Disruption Strategies for Online Child Pornography Networks

by

Kilauea Joffres

B.A. (Criminology), Simon Fraser University, 2010

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Arts

in the
School of Criminology

SIMON FRASER UNIVERSITY

Spring 2012

Approval

Name: **Kilauea Joffres**

Degree: **Master of Arts (Criminology)**

Title of Thesis: ***Disruption Strategies for Online Child Pornography Networks***

Examining Committee:

Chair: Neil Boyd, L.L.M.

Martin Bouchard
Senior Supervisor

Eric Beauregard
Supervisor

William Glackman
Supervisor

Aili Malm
External Examiner
California State University Long Beach

Date Defended/Approved: April 26, 2012

Partial Copyright Licence

Abstract

The advent of the Internet has allowed for the creation of online child pornography networks, in which websites link to one another and facilitate access to child pornographic materials. This project seeks to use social network analysis tools to identify effective disruption strategies against online child pornography networks. For this purpose, four networks of child exploitation material were extracted using a specially designed web-crawler. These networks were then submitted to three different attack strategies (hub, bridge, and fragmentation attacks), the effects of which were measured on three outcome measures (density, clustering, and reachability). It was found that, to reduce density and clustering, hub attacks were generally the most effective strategy. Conversely, to reduce reachability, fragmentation attacks were the most successful strategy. In addition, fragmentation attacks are valuable for extremely large attacks across all outcome measures (e.g., when over half of the network nodes are removed). Variables such as network size and type did not appear to influence the effectiveness of attack strategies. Implications for law enforcement practice and future research are examined.

Keywords: Child Pornography; Social Network Analysis; Online Network

Dedication

To myself.

And to everyone in my life, past, current, and future

Acknowledgements

My eternal gratitude to Dr. Martin Bouchard, whose inhumane patience and support made this project possible.

Thank you to my family who quietly assumed I could accomplish anything.

Thank you to my border collies, Harley and Chase, whose unconditional love and loyalty make life all the more awesome.

Table of Contents

List of Tables

List of Figures

1. Introduction

As early as the 18th century, academics have been interested in networks as purely theoretical objects (Newman, 2008). Networks have since emerged as a practical tool for representing real world systems of interacting components, in which pathways connect nodes that represent the components of the system (Newman, 2008). Many systems present in nature form intricate web-like structures. For instance, cells can be characterized as networks of chemicals linked by chemical reactions; ideas travel through social networks in which humans are the nodes connected by social relationships; the Internet involves a network of wireless or physical connections between various computers and routers; and the World Wide Web is a network of hyperlinks that connect billions of websites (Newman, 2008). Networks such as these generally fall into one of four types: (1) social networks (e.g., friendship, affiliation or sexual networks), (2) technological or physical networks (e.g., power grids, Internet routers, airlines or railway networks), (3) biological networks (e.g., metabolic reactions, neural networks, food webs, protein networks), and (4) information or knowledge networks (e.g., citation, semantic, or online networks) (Newman, 2003). This project focuses on a particular information network: the Web. The purpose is to examine child pornography networks on the Web and explore various disruption strategies by using a social networks perspective (SNA).

1.1. Social Networks Analysis

According to Wasserman and Faust (1994), a distinctive characteristic of SNA is its focus on relationships between social entities, the patterns of these relationships, and their implications. Rather than examining behaviours, beliefs, and attitudes from an individualistic standpoint, the focus of SNA is on the interactions between actors and the manner in which these interactions form a structure whose properties can be analyzed and studied (Knoke & Yang, 2008; Wasserman & Faust, 1994). This makes SNA a research approach that distinguishes it from others (Knoke & Yang, 2008).

The defining characteristics of networks are nodes and ties, or actors and relations. It is this combination that forms a network. Actors can be individuals (e.g., students, terrorists, employees), collectivities (e.g., informal groups or corporations), or objects (e.g., websites, Internet routers, railway lines). A relation or a tie is a particular kind of contact or connection between two actors. Knoke and Yang (2008) specify that rather than being an attribute of an actor, a relation is "a joint dyadic property that exists only so long as both actors maintain their association" (p. 5). Ties between actors can assist or hinder the transfer of resources; this can include material resources such as money and other items, or nonmaterial resources, such as political support, information, respect and friendship (Wasserman & Faust, 1994). In addition, ties can be directed or nondirected. With directed ties, an actor instigates a connection and another receives it (e.g., a website links to another website). With nondirected ties, there is mutuality of connection (e.g., two websites link to each other). In essence, networks, which can be social, political, economical, etc., in nature, are expressed as patterns of relationships among actors.

Large variation exists among network structures: some may be isolated where few or no actors are linked, whereas others may involve a dense web of interconnections between actors. Knoke and Yang (2008) note that real networks tend to have intermediate structures wherein certain actors have more ties than others. Another characteristic of networks is their dynamic nature. Network structures change and evolve through the interactions of nodes, and as new nodes and ties are added or removed (Knoke & Yang, 2008). However, the analysis of social networks moves beyond simple representations of actors and their ties. Underlying SNA is the idea that network structures influence individuals and systemic levels of analysis (Knoke & Yang, 2008). Actors in a network are not seen as independent or autonomous; rather, they and their actions are interdependent on other actors and their connections. Similarly, the structural environment of a network is understood to either constrain or provide opportunity for individual behaviour (Wasserman & Faust, 1994), and in explaining behaviour, structural relations are considered more important than personal attributes such as gender, age, ideology, and values (Knoke & Yang, 2008).

SNA has been employed in many fields including communications, biology, economics, geography, information science, social psychology, sociolinguistics, criminology, and others (Newman, 2003). In the area of criminology, SNA has previously provided a valuable means of studying networks of street gangs (McGloin 2005; Papachristos, 2009; Xu & Chen, 2008), drug trafficking groups (Malm & Bichler, 2011; Morselli & Petit, 2007; Natarajan, 2006), adolescent offenders (Haynie, 2001), automobile theft rings (Morselli & Roy, 2008), and terrorist organizations (Krebs, 2002; Xu & Chen, 2008). Other research has explored how particular network structures are amenable to specific intervention efforts (Frank, Westlake, & Bouchard, 2010; Malm & Bichler, 2011; Westlake, Bouchard, & Frank, 2011; Xu & Chen, 2008).

1.2. SNA History

In a fairly non-technical form, social network analysis originally emerged from Radcliff-Brown's work on the notion of "social structure," which was defined as a set of patterned social arrangements (Scott, 2000). Between the 1930s and the 1970s, several sociologists and social anthropologists expanded on Radcliff-Brown's idea of "social structure," leading them to consider metaphors such as the "fabric" and "web" of social life (Scott, 2000). These metaphors were designed to convey the "interlocking" and "interweaving" nature of relationships that characterize social action; from this, the metaphor of a social *network* gained momentum and researchers began studying the "texture" and "density" of these networks (Scott, 2000). Moreno was part of a group of German psychologists who initially developed the concept of social networks in 1937. He believed that social relations had specific structures that could be mapped into a "sociogram". This diagrammatically illustrated the channels through which information traveled and through which people exerted influence. These diagrams involved nodes or dots that represented people and lines that expressed the connections or interactions among them. Moreno (1937) emphasized that such sociograms could locate isolated individuals and leaders, expose the degree of asymmetry and reciprocity in a network, and clarify the chains of connection between individuals.

Following Moreno's work, early sociological studies involved distributing questionnaires where participants discussed their interactions with others and their responses were subsequently mapped out into a network (Newman, 2003). Such early studies often examined issues of centrality (which identified the most connected or influential actors) and connectivity (which examined the degree to which the actors were connected to each other) (Newman, 2003). More recently, the focus has shifted from studying small graphs and the characteristics of their

individual nodes and pathways to examining the statistical properties of large-scale graphs (e.g., Adamic & Huberman, 2000; Barabási, 2003; Jones & Handcock, 2003; Pastor-Satorras *et al.* 2001; Vázquez, Pastor-Satorras & Vespignani, 2002). In part responsible for this shift were advances in computer technology and communication networks which allowed for the gathering and analysis of data on a much larger scale (Newman, 2003). While past studies might have involved networks with a dozen or, in some cases, hundreds of nodes, more recent studies can easily examine networks with millions of nodes (e.g., Adamic & Huberman, 2000; Barabási, 2003).

1.3. Online Child Pornography

With these advances in technology, the Web has fallen under intensive scrutiny, with studies by Barabási and Albert (1999), Barabási (2001), Albert, Jeong and Barabási (1999, 2000), and Broder *et al.* (2000) achieving prominence. The growing research focus on the Web follows from its worldwide popularity, stemming from its tremendous impact on the manner in which people access and distribute information. It is estimated that there are over 2,267 million Internet users worldwide, approximately 488 thousand of which are from North America (Internet World Stats, 2011). Moreover, the total user figure is constantly increasing, with a 528% growth since 2000 (Internet World Stats, 2011). As previously mentioned, the Web is a network of information stored on websites that are connected through hyperlinks; it is distinct from the Internet, which is a network of computers connected by optical fiber and other data connections. Since the inception of the Web, new networks have proliferated, the properties of which have been investigated by academics. For instance, researchers have explored the properties of peer-to-peer networks (Adamic, Lukose, & Huberman, 2003; Aiello, Chung, & Lu,

2000; Kaafar, Mathy, Turletti, & Dabbous, 2006; Iamnitchi, Ripeanu, & Foster, 2002), online communities (De Laat, 2002; Ellison, Steinfield, & Lampe, 2007; Preece, Maloney-Krichmar, & Abras, 2003), and networks emerging from online search engines (Adamic, 1999; Albert *et al.*, 1999; Broder *et al.*, 2000).

However, the Web's structure has allowed for the emergence of "dark networks," which are webpages that encourage or participate in illicit behaviour. Child pornography features among these "dark networks," and is an issue that has received increasing amounts of public attention. As O'Donnell and Milner (2007) note "In recent years, few issues have inflamed public passions like child pornography" (p. 64). In an archival examination of a global English-language newspaper, Krone (2005) reported that the degree of media attention on child pornography had increased dramatically: there were 4,573 articles that referenced child pornography between 1990 and 1994 compared to 51,270 between 2000 and September 2004. This suggests that increasing amounts of concern over the issue is being generated.

Problematically, attempts to define child pornography have encountered many difficulties. Healy (1996), at the First World Congress Against the Commercial Sexual Exploitation of Children, emphasized the extent of this problem:

> The question of what constitutes child pornography is extraordinarily complex. Standards that are applied in each society or country are highly subjective and are contingent upon differing moral, cultural, sexual, social and religious beliefs that do not readily translate into law. Even if we confine ourselves to a legal definition of child pornography, the concept is elusive. Legal definitions of both "child" and "child pornography" differ globally and may differ even among legal jurisdictions within the same country (p.2).

Nevertheless, the Canadian Parliament enacted s.163.1 to make child pornography illegal. S. 163.1(1) defines child pornography as:

> **(a)** a photographic, film, video or other visual representation, whether or not it was made by electronic or mechanical means,
>
> **(i)** that shows a person who is or is depicted as being under the age of eighteen years and is engaged in or is depicted as engaged in explicit sexual activity, or
>
> **(ii)** the dominant characteristic of which is the depiction, for a sexual purpose, of a sexual organ or the anal region of a person under the age of eighteen years; or
>
> **(b)** any written material or visual representation that advocates or counsels sexual activity with a person under the age of eighteen years that would be an offence under this Act.

R. v. Sharpe has since clarified that 163.1(1)(a) includes "both actual and imaginary human beings". Possession of child pornography carries a maximum of 5 years imprisonment. Making, publishing or possessing child pornography for the purpose of publication is punishable with a maximum of ten years imprisonment, as is distributing, selling, or possessing child pornography for such purposes. These offences are also punishable on summary conviction.

The Internet has revolutionized the process of accessing and sharing child pornography. Adler (2001) notes that, despite exhaustive attempts to eliminate child pornography, it has increased exponentially with the Web. As Krone (2005) observed, "It appears that a once limited trade has seen remarkable growth, with the potential to intrude into the homes and workplaces of all those connected to the Internet" (p. 1). The Web has assisted in the spread of and access to child pornography in many ways. For instance: the Web allows for global accessibility of materials; it decreases transaction costs; it provides an atmosphere of anonymity; it lacks a central, governing authority; it facilitates direct communication and image sharing between users; it offers images of high digital quality that do not deteriorate and can be conveniently stored; it allows for a variety of formats (pictures, videos, sound); and it facilitates the formation of alliances and communities (Spink, Ozmutlu, & Lorence, 2004; Wortley & Smallbone, 2006; Young, Griffin-Shelley, Cooper, O'Mara, & Buchanan, 2000).

In addition to allowing the creation of websites that host child pornography, the Internet has also facilitated communication and file-sharing between users. For instance, newsgroups have emerged; these are discussion forums that allow people with common interests to communicate and upload pictures to a post. Akdeniz (2001) suggests that there are around 200 sex-related groups, some of which involve child pornography. The Internet also provides IRC (Internet Relay Chat) services, which are chat rooms that provide for real-time communication; these can connect individuals to other offenders and potential victims. Peer-to-peer file sharing technologies have also emerged and these permit individuals to connect to the computers of others without using a third party, such as an Internet Service Provider, and to download files from them. While IRC and peer-to-peer networks have facilitated access to child pornography, the focus of this study is on child pornographic websites, and the networks these create through incoming and outgoing links to other such sites. For this project, websites, rather than Internet users, form the nodes within the networks that are examined.

Due to the covert nature of child pornography, it is difficult to accurately describe the number of child pornographic videos and images on the Web. However, various attempts have been made. The *US Customs Today* (2001) found that over 100,000 websites were involved with child pornography in some manner. More recently, the United Nations has suggested a much higher estimate of over four million child pornographic websites (Engeler, 2009). In addition, the Internet Watch Foundation, a 'hotline' in the United Kingdom, has estimated that over one million images of child abuse are being circulated online, with approximately 200 new images appearing every day (Robbins & Darlington, 2003). Furthermore, online child pornography is believed to produce over three billion American dollars each year in revenue (TopTenREVIEWS, 2004).

In terms of content, one study found that 83% of child pornography sites contained material involving children ranging from 6-12 years old, 39% had images of children ranging from 3-5 years old, and 19% had images of children under 3 years old (Engeler, 2009). In addition, 35% of these images depicted serious sexual assault (Engeler, 2009). It should be noted that child pornography is a virtually permanent record of a child's abuse, with the victim continuing to be exploited so long as the images remain on the Web (O'Donnell & Milner, 2007). Taylor and Quayle (2003) have argued that victims struggle to live with the psychologically painful consequences of their abuse being constantly circulated.

Notably, prior to the Internet, accessibility of child pornographic materials was more difficult and incurred greater risks; materials had to go through the slow, uncertain process of the postal service (O'Donnell & Milner, 2007). In addition, images tended to be locally produced and of poorer quality (Wortley & Smallbone, 2006). Furthermore, networks were sparser, as it was more complicated and riskier to locate other similarly minded people (Beech, Elliot, Birden, & Findlater, 2008; Wortley & Smallbone, 2006). The Web's structure has permitted motivated individuals to bypass these problems, partially through its web-like structure. That is, the Web's network arrangement allows individuals to travel through it, accessing and collecting more material and building a larger social network without the hassle of the postal service. It then becomes important to examine the most effective strategies for disrupting such a network, thereby inhibiting the ability of child pornography consumers to access and distribute the material.

1.4. Law Enforcement Issues

Current attempts to limit child exploitation have often ignored the networked nature of the Web. Enforcement efforts have tended to focus on chat room stings, injunctions against websites hosting child pornography, and establishing hotlines and complaint sites (Stanley, 2001; Wortley & Smallbone, 2006). Several image databases have also been established to identify child pornography websites, including the National Child Victim Identification Program (with over 520,000 images of children), the United States Division of Criminal Justice Services Database (with over 8,000 images), and the International Child Sexual Exploitation Image Database (which is one of the largest databases). Prominent companies such as Google© and Microsoft© have further developed tools to assist in locating child pornography. Google© has used Youtube's pattern recognition program to detect child pornography images while Microsoft© and NCMEC have developed a program that can quickly examine large numbers of image files and identify known and modified child pornography images (Shiels, 2008).

While these efforts have, to some extent, impeded the spread of and access to child pornography, they are not necessarily the most effective means of doing so. Specifically, two problems arise from such intervention strategies. First, there tends to be an overreliance on investigating and targeting websites in isolation. As a more effective approach, it has been argued that law enforcement should focus on the links between websites and the reliance of individuals on these networks (Krone, 2004). This approach acknowledges that the connections between child pornography sites, and the networks they form, are important to consumers, and as such, they are a valuable focus for intervention (Krone, 2004). Second, current enforcement efforts have been met with limited success. For instance, it is estimated that less than 1% of

online pedophiles are caught (McLaughlin, 2008) and at least 2% of child pornographic websites in Britain remained online for more than a year after identification (Johnson, 2008).

As such, more effective strategies for disrupting online child pornography sites must be explored. Frank *et al.* (2010) note that:

> With so many websites containing child sexual abuse images (and videos), and the limited resources available to various organizations to combat the problem, there needs to be continued efforts to automate and simplify the process of selecting and prioritizing targets for the purpose of criminal investigation (p. 2).

Frank *et al.* (2010) used social network analysis tools to focus on removing sites with the most severe child pornography content. Similar to Frank *et al.*'s (2010) work, the aim in this project is to use SNA tools to identify priority targets for law enforcement. However, the focus here is to determine priority targets according to a website's position in a network rather than its content.

2. LITERATURE

2.1. Network Characteristics

This project involves networks with two important characteristics: 1) they are real networks (i.e., they have not been artificially produced) and; 2) they are online networks (of websites rather than Internet users). These fundamental characteristics have implications for the networks' topology (that is, the arrangement of connections between its nodes). A network's topology is important to examine as it guides the selection of the appropriate network disruption strategies. For instance, it has been found that real-world networks differ from random ones in ways that make them resilient to random attacks (Albert *et al.,* 2000; Barabási, 2003). The nature of the Web also impacts the structure of networks in ways that provide further direction for selecting effective disruption strategies. Specifically, online networks have been found to possess small-world and scale-free properties (Albert *et al.,* 1999; Barabási, 2003; Broder *et al.,* 2000), which are linked to the effectiveness of particular disruption measures such as hub and bridge attacks (Barabási, 2003; Malm & Bichler, 2011; Xu & Chen, 2008). Much theoretical work has been conducted on the properties of various types of networks and the following section examines such research and its implications for the networks employed in this study.

2.1.1. Random and Non-random Networks

One of the simplest and most studied models of a network is the random graph (Bollobás,1985; Bollobás, Riordan, Spencer & Tusnády, 2001; Janson, Luczak & Rucinski,

2000; Durrett, 2007), which has been examined in depth by Rapoport (1961; 1968), Solomonoff and Rapoport (1951), and Erdős and Rényi (1959; 1960; 1961). More recently, researchers have explored real-life networks and the ways in which these networks and their properties are unlike those of random graphs. Importantly, real networks differ from random ones in ways that indicate how the network was formed and how it might be exploited.

One of the earliest attempts to create a model for large, random networks was Solomonoff and Rapoport's "random net" (1951), which was later rediscovered by Erdős and Rényi (1959), who named it the "random graph". This type of network is typically created by randomly attaching links to a static set of nodes (Erdős & Rényi 1959; 1960; 1961). Research has since established an interesting phenomenon with respect to random networks. That is, if a few links are randomly added to a large number of unconnected nodes, the only effect is that some nodes will form pairs. When more links are added, some of these pairs will form clusters; however, when each node eventually gains an average of just one link, a unique giant cluster is created (Barabási, 2003). Specifically, after a critical number of links are added to a network (in the case of random networks, 1 for every node), it undergoes a dramatic change; prior to this critical number, there exists only several islands of isolated clusters and after, there is one giant cluster in which most of the nodes are connected (Barabási, 2003). This phenomenon is described as the emergence of a giant component by mathematicians, the formation of a community by sociologists, and a phase transition in percolation by physicists.

