvan and Newman, 2002; Newman and Girvan, 2004). Although both methods seek to find distinct sub-groups within a larger network, by maximizing the density[2] between group members and minimizing density between non-group members, they differ with how they identify the communities. The Girvan-Newman method identifies cohesive sub-communities by generalizing Freeman's (1979) concept of betweenness centrality to all edges in a network—what Girvan and Newman termed edge betweenness. The sequential removal of edges with high betweenness centrality effectively separates groups of actors that are more cohesive. As a result, Girvan-Newman is less adept at handling directed (non-reciprocal connections) networks such as those studied here. Our

[1]Although alternative method allow for the identification of nested subgroups allows for multiple group memberships (see Moody and White, 2003), we chose to focus on the "home base" community of these websites as a first step to understanding connections within this environment.

[2] Density is the proportion of direct connections found between nodes in relation to all possible connections between nodes (Garton, Haythornthwaite, and Wellman, 1997).

comparative analysis confirmed this limitation, as the Girvan-Newman did not produce satisfactory results (i.e., low modularity and similarity within communities).

The faction analysis algorithm we used begins with a random partition and tries to build the solution that maximizes the density in a number of groups selected in advanced by the researcher, moving individuals from one group to another until a solution deemed optimal is determined (de Amorim, Barthelemy, and Ribeiro, 1992; Glover and Laguna, 2013; Zhao, et al., 2011). Because faction analysis begins with a random partition, it is possible to obtain different results each time the method is conducted. Using a range between 2 and 20 communities, faction analyses were conducted multiple times, on each network, at each data collection point (wave), to ensure consistency. Based on goodness of fit models and visual inspection, optimal community configurations were selected and used for subsequent analyses. Goodness of fit was determined through the 'final proportion correct', which is the sum of the number of ties between websites in different factions divided by the total number of ties, and 'Q value', which provide the percentage of network ties that are within a community (Carolan, 2013). The average goodness of fit, across all ten waves of data collection, ranged from 0.81 and 0.85. Within each network, there was minimal between-wave variance (<0.01) in goodness of fit values.

Homophily

The tendency for two, similar, entities to associate with one another is called homophily (McPherson and Smith-Lovin, 1987). In delinquency research, the study of homophily often refers to similarities between co-offenders on attributes such as sex and age (Carrington, 2014; van Mastrigt and Carrington, 2014). However, the probability that two connected entities (e.g., people or websites) will be similar on an attribute is dictated by the demographics of the subject pool. For example, in a subject pool consisting of five boys and two girls, the baseline probability that a boy will randomly be connected with another boy is greater than the probability they will be connected with a girl. Therefore the preference for homogeneity needs to account for the connections that exist and compare them to the expected connections, based on the subject pool structure.

To determine whether websites were more likely to connect with similar websites, we used a formula adapted by van Mastrigt and Carrington (2014), to measure expected and observed homophily. Expected homophily is the number of connections expected between similar websites, based on a random distribution of connections. As the probability of homogeneity *and* heterogeneity need to be determined, expected homophily is calculated using combinatorics. The formulas for expected homophily are as followed: $EH_{1n}=p^n$; $EH_{2n}=q^n$; $EH_{3n}=1-(p^n+q^n)$ and explained through the following example, where *p= proportion of explicit*, *q= proportion of non-explicit*, and *n=2* (number of nodes in partnership). The probability that explicit websites are connected is determined by the explicit/non-explicit composition of the network's websites and the size of the desired homogeneous group. In a network with 75% explicit websites, the probability of a dyad explicit partnership is 0.75^2 and 0.25^2 for a non-explicit dyad. Therefore, the probability of an explicit/non-explicit dyad is $1-(0.75^2+0.25^2)$. These probabilities are then compared to observed homophily[3] using a X^2. We measured homophily across the website type of focus: *sex of victim, content,* and *medium*.

RESULTS

The 10 networks generated very similar social structures. Across the networks, Wave 1 consisted of two large, central, communities –accounting for 77.97% of websites– and three to five smaller, surrounding, communities. Average community density for site networks was 0.445, and 0.341 for blog networks. Network websites were overwhelming boy-oriented (79.2%), non-

---

[3] Observed homophily is simply the *actual* number of ties between explicit and non-explicit websites

explicit (86.7%) in their sexual content, and image-focused (81.6%). As shown by Table 2, predominant network characteristics appeared influenced by the seed website. All 10 seed websites were non-explicit; the two networks beginning with a girl-oriented seed contained the highest percentage of girl-oriented websites (89.0% and 71.4% compared to 5.9% for the other eight networks); and the two video-focused seeds had above average counts of video-focused websites (7.2% and 11.2% compared to 6.7% for the other eight networks). With the exception of blog network 4 and 5, each network's seed was located within one of the two central communities. Finally, all but blog seed 3 was active at the conclusion of the study.

Table 2. Characteristics of the seed, and of the network of hyperlinked websites around it

	# of CE Images	# of CN Images	# of Collateral Images	% Explicit Focused	% Boy Oriented	% Image Focused	% Video Focused
Site 1	342	24	329	8.9	96.4	76.7	17.7
Seed	4	0	0	Non-Explicit	Boy	Image	
Site 2	3112	47	2544	12.0	95.5	88.2	4.2
Seed	0	0	0	Non-Explicit	Boy	Image	
Site 3	516	161	414	6.6	97.4	85.5	5.3
Seed	27	0	27	Non-Explicit	Boy	Image	
Site 4	1292	1	1186	14.8	99.0	87.5	7.2
Seed	0	0	0	Non-Explicit	Boy	Video	
Site 5	0	17	0	12.3	11.0	72.6	8.4
Seed	0	0	0	Non-Explicit	Girl	Image	
Blog 1	5255	125	3146	23.9	74.5	85.3	2.3
Seed	0	0	0	Non-Explicit	Boy	Image	
Blog 2	1308	14	904	12.0	98.4	81.2	6.8
Seed	2	0	14	Non-Explicit	Boy	Image	

Blog 3	610	70	305	7.5	97.7	86.9	5.2
Seed	0	0	0	Non-Explicit	Boy	Image	
Blog 4	898	9	351	11.2	93.7	75.6	11.2
Seed	0	0	5	Non-Explicit	Boy	Video	
Blog 5	1	13	0	24.2	28.6	82.7	3.3
Seed	0	0	0	Non-Explicit	Girl	Image	

It is useful to focus on ‘representative’ networks (closest to the average across network characteristics) in order to illustrate some of the more detailed findings. Figure 1 displays representative site network 1 (Sweet Love[4]), and Figure 2, blog network 2 (Teddy Bear), with the seed-website circled. To aid in identification and subsequent recall of communities being described, their names are derived from the most prominent characteristic(s). Although the networks are large enough that the specific connections between websites cannot be emphasized, the figures still provide an important visual display on how dense are the hyperlinked networks formed and how the pockets of websites, cohesive enough to be identified as communities, form around the core. Figure 1, for example, shows different communities forming around the distribution of videos, despite the seed being image focused (only 0.08 videos per webpage). Figure 2, instead, shows a pattern around sub-communities specializing in the distribution of images, much like the seed. The first network (Figure 1), which started from a boy-oriented illegal website, even included a community of adult gay videos (right side of the figure). That community is slightly more isolated from the rest of the network, and as shall be seen below, is the only one without a trace of illegal or grey area material.

[4] To preserve the anonymity of the websites, we use fictitious web domain names for the purpose of this study.

Table 3 summarizes the community characteristics of the two network representatives Sweet Love and Teddy Bear. As was the case across all ten networks, websites were predominantly boy-oriented; 96.4% (294/305) in Sweet Love and 98.4% (304/309) in Teddy Bear. The two core communities within each network differed from the surrounding communities in two key ways. First, the majority of illegal images were within the two core communities (Table 3a). Within Sweet Love, the community *Network Core* (n=142) contained 95.3% of CE images and 94.5% of Collateral images. Within Teddy Bear, the community *CE Image Core* (n=117) contained 72.9%

Figure 1a: Network of site Sweet Love (bold white circle) at Wave 1, displaying its six communities

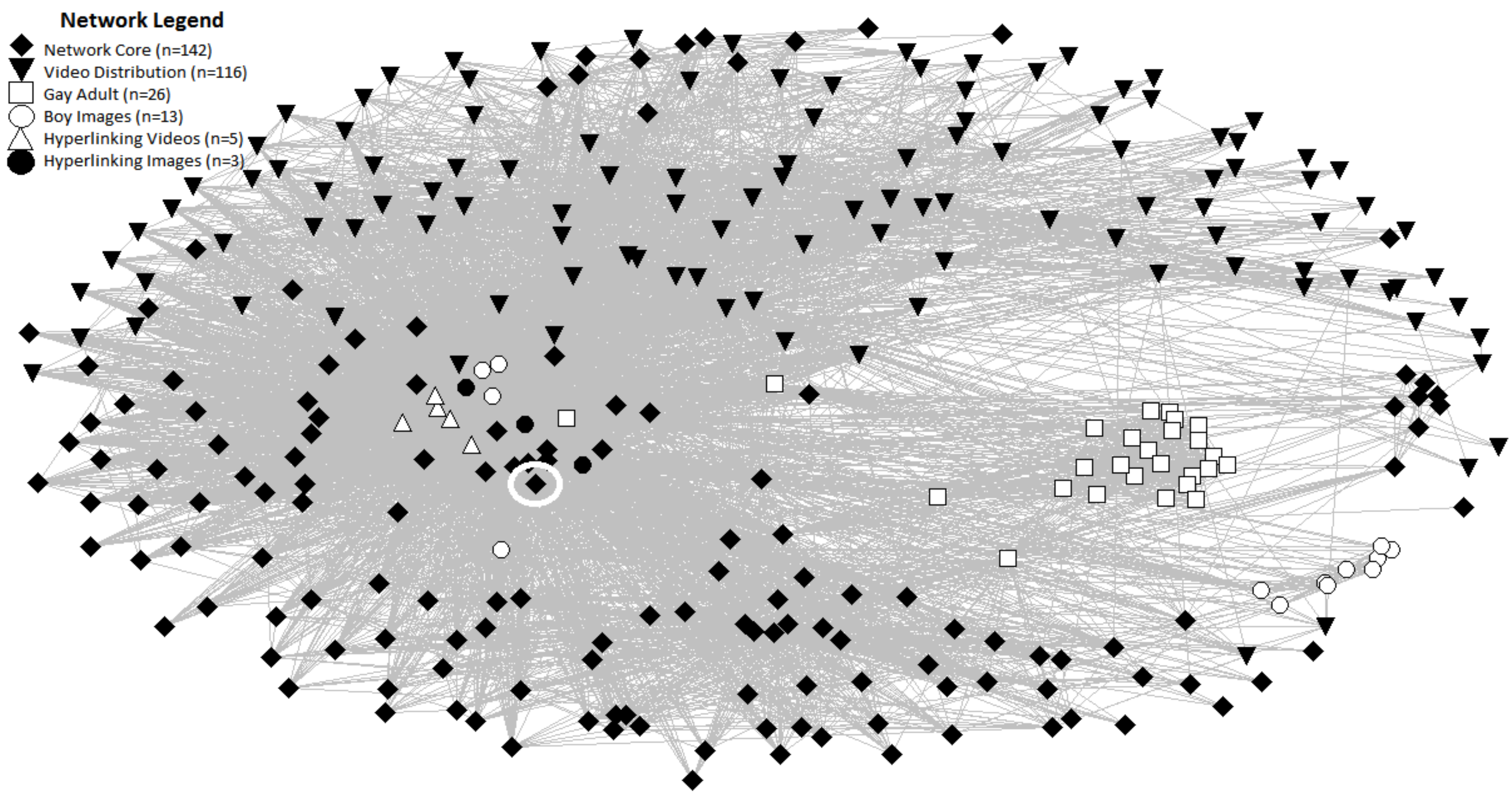
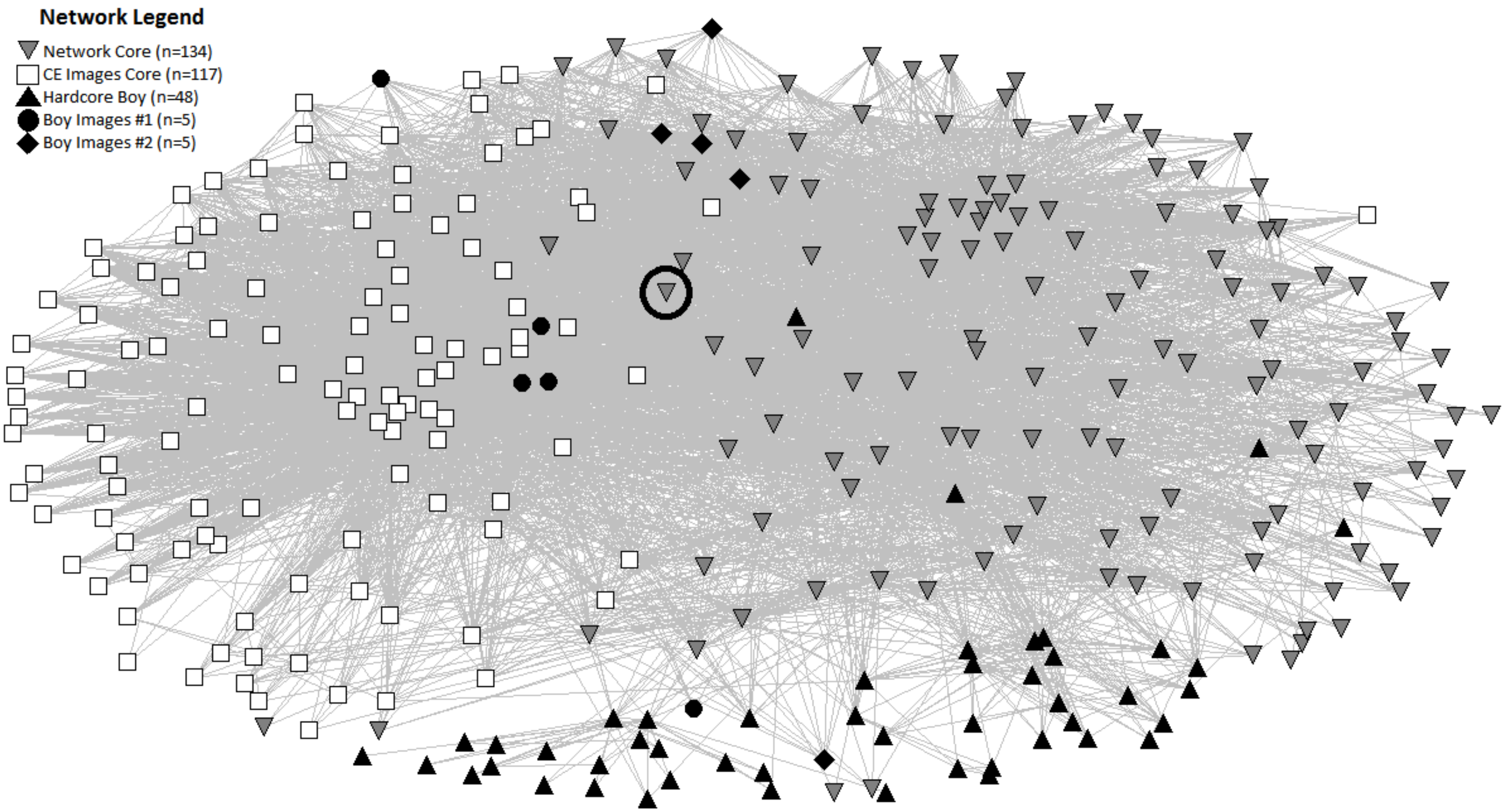