Another important characteristic of random networks is that they tend to be egalitarian. Since all nodes have an equal chance of receiving a link, nodes in these networks will end up with roughly the same number of links (Barabási, 2003). Random networks thus follow a Poisson distribution; this distribution is marked by a pronounced peak, signifying that most

nodes will have as many links as the average node. The two sides of the peak then decrease quickly, indicating that departures from the mean number of links are very unusual. As a result, Erdős and Rényi's model would predict that the majority of people have around the same number of friends, that most papers receive about the same amount of citations, and that most websites are linked to by an approximately equal number of websites. However, it has been found that the degree distributions for real-world networks differ greatly from those of random graphs. Rather than following a Poisson distribution, many follow instead a power-law distribution. This is a distribution in which larger events have a high likelihood of occurring alongside smaller ones.

One of the first published instances of a scale-free network following a power-law distribution was Price's (1965) work on citation networks among scientific papers. Since then, power law degree distributions have been found in numerous other networks, such as other citation networks (Redner, 1998; Seglen, 1992), the Web (Adamic & Huberman, 2000; Albert *et al.*, 1999; Barabási, Albert, Jeong, & Bianconi, 2000; Barabási, 2003; Broder *et al.*, 2000**;** Kumar, Rajalopagan, & Tomkins, 1999), the Internet (Chen, Chang, Govindan, Jamin, Shenker, & Willinger, 2002; Faloutsos, Faloutsos, & Faloutsos, 1999; Vázquez, Pastor-Satorras, & Vespignani, 2002), metabolic networks (Jeong, Néda, & Barabási, 2003; Jeong, Tombor, Albert, Oltvai, & Barabási, 2000), and the network of human sexual contacts (Jones & Handcock, 2003; Liljeros, Edling, Amaral, Stanley, & Aberg, 2001).

One of the earlier studies examining power-law distributions in the Web was conducted by Albert *et al.* (1999), who uncovered a high degree of unevenness in the Web's topology. Using a web-crawler that returned 325,000 pages on the University of Notre Dame's domain, they found that around 82% of the websites had 3 or less incoming links, whereas the remaining

had over 1,000 incoming links (Albert *et al.*, 1999). Extrapolating from these results, it was determined that the Web was characterized by: (a) many websites with few links and (b) a few sites with many links (Albert *et al.*, 1999). This suggested that the Web is distinguished by a few very highly connected nodes, or hubs. A later study also confirmed this; using a network of 208 million websites, Broder *et al.* (2000) determined that approximately 90% of the sites had 10 or less links directed to them while around 3% of the sites were referenced by nearly a million other websites (Broder *et al.*, 2000.). Hubs include popular sites such as Facebook.com, Amazon.com, and CNN.com.

These, along with later studies, established that the degree distribution of nodes in the Web (both in- and out-degree) followed the mathematical expression of a power law (Adamic & Huberman, 2000; Albert *et al.*, 1999; Barabási *et al.*, 2000; Barabási, 2003; Broder *et al.*, 2000**;** Kumar *et al.*, 1999). Power law distributions differ from ones produced by random graphs in that they are a continuously decreasing curve, which indicates that many small events occur alongside a relatively few large events. In other words, while the majority of websites or nodes have only a few links, there are also a few large hubs in which sites have an unusually high number of links. There is no characteristic or average node in power law distributions (as in random graphs); as such, networks with a power-law degree distribution are referred to as scale-free networks (Albert *et al.*, 1999).

Notably, each power law has a unique exponent; in the case of the Web, this exponent describes the number of very popular sites compared to less popular ones (Barabási, 2003). Given that the power law in networks describes degree distributions, the exponent is typically labelled the degree exponent; researchers have found that the distribution of incoming links on a website follows a well-defined degree exponent of approximately two, with a slightly larger

degree exponent for outgoing links. For instance, Albert *et al.* (1999) found that, in their 325 729 node subset of the Web, the distribution of incoming links on a website had an exponent of 2.1, with an exponent of 2.45 for outgoing links. In Kumar *et al.*'s (1999) Web sample of 40 million documents, they found an in-degree exponent of 2.1 and an out-degree exponent of 2.38 (see also Kleinberg *et al.*, 1999). Broder *et al.* (2000) later used two Altavista crawls with 200 million documents in total and uncovered a consistent in-degree exponent of 2.1 and an out-degree of 2.72. This has implications for a network's error tolerance, which is discussed in later sections.

As noted, power laws mathematically describe how, in many real-world networks, most of the nodes only have a few links and a small percentage (hubs) have a vastly larger number of links. The few pathways that connect the many small nodes to each other are insufficient to create a completely connected network; this occurs through the existence of the relatively uncommon hub and its many connections, which prevents the network from forming separate clusters. For instance, most websites have an average of seven links; however, there are a few websites with thousands of links (Barabási, 2003). These hubs help maintain the giant component, which has repercussions for the network's robustness. Of note is that scale-free networks can be assortative or disassortative. Social networks are generally assortative, wherein high-degree nodes tend to connect with other high-degree nodes (Maslov, Sneppen, & Zaliznyak, 2004). In contrast, information networks such as the Web and biological networks, such as protein interaction networks, are disassortative, wherein high-degree nodes generally link to low-degree nodes (Maslov *et al.*, 2004). This means that fairly unpopular sites can still be connected to high-degree nodes by providing a hyperlink to them.

The finding of power laws in the Web has specific implications in terms of a network's dynamic behaviour, its robustness to random errors, and its vulnerability to targeted attacks. This will be expanded on in later sections; first, the formation of power laws will be examined.

2.1.2. The Emergence of Scale-Free Networks

As mentioned, Price's (1965) work on networks of citations between papers was one of the earliest examples of a scale-free network. Shortly after, based on Simon's work (Bornholdt & Ebel, 2001; Simon, 1955), Price (1976) developed a model to account for the emergence of power laws in networks. Price (1976) indicated that power laws emerge when "the rich get richer"; that is, when the amount of links a node has increases according to the number of links it already has, such that those nodes with many links will tend to accumulate more new ones than nodes with few links. Price (1976) dubbed this occurrence "cumulative advantage". He suggested that the rate at which an article receives citations is proportional to the amount it already has (Price, 1976). This appears intuitive; as Newman (2003) notes, the likelihood that an individual notices a paper while scouring the literature will likely increase with the number of times it is cited in various papers. The probability that the individual will then cite the paper in their own work increases in turn. This line of reasoning might be extended to the Web, although it is not obvious that there is a strictly linear relationship between the probability of a site linking to another site and that latter site's popularity, as Price (1976) assumed of citation networks. That is, a website is more likely to be noticed if it is popular (linked by many other websites); visible websites are then more likely to receive additional links.

Barabási and Albert (1999) later followed with their influential model to explain the creation of scale-free networks, which has subsequently driven much of the research on the

subject. Their model incorporates both preferential attachment and the notion of network growth. Erdős and Rényi and others (e.g., Watts-Strogatz) created models which assumed a fixed number of nodes that were connected together in a particular manner. Such models produced static networks in which the number of nodes is the same throughout the network's life. However, this is not a realistic assumption for most real-world networks and as such, the notion of growth must be accounted for. Barabási and Albert (1999) recognized that networks are open and created by the continuous addition of new nodes. They note that these new nodes are preferentially attached to nodes rich in connections (Barabási & Albert, 1999). This acts to form the hubs seen in power law networks (Jeong *et al.*, 2001). As Barabási and Albert (1999) explain, early nodes in a network have more time to accumulate links and because new nodes connect to those with more links, early nodes with more connections will be linked to more often and, as such, will grow faster than the younger and less connected nodes. In terms of the Web, Barabási (2003) notes that "when choosing between two pages, one with twice as many links as the other, about twice as many people link to the more connected page" (p. 85). In this sense, popularity is attractive, and websites such as Facebook and CNN are more likely than other, less popular sites, to accumulate links. Barabási and Albert (1999) note that growth and preferential attachment are both necessary to reproduce the power laws.

However, Adamic and Huberman (2000) argued that this explanation does not accurately describe existing data, since older sites in their study did not gain significantly more links than newer sites. To account for this discrepancy, Pennock, Flake, Lawrence, Glover and Giles (2002) combined power law with random linking to better account for the fact that new sites may also gain large numbers of links (i.e., become suddenly popular). In doing so, this model helps explain the partial non-linearity in linking behaviour for several power law networks.

2.1.3. The Small World Phenomenon

In addition to scale-free properties, the Web has another important property in common with many other real-world networks: it is a small-world. In a 1967 study on human interconnectivity, Milgram highlighted the small-world phenomenon by reporting that any two Americans could be reached through an average path of 6 people. Milgram (1967) argued that this small-world phenomenon characterizes such social networks as a natural consequence of society's dense web of connections. Other networks including the Web, the Internet, and Hollywood actors have since been identified as small-worlds (Watts, 1999; 2003).

This finding is meaningful given the Web's vast size. Measurements made by Lawrence and Giles (1999) over a decade ago indicated that the Web had approximately a billion documents; more recent estimates put the Web at over 11.5 billion pages (Gulli & Signorini, 2005). Moreover, the Web is constantly expanding with the addition of new sites. Despite this, the Web still displays small-world characteristics. This was initially discussed by Albert *et al.* (1999) who reported that the average path length in a sample of 325 729 webpages was 11.2; using finite size scaling, they estimated that for a network of 800 million websites, the average path length would be approximately 19. Later, a collaborative study conducted by AltaVista, IBM, and Compaq using a 50-million node sample of the Web, found an average path length of 16 (Broder *et al.*, 2000). At the domain-level, Adamic (1999) reported that the network has an average path length of 3.1 (the domain level refers to the set of network addresses that specify a type of institution, such as those sites falling under the common *.edu, .org .net, etc.* addresses). Additional research has indicated that there is a high degree of clustering in the Web; that is, the likelihood that two sites, which are connected to a common neighbour, are also linked to one

another was much greater than expected from a random network (Barabási, 2001; Pastor-Satorras *et al.* 2001; Yook, Jeong, & Barabási, 2001).

Networks such as the Web, in spite of containing billions of nodes, consistently display short path lengths due to their highly interconnected structure (Barabási, 2003). Recall that random networks need only one link every node to create a giant cluster. At this critical point, where each node averages one connection, the separation between nodes can remain relatively large; however, the act of adding more links to each node causes the distance between nodes to decrease drastically. As previously noted, websites have an average of seven links on them, and some have thousands; it is this relatively large degree of connectivity that assists in creating the small-world effect (Barabási, 2003). Thus, traveling through online networks can be done with relative ease, suggesting that offenders can find new materials without much difficulty given its small-world characteristics. Once an individual accesses a single child pornography site, the remainder of the network can easily be reached as a result. This stresses the importance of targeting sites, not in isolation, but with regard to the underlying network structure.

Interestingly, random graphs share a vital feature of small-worlds: short average path lengths (Barabási, 2003; Watts, 1999). However, the random graph does not adequately represent real-world networks on most other dimensions. Other than its non-representative Poisson distribution, it also has a low clustering coefficient (that is, the probability that two nodes are connected if they have a common neighbour) (Watts, 1999). Thus, while random and small-world networks may have similarly short average path lengths, real small-world networks also exhibit a higher degree of clustering (Watts, 1999).

The structure of a network determines how amenable it is to different types of attacks (random or targeted). The topology of online networks is characterized by a power-law distribution and small-world features; as will be seen, these have important implications in terms of what attack strategies are most successful against such networks.

2.1.4. Error and Attack Tolerance

Cohen, Erez, ben-Avraham, and Havlin (2000) sought to determine the number of nodes that would have to be removed from a random or a scale-free network in order to effectively destroy it. It is clear that the probability of a network collapsing into small, isolated clusters of nodes increases with number of nodes that are eliminated. Yet, years of research on random networks have demonstrated that the disintegration of a network is not a gradual process (Barabási, 2003). While eliminating only some of the nodes will not produce much impact on the network's overall structure, should a critical number of nodes be removed, the network will suddenly break into small, non-communicating components (Barabási, 2003). Anything below this critical threshold will produce little effect on the network's integrity; this is referred to as an inverse phase transition.

In their work, Cohen *et al.* (2000) confirmed that anything above this critical threshold will cause the network to disintegrate. However, when nodes are randomly eliminated in scale-free networks, they discovered that this critical threshold vanishes when the degree exponent is smaller than or equal to three (Cohen *et al.*, 2000). As previously noted, the Web displays an in- and out-degree of less than 3. As Barabási (2001) notes, when randomly attacked, "these networks break apart only after *all* nodes have [randomly] been removed –or, for all practical purposes, never" (p. 115). Many networks demonstrate a relatively high degree of tolerance for

(random) errors (Albert *et al.*, 2000). For instance, fairly simple organisms can survive radical environmental or pharmaceutical interventions due to the robustness of their genetic and metabolic network (Jeong *et al.*, 2000; Jeong, Mason, Barabási, & Oltvai, 2001). Communication networks also have a high degree of robustness, as seen by the fact that while important elements routinely fail, local malfunctions seldom result in the collapse of a network's global information-carrying ability (Barabási, 2003). Such stability is often credited to redundant wiring within a network; however, beyond such redundancy, the topology of a network is also important in determining its error tolerance. A network is considered robust or error tolerant only when a giant cluster which contains most of the nodes persists even after a number of nodes have been eliminated (Cohen *et al,* 2000).

Research has found a strong association between a network's topology and its error tolerance or robustness. An important finding in this area is that scale-free networks are more resistant to random failures than random networks, yet they are more vulnerable to targeted attacks that eliminate the most connected nodes (Albert, *et al.* 2000). As Barabási (2003) notes, "a significant fraction of nodes can be randomly removed from *any scale-free network* without it breaking apart" (p. 113; emphasis in original). This occurs because random failures affect large nodes and small nodes with equal probability; however, because there are many more small nodes in scale-free networks, these will be disproportionately affected. Given that the contribution of these nodes to the integrity of the network is relatively insignificant, the network remains connected (Barabási, 2003). That is, this robustness emerges from the network's heterogeneous topology. In contrast, in homogeneous networks like the random graph, all nodes share roughly the same number of links, which all contribute similarly to the network's diameter; as such, the deletion of any node produces comparable amounts of network fragmentation

(Barabási, 2003). However, the robustness of scale-free networks comes at a cost: repeated attacks targeted at the hubs can effectively disconnect a network (Albert *et al.,* 2000). As previously mentioned, networks such as the Web are held together by the hubs and this reliance on highly connected nodes causes power law networks to break down sooner than random networks in targeted attack. This vulnerability characterizes all scale-free networks (Barabási, 2003).

For instance, in a study of the Internet's resilience to router failures, it was discovered that, even by randomly removing as many as 80% of all nodes, the surviving 20% still formed a tightly linked cluster (Albert, Jeong, & Barabási, 2000). Similarly, the average path length on the Internet is remains largely unchanged when 60% of its nodes are randomly removed (Albert *et al.,* 2000). Another Michigan-Ann Arbour study reported that hundreds of Internet routers are malfunctioning at any moment in time (at least 0.3%); yet Internet users rarely discern any disruption in their services (Labovitz, Ahuja, & Jahanian, 1999).

Albert, Jeong and Barabási (1999; 2000) also examined the error and attack tolerance of the Web, using their subset of 325 729 nodes. They began by removing the most connected hubs. The elimination of the first hub did not break the cluster, as the remaining hubs were successful in maintaining the network. However, the removal of several hubs produced various isolated islands of nodes; when they continued to remove more connected nodes, the critical point, which did not emerge with random attack, now surfaced, and the system dramatically fell apart (Albert *et al.,* 2000). In order to achieve this, the largest hubs must be simultaneously removed, and generally, 5 to 15 percent of the hubs must be removed (Albert *et al.,* 2000; Barabási, 2003). Such studies on the attack tolerance of online networks do not involve "dark networks." As such, no research has yet examined the attack tolerance of online child pornography networks, nor

have measures been specifically introduced to measure tolerance in these networks. This study seeks to extend current research by conducting attacks to online child pornography networks.

2.2. Key Players

As noted, the aim of this project is to identify strategies that will maximally disrupt a particular kind of network: online child pornography networks. This requires the removal of nodes, and the question becomes how to identify those nodes that will most successfully break apart a network and the criteria by which this identification process should be guided. For instance, in their work, Easton and Karaivanov (2009) show that targeting particular individuals is more effective at reducing crime than targeting random or repeat offenders. This notion contributes to arguments that the structural characteristics of a network will influence its resilience (Chase-Dunn, Kawano, & Brewer, 2000; Barabási, 2003; McGrath & Krackhardt, 2003). Examining a network's structure, and the arrangement of connections that shape it, will assist in understanding how a network operates (Raab & Milward, 2003); this will then allow for the identification of important "pressure points" that will maximize network disruption (Malm & Bichler, 2011; Xu & Chen, 2008). Some researchers have identified central actors as the adequate pressure points, and have further shown that targeting such actors can decrease the risks posed by criminal networks (Morselli, 2009; Xu & Chen 2005). Borgatti and Everett (2006) have also argued that disruption of a network is best achieved by identifying key or central players. However, the extent to which a network demonstrates scale-free, small-world and other characteristics will impact the "pressure points" or central players that are appropriate to target. Recent work by Malm and Bichler (2011) and Xu and Chen (2008) on dark networks has expanded on these points.

2.2.1. SNA Disruption Strategies

In their work, Malm and Bichler (2011) studied which law enforcement strategies worked best for drug trafficking-related networks. Specifically, they explored networks of activities within the drug market commodity chain by using data drawn from various police data measures, including arrest and incident reports, intelligence files, police investigation reports, and surveillance records. Networks were formed for those involved in the production, transportation, supplying, and retail areas of drugs, as well as those described as feeders (who are not directly involved in the distribution process). These networks were then examined to determine the degree to which they demonstrated small-world characteristics (e.g., an average of 6 paths separating nodes), scale-free properties (e.g., contained hubs in which certain nodes had many connections), and vulnerability features (e.g., had high fragmentation scores). They found that these network features affected in specific ways which targeting strategies (e.g., a hub attack, a repeated hub attack, a bridge attack, or a combination thereof) best disrupted the network. They argued that, for small-world networks, which have high levels of clustering and thus leave nodes in a position to replace others, repeated attacks on multiple nodes will most successfully disrupt the network (Malm & Bichler, 2011). In contrast, scale-free networks will be best disrupted through hub attacks, a finding corroborated by Albert *et al.* (2000) and Xu and Chen (2008). Finally, networks with high vulnerability, characterized by many actors who bridge together subgroups, were found to be susceptible to attacks that disrupt bridges which sever the flow of information. This result is consistent with McGloin's (2005) work on police interventions for street gangs.

In another study, Xu and Chen (2008) examined two types of terrorist networks and two criminal-related networks. The terrorist networks included the Global Salafi Jihad, a terrorist

network of 366 members, and an online terrorist network that included 104 websites managed by four major international terrorist groups. The criminal-related networks included a methamphetamines trafficking network of 1,349 criminals and a network of 3,917 criminals convicted of gang-related crimes. Xu and Chen (2008) also determined the extent to which their networks exhibited scale-free and small-world properties; they found that pure scale-free networks were vulnerable to both hub and bridge attacks, while small-world networks were more vulnerable to bridge attacks.

Malm and Bichler (2011) and Xu and Chen (2008) thus introduced broad strategies for attacking networks. These strategies identify nodes with a particular type of centrality in a network. As a result, they can be translated into specific suggestions for key players to target in a network. Diverse measures of centrality have been examined in the literature, the most common including measures of degree (the number of ties a node has) and betweenness (the extent to which a node brokers between others) (Freeman, 1979; Wasserman, Faust, Iacobucci, & Granovetter, 1994). Hub attacks target those nodes with many links to and from other nodes in a network. In this sense, hub attacks remove those nodes high in degree centrality. Conversely, bridge attacks sever those nodes that connect other nodes in a network. In this way, bridge attacks eliminate those nodes high in betweenness.

Degree centrality has previously been described as a useful measure to identify prominent nodes. For instance, Baker and Faulkner (1993) focused on degree centrality, among other measures, to uncover the central individuals in a price-fixing conspiracy network in the electrical equipment industry. In his work, Krebs (2002) found that in a network of 19 hijackers, the member commonly identified as the ring leader had the highest degree score. Betweenness is also a commonly acknowledged measure of centrality. Burt (1992) characterized actors who

bridge gaps as structural holes in a network; such actors are exposed to greater diversity, and in the case of online networks, expose others to more diversity. Morselli and Tremblay's (2004) work on drug trafficking groups found that those in brokerage-type positions benefit from certain advantages in a network.

However, identifying key players to target in a network is not necessarily obvious. In Borgatti's (2003) approach to centrality, he argues that how centrality is measured depends on why it is important. He recognizes that there are two reasons for targeting central actors: 1) central actors can be targeted in order to maximally disrupt the network and 2) central actors can be targeted to maximally collect information on networks. For this project, those measures that produce the greatest network disruption are of interest. Borgatti (2003) cautions that that traditional measures of centrality cannot always "optimally solve the key player problem" (p. 127). It is possible for traditional measures to identify a node that, while central in a network, will cause little disruption if removed. This would occur if, for example, a node is linked to many actors, but these actors can still reach each other through alternative ties when this central node is removed. Conversely, if many actors in a network rely on a particular node to reach each other, its removal would have a more significant impact on the network. Instead of being redundant, this node is integral to the flow of information in the network, making it a valuable law enforcement target. To resolve the problem of redundancy, Borgatti (2003) develops the measure of fragmentation to identify those actors whose removal would most disrupt the network by isolating various nodes and network sections.

The current study examines various strategies identified as important in the literature in order to determine which will produce the greatest disruption in four different online child pornography networks. This may then allow law enforcement to select strategies that will cause

the most disruption to online child pornography networks while expending the fewest amounts of resources.

2.3. Current Study

Aided by the Internet, which has facilitated the distribution and access to information, child pornography continues to be a prevalent problem. Law enforcement attempts to curb child pornographic websites have generally involved individually investigating and shutting down websites, chat room stings, complaint hotlines, and other strategies that often ignore the networked nature of the issue. Unlike an individual focus on websites, a network approach acknowledges the importance of the structure of connections between groups of websites. Websites may then be removed in a way that maximizes their disruption to the entire network, rendering it more difficult for individuals to traverse the network and access materials. This strategy may prevent police from using limited resources to remove websites that produce little impact to the overall network. To accommodate the reality of online child pornography, a social network perspective is warranted.

This perspective has grown more prominent in the literature with the acknowledgement that many objects of interest, including websites, form networks. With the considerable popularity of the Internet, various researchers have applied the study of networks to websites (Albert *et al.*, 1999, 2000; Barabási, 2000, 2003; Broder *et al.*, 2000). It has since been established that many networks (online or otherwise) differ from random ones in important structural ways: they follow a power-law distribution, they form small-worlds, and they have higher network clustering. As a consequence, various disruption strategies may be differentially effective. For instance, it has been found that networks with a power-law distribution are resilient

to random attacks and vulnerable to certain targeted attacks, such as hub and bridge attacks (Barabási, 2003; Malm & Bichler, 2011; Xu & Chen, 2008).

Much of the past research on online networks has focused on their topology, and much of this research has used networks that do not necessarily contain illegal elements. Little research has been conducted on *disrupting* criminal networks using a social networks perspective (the main exception being Malm and Bichler, 2011). Even less research has been done on disrupting *online* criminal networks using this perspective (the main exception being research by Xu and Chen, 2008). Finally, even less research has been conducted on disrupting online *child pornography* networks (the exception being work by Frank *et al.*, 2010). This is despite the serious nature of the crime and the large amount of online child pornography. The problems with current online law enforcement efforts, combined with a lack of knowledge surrounding appropriate online child pornography disruption strategies, produce an important research gap to fill. This study seeks to fill that gap by testing certain social networks-guided attack strategies against four online child pornography networks to determine which strategies are more effective in what circumstances. This will ideally generate an understanding of the structure of certain online child pornography networks while providing law enforcement with guidelines to most effectively attack such networks.

More specifically, by using four different networks of online child pornography, this study has the following goals:

1) To explore the structure of the online child pornography networks to determine whether they exhibit the small-world and scale free properties identified as basic properties of online networks.

2) To create an understanding of the disruption that various attack strategies (fragmentation, bridge, hub, and random) produce on these networks.
    a. To examine which of these strategies is most the effective at disrupting the networks:
        i. for different outcome measures (density, clustering, reachability, and cohesion),
        ii. for different sizes of attack (when 10%, 20%, 30%, and 40% of the nodes are removed),
        iii. for different network sizes (small or larger), and
        iv. for different network types (with a blog-seed or a site-seed)

To evaluate different strategies of disrupting child exploitation networks, the method presented in this paper first extracts a sub-network from the Web that deals with child exploitation material and then uses established SNA tools to guide attacks against the network. In order to do so, a web-crawler called the Child Exploitation Network Extractor (CENE) was used to produce four networks of existing child pornography websites. This web-crawler has been used in past research to extract child exploitation networks (see Frank *et al.,* 2010; Westlake *et al.,* 2011).Two of the four networks for this project emerged from a typical website with interlinking sites, while the other two emerged from a blog-type website. These networks were then submitted to four different attack strategies previously used in network disruption studies: a fragmentation attack, a bridge or betweenness attack, a hub or degree attack, and a random attack. The effects of these attacks were subsequently examined at different levels of attack. That is, the networks' structure was re-examined after 10%, 20%, 30%, and 40% of the nodes identified by each attack strategy were removed. Damage to the network was examined

based on specific outcome measures that originate from the social networks perspective: density, clustering, and reachability. Decreasing each of these measures diminishes accessibility of information in a network in various significant ways. Using a software program (UCINET) designed for social network analysis, the goal of this study is to determine which attack strategy most effectively reduces certain outcome measures in the (four) online child pornography networks sampled.

3. DATA AND METHODS

3.1. Child Exploitation Network Extractor (CENE)

As noted, four online networks were used in this project; they were produced by CENE, a custom-written web-crawler designed by computer scientist Richard Frank and colleagues (Frank *et al.,* 2010). This crawler is designed to recursively follow links from a starting website until it meets specific termination criteria (i.e., a certain number of pages and websites). As the crawler does this, it collects statistics on the number of keywords, images and videos on each of the webpages stemming from that particular website. This information is then aggregated at the website level, and links to and from relevant sites are reproduced. The product is a mapped network of websites with information on the content within, and the directed links between, these websites.

Three limits were imposed on CENE to prevent it from perpetually crawling the Internet. First, a limit of 250,000 webpages retrieved was included to keep the extraction process time bounded. Second, network size was limited to 200 websites, with webpages sampled as equally as possible between websites. Due to this limit, the network is not a complete one; rather, it is a subset of a larger network. Third, a set of keywords were defined in an attempt to ensure that the websites extracted were topic relevant. This set includes 63 child pornography related words, many of which were (a) commonly used by the Royal Canadian Mounted Police (RCMP) to locate illegal child-related content and (b) used in other studies of online child pornography (Grand, Guillaume, Latapy, & Magnien, 2009). The web-crawler included ‘softcore’ words such

as girl, boy, love, child, teen, variations of Lolita, young, bath*, twink, pre/post pubescent, innocent, smooth and hairless. It also included a set of 'hardcore' words, such as penis, cock, vagina, pussy, anus, anal, sex, pedo/paedo, oral, virgin, naked and nude.

To be included into the network, a webpage had to have at least seven of the 63 keywords. If it failed to meet this criterion, the webpage was discarded and no links were followed from it. It was determined through manual verification that seven keywords reliably distinguished between child exploitation webpages and unrelated ones. The web-crawler also discarded broken links or websites inaccessible for other reasons (including timeouts, sign-in requirements, and password barriers). Videos and images from each webpage were also recorded. In order to avoid including very small images such as logos and emoticons, images were recorded only if they were 150x150 pixels or larger. No requirement was imposed on videos. However, the content of these websites, beyond the fact that they contain child exploitation materials, is not of particular interest to this study.

While CENE provides a useful way of uncovering online child pornography networks, it has limitations. For instance, given some of the keywords, there is the potential for false positives (for example, it is possible for a website to include words such as child, girl, boy, teen, and innocent, and not be about child exploitation). Nonetheless, these false positives may link to child pornography or vice-versa. In this way, they play a role in the network that may also be relevant to examine. A further limitation of the web-crawler is its inability to analyze content from, and follow links out of, password protected websites. Consequently, these websites were not captured in the networks. Nonetheless, CENE remains a helpful method for extracting networks of child exploitation websites.

3.1.1. Network Starting Points

Four websites containing child exploitation material were selected as the starting point for the network extraction process. These were selected from two locations: the RCMP's Integrated Child Exploitation (ICE) unit, which provided a record of child pornography websites, and Google© searches, which provided a list of websites following a search of specific child pornography-related words (such as Lolita, pthc, realkiddy, and nymphet). The top search returns were manually screened to ensure that the websites selected were topic relevant; if a website contained child exploitation materials, it was included as a starting point. Websites included did not necessarily have to feature "hardcore" content, although children had to be depicted in a sexually provocative manner. Softcore videos and images, instances of sexual objectification, and written materials involving sexual activities with children were all materials that could potentially make a website topic-relevant and be included as a starting point.

According to Westlake (2011), "[These] two methods of website selection … mimic the process a person might take searching for child pornography" (p. 36). The websites gathered from the ICE unit represent an individual who may gather websites from specific sources, such as friends or pedophilic others. Individuals may also use Google© to search for child pornography if they have no such sources or they seek to expand their collection. The web-crawler is then designed to extract networks that include all potential paths an individual might take when searching for more material. An individual might begin with a particular website (obtained from Google© or another source) and follow the links out of that website to reach other websites, from which the individual can access further information. Using specific starting websites, the web-crawler thus produces a network that emerges from the connections between child pornography websites.

3.1.2. Network Type

The four networks were extracted using different types of websites as starting points. Two of these had a blog as a starting point or seed (and are referred to as Blog-A and Blog-B) while the other two emerged from a traditional site-seed (referred to as Website-A and Website-B). The site-seed networks were produced using as starting point a typical website format, which includes a home page with many interlinking pages. Such sites may feature galleries of pictures, videos, text and unrestricted chat forums. The other two networks grew from a single blog site (i.e., a blog-seed). Blogs are defined as a type of journal published online in which users can describe their thoughts, emotions, and post images (Gruhl, Guha, Liben-Nowell, and Tomkins, 2004). These blogs can be integrated into other websites including Facebook or Google or may function primarily to provide blogging services, such as Tumblr and Livejournal.

The decision to include a blog as a starting site was made due the increasing popularity of these sites (Furukawa *et al.*, 2007; Mitchell, Wolak, Finkelhor, & Ybarra, 2008), as well as their potential implications for the content of child pornography sites and the structure of these networks. Blogs are advantageous in at least two ways: 1) they are a more efficient and less expensive means of hosting child pornography, and 2) they increase the appearance of anonymity (Westlake, *et al.*, 2011). Regularly updated journals often contain hyperlinks and photographs that can be quickly and easily shared online; on account of their structure, blogs may therefore link to other sites and blogs at a greater frequency than traditional websites (Ali-Hasan, & Adamic, 2007). Many blogging websites also allow users to engage their services free of charge, removing the need to pay for a host site. In addition, blog websites may provide journal templates, negating the need for HTML knowledge (Westlake *et al.*, 2011). These blog features increase the ease of distributing and accessing materials. As mentioned, the apparent

anonymity for blogs is also greater. Blogs can be created using only an e-mail address and a username. This is unlike websites, for which a name and address may be required for payment purposes. Moreover, if a blog is shut down on account of hosting illegal content, it is relatively simple for a user to create a new account and resume posting (Westlake *et al.*, 2011).

However, the type of starting point (blog or site) does not imply that all websites in the network will be of the same type. Rather, a blog used as a starting point may link not only to other blogs, but also to various sites. The same can be said of a network with a site-seed. The difference in network type is mostly related to the starting point rather than the entire content of the network, although this may be influenced by the starting point.

3.1.3. Network Size

The four networks for this study were selected so that each pair would include a smaller and a larger network. For instance, Blog-A is a smaller network of 111 nodes while Blog-B is a larger network of 163 nodes. Similarly, Website-A has 46 nodes while Website-B has 162 nodes. The networks are different sizes due to the nature of the web-crawler. The web-crawler is designed to visit all links outside of a website and subsequently analyze its content to determine whether or not it will be included in the network. If the website does not meet the appropriate criteria, it was not included in the network; however, it is still included as part of the 200 website limit imposed on the web-crawler. For example, in Blog-A, 111 websites were considered relevant and included in the network, while 89 were considered irrelevant and excluded. Selecting networks of different sizes allows, to some extent, for the determination of whether the effectiveness of attack strategies differs with network size. If this is the case, then law enforcement efforts may have to be adjusted to the size of the network.

3.2. Network Attacks

3.2.1. Social Network Attack Strategies

As previously discussed, this project seeks to identify the most effective social network analysis measures to disrupt online child pornography networks. For this purpose, various attack strategies were used to identify particular sites whose elimination would have the largest impact on specific outcome measures. These attack strategies involve hub attacks (using the measure of degree centrality), bridge attacks (using the measure of betweenness), fragmentation attacks (using the measure developed by Borgatti, 2003), and random attacks (where each node has an equal chance of being targeted).

With the exception of random attacks, each of these network disruption strategies identifies key players who are central to the network in varying ways. For hub attacks, the degree centrality measure examines the number of ties that a website has to other websites (Freeman, 1979). The underlying assumption of this measure is that nodes with many connections are more likely to be powerful since they can directly influence more actors, access more resources in a network, and are less dependent on other actors since they have alternative means for fulfilling their needs (Hanneman & Riddle, 2005). The networks for this project have directed ties; some websites link to others (out-degree ties) while some websites are linked to by others (in-degree ties). Websites with many in-degree ties may be considered more important or prominent; a website can easily link to others, but it may not be relevant or interesting enough to receive links from other websites. By virtue of their ability to attract traffic, popular sites may be important law enforcement targets. Websites with out-degree ties are also valuable to consumers, as they may connect them with many other websites, thus providing them with abundant access to materials in the network.

For bridge attacks, betweenness centrality identifies those websites that fall on the shortest path between other websites in a network (Freeman, 1979). It describes the extent to which a website 'brokers' between other websites. In a network, this position can be advantageous, as it allows certain websites to bridge groups and control the flow of information between actors (Malm & Bichler, 2011). For interested individuals, these websites are important insofar as they provide access to various parts of a child pornography network that would otherwise be more difficult to reach.

Key players were also identified through a fragmentation analysis. This measure indicates the proportion of sites that would not be able to reach each other if any particular site was removed (Borgatti, 2003). That is, it reflects the extent to which a user's ability to access other websites will be affected by the removal of certain sites. If most sites produce little fragmentation, the network will largely maintain its structure even after some of the nodes have been removed, allowing users to continue traveling through the network with relative ease. However, if many sites have high fragmentation scores, their removal will likely break the network apart in ways that will severely limit a user's ability to effectively make use of the network. By removing nodes with high fragmentation scores, the greatest number of isolates (individual nodes disconnected from the network) and islands (clusters of nodes disconnected from the network) will be produced.

The removal of websites identified by these various measures followed a sequential process. This involved (a) identifying the website that scored highest for one measure, (b) removing it, and (c) reanalyzing the network to identify the next top website. This strategy avoids the potentially redundant effect of eliminating certain websites simultaneously (Borgatti, 2003; Xu & Chen, 2008; Schwartz & Rouselle, 2009). For instance, a site's high score on a

measure may be dependent on a higher scoring node whose removal would then mitigate this first node's score. The sequential method of removing nodes avoids such problems. This process of eliminating nodes is repeated until the desired number of nodes has been removed.

3.2.2. Attack Sizes

In attempting to identify effective attack strategies, it is important to do so in a manner that highlights the potential nuances of each attack. As such, four waves of attacks were directed against each network. At the end of each attack, network structure was re-analyzed to determine the effectiveness of various attack strategies. In the first wave of attack, 10% of the nodes that ranked highest on a particular measure (fragmentation, betweenness, and centralization) were removed sequentially. In the second attack wave, 20% of websites were removed; in the third, 30% of websites were eliminated; and, finally, in the final attack 40% of the nodes in the network were deleted. The attacks were only extended to the removal of 40% of the nodes, as the smaller networks tended to disintegrate completely at this point (i.e., only a few pairs of nodes remained). The effects of each attack were subsequently examined on the outcome measures of density, clustering, reachability. Employing different attack sizes will provide an idea of what the networks look like following small attacks (when 10% of the nodes are removed), mid-scale attacks (when 20% and 30% of the nodes are removed), and large attacks (when 40% of the nodes are removed). It will also determine whether the effectiveness of certain attack strategies change as attacks are increased. Due to the different network sizes, varying numbers of nodes were removed at every attack stage for each network (see Table 1). For example, in the first wave of the attack, where 10% of the nodes were removed, 11 of 111 nodes were removed from Blog-A; 16 of 163 nodes were eliminated from Blog-B; 5 of 46 nodes were removed from Website-A; and, 16 of 162 nodes were eliminated from Website-B.

Table 1 - Number of Nodes Removed for Each Attack Size for all Networks

Nodes Removed (%)	Blog-A	Blog-B	Website-A	Website-B
10	11	16	5	16
20	22	33	9	32
30	33	49	14	49
40	44	65	18	65

3.2.3. Outcome Measures

The impact of fragmentation, bridge, hub, and random attacks websites was then examined on several outcome measures including density, clustering, and reachability. Density is calculated by dividing the number of existing ties in a network with the number of possible ties (Hanneman & Riddle, 2005). Assessing the changes in density is valuable, since it examines the changes in the amount of ties. The more ties that are eliminated in a network, the more difficult it is for individuals to reach other websites.

Change in the overall clustering of the network was also assessed. The clustering coefficient is the average density of the neighbourhoods of the websites in a network (Hanneman & Riddle, 2005). In other words, it examines the likelihood that two websites, which are linked to one particular website, are also linked to one another. As with the overall network density, by eliminating certain websites, and therefore certain ties within a cluster or a neighbourhood, access to materials within a network becomes more difficult. In addition, this prevents consumers from becoming embedded in a tightly-knit community that promotes their views and interests. When the clustering coefficient is examined in relation to density, it provides information on whether there is an even distribution of ties in the network or whether there are clusters of websites that link mainly to each other.

A final measure of network structure was examined: network reachability. This calculates all the paths that exist in the network between each node. Even if network density is reduced, and

certain pathways are eliminated, alternative routes may exist to a particular node. Reduced density may make it more difficult, but not impossible, to access certain websites. By decreasing reachability, websites can become increasingly isolated from the network until they can no longer be reached. As such, reductions in this measure may make it not only increasingly difficult to access materials in the network, but potentially impossible by simply surfing through a network.

In sum, the effectiveness of each attack strategy (hub, bridge, and fragmentation) was assessed on particular outcome measures (density, clustering and reachability). The effectiveness of these attack strategies on the outcome measures was also examined against three variables: 1) attack size (small, medium, and large), 2) network type (blog and website), and 3) network size (small and large). Should different attack sizes lend themselves to different attack strategies, law enforcement may have to consider the resources available prior to designing and launching an attack. If attack strategies differ with the type of network, it may then be important for law enforcement to first determine whether the network of interest falls in one category or the other before proceeding with an attack. Finally, if network size affects the success of certain attack strategies, then law enforcement may also have to first consider the size of the network.

This study's findings, expanded on in the following sections, indicate that specific attack strategies are most appropriate for certain outcome measures as well as attack sizes. For most attack sizes, hub attacks tend to be the most effective strategy at disrupting a network's density and clustering. However, as attacks become larger in size, the fragmentation measure grows increasingly effective. In fact, when nearly half of the nodes are removed from a network, fragmentation attacks sometimes become more effective than hub attacks. In terms of reducing reachability, fragmentation and betweenness attacks tended to be the most effective strategies

across different attack sizes. No clear differences in the effectiveness of attack strategies emerged for network type; the effectiveness of these strategies appeared to depend more on the particular arrangement of nodes and ties in a network, which did not appear to differ in obvious ways between network type. Similarly, while smaller networks were easier to disrupt than larger networks, size did not appear to determine the effectiveness of particular attack strategies.

4. RESULTS

4.1. Network Descriptives

The following chapter details the results of the study. They are structured in the following manner. Descriptives of the networks are first presented, followed by a detailed examination of which attack strategies are best for each outcome measure in the first, smallest wave of attack (where 10% of the nodes in each network are removed). Change in the effectiveness of these strategies for each outcome measure is then examined for each progressively larger attack wave.