Figure 1b: Network of blog Teddy Bear (black bold circle) at Wave 1, displaying its five communities



of CE images and 88.8% of Collateral images. Second, the two core communities within each network contained the vast majority of websites publishing videos.

Within each network, websites that clustered together often shared some prominent characteristics. In the two networks where the seed-website was not located in one of the primary two communities, the seed did not appear to influence the community characteristics. For example, blog seed 4 was the *only* video-focused website within their community while blog seed 5's community was more than 40% boy-oriented; despite the seed being one of two girl-oriented seeds. Focusing on the two representative networks, within Sweet Love, communities *Hyperlinking Videos* (n=5) and *Hyperlinking Images* (n=3) contained websites that could be characterized as brokers; hyperlinking out to between 30 and 50% of all the websites within the network. *Gay Adult* (n=26) was comprised of adult gay content -based on website names and type of content- while *Video Distribution* (n=116) was core to the network but distributed video content. In Teddy Bear, similar patterns emerged. Communities *Boy Images #1* (n=5) and *Boy Images #2* (n=5) distributed boy images while *CE Images Core* (n=117) predominantly distributed known illegal images. Together, these communities, especially Sweet Love's *Video Distribution* (#6) and Teddy Bear's *CE Images Core* (#2), suggest that some websites will become close-knit with websites that they share important similarities.

Table 3a: Community size and CE image composition for representative networks Sweet Love (site) and Teddy Bear (blog).

	Community Name	Count (% of Network Websites)	Avg. Nb. of Pages per Website	CE Images (% of Websites)	Child Nudity Images (% of Websites)	Collateral Images (% of Websites)
Sweet Love	Network Core	142 (46.6)	917.8	326 (7.8)	24 (16.9)	311 (6.3)
	Video Distribution	116 (38.0)	1671.0	1 (0.9)	0 (0.0)	10 (0.9)
	Gay Adult	26 (8.5)	299.2	0 (0.0)	0 (0.0)	0 (0.0)
	Boy Images	13 (4.3)	243.5	13 (30.8)	0 (0.0)	6 (23.1)
	Hyperlinking Video Sites	5 (1.6)	1864.4	1 (20.0)	0 (0.0)	1 (20.0)
	Hyperlinking Image Sites	3 (1.0)	1866.7	1 (33.3)	0 (0.0)	1 (33.3)
Teddy Bear	Network Core	134 (43.4)	972.7	233 (2.2)	1 (0.8)	14 (0.8)
	CE Images Core	117 (37.9)	521.5	953 (10.3)	13 (11.1)	803 (6.8)
	Hardcore Boy	48 (15.5)	4298.7	43 (2.1)	0 (0.0)	42 (2.1)
	Boy Images #1	5 (1.6)	617.0	66 (40.0)	0 (0.0)	32 (40.0)
	Boy Images #2	5 (1.6)	2695.4	13 (40.0)	0 (0.0)	13 (40.0)

Table 3b: Content and connectivity descriptives, by community, for representative networks Sweet Love (site) and Teddy Bear (blog).

	Community Name (Count)	% Explicit	% Boy	% Video	% Image	% Story	Avg. Outgoing	Avg. Incoming
Sweet Love	Network Core (142)	4.9	93.0	19.0	71.8	9.2	21.8	26.4
	Vid. Distribution (116)	17.2	99.1	21.6	75.0	3.5	18.4	23.5
	Gay Adult (26)	0.0	100.0	0.0	100.0	0.0	29.7	29.4
	Boy Images (13)	0.0	100.0	7.7	92.3	0.0	48.3	17.2
	Hyperlinking Video Sites (5)	0.0	100.0	20.0	80.0	0.0	109.6	23.4
	Hyperlinking Image Sites (3)	0.0	100.0	0.0	100.0	0.0	157.3	21.7
Teddy Bear	Network Core (134)	16.4	99.3	9.0	76.9	14.2	23.8	30.0
	CE Images Core (117)	4.3	97.4	6.8	88.0	5.1	32.0	32.0
	Hardcore Boy (48)	20.8	97.9	2.1	72.9	25.0	17.3	10.7
	Boy Images #1 (5)	0.0	100.0	0.0	100.0	0.0	65.8	16.2
	Boy Images #2 (5)	0.0	100.0	0.0	100.0	0.0	64.6	12.4

Central to the notion of community cohesion is mutual acknowledgement. In network terms, mutual acknowledgement can be categorized as the percentage of connections (i.e., hyperlinks) that are reciprocated amongst community members. Across the ten networks, reciprocity was 23.0% (22.6% for site networks and 23.4% for blog networks). However, we observed higher cohesion when examining within-community reciprocity. As would be expected, smaller communities were substantially more cohesive than larger communities. For example, weighting each community equally, communities with more than 100 websites had a reciprocity of 27.4%, compared to 53.1% within communities with less than 100 websites. For communities with less than 30 websites, reciprocity averaged 47.7%. While smaller communities were more reciprocal, the presence of known CE images did not enhance or decrease reciprocity. This finding held at multiple cut-off points for both number of CE images and websites.