Table 1 introduces the major characteristics of the networks, while Figure 1 illustrates each network. For instance, it can be seen from Table 2 that Blog-A has 663 ties, a density of 0.0543, a clustering coefficient of 0.424, a network reachability of 3054, an out-degree centralization of 21.124%, and an in-degree centralization of 22.041%. The graph in Figure 1 shows that the left side of the network tends to be denser, indicating that many of the network's hubs are located there.

When the networks are compared, it can be seen that the larger networks demonstrate greater density (over 0.1) than the smaller networks (which have a density of 0.0543 and 0.0725). This indicates that there is a higher ratio of pathways to nodes in the larger networks. In this sense, more connections exist between the nodes in larger networks. As such, along with having a greater number of nodes, the nodes in larger networks are also more reachable. This suggests that not only is there more information to access in larger networks, but that it is easier

to access. The difference in density of the networks is illustrated in Figure 1; due to the fewer number of ties, the smaller networks are much sparser.

Overall, the networks have roughly similar amounts of clustering; that is, the density of the neighbourhoods of nodes in the networks are fairly similar. The two website-seed networks have a clustering of 0.441 (Website-B) and 0.442 (Website-A), while Blog-B network has a larger clustering of 0.525 and Blog-A has the smallest clustering coefficient of 0.424. This indicates that the extent to which the networks have areas in the network where groups of nodes all link to each other is comparable. As such, networks with site-seeds and blog-seeds may have similar potential for the formation of communities. Nevertheless, there exists greater range in the clustering between the Blog-seed networks than the site-seed ones; some Blogs or journals may be particularly effective at fostering network community, while others may be less so.

Due to the greater number of nodes and ties, the larger networks have higher reachability scores (20145 for Blog-B and 22059 for Website-B compared to 3054 for Blog-A and 1021 for Website-A). The higher density of larger networks results in more ties, which provide more alternative pathways to nodes within the network and increases reachability. As such, greater reachability indicates that even when one path to a node is removed, there is a higher likelihood for the larger networks to have a substitute route to that node. Unless the appropriate attack strategies are selected, this may render larger networks more difficult to disrupt, and require the elimination of more nodes.

The larger networks also differ from the smaller ones in terms of their out- and in-degree centralization scores. This measure expresses the overall degree of variance in network centrality as a percentage. In the bigger networks, there is a much larger difference between the out- and

in-degree centralization than in the smaller networks. Blog-B has an out-degree centralization of 70.725% and an in-degree centralization of 26.627%. Similarly, Website-B has an out-degree centralization of 81.089% and an in-degree centralization of 31.716%. This indicates that there are certain nodes which contribute a large percentage of the network's out-going links, thus forming prominent hubs. In contrast, the in-degree links are more evenly distributed across the network, although there is still some concentration among particular nodes. In terms of the smaller networks, Blog-A has an out-degree centralization of 21.124% and an in-degree centralization of 22.041% while Website-A has an out-degree centralization of 19.852% and an in-degree centralization of 13.037%. For these networks, in-degree and out-degree links are similarly distributed among nodes in the network. There is an absence of disproportionately large out-degree hubs present in the larger networks where some of the nodes dominate in terms of receiving out-degree links.

Overall, there appears to be no substantially descriptive differences between networks that have a blog or a traditional website as a starting point. They both share similarities and differences that do not appear to be divided according to network type (save perhaps for clustering). This may be due to the fact that all networks undoubtedly contain both websites and blogs that link to each other, although the proportion of websites to blogs may differ across networks. Because not all sites were manually verified, the number of blogs and websites in each network is unknown. As such, it cannot be said that blog-seed and website-seed networks differ in any meaningful manner based on the extracted descriptive statistics.

Notably, the networks all display characteristics identified in real-world, online networks: small-world tendencies, high clustering, and a power-law distribution. For instance, they all have an average path length of less than 3.5, which is shorter than the path length of 6 that marks

small worlds (Milgram, 1967). The networks also demonstrate higher clustering than randomly generated networks of matching size and density. For example, Website-B has a clustering coefficient of 0.441 while its random counterpart has one of 0.108. Conversely, the random networks have greater reachability and are more compact than the real-world networks; this may be due to the equal distribution of ties in these random networks.

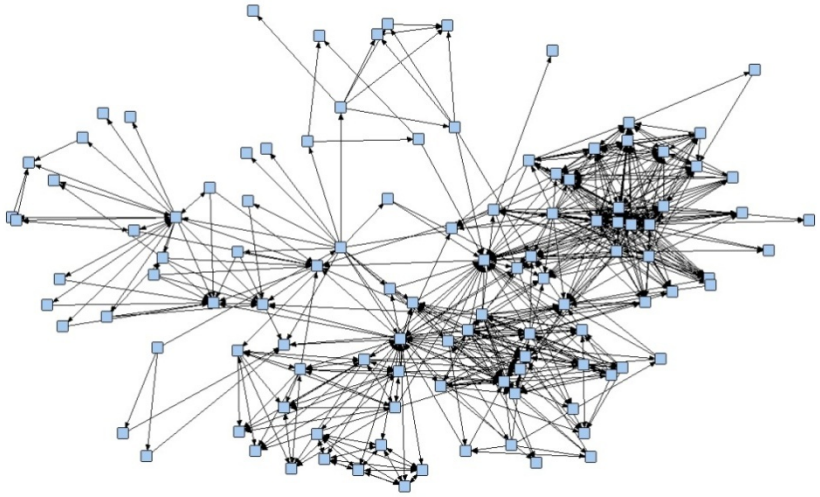
The real-world networks also show features of scale-free networks. That is, these networks tend to have much greater centralization than the randomly generated ones. This difference is most marked in the larger networks. For example, Blog-B has an out-degree centralization of 70.725% and an in-degree centralization of 26.627% while its random counterpart has an out- and in-degree centralization of just 8.615%. The same trend holds for the other network. The smallest difference occurs between Website-A in-degree centralization of 13.037 and its random counterpart's in-degree centralization of 10.765%. In sum, this indicates that the networks used in this project have more hubs than would be expected from random networks; this is a property identified in networks following a power-law distribution. As such, these networks tend to reproduce qualities previously found in other online networks.

Table 2: Descriptive Measures for the Four Online Child Pornography Networks and Four Random Networks

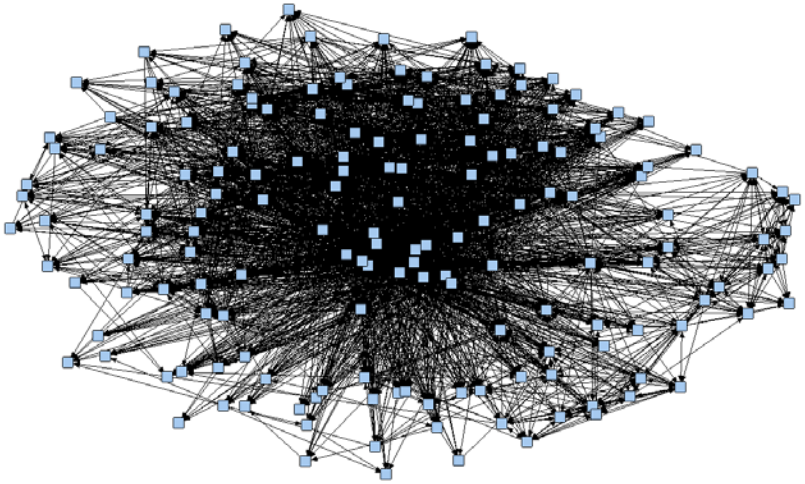
	Network							
Measure	Blog-A	Random Blog-A	Blog-B	Blog-B Random	Website-A	Random Website-A	Website-B	Website-B Random
Nodes	111	111	163	163	46	46	162	162
Ties	663	663	4096	4096	150	150	2795	2795
Density	0.0543	0.0543	0.1551	0.1551	0.0725	0.0725	0.1072	0.1072
Clustering Coefficient	0.424	0.055	0.525	0.152	0.442	0.063	0.441	0.108
Average Path Length	2.409	2.809	1.975	2.245	3.491	3.172	2.352	2.149
Reachability	3054	12100	20145	26406	1021	1981	22059	26082

Centralization									
	Out	21.124%	5.562%	70.725%	8.615%	19.852%	8.494%	81.089%	6.713%
	In	22.041%	9.231%	26.627%	8.615%	13.037%	10.765%	31.716%	8.588%

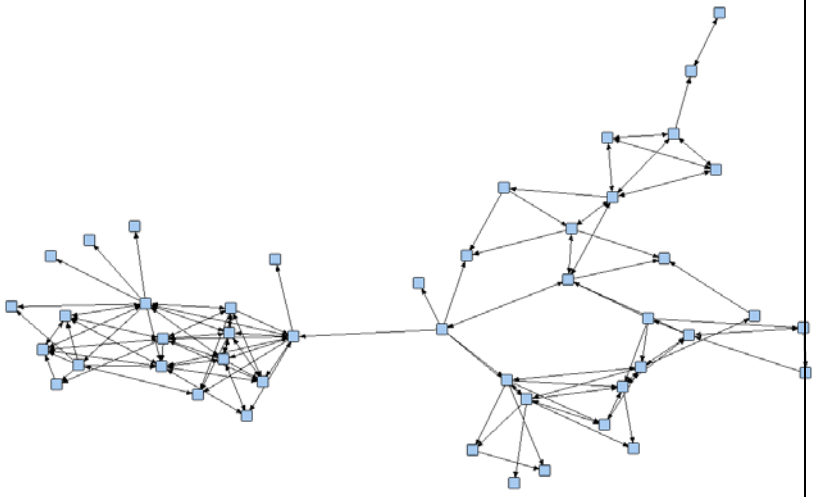
Figure 1. The Four Online Child Pornography Networks



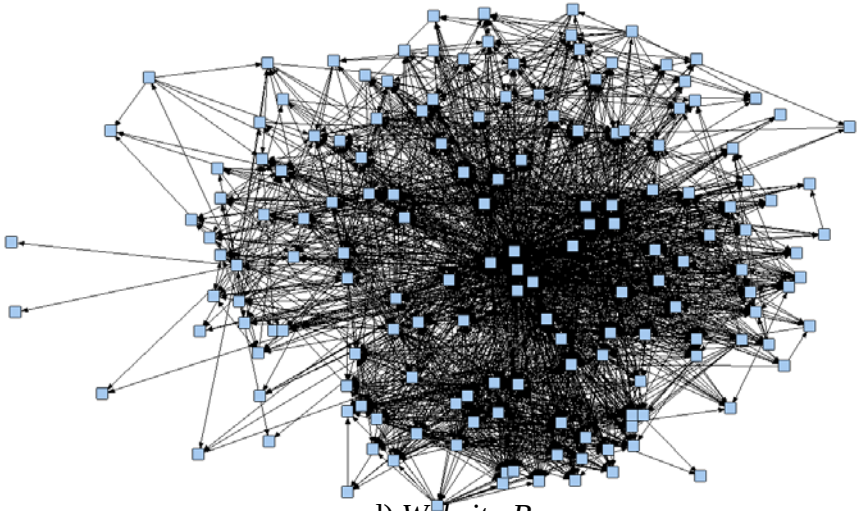
a) *Blog-A*



b) *Blog-B*



c) *Website-A*



d) *Website-B*

4.2. The Impact of Attacks on Outcome Measures

Four waves of attacks were conducted against the networks. These attacks ranged from small, mid-sized, to large. The attacks provide information on which disruption strategy is the most effective for different outcome measures; the sequential nature of the attacks provides information on whether certain strategies are more effective for different attack sizes. The most effective disruption strategy for each outcome measure is initially examined for small attacks and then changes in the effectiveness of these attack strategies is examined as the attacks are extended. Results suggest that overall, out-degree attacks are more effective at reducing density and clustering for smaller to larger attacks, although fragmentation attacks slowly become more effective for larger attacks. Fragmentation and betweenness attacks are more effective at decreasing reachability for all attack sizes.

4.2.1. Density

Density represents the proportion of ties present to the number of ties possible. By reducing density, the accessibility of information in the network is impeded as fewer ties exist between websites in the network. Successful attack strategies for the smallest attack are first examined (when 10% of the nodes have been removed). Two attack strategies were found to be effective across networks: out-degree and betweenness attacks (see Table 3). The former attack strategy was most effective against Blog-A and Website-B, while both attacks produced similar amounts of disruption against Blog-B and Website-A.

For Blog-A, the out-degree attack produced the greatest reduction in density: it fell from 0.0543 to 0.0333 (a 38.7% reduction) while the number of ties dropped from 663 to 330. Out-

degree attacks were also most effective for Website-B, producing a 45.62% reduction in density, from 0.1072 to 0.0583. The number of ties dropped from 2795 to 1235. For Website-A and Blog-B, both out-degree and betweenness attacks were similarly successful at reducing density. For Website-A, the out-degree attack produced a 31.034% reduction in density, from 0.0725 to 0.0500, while the number of ties fell from150 to 82. The betweenness attack decreased density by 30.207% to 0.0506 while 83 ties remained. For Blog-B, the out-degree attack reduced density from 0.1551 to 0.0968 for a 37.59% reduction, while the number of ties fell from 20145 to 2078. However, the betweenness attack was slightly more effective; density fell to 0.0958 for a 38.2% reduction while the number of ties fell to 2055.

Random attacks against the networks were the least effective disruption strategy. In all cases except for Blog-A, the attacks slightly increased density. This indicates that while fewer nodes in the network were available to access (as these were removed), the attacks failed to target nodes those nodes with a disproportionately high amounts of ties. Rather, more peripheral (and potentially less important) nodes with fewer ties may have been removed.

In sum, out-degree attacks tended to be the most effective density-reduction strategy in a small attack. However, betweenness attacks may be just as effective or even slightly more effective in certain cases. This suggests that nodes with high out-degree scores provide many of the network's ties and consequently, their elimination produces notable reductions in density. Some nodes have up to 147 out-going links in the larger networks. Indeed, the prominence of these nodes is expressed in the large out-degree centralization for some of the networks. Yet, the similar effectiveness of betweenness attacks may indicate that bridging nodes also have many out-going and/or in-going links. Such nodes may be particularly important to target as they are not only notable contributors to the density of the network, but they are also useful for

connecting different parts of the network together. For instance, the website that scored highest in betweenness in Blog-B also had 136 out-going links and 56 in-coming links, ranking 3rd highest for the number of out- and in-degree ties in the network. At the same time, it is the most significant bridge in the network. This suggests that in some networks, certain hubs might also function as important bridges.

As noted, in the first wave of attacks, removing hubs was the most effective strategy for reducing density overall. To a large extent, this continued to be the case as attacks were expanded; however, with each increasingly large attack, the effectiveness of fragmentation strategies increased concomitantly (see Tables 3-5). For certain networks, fragmentation attacks eventually became more effective than hub attacks.

In the second attack wave, where 20% of the nodes were removed, hub attacks remained the most successful strategy (see Table 4). In Blog-B and Website-B, out-degree attacks produced a 61% and a 62.7% reduction in density respectively; in Blog-B, 1015 ties remained while 670 remained for Website-B. However, for certain networks (Blog-A and Website-A), more than one strategy produced similar amounts of disruption. For Blog-A, the out-degree attack reduced density by 55.8% with 188 ties remaining while the fragmentation attack was slightly more effective, producing a 56.7% reduction in the measure with 184 ties remaining. For Website-A, the out-degree attack targeted many of the same nodes as the fragmentation and betweenness attacks, and as such, they all produced the same impact. All three measures reduced density by 44.1% and the number of ties fell to 54.

In the third attack wave where 30% of the nodes were removed, hub attacks continued to be an effective strategy at reducing density (see Table 5). However, fragmentation attacks

became an increasingly successful strategy. Hub attacks were most effective against two networks: Blog-B and Website-B, producing a 72.0% and a 76.9% reduction in density respectively. Conversely, fragmentation attacks were more effective for another network (Blog-A), and both out-degree and fragmentation attacks were equally effective for the remaining network (Website-A). For Blog-A, density fell by 73.7% following a fragmentation attacks compared to 68.1% following a hub attack. For Website-A both out-degree and fragmentation attacks reduced density by 65.2%.

For the last wave of attack, in which 40% of the nodes were eliminated, hub attacks became effective at disrupting network density in only half of the networks (Website-A and Blog-B), whereas fragmentation attacks became the most successful strategy for the other networks (Blog-A and Website-B) (see Table 6). Furthermore, even for Website-A and Blog-B, the disruption caused by the fragmentation attacks was similar to that caused by the out-degree attacks. For Website-A, the out-degree attack produced a 79.9% reduction in density while the fragmentation attack produced a 76.3% reduction. For Blog-B, density was also reduced by 79.9% following an out-degree attack and by 79.1% following a fragmentation attack. Fragmentation attacks were more successful for Blog-A and Website-B. For Blog-A, this attack produced an 85.1% reduction (compared to 77.2% for an out-degree attack). For Website-B, the fragmentation attack caused density to drop by 87.4% while the out-degree attack reduced density by 85.2%.

In essence, out-degree (or hub) attacks tend to be the most effective strategy for reducing density, while fragmentation attacks become increasingly useful for larger attacks. Given the scale-free nature of the network, certain nodes with a disproportionately large number of links exist within the network; eliminating the largest hubs thus maximizes the number of links that

are removed while minimizing the number of nodes removed. The centralization of the networks is such that hubs attacks continue to be a viable option for the mid-scale attacks in which 20% to 30% of nodes are eliminated. It is in the largest wave of attacks that fragmentation attacks become more prominent. This is due to the fact that hub attacks tend to lose power as attacks are extended. An important characteristic of scale-free networks is that there are a few nodes with many links and many nodes with few links. After a certain point, those few nodes with many links have been removed and the many nodes with few links remain. These nodes are less important to the structural integrity and density of the network, and so their removal produces less network disruption. In contrast, fragmentation attacks gain power as attacks are extended. The goal of this attack is to completely fragment the network so that only isolated nodes or pairs remain. Thus, the network is attacked in such a way that progressively more nodes become disconnected or grow isolated from the network. This manner of targeting nodes in the fragmentation attack is very different from hub attacks: its effects are less obvious in early attacks, but the benefits continue to grow rather than diminish, as in the case of hub attacks. Consequently, hub attacks are preferable for smaller to mid-sized attacks, while fragmentation attacks may be best for very large attacks where its effects have an opportunity to flourish. For a summary of the most effective attack strategies across attack sizes, refer to Table 7.