Change in Community Composition

The stability, or volatility, of virtual communities is important for understanding how quickly online networks evolve. If connections between websites are transient, in place only to serve a specific purpose and then removed, then distribution networks can be seen as operating closer to a pure availability framework. However, if the connections are long-lasting, this suggests that relationships are more likely to form between websites and their operators. That is, they are in a better position to develop a certain level of trust with specific websites and work together to achieve a common goal. To measure stability, we had to force select the number of factions at each time point to allow for cross-wave comparison[5]. Upon selecting the *best* stable community

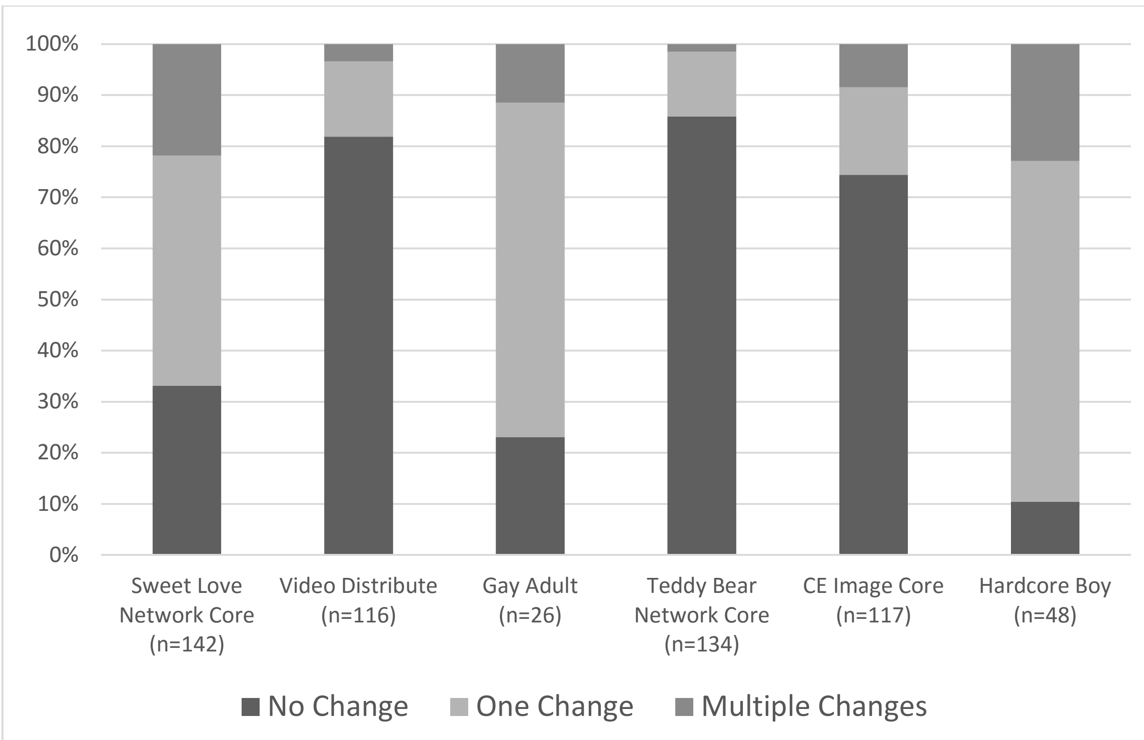
[5] For each network, faction analyses were conducted at each wave for four to 18 clusters. The corresponding *final proportion correct* and *q-value* were compared within and between waves. The cluster formation that corresponded to a high final proportion correct and q-value, compared to the other cluster formations, was selected. In situations where the highest final proportion correct and q-value was not selected, these values did not differ from the highest by more than 0.1%.

structure, across the duration of the study, community movement was calculated at each wave. If a website moved from one community to another at any point during the study period, it was classified as being dynamic.

Results of the community stability analyses reveal equal consistency and instability in group membership. Nearly half (48.86%) of all websites did not transition to another community within their network, with less between-network variability amongst blog seed networks. Within blog networks stability ranged from 44.1% to 67.0% (s.d. =9.9%), while community stability ranged from 14.8% to 76.0% (s.d. =20.8%) in site networks. Amongst the ten seed websites, three remained in the same community throughout while five switched communities three or more times. The five most stable seeds remained in the two largest communities. Where community stability was the lowest –site network 5, blog networks 3 and 5– seed-websites experienced the highest rates of community movement. Figure 2 displays stability for the three largest communities –where stability could best be measured– in Sweet Love and Teddy Bear. We also identified the number of websites that switched between two communities and more than two.

Community transition most often involved websites from smaller communities moving to one of the two core communities; potential evidence of smaller websites being 'accepted' into the larger community. In some situations, community transition was cluster-based. For example, in Sweet Love, 14 websites moved from *Gay Adult* to *Boy Images* at Wave 4. However, at Wave 5, all 14 rejoined *Gay Adult.* Finally, some websites oscillated between the two core communities. Within each network, the stability of the two core communities was higher than the surrounding, smaller, communities. Community movement was dominated by websites within the smaller communities transitioning to the core communities. There were some exceptions to this such as *Boy Images* from Sweet Love, where 11 out of the 13 websites remained in the same community throughout the study. *Boy Images* community was also one of only a handful that saw an increase in the quantity of known child exploitation images from Wave 1 to Wave 10. Smaller stable communities might be evidence of some websites insulating themselves from potential infiltration and/or detection.

Figure 2: Community stability in the three largest communities in site network Sweet Love and blog network Teddy Bear.



Homogeneity between Connected Websites

As entities, websites do not possess any personality characteristics to examine as predictors of websites selecting similar others. In order to determine homophily amongst connected websites, website content characteristics are logical choices for analyses. Using two of the composite variables[6] created from our keywords –severity (explicit/non-explicit), and medium

[6] The boy/girl orientation of a website was not analyzed for homophily because the vast majority of websites examined were boy focused (over 95% in most networks). We did run the sex homophily analysis on the two networks with a majority of girl-oriented websites, and that although not statistically significant girl-oriented websites did cluster at higher than expected rates. In Blog 5, 68% versus 38% while in Site 5, 84% versus 78%.

(image/video/story) – we explored whether websites purposely connected to websites with the same composition. The results suggest that websites do not show a strong tendency for homophily in terms of severity, or the medium predominantly used. Sexually explicit websites were just as likely to connect to non-explicit websites, and rarely did we find the medium used on websites leading to these websites narrowing down their hyperlinks to websites emphasizing similar media. Diversity, rather similarity, appears to drive the hyperlinking among the websites analyzed in this study.

DISCUSSION

Growth in the prevalence of cybercrime has necessitated increased research into criminal processes conducted online. For many cybercrimes, websites play a crucial role in facilitating criminal activities. While initially produced by individual offenders, the distribution of child sexual exploitation (CE) material is heavily influenced by public websites that host and disseminate images, videos, and other content, and connect individual offenders with each other. The vast amount of data available on the Internet, and public websites especially, provides an opportunity for innovative conceptual and methodological approaches to studying online (and offline) criminal processes. In the current study, we designed an automated tool to collect data on websites involved in the dissemination of CE material and the surrounding hyperlinked websites. From this examination we can assess how criminal-based website communities structure and function within the larger network, and how they evolve over time. We focus on two main findings from this study and conclude with a discussion of the implications for subsequent research into co-offending selection processes.

First, we found that despite having stronger ties (more reciprocated hyperlinks), communities were not comprised of websites focused on specific types of content. This finding extended beyond the macro, community, level to the micro, website, level, as our homophily analysis revealed that websites focused on the same type of content did not connect with homogeneous others at higher than expected rates. Although prior research on these issues is scant, this finding contrasts with those of Burris et al. (2000) who found that the online White Supremacists community showed a clear preference for connecting to websites with similar interests. The lack of content-based clustering found in our study may be indicative of the multiplicative nature of the Internet. In cyberspace CE material can easily be copied from one website to another. This easy access to content means that websites do not need to be closely connected to similar websites. In fact, the nature of the Internet may mean that multiple connections to similar websites may have negative consequences. While connecting with homogeneous others may provide access to preferred content, the wide distribution of material might mean that the law of diminishing returns applies –being connected to many similar websites may provide access to competitors, leading to lost traffic (e.g., visitors preferring the 'new' website). This conclusion is supported, in part, by Burris, Smith, and Strahm's findings that among commercial White Supremacist enterprises, network cohesion was lower, as they competed for customers.

On the other hand, actively connecting to heterogeneous others may provide the greatest array of opportunities for CE-related websites. Like the experienced offender who diversifies their connections, in part, to diversify criminal opportunities (Tremblay, 1993), a website's ability to connect to a variety of CE content types maximizes their potential traffic and partners for exchanging material. A website potentially increases its appeal by providing one type of content directly (via text, images, or videos), and a variety of types indirectly (via hyperlinks). As our analyses were focused on the content distributed, it remains possible that websites do connect to homogeneous others based on other, unmeasured, characteristics, such as the type or quantity of community members, the ability to acquire and distribute new material, the personal relationships between website operators, or the aforementioned age of the victim depicted.

It is important that we also acknowledge the possibility that, outside of a few choice connections, the communities that form around public websites disseminating CE material base their hyperlinking practices on pure availability. Public websites may operate as catch-all marketplaces with their primary purpose being to attract and connect as many people as possible, regardless of their primary material of interest. This possibility is reinforced by the similarities between the distribution mechanisms of piracy (e.g., Drachen and Veitch, 2013; Qu, et al., 2013) and CE material. At some point a website provides a new piece of content. Shortly after, every other website copies that piece of content from one another, ignoring the original source, and integrates it into their community. This method allows for the quickest access of material, through division of labor, and minimizes the risk for the originating source as it becomes nearly impossible to identify the original due to slight modifications during each replication (e.g., resizing of the image). If this is true, that large illicit networks improve efficiency, increase deindividuation, and reduce risk (McGloin and Piquero, 2010), this diffusion method provides a suitable online adaptation. Contrary to many illicit enterprises that strive to confirm the adage that ‘small is beautiful’ (e.g. Bouchard and Ouellet, 2011; Reuter, 1983), *public* CE-related communities may function under the premise that any connection is a good connection.

Are online hyperlinking ties so different from offline ties, and should we expect higher numbers than the ones we found in this study? Our second main finding was that community stability varied across networks and no consistent trends could be uncovered. Unlike connections in offline criminal communities, which can easily be severed through inactivity, the removal of a hyperlink between two websites requires a conscious effort on the part of the operator(s) to sever the connection. That is, hyperlinks between two websites do not have natural decay periods. This same phenomenon can be extended to all cyber-relationships. For example, friendship ties on social networks websites (e.g., Facebook©) are only severed when a person consciously ‘unfriends’ a relationship. With this in mind, it is important to note that the presence of a hyperlink between two websites should not be viewed as evidence of a strong connection between the websites. Rather, the presence of the hyperlink is simply an acknowledgement of awareness and that the connected websites share a common interest.

Given the lack of research on criminal community stability it is difficult to say definitively whether our finding of 49% community stability is different from offline criminal communities. While the level of care and nurturing required for a virtual connection may be lower than for an offline connection, it is worth noting that Kreager and colleagues (2011) reported a group stability of 35% between grade 8 and 9 for offline adolescent friendships, slightly lower than our virtual communities followed for 60 weeks. One key difference between offline and online criminal communities is that the lower amount of efforts required to maintain a connection in cyberspace may facilitate the creation of larger criminal communities, with potential implications for the creation of offending opportunities.

Research into gang member activities online has suggested the potential for increased offending opportunities through the collective identity that forms between users on websites (Pyrooz et al., 2015). Pyrooz et al. note that while the collective identity may enhance offline connections, it may also supplement the offline identity, extending personal offending opportunities. Translated to CE, the collective identity formed through membership to the same CE-related website can facilitate offending opportunities between users, assisting in the development of larger, criminal diverse, offender networks with few requirements for upkeep beyond website membership. Tremblay (2006) demonstrated as much in his study of “boy lover” online communities which provided the kind of identity support that some pedophiles were lacking prior to the Internet.

While the website communities formed provide individual offenders with the opportunity to acquire criminal capital (McAndrew, 2000; McCarthy and Hagan, 1995; McGloin and Piquero, 2010), they also provide opportunities for non-criminal physical, economic, social, or psychological exchanges which may otherwise be unattainable. These added benefits have been found in both offline (Morselli, Tremblay, and McCarthy, 2006; Weerman, 2003) and online contexts (Taylor and Quayle, 2003; Tremblay, 2006). As such, it is important that co-offending research goes beyond the immediate criminal event, and examines the role of the broader criminal community in which offenders are embedded, and how its structuring may impact co-offending opportunities. Online, this means examining the communities formed between websites and even within a website, as they provide information on how partnerships materialize, what steps are taken to nurture said partnerships, and the success of the partnership.

Cyberspace has evolved to play a central role in today’s society. However, the fast-pace and abundant amount of data available has presented challenges to social science researchers studying online phenomena. The emergence of the subfield ‘big data’ addresses some of these challenges. As shown in this study, online communities are part of a larger global network, with many interconnected pieces. Big data research fills an important gap as previously ‘un-researchable’ questions of criminological and sociological interest can now be answered (e.g., How large and how homogeneous are online communities? What drives online connections and what content do we find in these millions of connected webpages?). Progress in data collection techniques (e.g., webcrawlers) and analytic methods are changing the kinds of questions we can pose with clear implications for theoretical advances in cyberspace studies.

Limitations

Our study was subject to three primary limitations that we feel deserved specific attention. First, identifying communities within a network is a subjective process. If tasked with deriving a specific number of communities from any group of people, entities, or objects, a solution can be found. However, it does not necessarily mean that the solution found is the most suitable choice nor that it has any significant meaning. Therefore, it is important that we acknowledge the possibility that the communities we identified were not the most accurate representation of the true nature of the larger network. To best address this possibility, we undertook the community analysis with no preconceived expectations and allowed the *final proportion correct* (goodness of fit) and community densities dictate the number and size of each community. In addition, we tested goodness of fit for between four and 18 communities and conducted our analysis across ten networks, starting from two different website types (blog/site). As our final proportion correct was high (0.80 to 0.85), within-network variance was low (<0.01), and each network was similar in the number of communities, we believe that our interpretation of the community structure of our networks was valid.

Second, faction analysis functions under the notion that nodes, in this case websites, can belong to *only* one community. Research into multiplexity offers plenty of theoretical and empirical support for how individuals belong to multiple, partially overlapping, communities (e.g., Feld, 1981; Krohn Massey, and Zielinksi, 1988; Papachristos and Smith, 2014). Although a nested community detection analysis (see Moody and White, 2003) would have allowed for multiple group memberships, we elected to focus on each website's 'home base' as allowing 300+ websites to each belong to multiple communities would have made the analysis unnecessarily complicated and removed some of the meaningfulness of a website's primary community attachment.