Table 3 – Density after 1st Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation	0.0379 (↓30.2%)	0.0979 (↓36.9%)	0.0561 (↓22.62%)	0.0686 (↓36.01%)
Ties	375	2101	92	1452
Betweenness	0.0382 (↓29.65%)	0.0958 (↓38.2%)	0.0506 (↓30.21%)	0.0642 (↓40.11%)
Ties	378	2055	83	1360

Degree	0.0333	0.0968	0.0500	0.0583
Out	(↓38.7%)	(↓37.6%)	(↓31.03%)	(↓45.62%)
Ties	330	2078	82	1235
In	0.0386	0.1164	0.0506	0.0747
Ties	(↓28.91%)	(↓24.95%)	(↓30.21%)	(↓30.32%)
	382	2499	83	1582
Random	0.0518	0.1553	0.0744	0.1101
	(↓4.6%)	(↑0.1%)	(↑2.6%)	(↑2.7%)
Ties	513	3331	122	2330

Table 4 – Density after 2nd Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation	0.0235	0.0661	0.0405	0.0526
	(↓56.7%)	(↓57.4%)	(↓44.1%)	(↓50.9%)
Ties	184	1108	54	882
Betweenness	0.0250	0.0630	0.0405	0.0466
	(↓54.0%)	(↓59.4%)	(↓44.1%)	(↓56.5%)
Ties	196	1057	54	782
Degree	0.0240	0.0605	0.0405	0.0400
Out	(↓55.8%)	(↓61.0%)	(↓44.1%)	(↓62.7%)
Ties	188	1015	54	670
In	0.0301	0.0891	0.0368	0.0555
Ties	(↓44.6%)	(↓42.6%)	(↓49.2%)	(↓48.2%)
	236	1495	49	930
Random	0.0492	0.1618	0.0751	0.1120
	(↓9.4%)	(↑4.3%)	(↑3.6%)	(↑4.5%)
Ties	385	2714	100	1878

Table 5 – Density after 3rd Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation	0.0143	0.0447	0.0252	0.0278
	(↓73.7%)	(↓71.2%)	(↓65.2%)	(↓74.1%)
Ties	86	576	25	352
Betweenness	0.0180	0.0468	0.0262	0.0317
	(↓66.9%)	(↓69.8%)	(↓63.9%)	(↓70.4%)
Ties	108	603	26	401
Degree	0.0173	0.0435	0.0252	0.0248
Out	(↓68.1%)	(↓72.0%)	(↓65.2%)	(↓76.9%)
Ties	104	560	25	314

In Ties	0.0256 (↓52.9%) 154	0.0746 (↓51.9%) 978	0.0272 (↓62.5%) 27	0.0473 (↓55.9%) 598
Random Ties	0.0493 (↓9.2%) 296	0.1663 (↑7.2%) 2142	0.0665 (↓8.3%) 66	0.1100 (↑2.6%) 1467

Table 6 – Density after 4th Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation Ties	0.0081 (↓85.1%) 36	0.0315 (↓79.1%) 299	0.0172 (↓76.3%) 13	0.0135 (↓87.4%) 126
Betweenness Ties	0.0226 (↓58.4%) 173	0.0332 (↓78.6%) 316	0.0209 (↓71.2%) 17	0.0186 (↓82.6%) 173
Degree Out Ties	0.0124 (↓77.2%) 55	0.0311 (↓79.9%) 296	0.0146 (↓79.9%) 11	0.0159 (↓85.2%) 148
In Ties	0.0174 (↓68.0%) 77	0.0563 (↓63.7%) 546	0.0198 (↓72.7%) 15	0.0387 (↓63.9%) 360
Random Ties	0.0513 (↓5.5%) 227	0.1765 (↑13.8%) 1678	0.0661 (↓8.8%) 50	0.0925 (↓13.7%) 861

Table 7 – Summary of the Most Effective Attack Strategies Against Density

Attack Size	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
1	Out-Degree	Betweenness	Out-Degree	Out-Degree
2	Fragmentation	Out-Degree	Out, Frag & Betw	Out-Degree
3	Fragmentation	Out-Degree	Out & Frag	Out-Degree
4	Fragmentation	Out-degree	Out-Degree	Fragmentation

4.2.2. Clustering

Clustering examines the extent to which nodes with a common neighbour are connected to one another. More specifically, it calculates the average density of neighbourhoods of websites. Reducing clustering eliminates those websites that might be in ideal positions to

replace other websites while also potentially preventing the formation of tight-knit communities. Results for this measure were similar to those for density: out-degree attacks are favoured in general, while fragmentation attacks become progressively more powerful. However, in certain cases, betweenness attacks are the most effective disruption strategy.

For small attacks, hub attacks tended to be the most successful strategy (see Table 8). In Blog-A, clustering fell from 0.424 to 0.365 (a 13.92% reduction) following the out-degree attack. For Website-B, this attack reduced clustering from 0.108 to 0.280, producing a 36.5% reduction. For Website-A, the in-degree attack was instead the most effective: clustering fell from 0.442 to 0.415 for a 6.1% reduction in the measure. Given that, in this network, out-degree and in-degree centralization were relatively similar, it is not surprising that nodes with high numbers of in-degree ties similarly contributed to the structure of the network in certain ways (i.e., clustering). For Blog-B, the betweenness attack was the most effective; clustering fell from 0.525 to 0.348, a 33.71% reduction in the measure. This re-iterates the notion that nodes in certain networks may act as both hubs and bridges and effectively undermine both the overall density of a network and the density of the neighbourhoods within it.

As was the case with density, random attacks produced little disruption to clustering (less than 0.10%) and in the case of Website-B, acted to increase clustering. The danger of random attacks is that they may actually target websites of little importance to the network (and potentially to users), thus eliminating websites that are mostly “white noise” or of little relevance to users. In a sense, these attacks might “clean up” the network for such users, making their access of relevant information in the network more efficient; less time is wasted visiting or exploring unimportant websites.

In sum, clustering was most effectively reduced by hub attacks in the first round of attacks. This was also the case in the second and third wave of attacks (see Tables 9 and 10). However, in certain cases, betweenness attacks were similarly effective. In addition, fragmentation attacks were favoured by one network. Following the removal of 20% of the nodes, clustering in Blog-A fell by 18.4% after an out-degree attack. For Website-B, clustering decreased by 49.4% following the same kind of attack. For Blog-B, out-degree and betweenness attacks resulted in similar disruption. Clustering experienced a 37.9% reduction from the out-degree attack and a 36.8% reduction from the betweenness attack. Finally, for Website-A, clustering fell by 9.5% following both betweenness and fragmentation attacks.

In the third wave of attacks, hub attacks remained the most successful disruption strategy (see Table 10). The exception was for Website-B, where the fragmentation attack was slightly more effective. In Blog-A, the out-degree attack produced a 36.6% reduction in clustering. For Blog-B, this attack was also the most effective attack strategy, producing a 43.6% reduction in clustering. For Website-A, out-degree attacks produced a 52.0% reduction in clustering. However, for Website-B, fragmentation was the relatively more effective strategy; clustering fell by 61.2% compared to 58.5% for the out-degree attack.

While different attack strategies emerged as effective for different networks in the fourth attack wave, both out-degree and fragmentation attacks tended to be the most successful disruption strategies (see Table 11). The exception was for Website-B, in which the betweenness attack was the most effective. For Blog-A, fragmentation attacks were the most effective, closely followed by out-degree attacks. The fragmentation attack produced a 53.3% reduction clustering while the out-degree attacks produced a 52.8% reduction in the measure. For Blog-B, out-degree attacks were the most effective; clustering dropped by 45.1%. For Website-A, three strategies

were equally effective: fragmentation, out-degree and in-degree attacks all reduced clustering to 0. Only pairs of attached nodes remained. For Website-B, the most successful attack strategy was that of betweenness, which reduced clustering by 68.9%.

For the clustering measure, greater variety existed in terms of which strategies were most effective. Hub attacks tended to be effective throughout, with fragmentation measures becoming more effective with larger attacks. However, betweenness attacks could be effective in both early and later attacks for different networks. It appears that particular differences in the network structure are more important for the measure of clustering than for density in determining which strategies are most effective at various attack sizes. While it is safe to select hub attacks for early attacks and fragmentation attacks for much larger attacks, knowing the structure of the network and the manner in which the network is clustered may provide for more nuanced attack methods (for instance, it may identify cases where betweenness attacks are more advantageous). Table 12 summarizes the most effective attack strategies for each network at each attack size.

Table 8 – Overall Clustering Coefficient after 1st Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation	0.441 (↑4.01%)	0.390 (↓25.71%)	0.514 (↑16.289%)	0.304 (↓31.1%)
Betweenness	0.439 (↑3.45%)	0.348 (↓33.71%)	0.438 (↓0.09%)	0.297 (↓32.7%)
Degree Out	0.365 (↓13.92%)	0.387 (↓26.29%)	0.429 (↓2.94%)	0.280 (↓36.5%)
In	0.495 (↑16.75%)	0.495 (↓5.71%)	0.415 (↓6.1%)	0.369 (↓16.3%)
Random	0.422 (↓0.5%)	0.522 (↓0.6%)	0.438 (↓0.9%)	0.453 (↑2.7%)

Table 9 - Overall Clustering Coefficient after 2nd Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	0.412 (↓2.8%)	0.340 (↓35.2%)	0.400 (↓9.5%)	0.284 (↓35.6%)
Betweenness	0.421 (↓0.7%)	0.332 (↓36.8%)	0.400 (↓9.5%)	0.291 (↓34.0%)
Degree Out	0.346 (↓18.4%)	0.326 (↓37.9%)	0.462 (↑4.5%)	0.223 (↓49.4%)
Degree In	0.388 (↓8.5%)	0.470 (↓10.5%)	0.431 (↓2.5%)	0.332 (↓24.7%)
Random	0.434 (↑2.4%)	0.526 (↑0.2%)	0.492 (↑11.3%)	0.461 (↑4.5%)

Table 10 – Overall Clustering Coefficient after 3rd Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	0.325 (↓23.3%)	0.320 (↓39.0%)	0.319 (↓27.8%)	0.171 (↓61.2%)
Betweenness	0.398 (↓6.1%)	0.322 (↓38.7%)	0.403 (↓8.8%)	0.264 (↓40.1%)
Degree Out	0.269 (↓36.6%)	0.296 (↓43.6%)	0.212 (↓52.0%)	0.183 (↓58.5%)
Degree In	0.432 (↑1.9%)	0.472 (↓10.1%)	0.482 (↑9.0%)	0.367 (↓16.8%)
Random	0.434 (↑2.4%)	0.537 (↑2.3%)	0.456 (↑3.2%)	0.464 (↑5.2%)

Table 11 – Overall Clustering Coefficient after 4th Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	0.198 (↓53.3%)	0.322 (↓38.7%)	0.000 (↓100%)	0.183 (↓58.5%)
Betweenness	0.411 (↓3.1%)	0.321 (↓38.9%)	0.500 (↑13.1%)	0.137 (↓68.9%)

Degree Out	0.200 (↓52.8%)	0.288 (↓45.1%)	0.000 (↓100%)	0.167 (↓62.1%)
In	0.330 (↓22.2%)	0.426 (↓18.9%)	0.000 (↓100%)	0.369 (↓16.3%)
Random	0.424 (0%)	0.538 (↑2.5%)	0.497 (↑12.4%)	0.446 (↑1.1%)

Table 12 – Summary of the Most Effective Attack Strategies Against Clustering

Attack Size	Blog-A	Blog-B	Website-A	Website-B
1	Out-Degree	Betweenness	In-Degree	Out-Degree
2	Out-Degree	Out-Degree	Frag & Betw	Out-Degree
3	Out-Degree	Out-Degree	Out-Degree	Fragmentation
4	Out-Degree	Out-Degree	Out, Frag & Betw	Betweenness

4.2.3. Reachability

Reachability examines whether a path exists between any two nodes in a network. In a network with high reachability, users can travel from one site to almost any other, and when one path is removed, alternative ones remain in place. Reducing reachability eliminates routes between websites and may either isolate certain websites or increase the path lengths between them. In order to reduce reachability, the best strategies across networks were fragmentation and/or betweenness attacks.

For Website-B, when 10% of the nodes are removed, the fragmentation attack was the most effective (see Table 13). It acted to reduce reachability from 22059 to 8721 (a 60.5% reduction). For Blog-A, both fragmentation and betweenness attacks had similar effects. The fragmentation was slightly more effective, reducing reachability from 3054 to 698 for a 77.1% reduction. Betweenness attacks managed a similar 76.8% reduction in reachability. For Website-A and Blog-B, betweenness attacks produced the greatest reduction in reachability. For, Website-A reachability fell from 1021 to 9801 (a 78.5%% reduction) and in Blog-B, from 20145 to 220 (a 51.3% reduction).

Random attacks were comparatively ineffective. However, for Blog-B and Website-A, random attacks performed almost as well as some of the less effective targeted attacks. In fact, in Website-A, the random attack is more successful than the in-degree attack (reducing density by 62.2% compared to 55.4% for the in-degree attack). This indicates a) that the benefits of targeted attacks are less significant for the measure of reachability than for other measures, and b) random attacks can occasionally target important nodes by chance.

Betweenness attacks target those websites that act as bridges. When these bridges are severed, distances between websites increase as access points and short-cuts no longer remain between nodes. Conversely, the measure of fragmentation seeks to isolate the largest number of websites. By removing websites that act to preserve those pathways that connect the most websites to the network, parts of the networks become more isolated or important bridges may have become disconnected from the network. Through both of these attack strategies, network reachability effectively decreases.

Betweenness and fragmentation attacks were thus the most effective disruption strategies for small attacks. Both measures continue to be similarly effective for larger attacks, with fragmentation attacks become more effective than betweenness attacks when larger amounts of nodes are removed (see Tables 14-16).

In the second wave of attack, fragmentation attacks were slightly more successful than betweenness attacks at disrupting reachability attack for Blog-B, they produced the same disruption results for Website-A, and betweenness attack were slightly more effective for the remaining networks (Website-B and Blog-B) (see Table 14). For Blog-A, reachability fell by 91.6% following a fragmentation attack, and by 91.3% following a betweenness attack. For

Blog-B, fragmentation attacks produced a 90.3% reduction in reachability while betweenness attacks produced a 91.5% reduction. For Website-A, fragmentation and betweenness attacks were equally effective: reachability fell by 91.7% for both attacks. Finally, for Website-B, the betweenness attacks most effectively reduced reachability, causing a 91.5% reduction in the measure.

Betweenness and fragmentation attacks were similarly effective strategies in the third attack wave as well (see Table 15). For Blog-A, both fragmentation and betweenness attacks produced the greatest reduction in reachability. Reachability fell by 96.6% following a fragmentation attack and by 96.2% following a betweenness attack. For Website-A, reachability fell by 97.2% following both fragmentation and betweenness attacks. For Blog-B, fragmentation and betweenness attacks were also somewhat similar, with an 88.5% reduction after a betweenness attack and an 87.5% reduction following a fragmentation attack. For Website-B, fragmentation attacks alone were the most effective (rather than the betweenness attack in the 2nd wave), reducing reachability by 97.3%.

For the last attack wave, fragmentation and betweenness attacks continued to produce close amounts of network disruption (see Table 16). The exception was for Website-A, in which an out-degree attack became the most effective attack strategy (by very little). For Blog-A, the fragmentation attack produced a 98.7% reduction in the measure while the betweenness attack produced a 98.5% reduction. For Blog-B, reachability fell by a 97.2% from both the fragmentation and the betweenness attack. For Website-A, out-degree attacks were the most effective strategy, reducing reachability by 98.8%, closely followed by fragmentation attacks, which reduced the measure by 98.7%. Finally, for Website-B, the fragmentation attack produced a 99.1% reduction in reachability while the betweenness attack produced a 98.8% reduction.

For the other outcome measures (density and clustering), hub attacks were effective strategies particularly for small to mid-sized attacks, while fragmentation attacks became increasingly effective for larger attacks. In the case of reachability, both fragmentation and betweenness attacks tended to produce similar results across attack sizes, although fragmentation attacks did become slightly more effective in the last attack wave. Fragmentation attacks remove nodes so as to maximize the isolation of other nodes, whereas betweenness attacks target those nodes that act as a bridge between the greatest number of other nodes. Both of these measures act to eliminate pathways between certain parts of the network, and so it is not surprising that both reduce network reachability in similar ways. However, betweenness attacks will reach a certain point where all bridges have been removed, much like there becomes a certain point in a network when all the prominent hubs have been removed, and yet clusters of nodes will still remain. In this case, fragmentation attacks continue to identify nodes that can be removed to more fully disrupt the network. As such, while both attack strategies produce fairly similar results when up to 40% of the nodes are removed, if the attacks are extended any further, fragmentation attacks may become preferable. A summary of the most effective attacks can be found in table 17.

Table 13 – Reachability after 1st Attack

	Network			
Measure	**Blog-A**	**Blog-B**	**Website-A**	**Website-B**
Fragmentation	698 (↓77.1%)	15186 (↓24.6%)	241 (↓76.4%)	8721 (↓60.5%)
Betweenness	709 (↓76.8%)	9801 (↓51.3%)	220 (↓78.5%)	9801 (↓55.6%)
Degree Out	849 (↓72.2%)	15330 (↓23.9%)	332 (↓67.5%)	14842 (↓32.7%)
In	1224 (↓59.9%)	16112 (↓20.0%)	455 (↓55.4%)	12270 (↓44.4%)
Random	2417 (↓20.9%)	16673 (↓17.2%)	386 (↓62.2%)	17837 (↓19.1%)

Table 14 – Reachability after 2nd Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	256 (↓91.6%)	5607 (↓72.2%)	85 (↓91.7%)	2147 (↓90.3%)
Betweenness	267 (↓91.3%)	5719 (↓71.6%)	85 (↓91.7%)	1875 (↓91.5%)
Degree Out	417 (↓86.3%)	8438 (↓58.1%)	151 (↓85.2%)	8798 (↓60.1%)
In	510 (↓83.3%)	12296 (↓39.0%)	99 (↓90.3%)	9798 (↓55.6%)
Random	1790 (↓41.4%)	13433 (↓33.3%)	290 (↓71.6%)	14067 (↓36.2%)

Table 15 – Reachability after 3rd Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	104 (↓96.6%)	2515 (↓87.5%)	29 (↓97.2%)	593 (↓97.3%)
Betweenness	116 (↓96.2%)	2322 (↓88.5%)	29 (↓97.2%)	5941 (↓73.1%)
Degree Out	198 (↓93.5%)	4116 (↓79.6%)	36 (↓96.5%)	5941 (↓73.1%)
In	328 (↓89.3%)	8915 (↓55.7%)	37 (↓96.4%)	9102 (↓58.7%)
Random	1275 (↓58.3%)	10514 (↓47.8%)	191 (↓81.3%)	10855 (↓50.8%)

Table 16 – Reachability after 4th Attack

Measure	Network Blog-A	Blog-B	Website-A	Website-B
Fragmentation	39 (↓98.7%)	566 (↓97.2%)	13 (↓98.7%)	191 (↓99.1%)
Betweenness	47 (↓98.5%)	567 (↓97.2%)	17 (↓98.3%)	192 (↓99.1%)

Degree		91 (↓97.0%)	1711 (↓91.5%)	12 (↓98.8%)	274 (↓98.8%)
	Out				
	In	161 (↓94.7%)	3711 (↓81.6%)	18 (↓98.2%)	3331 (↓84.9%)
Random		954 (↓68.8%)	7668 (↓61.9%)	132 (↓87.1%)	7062 (↓68.0%)

Table 17 – Summary of the Most Effective Attack Strategies Against Reachability

Attack Size	Blog-A	Blog-B	Website-A	Website-B
1	Fragmentation	Betweenness	Betweenness	Fragmentation
2	Fragmentation	Fragmentation	Frag & Betw	Betweenness
3	Fragmentation	Betweenness	Frag & Betw	Fragmentation
4	Fragmentation	Frag & Betw	Out & Frag	Fragmentation

4.3. Bivariate Analysis of the Targeted Attack Strategies

While certain attack strategies emerged as most effective against particular outcome measures, the differences in the effectiveness of various attack strategies were generally quite small. The amount of disruption produced by the targeted attacks often differed by 10% or less between the strategies. As previously noted, certain nodes in a network tend to occupy multiple positions within a network; for instance, certain nodes may act as both prominent hubs and bridges. To explore the extent of overlap between network positions, a bivariate analysis was conducted; this examined the correlation between each of the targeted attacks for all networks. The outputs for this analysis can be found in Tables 18-21. In all cases save one, the relationships between each attack strategy was significant ($p < 0.05$), and the strength of the association generally ranged from moderate to strong. In particular, the correlation between bridge and fragmentation attacks tended to be relatively high for all networks. For instance, in Website-A, $r = 0.977$ for the bridge and fragmentation measures, with $p < 0.01$. This indicates that certain nodes are not only important bridges that connect other nodes to each other, but they

are also bridges that connect nodes to the network in general; without such nodes, not only would certain shortcuts in a network be eliminated, but access to particular nodes would be as well. Significant associations ($p < 0.01$) between the measures of betweenness and fragmentation also appeared for the other networks (Website-B, $r = 0.761$; Blog-A, $r = 0.929$; and, Blog-B, $r = 0.696$). In addition, for Blog-B, the correlation between out-degree attacks and fragmentation attacks was quite high ($r = 0.815$, $p < 0.01$). This strong association was represented in the finding that both out-degree and bridge attacks tended to produce similar amounts of disruption against network density and clustering across attack sizes. The correlations between the targeted attack strategies employed in this project indicate that nodes often occupy several positions in a network. This suggests that, in the future, such attack strategies might be combined into a single strategy designed to identify nodes that are important to a network in multiple ways.