Third, as the data collection process was automated and not exhaustive, it is unclear whether each website included in a network was directly involved in the dissemination of CE material. As offline delinquent peer groups are interspersed with non-delinquent peer groups (Kreager, Rulison, and Moody, 2011; Haynie, 2002), we would expect that pattern to exist online, with delinquent websites being connected to non-delinquent websites. However, a website does not need to be directly distributing CE material to be CE-related. Websites that connect to other websites distributing CE material can act as gatekeepers to material and perpetuate the distribution process by providing direct access. Although, we included those not directly distributing CE material, to preserve the reality of the larger online social structure, our inclusion criteria (images or keywords) should have minimized false positives, or more colloquially guilty-by-association.

CONCLUSION

Using a repeated measures design, to study the communities surrounding websites distributing child sexual exploitation (CE) material, we found that networks were comprised of two, large, core communities surrounded by a series of smaller communities. Known CE images were rarely found once we moved away from the immediate communities surrounding the seed website. Homophily within various content types was no greater than would be expected had websites been connected at random. Findings from this study shed light on the communal nature of CE distribution in cyberspace and how CE-related website connect. It also provides a framework for future research into co-offender selection processes, as the opportunities afforded to individual offenders are influenced by the macro network structure they reside within.

The public nature of the Internet provides an ideal setting for exploring social science questions that may otherwise be difficult to measure offline. The key is whether the processes identified online can be applied offline, and vice-versa. While outwardly different, online and offline criminal practices have been suggested as overlapping extensively (Grabosky, 2001). Moreover, online practices have been shown to influence offline criminal networks (Moule, Pyrooz, and Decker, 2014; Pyrooz, Decker, and Moule, 2015). It is conceivable that how other social problems manifest online may be indicative of how they function offline. Therefore, the Internet, and the use of automated data collection tools, may provide a huge advantage for accessing hard-to-reach populations, answering research questions that require additional anonymity, and examining how the Internet has modified societal processes.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Richard Frank for his work on creating the web-crawler used for this study, as well as the contributions of research assistant Ashleigh Girodat for her contributions to data collection and analysis. Finally, the authors would like to thank the Social Science and Humanities Research Council for funding this project (Grant Nb. 435-2012-0336).

REFERENCES

Ackland, Robert and Jamsheed Shorish. 2009. “Network Formation in the Political Blogosphere: An Application of Agent Based Simulation and E-research Tools.” *Computational Economics* 34(4): 383-398.

Adamic, Lada A and Natalie Glance. 2005. “The Political Blogosphere and the 2004 US Election: Divided They Blog.” In *Proceedings of the 3rd International Workshop on Link Discovery* 36-43. doi: 10.1145/1134271.1134277.

Basamanowicz, Jonathan and Martin Bouchard. 2011. “Overcoming the Warez Paradox: Online Piracy Groups and Situational Crime Prevention.” *Policy and Internet,* 32(1), Article 4. doi: 10.2202/1944-2866.1125.

Beech, Anthony R, Ian A Elliott, Astrid Birgden, and Donald Findlater. 2008. "The Internet and Child Sexual Offending: A Criminological Review." *Aggression and Violent Behavior* 13(3): 216-228.

Borgatti, Steven P., Martin G. Everett, and Linton C. Freeman. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Bouchard, Martin and Richard Konarski. 2014. “Assessing the Core Membership of a Youth Gang from its Co-offending Network.” Pp. 81-93 in Crime and Networks, edited by C. Morselli. New York, NY: Routledge.

Bouchard, Martin, and Frederic Ouellet. 2011. “Is Small Beautiful? The Link between Risks and Size in Illegal Drug Markets.” *Global Crime* 12(1): 70-86.

Bright, David A. and Jordan J. Delaney. 2013. “Evolution of a Drug Trafficking Network: Mapping Changes in Network Structure and Function across Time.” *Global Crime* 14(2-3): 238-260.

Burris, Val, Emery Smith, and Ann Strahm. 2000. “White Supremacist Networks on the Internet.” *Sociological Focus* 33(2): 215-235.

Burt, Ronald S. 2004. “Structural Holes and Good Ideas.” *American Journal of Sociology* 110(2): 349-399.

Carolan, Brian V. 2013. Social Network Analysis and Education. Thousand Oaks, CA: Sage Publication, Inc.

Carr, A. 2004. *Internet traders of child pornography and other censorship offenders in New Zealand.* Wellington, NZ: Department of Internal Affairs. Retrieved from http://www.dia.govt.nz/Pubforms.nsf/URL/entirereport.pdf/$!le/entirereport.pdf

Carrington, Peter J. 2002. "Group Crime in Canada." *Canadian Journal of Criminology and Criminal Justice* 44(3): 277-315.

Carrington, Peter J. 2014. The Structure of Age Homophily in Co-offending Groups. *Journal of Contemporary Criminal Justice*, online first. doi: 10.1177/1043986214553376.

Carrington, Peter J, Shannon Brennan, Anthony Matarazzo, and Marian Radulescu. 2013. Co-offending in Canada, 2011. *Juristat 333.* Statistics Canada Catalogue no. 85-002-X. Ottawa, ON; Statistics Canada.

Chau, Michael, Boby Shiu, Ivy Chan, and Hsinchun Chen. 2007. "Redips: Backlink Search and Analysis on the Web for Business Intelligence Analysis." *Journal of the American Society for Information Science and Technology* 58(3): 351-365.

Chau, Michael and Jennifer Xu. 2008. "Using Web Mining and Social Network Analysis to Study the Emergence of Cyber Communities in Blogs." *Terrorism Informatics* 18(1): 473-494.

Chow-White, Peter A. 2006. "Race, Gender, and Sex on the Net: Semantic Networks of Selling and Storytelling Sex Tourism." *Media Culture Society* 28(6): 883-905. doi: 10.1177/0163443706068922.

De Amorim, Saul G, Jean-Pierre Barthelemy, and Celso C. Ribeiro. 1992. "Clustering and Clique Partitioning: Simulated Annealing and Tabu Search Approaches." *Journal of Classification* 9(1): 17-41.

Décary-Hétu, David and Benoit Dupont. 2012. “The Social Network of Hackers.” *Global Crime* 13(3): 160-175.

Décary-Hétu, David, Carlo Morselli, and Stephane Leman-Langlois. 2012. “Welcome to the Scene: A Study of Social Organization and Recognition among Warez Hackers.” *Journal of Research in Crime and Delinquency* 49(3): 359–382.

Dolliver, Diana S. 2015. “Evaluating Drug Trafficking on the TOR Network: Silk Road 2, the Sequel.” *The International Journal of Drug Policy,* online first. doi: 10.1016/j.drugpo.2015.01.008.

Drachen, Anders, and Robert WD Veitch. 2013. “Patterns in the Distribution of Digital Games via BitTorrent.” *International Journal of Advanced Media and Communication* 5(1): 80-99.

Dupont, Benoit. 2013. “Skills and Trust: A Tour Inside the Hard Drives of Computer Hackers.” Pp. 195-217 in Crime and Networks, edited by C. Morselli. New York, NY: Routledge.

Estes, Richard J. 2001. The Sexual Exploitation of Children: A Working Guide to the Empirical Literature. Philadelphia, PA: National Institute of Justice.

Feld, Scott L. 1981. “The Focused Organization of Social Ties.” *American Journal of Sociology* 86(5): 1015-1035.