Table 18 – Bivariate Analysis of Targeted Attack Strategies for Blog-A

Correlations

		OutDegree	InDegree	Between	Frag
OutDegree	Pearson Correlation	1	.236*	.357**	.415**
	Sig. (2-tailed)		.013	.000	.000
	N	111	111	111	111
InDegree	Pearson Correlation	.236*	1	.429**	.382**
	Sig. (2-tailed)	.013		.000	.000
	N	111	111	111	111
Between	Pearson Correlation	.357**	.429**	1	.929**
	Sig. (2-tailed)	.000	.000		.000
	N	111	111	111	111
Frag	Pearson Correlation	.415**	.382**	.929**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	111	111	111	111

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 19 – Bivariate Analysis of Targeted Attack Strategies for Blog-B

Correlations

		OutDegree	InDegree	Between	Frag
OutDegree	Pearson Correlation	1	.403**	.815**	.793**
	Sig. (2-tailed)		.000	.000	.000
	N	163	163	163	163
InDegree	Pearson Correlation	.403**	1	.552**	.504**
	Sig. (2-tailed)	.000		.000	.000
	N	163	163	163	163
Between	Pearson Correlation	.815**	.552**	1	.696**
	Sig. (2-tailed)	.000	.000		.000
	N	163	163	163	163
Frag	Pearson Correlation	.793**	.504**	.696**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	163	163	163	163

**. Correlation is significant at the 0.01 level (2-tailed).

Table 20 – Bivariate Analysis of Targeted Attack Strategies for Website-A

Correlations

		OutDegree	InDegree	Between	Frag
OutDegree	Pearson Correlation	1	.708**	.471**	.559**
	Sig. (2-tailed)		.000	.001	.000
	N	46	46	46	46
InDegree	Pearson Correlation	.708**	1	.268	.340*
	Sig. (2-tailed)	.000		.071	.021
	N	46	46	46	46
Between	Pearson Correlation	.471**	.268	1	.977**
	Sig. (2-tailed)	.001	.071		.000
	N	46	46	46	46
Frag	Pearson Correlation	.559**	.340*	.977**	1
	Sig. (2-tailed)	.000	.021	.000	
	N	46	46	46	46

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 21 – Bivariate Analysis of Targeted Attack Strategies for Website-B

Correlations

		OutDegree	InDegree	Between	Frag
OutDegree	Pearson Correlation	1	.412**	.582**	.405**
	Sig. (2-tailed)		.000	.000	.000
	N	162	162	162	162
InDegree	Pearson Correlation	.412**	1	.548**	.454**
	Sig. (2-tailed)	.000		.000	.000
	N	162	162	162	162
Between	Pearson Correlation	.582**	.548**	1	.761**
	Sig. (2-tailed)	.000	.000		.000
	N	162	162	162	162
Frag	Pearson Correlation	.405**	.454**	.761**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	162	162	162	162

**. Correlation is significant at the 0.01 level (2-tailed).

4.4. The Impact of Attack Sizes

Four progressively larger attacks were conducted against each network. However, with limited law enforcement funding, the issue of diminishing returns for larger attacks becomes a concern. It then becomes important to examine such questions as whether disruption to a network is doubled when the attack size is doubled. It was found that larger attacks against the measure of reachability produced the smallest increasing increments in disruption. However, for density and clustering, the amount of disruption for each attack was generally consistent, and network disruption would often double only between the 1st and 3rd attack. Table 22 provides details on the amount of network disruption by attack size.

For density, diminishing returns were evident as attacks were extended; nevertheless, for each attack, network disruption tended to increase between 10% to 20% depending on the

network (slowly decreasing as attacks grew larger). Doubling the attacks did not double the amount of disruption; this occurred only between the 1st and 3rd attack, with network disruption increasing around 50% for most networks (40% for Website-B). In terms of clustering, disruption would sometimes more than double between the 1st and 3rd attack and the 2nd and 4th (i.e., for Website-A), while it would sometimes increase by around 5% or 10% for each attack (i.e., for Blog-B and Website-B), or between 5% to 10% for each attack (i.e., for Blog-A). This indicates that for some networks, the returns can increase tremendously when an attack is doubled, while for other networks, the return is fairly consistent for each attack. Finally, for reachability, diminishing returns were more evident for later attacks: for all networks, the amount of disruption increased by 20%-30% between the first and second attack. Between the second and third attack, disruption increased instead by around 10%. For the last attack, disruption increased only by around 2% from the 3rd attack for most networks.

These findings indicate that often, if network disruption is sought to be doubled from the first attack, stopping once 30% of the nodes in a network have been removed is sufficient. While the increase in disruption for outcome measures such as density and clustering tend to be consistent for each attack, diminishing returns become evident as attacks are extended. This is particularly the case for reachability, in which very few increases to disruption appear for larger attacks.

Table 22 – Amount of Network Disruption by Attack Size

Attack Size	Blog-A	Blog-B	Website-A	Website-B
Density				
10%	38.7%	38.2%	31.03%	45.62%
20%	56.7% (↑18%)*	61.0% (↑23%)	49.2% (↑18%)	62.7% (↑18%)
30%	73.7% (↑17%)	72.0% (↑11%)	65.2% (↑16%)	76.9% (↑14%)
40%	85.1% (↑11%)	79.9% (↑8%)	79.9% (↑14%)	87.4% (↑10%)
Clustering				

10%	13.92%	33.71% (↑ %)	6.1%	36.5%
20%	18.4% (↑4%)	37.9% (↑4%)	9.5% (↑4%)	49.4% (↑12%)
30%	36.6% (↑19%)	43.6% (↑6%)	52.0% (↑42%)	61.2% (↑12%)
40%	53.3% (↑16%)	45.1% (↑1%)	100% (↑48%)	68.9% (↑8%)
Reachability				
10%	77.1%	51.3%	78.5%	60.5%
20%	91.6% (↑15%)	72.2% (↑21%)	91.7% (↑13%)	91.5% (↑31%)
30%	96.6% (↑5%)	88.5% (↑17%)	97.2% (↑5%)	97.3% (↑5%)
40%	98.7% (↑2%)	97.2% (↑8%)	98.8% (↑2%)	99.1% (↑2%)

*Denotes the increase in the percentage of disruption between attack sizes (e.g., 10-20%; 20-30%; and 30-40%)

4.5. The Impact of Network Type and Size

The effectiveness of attack strategies was also examined against network type and size. It was found that particular attack strategies were not favoured by certain networks, nor did certain disruption strategies align themselves to networks of particular sizes. This is likely due to the fact that the network structures did not vary greatly between the blog-seed and website-seed networks. Rather, all networks were somewhat different, while sharing basic characteristics such as power-law distributions and small path lengths. This means that overall, certain strategies were more or less effective across network type, although differences in network structures produced slight differences in terms of which disruption measure was favoured for particular outcome measures.

4.5.1. The Impact of Network Type

Two findings suggested that network type contributed little to the differences in strategy effectiveness between networks. First, networks of the same type (blog-seed and site-seed) sometimes differed in terms of their vulnerability to particular strategies, and second, networks of different types sometimes shared vulnerabilities to the same strategy. Table 23 provides a summary of the most effective strategies for each network. When examining the outcome

measure of density, Blog-A and Blog-B differ from one another in terms of the most effective strategy: Blog-A favours fragmentation attacks while Blog-B is more vulnerable to out-degree attacks. Conversely, Blog-B is similar to Website-B insofar as both are disrupted mainly through out-degree attacks. Because of Website-A's structure, many of the same nodes were identified for the different attack strategies, and as such, certain attacks produced the same results. In this way, Website-A shares common attack strategies with all networks, but is unique in terms of favouring multiple attack strategies.

For clustering, Blog-A and Blog-B both favoured out-degree attacks. Website-A, like the blog-seed networks, also favoured out-degree attacks. However, the site-seed networks were also similar in terms of being vulnerable to fragmentation and betweenness attacks. This difference in effectiveness between network types may be due to the fact that the website-seed networks had very similar clustering, while the clustering for blog-seed networks varied more.

Finally, for reachability, Blog-A and Website-B favoured fragmentation attacks while betweenness attacks featured prominently in both Blog-B and Website-A. Nevertheless, fragmentation attacks appeared effective for all networks in many instances. In addition, Blog-B and Website-A are similar in that both are susceptible to betweenness attacks across different outcome measures and attack sizes.

As such, there appears to be no clear difference in terms of which network-types favour which attack strategies; in many instances, a blog-seed network will share an effective strategy with one or two website-seed networks, and will differ from its blog-seed counterpart. However, in some instances, blog-seed networks are more similar to each other than to other networks in terms of favoured attack strategy. Overall, no clear pattern of strategy preference emerged between or within network types.

Table 23 – Summary of Effective Disruption Strategies

Attack	Blog-A	Blog-B	Website-A	Website-B
Density				
1	Out-Degree	Betweenness	Out-Degree	Out-Degree
2	Fragmentation	Out-Degree	Out, Frag & Betw	Out-Degree
3	Fragmentation	Out-Degree	Out & Frag	Out-Degree
4	Fragmentation	Out-degree	Out-Degree	Fragmentation
Clustering				
1	Out-Degree	Betweenness	In-Degree	Out-Degree
2	Out-Degree	Out-Degree	Frag & Betw	Out-Degree
3	Out-Degree	Out-Degree	Out-Degree	Fragmentation
4	Out-Degree	Out-Degree	Out, Frag & Betw	Betweenness
Reachability				
1	Fragmentation	Betweenness	Betweenness	Fragmentation
2	Fragmentation	Fragmentation	Frag & Betw	Betweenness
3	Fragmentation	Betweenness	Frag & Betw	Fragmentation
4	Fragmentation	Frag & Betw	Out & Frag	Fragmentation

4.5.2. The Impact of Network Size

As with network type, no obvious pattern emerged for network size and preferred attack strategies. Blog-A and Website-A are the smaller networks (with 111 and 46 nodes respectively) while Blog-B and Website-B are the larger networks (163 and 162 nodes respectively). Website-A was vulnerable to multiple strategies for each outcome measure, while Blog-A tended to favour one strategy for the different out-come measures (fragmentation for density and reachability, and out-degree for clustering). In addition, for density, Blog-A was susceptible to fragmentation attacks, while Website-A was vulnerable mainly to out-degree attacks (although fragmentation attacks were sometimes equally effective). As such, there is no clear similarity between the small networks for effective attack strategies.

The larger networks, Blog-B and Website-B, were more similar: both tended to favour out-degree attacks for density and clustering, while acknowledging the effectiveness of betweenness attacks in certain instances. For reachability, both networks were vulnerable to

fragmentation attacks, and in some cases, betweenness attacks. However, because Website-A favours multiple attacks, the larger networks also share similarly effective attack strategies with this smaller network. Furthermore, like Blog-B and Website-B, Blog-A also favours out-degree attacks against clustering and fragmentation attacks against reachability. As such, much like with network-type, no clear pattern is evident in terms of which strategies are most effective for different network sizes. The smaller networks are dissimilar in some instances, while the large networks are similar; however, the small networks are similar to the larger ones in certain cases as well. This indicates that the structure of the network rather than its size may be more important for law enforcement to determine prior to conducting attacks.

4.6. Visualizing the Impact of Attack Strategies

The following sections use graphs to illustrate the impact of certain attacks in various ways using the different networks in this study. Figure 2 depicts the nodes that were targeted for different attack strategies to illustrate the differences and similarities. Figure 3 provides a contrast between the impact of the least effective disruption strategy and the most effective one for a network. Figure 4 demonstrates the progressively larger impact of increasingly large attacks. Finally, Figure 5 emphasizes the notion that certain networks are more easily disrupted than others.

4.6.1. Visualizing the Nodes Targeted by Various Attack Strategies

Figure 2 shows the top three websites that were targeted for the fragmentation, betweenness, and out- and in-degree attacks in Website-A. It can be seen that certain nodes were targeted for more than one attack strategy; this suggests that such websites hold a prominent

position in the network in terms of maintaining its structural integrity. For instance, the red triangle node in Figure 2 appeared among the top three targets for the out- and in-degree, betweenness, and fragmentation attack strategies. This indicates that the node is an important hub, bridge, and node connector (i.e., prevents network fragmentation). As a result, if only one node is to be removed from the network, this one would be an important target. Furthermore, the blue circle node represents a prominent hub in the network; this website was targeted in both out- and in-degree attacks. Other nodes, such as the pink and green circle ones, are strictly out- or in-degree hubs. That is, one node tends to receive links while the other tends to send them; this may potentially reflect different content in, or purpose of, the websites. In any case, the sites that are the most popular in terms of receiving links from others should receive attention from researchers, if only to identify the characteristics that set them apart. Conversely, the purple diamond shaped nodes are prime targets for both bridge and fragmentation attacks, as they are both in a position where they are important bridges which, if removed, would isolate various nodes or parts of the network. In this network, it can be seen that the hubs tend to be located in the left component of the network, while the bridge and fragmentation attacks tend to be found in the right component of the network, illustrating how certain parts of a network are structurally different than others. The main point remains that certain nodes may be targeted by more than one attack strategy (or in all), rendering them important targets if only a few nodes are to be removed from the network – a type of “key player” that are important for law enforcement agencies. However, when the attack strategies are prolonged, greater variation in the nodes targeted may appear, resulting in the differences in the effectiveness of various attack strategies.

4.6.2. The Least Effective Strategy vs. the Most Effective Strategy

Figure 3 uses Blog-A to illustrate the difference between the most effective and the least effective attack strategies. Figure 3(a) presents the original network prior to any attacks while Figure 3 (b) is Blog-A once 40% of the nodes have been randomly removed (the least effective attack). Given that almost half of the network is missing, it naturally looks a lot sparser. Although three nodes have been isolated from the network, a giant component persists. That is, the remainder of the network is still connected in some way. Figure 3(c) shows the network once 40% of the nodes have been removed according to the fragmentation attack (the most effective attack). Twenty three nodes have become isolated from the network and fourteen disconnected islands of 2-5 nodes remain. At this point, the network has effectively collapsed. Individuals seeking child pornography can no longer use one website to locate dozens of others (as in the original network or the one submitted to random attacks). Instead, they quickly reach dead ends, with websites that have only or mostly broken links. This demonstrates the importance of a social networks perspective. Without regard for the network, law enforcement agencies may be, in a manner of speaking, “randomly” targeting websites. That is, certain websites may come to their attention and be removed; such websites may or may not hold a key position in the network. A network perspective assists in ensuring that when a node is removed, it produces the maximum amount of disruption to the network possible. When a targeted strategy is adopted, a child pornography network may eventually come to look like that of Figure 3 (c), in which only a few, small, non-communicating groups of websites continue to exist. One of the key advantages of the internet –easy accessibility of information– may thus be diminished.

4.6.3. Visualizing the Progression of Attacks

Figure 4, which presents fragmentation attacks against Website-B, shows the progressive impact of the four attacks on the network. Although Website-B was chosen, the results were substantively similar for all other networks. In each attack, the network becomes sparser, with fewer connections and nodes present. In addition, a growing number of isolates are produced, as well as pairs of nodes that have become disconnected from the main component. The first attack (in which 10% of the nodes are removed) created 3 isolates and a disconnected pair or nodes; the second attack (where 20% of the nodes are eliminated) left 7 isolates and two disconnected pairs of nodes; the third attack (in which 30% of the nodes are deleted), 10 isolates remained along with the same two disconnected pairs of nodes; in the final attack (where 40% of the nodes are removed), 19 isolates were produced and three pairs of nodes were detached from the remainder of the network. At this point, most of the network connections have been eliminated, and fewer paths exist among the remaining nodes. What remains is a skeletal representation of the original network. The isolates have now become websites that are difficult to access unless an individual has the right keywords when making a search or is provided with the link by an independent source. Individuals can no longer find them by traveling (i.e., following the links) through the network. In addition, if an individual does manage to find an isolated website, it may no longer be used to locate other materials or access the remainder of the network since it is no longer attached to it. The same applies to the small islands produced by the attacks; these cannot be accessed through the larger network and they limit the information an individual may easily obtain. Moreover, even in the main component that is left, travel within it is greatly impeded. Online networks have directed links, meaning that while some websites link to others, these may not reciprocate links or link to other parts of the network. As such, someone who begins their search using a website that has mainly in-degree ties may not be able to access much of the

network. Certain areas of the network may then become very difficult to access (or inaccessible) depending on the website an individual uses as starting point or comes across while traveling through the network. That is, with fewer connections present, depending on the nature of these connections (in- or out-degree), parts of a network may become less accessible, and a greater number of dead-ends may appear. This problem is less prevalent prior to the attacks, when a network is larger and denser, as more in- and out-degree links exist for each node and multiple alternative pathways exist between websites.

4.6.4. The Impact of Similar Attacks on Different Networks

Figure 5 illustrates Website-A and Blog-B after 20% of the nodes with the highest out-degree scores have been removed. It can be seen that, for Website-A, this attack is sufficient to rupture the network into three separate islands and produce 7 isolates. In contrast, when the larger network is submitted to the same attack, the main component persists; just 5 nodes have become isolated and a cluster of four nodes has separated from the network. This demonstrates that, for certain networks (particularly ones with less density), smaller attacks may be enough to collapse the network into non-communicating parts. For denser networks, repeated attacks are necessary until the density of the network is such that it can easily be dismantled. Nevertheless, the last image shows that even with smaller attacks, the vast majority of connections can be eliminated in dense networks, rendering travel through the network more difficult. As such, before proceeding with an attack, it is important for law enforcement to determine how dense a network is, as it may determine the size of the attack and the resources required to dismantle the network.

Figure 2. Nodes Removed in Targeted Attacks

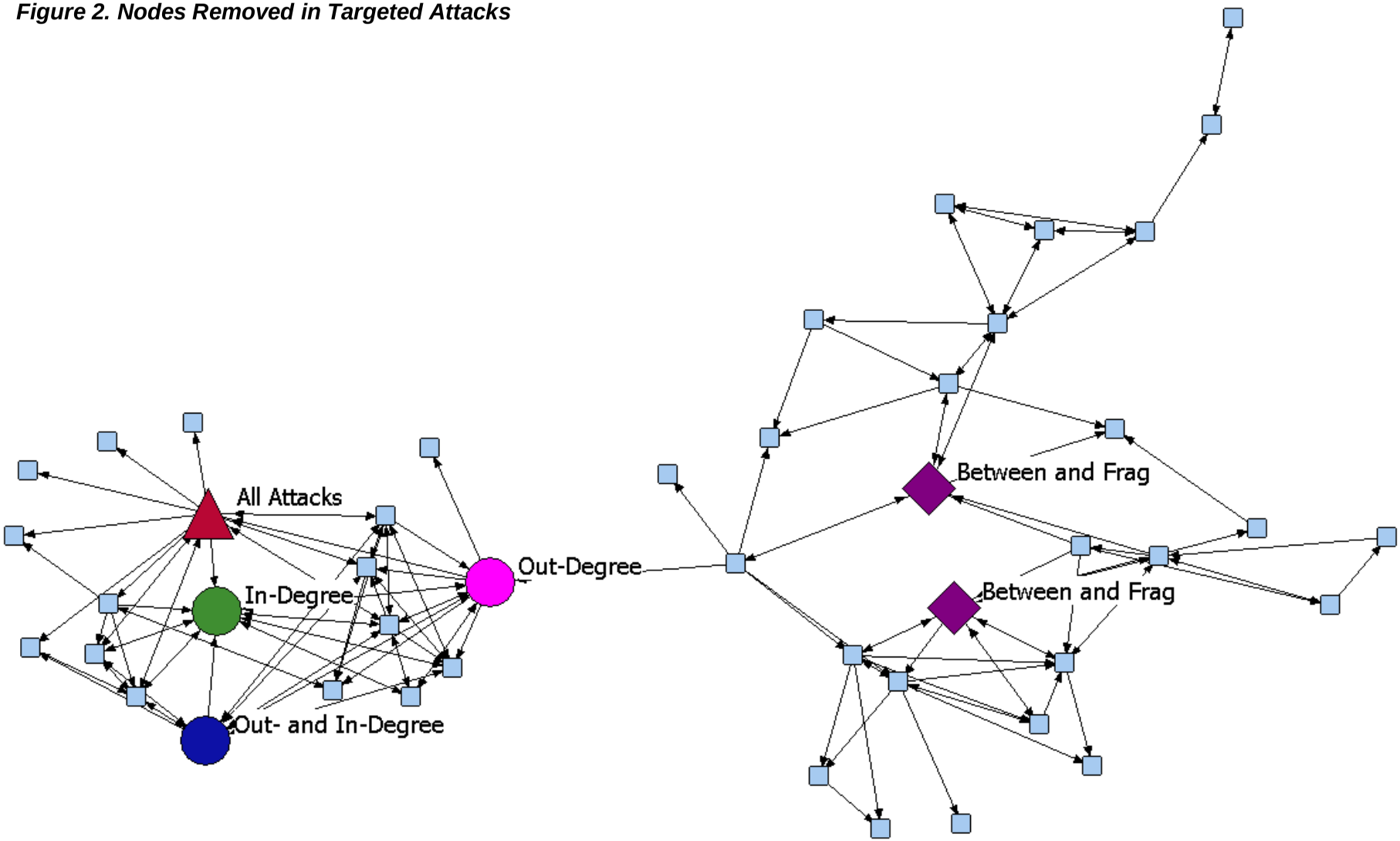


Figure 3. A Comparison of the Most and Least Effective Attack Strategies

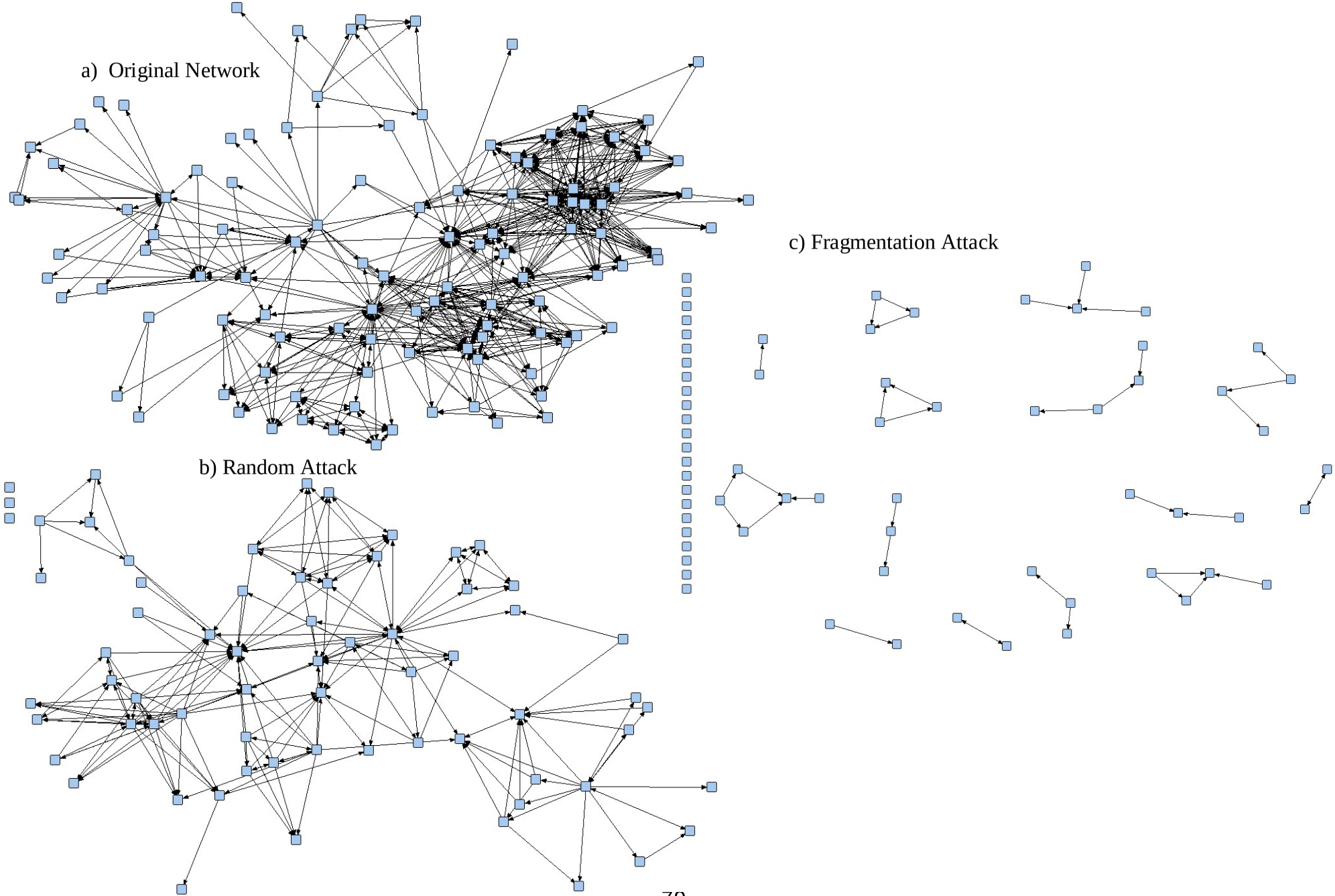
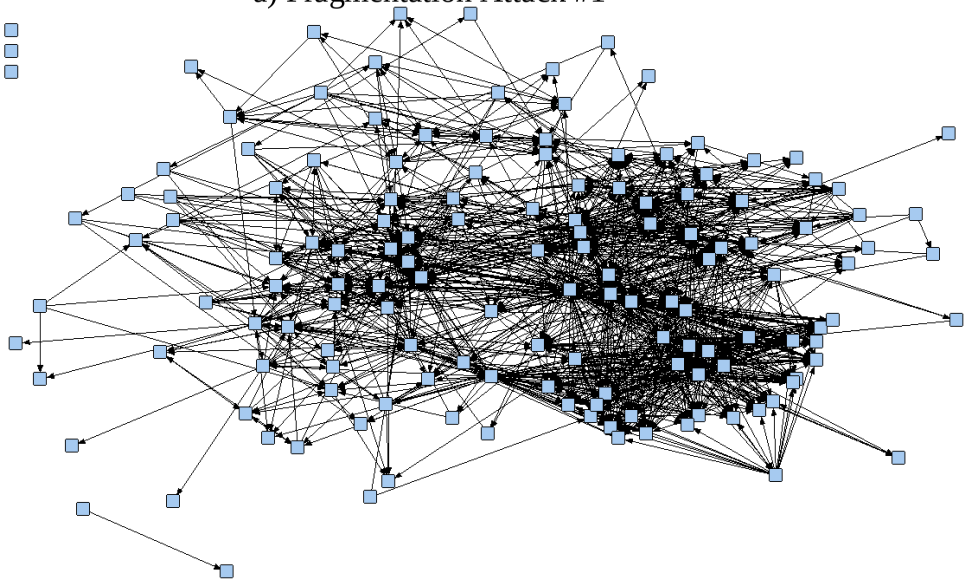
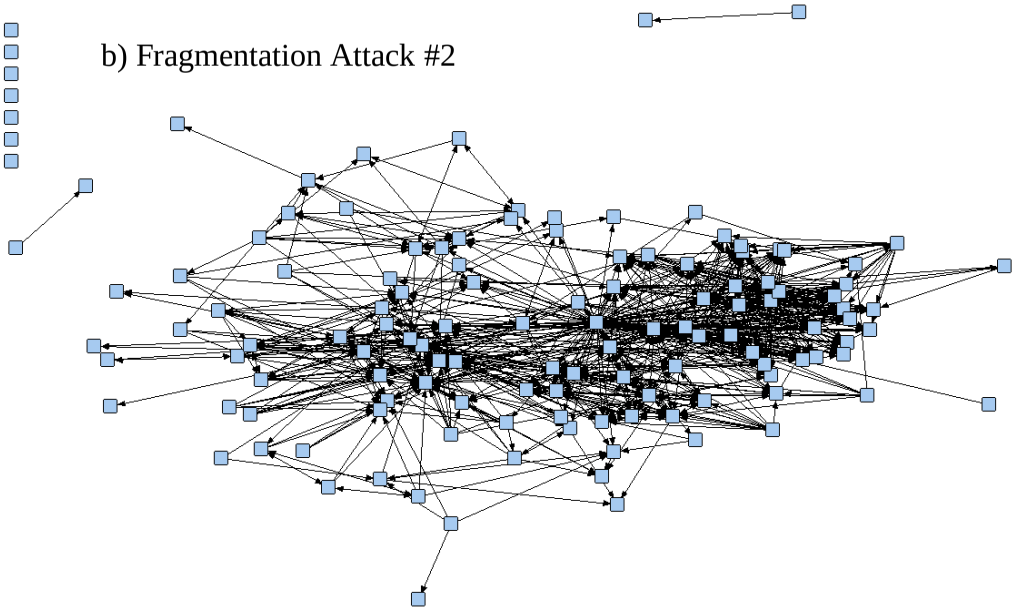


Figure 4. Progression of the Four Attack Waves

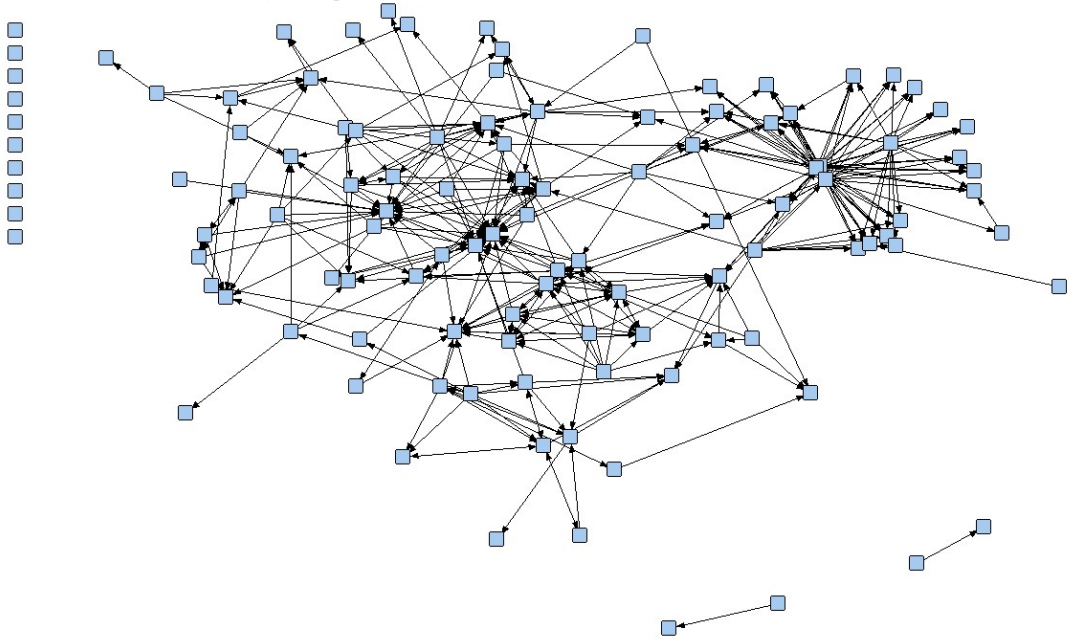
a) Fragmentation Attack #1



b) Fragmentation Attack #2



c) Fragmentation Attack #3



d) Fragmentation Attack #4

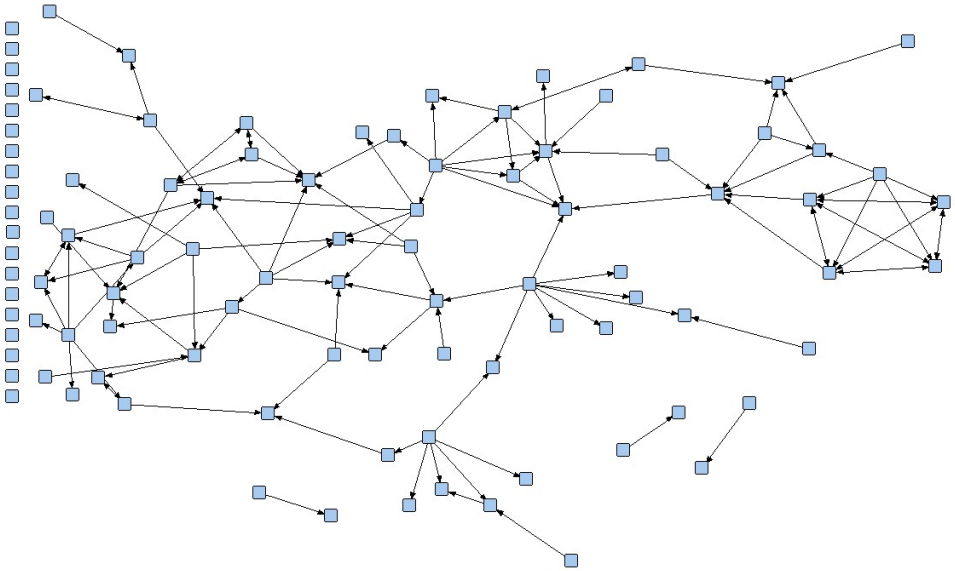
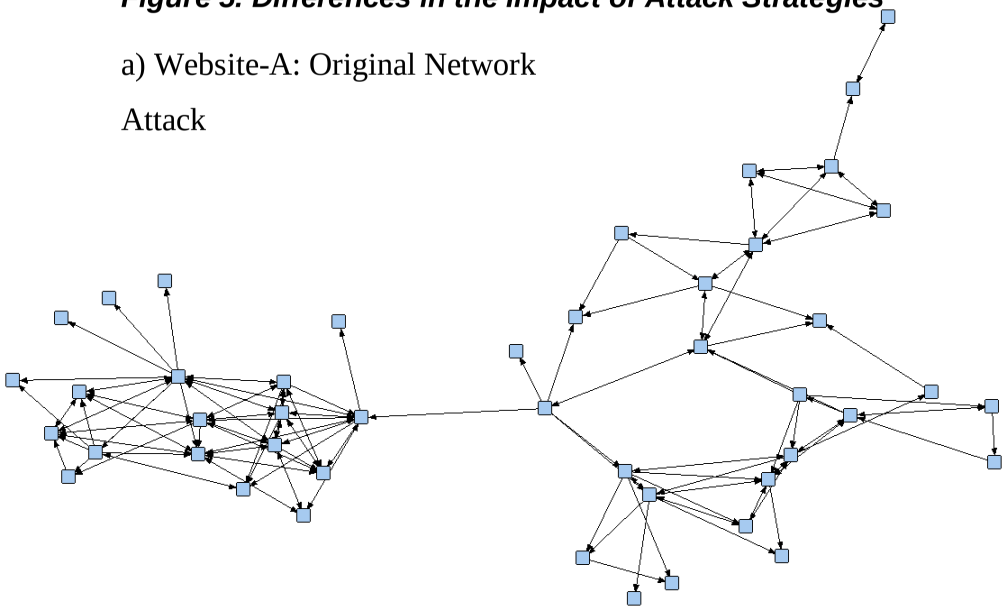
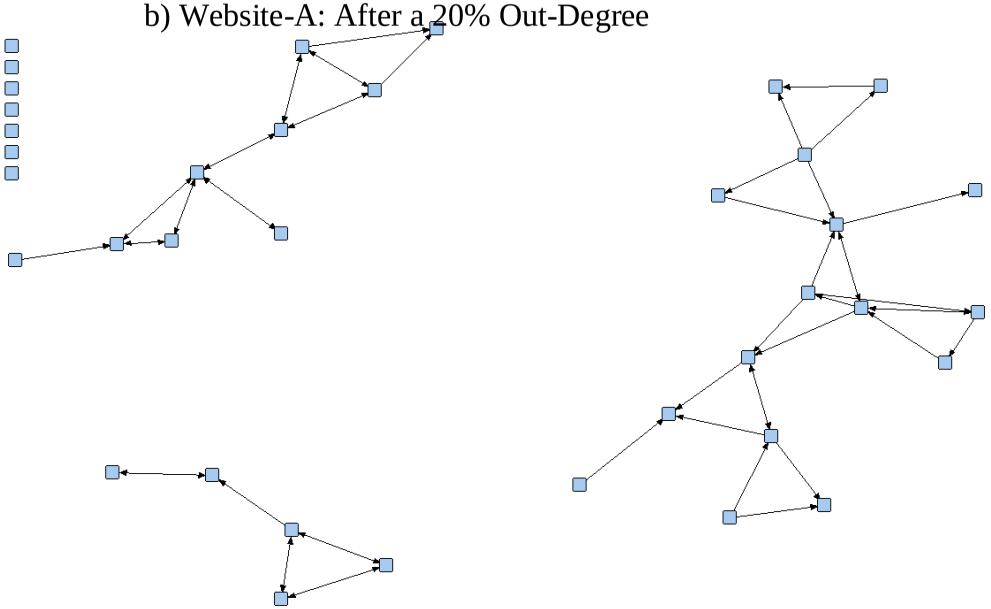


Figure 5. Differences in the Impact of Attack Strategies

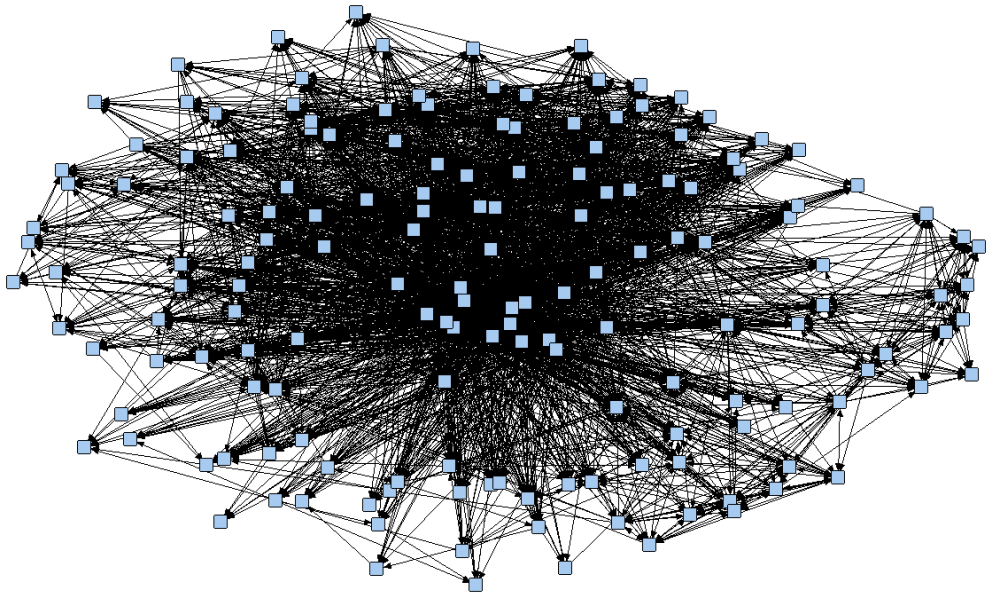
a) Website-A: Original Network



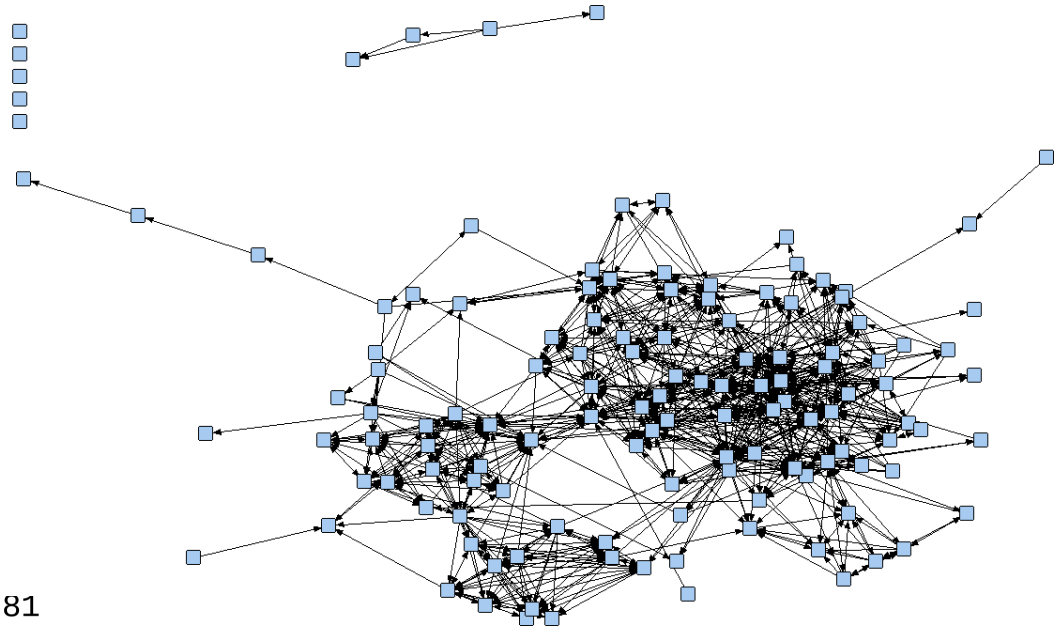
b) Website-A: After a 20% Out-Degree Attack



c) Blog-B: Original Network



d) Blog-B: After a 20% Out-Degree Attack



5. DISCUSSION

After child pornography had "been virtually stamped out" in the early 1980s, the advent of the Internet produced a resurgence of the material (Esposito, 1998, p. 3). Now, the Internet's constantly growing popularity is continuing to exacerbate and perpetuate the problem. Durkin (1997) and Esposito (1998) note that child sex offenders use the Internet in at least five ways: to access and distribute child pornography, traffic child pornography, locate child victims, communicate inappropriately with children, and communicate with other offenders. As "the most efficient pornography distribution engine ever conceived" (Frezza, 1997, p. 10), the Internet is becoming the main medium through which child pornography is transferred and is viewed as the "absolute best hunting ground (for a) pedophile" (Glaister, 1995, p. 1). It has made access to materials less risky, less time-consuming, and more convenient.