Felson, Marcus. 2003. “The Process of Co-offending.” Pp. 149-167 in Theory for Practice in Situational Crime Prevention, edited by M. J. Smith and D. B. Cornish. Monsey, NY: Criminal Justice Press.

Foot, Kirsten A, Steven M Schneider, Meghan Dougherty, Michael Xenos, and Elena Larsen. 2003. “Analyzing Linking Practices: Candidate Sites in the 2002 U.S. Electoral Web

Sphere." *Journal of Computer-Mediated Communities* 8(4). doi: 10.1111/j.1083-6101.2003.tb00220.x.

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. doi: 10.1145/1938606.1938609.

Freeman, Linton C. 1979. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1(3): 215-239.

Garton, Laura, Caroline Haythornthwaite, and Barry Wellman. 1997. "Studying Online Social Networks". *Journal of Computer-Mediated Communication* 3(1). doi: 10.1111.j.1083-6101.1997.tb00062.x

Gillespie, Alisdair A. 2012. Child Pornography: Law and Policy. New York, NY: Routledge

Girvan, Michelle and Mark EJ Newman. 2002. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences* 99(12): 7821-7826.

Glover, Fred and Manuel Laguna. 2013. "Tabu Search." Pp. 3261-3362 in Handbook of Combinatorial Optimization, edited by D. Z. Du and P. M. Pardalos. New York, NY: Springer.

Hanneman, Robert A and Mark Riddle. 2005. Introduction to Social Network Methods. Riverside, CA: University of Riverside.

Hargittai, Eszter, Jason Gallo, and Matthew Kane. 2008. "Cross-ideological Discussion among Conservative and Liberal Bloggers." *Public Choice* 134(1-2): 67-86.

Haynie, Dana L. 2002. "Friendship Networks and Delinquency: The Relative Nature of Peer Delinquency." *Journal of Quantitative Criminology* 18(2): 99-134.

Hillman, Henry, Christopher Hooper, and Kim-Kwang Raymond Choo. 2014. “Online Child Exploitation: Challenges and Future Research Directions.” *Computer Law and Security Review* 30(6): 687-698.

Holt, Thomas J. 2007. “Subcultural Evolution? Examining the Influence of On and Offline Experiences on Deviant Subcultures.” *Deviant Behavior* 28(2): 171-198.

Holt, Thomas J. 2009. “Lone Hacks or Group Cracks: Examining the Social Organization of Computer Hackers.” Pp. 336-355 in Crimes of the Internet, edited by F. Smalleger and M. Pittaro. Upper Saddle River, NJ: Pearson Prentice Hall.

Holt, Thomas J, Kristie R. Blevins, and Natasha Burkert. 2010. “Considering the Pedophile Subculture On-line.” *Sexual Abuse: Journal of Research and Treatment* 22(1): 3-24.

Holt, Thomas J, Deborah Strumsky, Olga Smirnova, and Max Kilger. 2012. “Examining the Social Networks of Malware Writers and Hackers.” *International Journal of Cyber Criminology* 6(1): 891-903.

Holt, Thomas J and Michael G Turner. 2012. “Examining Risks and Protective Factors of Online Identify Theft.” *Deviant Behavior* 33(4): 308-323.

Hurley, Ryan, 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*, pp. 631-642. Geneva, Switzerland.

Kreager, Derek A, Kelly Rulison, and James Moody. 2011. “Delinquency and the Structure of Adolescent Peer Groups.” *Criminology* 49(1): 95-127.

Krohn, Marvin D, James L Massey, and Mary Zielinksi. 1988. “Role Overlap, Network Multiplexity, and Adolescent Deviant Behavior.” *Social Psychology Quarterly* 51(4): 346-356.

Latapy, Matthieu, Clemence Magnien, and Raphael Fournier. 2013. “Quantifying Paedophile Activity in a Large P2P System.” *Information Processing and Management* 49(1): 248-263.

LeGrand, Benedicte, Jean-Loup Guillaume, Matthieu Latapy, and Clemence Magnien. 2009. Technical report on Dynamics of Paedophile Keywords in eDonkey Queries. Measurement and analysis of P2P activity against paedophile content project. Retrieved from: http://antipaedo.lib6.fr/.

Malm, Aili, Gisela Bichler, and Rebecca Nash. 2011. “Co-offending Between Criminal Enterprise Groups”. *Global Crime* 12(2): 112-128.

McAndrew, Meghan Anne. 2000. “Presidential Campaigns, Political Messages and the Internet: AN Analysis of Communication and Design Strategies on Web Sites with Political Content.” PhD dissertation, Georgetown University, Washington.

McCarthy, Bill and John Hagan. 1995. “Getting into Street Crime: The Structure and Process of Criminal Embeddedness.” *Social Science Research* 24(1): 63-95.

McCarthy, Bill, John Hagan, Lawrence E Cohen. 1998. “Uncertainty, Cooperation and Crime: Understanding the Decision to Co-offend.” *Social Forces* 77(1): 155-184.

McGloin, Jean M. and Holly Nguyen. 2013. “The Importance of Studying Co-offending Networks for Criminological Theory and Policy.” Pp. 13-27 in Crime and Networks, edited by C. Morselli. New York, NY: Routledge.

McGloin, Jean M., and Alex R Piquero. 2010. “On the Relationship between Co-offending Network Redundancy and Offending Versatility.” *Journal of Research in Crime and Delinquency* 47(1): 63-90.

McGloin, Jean M and Wendy P Stickle. 2011. “Influence or Convenience? Rethinking the Role of Peers for Chronic Offenders.” *Journal of Research in Crime and Delinquency* 48(3): 419-447.

McPherson, J Miller and Lynn Smith-Lovin. 1987. “Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups.” *American Sociology Review* 52(3): 370-379.

Milrod, Christie and Ronald Weitzer. 2012. “The Intimacy Prism: Emotion Management among the Clients of Escorts.” *Men and Masculinities* 15(5): 447-467.

Moody, James and Douglas R. White. 2003. “Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups.” *American Sociological Review* 68(1): 103-127.

Morselli, Carlo, Pierre Tremblay, and Bill McCarthy. 2006. “Mentors and Criminal Achievement.” *Criminology* 44(1): 17-43.

Moule, Richard K, David C Pyrooz, and Scott H Decker. 2014. “Internet Adoption and Online Behaviour Among American Street Gangs.” *The British Journal of Criminology* 54(6): 1186-1206.

Newman, Mark EJ and Michelle Girvan. 2004. “Finding and Evaluating Community Structures in Networks.” *Physical Review* 69(2): doi: 10.1103/PhysRevE.69.026.113.

Papachristos, Andrew V and Chris M Smith. 2014. “The Embedded and Multiplex Nature of Al Capone.” Pp. 97-115 in Crime and Networks, edited by C. Morselli. New York, NY: Routledge.

Park, Han Woo, Chun-Sik Kim, and George A Barnett. 2004. "Socio-communicational Structure among Political Actors on the Web." *New Media and Society* 6(3): 403-423.

Park, Han Woo and Mike Thelwall. 2003. "Hyperlink Analyses of the World Wide Web: A Review." *Computer-Mediated Communication* 8(4): doi: 10.1111/j.1083-6101.2003.tb00223.x.

Park, Han Woo, Mike Thelwall, and Randolph Kluver. 2005. "Political Hyperlinking in South Korea: Technical Indicators of Ideology and Content." *Sociological Research Online*, 10(3).

Pratt, Travis C, Kristy Holtfreter, and Michael D Reisig. 2010. "Routine On-line Activity and Internet Fraud Targeting: Extending the Generality of Routine Activity Theory." *Journal of Research in Crime and Delinquency* 47(3): 267-296.

Pyrooz, David C, Scott H Decker, and Richard K Moule. 2015. "Criminal and Routine Activities in Online Settings: Gangs, Offenders, and the Internet." *Justice Quarterly* 32(3): 471-499.

Qu, Benke, Wenjia Niu, Tianging Zhu, Lei Wu, Shijun Liu, and Na Wang. 2013. "Dynamic User Behavior-based Piracy Propagation Monitoring in Wireless Peer-to-Peer Networks." *Behavior and Social Computing: Lecture Notes in Computer Science* 8178(2013): 44-55.

Quayle, Ethel and Max Taylor. 2011. "Social Networking on a Nexus for Engagement and Exploitation of Young People." *Information Security Technical Report* 16(2): 44-50.

Reiss, Albert J and David P Farrington. 1991. "Advancing Knowledge about Co-offending: Results from a Prospective Longitudinal Survey of London Males." *Journal of Criminal Law and Criminology* 82(2): 360-395.

Reuter, Peter. 1983. Disorganized Crime: The Economics of the Visible Hand. Cambridge, MA: MIT Press.

Rheingold, Howard. 2006. “Social Networks and the Nature of Communities.” Pp. 47-75 in Networked Neighbourhoods, edited by P. Purcell. London, UK: Springer.

Rutgaizer, Moshe, Yuval Shavitt, Omer Vertman, and Noa Zilberman. 2012. “Detecting Pedophile Activity in BitTorrent Networks.” Pp. 106-115 in Passive and Active Measurement, edited by N. Taft and F. Ricciato. Berlin, GER: Springer Berlin Heidelberg.

Schaefer, David R. 2012. “Youth Co-offending Networks: An Investigation of Social and Spatial Effects.” *Social Networks* 34(1): 141-149.

Steel, Chad MS. 2009. “Child Pornography in Peer-to-Peer Networks.” *Child Abuse and Neglect* 33(8): 560-568.

Steinmetz, Kevin F and Kenneth D Tunnell. 2013. “Under the Pixelated Jolly Roger: A Study of On-line Pirates.” *Deviant Behavior* 34(1), 53-67.

Taylor, Max and Ethel Quayle. 2003. Child Pornography: An Internet Crime. New York, NY: Brunner-Routledge.

Tenti, Valentina and Carlo Morselli. 2014. “Group Co-offending Networks in Italy’s Illegal Drug Trade.” *Crime, Law, and Social Change* 62(1): 21-44.

Tremblay, Pierre. 1993. “Searching for suitable co-offenders.” Pp. 17-36 in Routine Activity and Rational Choice, edited by R. V. Clarke and M. Felson. New Brunswick, NJ: Transaction Publishers.

Tremblay, Pierre. 2006. “Convergence Settings for Nonpredatory 'Boy Lovers'.” Pp. 145-168 in Situational Prevention of Child Sexual Abuse, edited by R. Wortley and S. Smallbone. Monsey, NY: Criminal Justice Press.

Tretyakov, Konstantin, Sven Laur, Geert Smant, Jaak Vilo, and Pjotr Prins. 2013. “Fast Probabilistic File Fingerprinting for Big Data.” *BMC Genomics* 14(S2): S8. doi: 10.1186/1471-2164-14-S2-S8.

Van Hout, Marie Claire and Tim Bingham. 2013. “‘Silk Road’, the Virtual Drug Marketplace: A Single Case Study of User Experiences.” *International Journal of Drug Policy* 24(5), 385-391.

Van Mastrigt, Sara B and Carrington, Peter J. 2014. “Sex and Age Homophily in Co-offending Networks: Opportunity or Preference?” Pp. 28-51 in Crime and Networks, edited by C. Morselli. New York, NY: Routledge.

Van Mastrigt, Sara B and David P Farrington. 2011. “Prevalence and Characteristics of Co-offending Recruiters.” *Justice Quarterly* 28(2): 325-359.

Van Wijk, Anton, Annemiek Nieuwenhuis, and Angeline Smeltink. 2009. Behind the Scenes: An Exploratory Investigation into the Downloaders of Child Pornography. Arnhem, ND: Bureau Beke.

Vehovar, Vasja, Ales Ziberna, Matej Kovacic, Andrej Mrvar, and May Dousak. 2009. Technical report on An Empirical Investigation of Paedophile Keywords in eDonkey P2P Network. Measurement and analysis of P2P activity against paedophile content project. Retrieved from: http://antipaedo.lib6.fr/.

Weerman, Frank M. 2003. “Co-offending as Social Exchange: Explaining Characteristics of Co-offending.” *British Journal of Criminology* 43(2): 398-416.

Wellman, Barry, Jeffrey Boase, and Wenhong Chen. 2002. "The Networks Nature of Community: Online and Offline." *IT and Society* 1(1): 151-165.

Westlake, Bryce G, Martin Bouchard, and Richard Frank. 2011. "Finding the Key Players in Child Exploitation Networks." *Policy and Internet* 32(1), Article 6. doi: 10.2202/1944-2866.1126.

Westlake, Bryce G, Martin Bouchard, and Richard Frank. 2012. "Comparing Methods for Detecting Child Exploitation Content Online." In *Intelligence and Security Informatics Conference (EISIC), 2012,* 156-163.

Whittle, Helen C, Catherine E Hamilton-Giachritsis, and Anthony R Beech. 2014a. "Under His Spell": Victims' Perspectives of Being Groomed Online." *Social Sciences* 3(3): 404-426.

Whittle, Helen C, Catherine E Hamilton-Giachritsis, and Anthony R Beech. 2014b. "In Their Own Words: Young Peoples' Vulnerabilities to Being Groomed and Sexually Abused Online." *Psychology* 5(10), 1-12. doi: 10.4236/psych.2014.510131.

Whittle, Helen C, Catherine E Hamilton-Giachritsis, Anthony R Beech, and Guy Collings. 2013. "A Review of Online Grooming: Characteristics and Concerns." *Aggression and Violent Behavior* 18(1): 62-70.

Wolak, Janis, Marc Liberatore, Brian Neil Levine. 2014. "Measuring a Year of Child Pornography Trafficking by U.S. Computers on a Peer-to-Peer Network." *Child Abuse and Neglect* 38(2): 347-356.

Wortley, Richard K, and Steven Smallbone. 2006. Child Pornography on the Internet. Retrieved from: www.cops.usdoj.gov/publications/e04062000.pdf.

Wortley, Richard K, and Steven Smallbone. 2012. Internet child pornography: Causes, investigation, and prevention. Santa Barbara, CA: ABC-CLIO.

Zhao, Yunpeng, Elizaveta Levina, and Ji Zhu. 2011. "Community Extraction for Social Networks." *Proceedings of the National Academy of Science* 108(18): 7321-7326.

Zhou, Yilu, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai. 2005. "US Domestic Extremist Groups on the Web: Link and Content Analysis." *Intelligent Systems* 20(5): 44-51.