This ease of access is facilitated, in part, by the networked nature of the Internet, in which webpages link to and from each other. The Web has been described by one of its founders as "a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents" (Dern, 1994, p. 323). The web is formed of both documents and links; these links direct users to millions of other documents and resources (Dern, 1994, p. 323). Dern (1994) notes that one of the Web's goals is to allow individuals to use these links to "search, traverse, and use" information from multiple sites and in multiple formats (p. 324). With respect to child pornography, the existence of these links allows individuals to start with one child pornography website and use the network it is embedded in to locate and collect more

material. This material may then be used to groom children, lower their inhibitions, threaten or blackmail victims or other offenders, validate sexual preferences (Klein *et al.*, 2001; Tyler & Stone, 1983) or for the purposes of sexual fantasy, arousal and gratification (Lanning, 1992).

To date, law enforcement has largely ignored a network approach when addressing online child pornography (see Stanley, 2001; Wortley & Smallbone, 2006). Current efforts tend to target websites individually when they come to the attention of law enforcement in various ways, such as through investigations or complaint hotlines (Stanley, 2001; Wortley & Smallbone, 2006). This approach often overlooks a website's position in the network and the manner in which individuals might rely on its connections to other websites (see Krone, 2004). Using a network approach, child pornography websites can be targeted in ways that maximally disrupt the network and thereby limit the material an individual might access by following ties through a network. With this approach, developing a clear, systematic system for locating, prioritizing, and targeting child pornography sites is necessary to ensure that law enforcement (a) uses its resources wisely and (b) maximally reduces the ease with which offenders can access children pornography. By using the Child Exploitation Network Extractor (CENE), a web-crawler designed to extract online child pornography network, four networks were produced. Using these networks, this project had two major goals:

> (1) To examine which targeted attack strategy (degree centrality, betweenness, and fragmentation) produced the largest network disruption, as represented by specific outcome measures (density, clustering, and reachability) and,

(2) To examine whether the effectiveness of attack strategies changed depending on the attack size (small, medium, large), network size (small vs. large), and network type (blog-seed vs. site-seed).

Past research has made several observations about the topology of online networks. For instance, such networks follow a power-law distribution and are referred to as scale-free networks (Adamic & Huberman, 2000; Albert *et al.*, 1999; Barabási *et al.*, 2000; Barabási, 2003; Broder *et al.*, 2000**;** Kumar *et al.*, 1999). This indicates that large events are not necessarily atypical; rather, they have a high likelihood of occurring alongside small events. This means that while the majority of websites in a network receive and send a few links (i.e., less than 10), these coexist with hubs that may have up to a hundred or more links (Barabási, 2003). In addition, online networks demonstrate small-world characteristics: despite the large size of the Web, online networks tend to have a low average path length and high clustering (Albert *et al.*, 1999; Barabási, 2001; Broder *et al.*, 2000; Pastor- Satorras *et al.* 2001; Yook *et al.* 2001). These basic network characteristics were shared by the four online child pornography networks extracted for this project, all of which had low average path length (less than 4) and higher clustering and greater centralization (formation of hubs) than randomly generated networks of the same size and density. In this sense, online child pornography networks, despite their illegal content, are structurally similar to other online networks.

The topological qualities of online networks have specific implications for targeting nodes in order to disrupt such networks. Previous research has shown that targeted attacks against a network are more effective than random ones (Easton & Karaivanoc, 2009). This has been re-iterated by Frank *et al.* (2010), who have recognized that “removing a random website

does little to disrupt the network, however, targeting specific websites that link to a lot of other websites can result in larger impacts to the online network" (p. 6). Barabási (2003) has also indicated that, with scale-free networks, a large number of nodes can be randomly removed with relatively little impact to the structural integrity of the network. This finding was extended to this study: random attacks against the networks were ineffective compared to the targeted attacks, even the least effective targeted attacks, for all outcome measures. This makes online child pornography networks, like other online networks, resilient to random attacks. This raises concern about current law enforcement efforts: by disregarding the network, officers may, in a sense, be "randomly" targeting websites, leaving the surrounding network largely intact. This is not to suggest that law enforcement officials rarely shut down important hubs or websites with other key network positions, but they may not be doing so deliberately and thus, not consistently.

Past research has found that certain attack strategies are favoured by certain networks and network structures. For instance, Malm and Bichler (2011) found that the extent to which a network demonstrated small-world characteristics, scale-free properties, and vulnerability features affected the success of targeting strategies. When networks had high levels of clustering (thereby leaving nodes in a position to replace others), repeated attacks on multiple nodes would most successfully disrupt the network (Malm & Bichler, 2011). Scale-free networks were found to be best disrupted through hub attacks (Barabási, 2003; Medina & Hepner, 2008; Xu & Chen, 2008). Networks with high vulnerability, characterized by many actors who bridge together subgroups, were found to be susceptible to attacks that disrupt these bridges, thereby severing the flow of information (Malm & Bichler, 2011; McGloin, 2005). Borgatti (2003) introduced

another attack strategy, that of fragmentation, designed to target websites in order to produce the most network disruption by isolating the largest number of nodes possible.

Despite the prevalence and seriousness of online child pornography, its networked nature, and previously generated knowledge on network disruption strategies, little research has applied such knowledge to the disruption of online child pornography networks. The main exception is Westlake *et al.*'s (2011) work, in which they introduce the concept of network capital to the context of online child pornography, and define it as the amount of topic-relevant images, text videos, and ties to other websites that one website has. In this way, the authors combine measures of severity (harmful content) and connectivity (a website's network position). It was found that a website's severity score was not necessarily related to its connectivity. This study differs from Westlake *et al.,*'s (2011) work in that it focuses solely on a network perspective and seeks to establish effective and strictly network disruption strategies, as has been done in other areas (Barábasi, 2003; Malm & Bichler, 2011; Xu & Chen, 2008). While previous research has proposed broad strategies for disrupting networks, there is little information on how to best disrupt online child pornography networks. That is, there is a lack of information on which disruption measures to use, in which (network) context, and with which outcome measures.

To fill this gap in knowledge, four online child pornography networks were submitted to attack strategies previously identified in the literature (degree/hub, betweenness/bridge, and fragmentation attacks). Hub attacks target those nodes that receive and send the most ties; bridge attacks target those nodes that connect the most websites, and; fragmentation attacks seek to produce the greatest number of isolates in the network. The effects of these attack strategies was

subsequently calculated on network cohesion measures such as density, clustering, and reachability. These outcome measures examine the structure of networks in different ways, allowing for differences in the impact of various disruption strategies to be observed. For instance, a network's density represents the ratio of ties to nodes within a network. By reducing density, fewer ties remain in the network, impeding travel through a network, as some paths become unavailable and certain websites become isolated from the network. Clustering examines the density of neighbourhoods of websites in a network; it represents the presence of interlinked groups of websites within the network. By reducing clustering, these groups begin to disappear. This prevents the formation of tight-knit communities or, when similar websites link to one another, this may make it more difficult for an individual with a specific interest to locate groups of websites that cater to this interest. Finally, reachability examines the number of all possible pathways in a network. If one path to a website has been removed, alternative ones may remain. Reducing reachability limits accessibility in a network by removing whole pathways to various websites.

The network structure for all four networks was examined post-attack. The main findings of the project can be broadly summarized in the following manner:

1) Different attack strategies are effective for a) different outcome measures, b) different attack sizes, and c) different network structures;
2) The effectiveness of attack strategies did not appear to differ meaningfully between network type and size.

It was found, consistent with past research on scale-free networks, that hub attacks are often effective network disruption measure (Barábasi, 2003; Malm & Bichler, 2011; Xu &

Chen, 2008). However, it was found that a nuanced approach to hub attacks is warranted: in certain instances, and for certain networks, other attack strategies were more effective. For instance, if law enforcement's aim is to decrease network density and clustering, hub attacks are particularly effective. This is not the case for decreasing reachability. Furthermore, to decrease density, fragmentation attacks may also be effective for certain networks. Conversely, to reduce clustering, betweenness attacks may be more effective for certain networks. It may be that these networks have high vulnerability and as such, as Malm and Bichler (2011) and McGloin (2005) found, different strategies may be more effective against such networks. Networks with high vulnerability are those that are easily fragmented, in which attacks tend to easily produce isolates. Nevertheless, in general, when seeking to reduce the number of ties in the network overall (density) or in small groups within the network (clustering), hub attacks were found to be preferable for small to large attacks. Furthermore, the effectiveness of in- and out-degree attacks differs between networks; certain networks may have greater out-degree centralization and may thus be more vulnerable to such attacks. The common presence of hubs in a network, and the fact that they are the recipients and/or senders of a disproportionate amount of links, renders them particularly important contributors to the number of ties in a network; as such, their removal tends to maximize disruption to a network's density and clustering.

However, hub attacks are not universally effective. If law enforcement's aim is to reduce reachability, then fragmentation and betweenness attack strategies are preferable. Both tend to have a similar impact on reachability for various networks, as they each tend to isolate nodes and parts of the network from one another, making them unreachable. Fragmentation attacks may also be favoured for extremely large attacks against the networks for all three outcome

measures. This is due to the fact that the benefits of fragmentation attacks increase as larger attacks are conducted; more and more nodes are successfully isolated from the network as the attack progresses. In contrast, hub and bridge attacks gradually lose power once the major hubs and bridges have been removed. As such, for particularly large attacks (i.e., over half of the nodes), fragmentation attack may be preferable for reducing density and clustering in addition to reachability.

It was found that network type and size had no clear influence on the impact of the attack strategies, likely because they had no clear influence on the structure of the networks. For example, site-seed and blog-seed networks had structures that were both similar and different in various ways, negating any obvious differences. In addition, while larger networks were denser than smaller networks, network size did not appear to influence the effectiveness of attack measures in any distinguishing manner. However, network structure did differ between the networks and as found by Malm and Bichler (2011) and Xu and Chen (2008), it no doubt had an impact on the effectiveness of certain attack strategies. It may be important for law enforcement to analyze the structure of the network before proceeding with an attack to identify cases where less common strategies may be more effective. Future research in this area would be helpful for determining more concretely how certain network differences impact attack strategies.

It was also found that certain nodes in a network were targeted by the various attack strategies. A bivariate analysis of the various targeted attack strategies revealed significant correlations between these attack strategies, indicating overlap of network positions within certain nodes. For instance, in Website-A, a node acted as both a hub and a bridge in the network; removing it would not only eliminate a popular website, but also one that was

strategically placed to link other parts of the network. This makes such websites important to both the network structure and to users looking for child pornography. As a result, nodes that occupy multiple network positions, and are the top occupants of such positions (i.e., have the highest hub, bridge, and fragmentation scores), become “key players” within the network. When only a few websites are being removed, such key players should be prioritized in order to maximally disrupt the network in multiple ways.

Overall, this project determined that, in order to best reduce the number of ties in a networks and its neighbourhood, hub attacks are preferable. For extended attacks, when more than half of the nodes are removed, fragmentation attacks are likely to become more effective. Fragmentation attacks, along with bridge attacks, were also useful for reducing reachability in a network. The attack strategy implemented should depend on the outcome law enforcement seeks to achieve. In addition, certain nodes in a network were “key players” in virtue of occupying multiple positions in a network (e.g., a hub and a bridge), and such websites may also be prioritized for removal in cases when only a few websites are to be eliminated.

6. Conclusion

Due to the continued presence of child pornography online, and the networked manner in which the Web is designed, this study sought to apply social network analysis to the issue. In this way, different targeted attack strategies were conducted in an attempt to determine which attack strategies would be the most effective, when, for what purpose, and for what network. While the study does have various limitations, it also provides law enforcement with concrete suggestions on combating online child pornography, and can be extended in ways to generate new research on the topic.

6.1. Policy Implications

This project seeks to shift law enforcement focus from addressing child pornography websites in isolation to eliminating them with regard to the network structure they form part of. Social network analysis has shown itself to be useful for determining the structure of various illegal groups (e.g., Krebs, 2002; Morselli, 2009; Natarajan, 2006; Papachristos, 2009) and in disrupting networks (e.g., Frank *et al.,* 2010; Malm & Bichler, 2011; Xu & Chen, 2008; Westlake *et al.,* 2011). Despite its benefits, law enforcement has yet to fully embrace such an approach. While law enforcement officials should continue to take individual tips and act of these, and to conduct investigations of their own, a different, proactive approach to online child pornography is also recommended. More specifically, the following law enforcement process is recommended:

- That law enforcement employ a program similar to CENE in order to automatically extract online child pornography networks in a timely manner
- That law enforcement ascertain its goals in terms of network disruption (to reduce density and clustering or reachability)
- That the structure of the networks extracted by CENE be ascertained
- That the appropriate law enforcement strategies based on network structure and law enforcement goals be selected

A network approach is important not only because the Web is structured in this way, but also because individuals take advantage of this network structure. As previously mentioned, individuals searching for online child pornography might begin by using an Internet search engine, a website provided to them by another sources, or an old website they are familiar with. They may then expand their search for material by following the links to and from these websites. In the end, they may have amassed new child pornographic material and new sources of such material (which they can then provide to others or use as starting points for future searches). It should be noted that not all individuals search for materials in this way; it may be mainly novices who rely largely on such tactics. Nevertheless, for those who operate in this manner, the aim is to impede travel through a network, thereby decreasing the ease and amount of information an individual can access on the Internet. At present, this does not appear to be a prominent law enforcement goal. However, this project, along with research by Frank *et al.* 2010) and Westlake *et al.* (2011), emphasize the need to move in the direction of making it so.

The approach undertaken in this paper does not analyze website content; it focuses strictly on connections between websites in order to emphasize their potential relevance to law

enforcement goals. However, this is not to say that content is irrelevant to law enforcement purposes; rather, it is to suggest that other aspects of online child pornography (i.e., its networked nature) are also important to address for novices or others who search for online child exploitation materials.

While a network perspective provides law enforcement with a fresh, proactive approach to child pornography, the tools this approach uses (i.e., the web-crawler) are also beneficial in several ways. By design, the web-crawler is automatic, thereby limiting the need for human involvement. In this way, it reduces the need for law enforcement officials to make manual searches for child pornography, and frees up time that would be otherwise spent on such activities. The automated nature of the web-crawler may also reduce official's exposure to child pornography material, which may consequently reduce traumatic stress associated with the activity and the costs of subsequent therapy (Burns, Morley, Bradshaw, & Domene, 2008). The web-crawler is also beneficial in its efficiency: it has the ability to construct a network and analyze websites at a greater pace than an officer, thereby investigating more material at quicker speeds. Such time and money can then be spent elsewhere; for instance, to investigate incidents of child pornography, plan operations, and refine intervention strategies. In this way, a network approach provides law enforcement with a proactive, reliable intervention strategy that accounts for the online context of child pornography while the web-crawler used for such an approach further benefits law enforcement by increasing efficiency, reducing the amount of resources required, and decreasing officer exposure to child pornography.

Given the decentralized, and global nature of the Web, several of the child pornography websites in a network may be hosted in other countries. Esposito (1998) argues that, in order to

successfully regulate online child pornography, three steps must be taken: 1) universal standards must be established and adopted into law by all counties; 2) enforcement of these standards must be mandated at a national level, and; 3) a mechanism for global monitoring of national enforcement and a means of global enforcement must be established. The general consensus in policy recommendations appears to be that global partnerships and international law enforcement collaboration are required to deal with the issue (Wells, Finkelhor, Wolak, & Mitchell, 2007). In order to effectively dissolve networks, and shut down key player websites in other countries, it may be important for law enforcement to form partnerships with and gain the cooperation of foreign police forces. With such partnerships established, international networks can be more fully disrupted.

6.2. Limitations

Despite being helpful in terms of directing law enforcement focus, this study has several limitations. Specifically, concerns arise regarding the nature of social network analyses, the web-crawler, and policing the dynamic environment of the Internet. Some of these limitations provide suggestions for future research, where they can be addressed.

Sparrow (1991) outlines some problem areas in social network analyses. For instance, he notes that extracted networks are rarely complete and, as such, crucial elements may be missing. This is a constraint in the current project. Although the Internet allows for the development of networks with thousands of interlinked child pornography sites, the sample sizes in this study were fairly small. The smallest network involved 46 websites while the largest included 163 networks; therefore, they may represent only a mere fraction of a more complete network. While

it is possible that individuals travelling through a network may not visit 163 websites, or even 46, an individual starting at a particular point in the network may end up along a pathway, visiting various websites that were not mapped by CENE.

Nevertheless, this project corroborates other research in which general strategies (e.g., hub attacks) were found to be effective against online networks. Even if a complete network is extracted, it is likely that attack strategies like hub attacks would remain effective for particular outcome measures. This is supported by the fact that hub attacks were effective for both small and large networks, as were other attack strategies. This may be due to the fact that online networks tend to be scale-free with small-world characteristics, and this is reflected in all online networks of various sizes. For instance, even Website-A with 45 nodes contained hubs; these were simply smaller than the ones in larger networks. Moreover, while extending a network such as Website-A might have identified new hubs in the network, it is likely that the old hubs would remain important, with an extension of the network simply adding even more ties to these hubs (that is, popular hubs in a small network may remain popular in a larger network as well). As such, extracting more complete networks might increase confidence in this project's findings, without necessarily changing them.

However, limitations to the web-crawler may have inhibited the extraction of complete, accurate networks in other ways. For instance, many legal pornographic websites and forums require an account and a password to access; this is also (and perhaps especially) the case for child pornography. It is possible that some of the most relevant or severe child pornography sites were password protected. This would prevent the crawler from accessing them and as such, they (and the sites they link to) would not be included in the network. As a result, they would not

feature among the websites to be targeted and eliminated, despite their potentially important positions in the network. The impact of these websites on the structure and effectiveness of attack strategies is thus unknown. Notably, websites with passwords are not considered public; as such, accessing them would require a warrant, without which, access would become a breach of privacy. This becomes an issue beyond the scope of CENE.

Sparrow (1991) also highlights the problem of ambiguous, fuzzy boundaries. As previously noted, the sites included in the both networks were chosen on account of having 7 of 63 child pornography-related words. Yet, several of these words were not necessarily pornographic (e.g., young, love, boy, girl, etc.). This creates the possibility for false positives in the network, particularly for websites with large dictionaries of words. For instance, Website-A included several lyrics websites, which tend to contain a large variety of words. Nonetheless, it may be argued that the mere fact that these sites are linked by child pornography sites (and possibly link to them) makes them relevant to understanding and disrupting the network. They form, in some way, part of the network. Moreover, to reflect their irrelevant content, such websites may not hold particularly important network positions.

While certain websites may have been included in the network that did not contain child exploitation material, it is also possible that relevant websites were excluded. Websites that featured mainly child pornography images and videos would not have been captured under the 7 keyword requirement. In addition, some websites may not host child pornography, but they may support it, provide links to it, impart tips on avoiding police intervention, host chat rooms, and so on. Such websites may also be important to capture in the network in virtue of their relevance to individuals interested in child pornography.

There are also issues with the concept of "network type" employed in this paper. Two "types" of networks were identified for this project: those emerging from a blog-seed and those emerging from a site-seed. However, only the starting website was guaranteed to be a particular type of website. The degree to which this starting point influenced the remainder of the websites in the network is unknown. While Westlake (personal communication, March 21, 2001) noted that starting with a particular website-type tended to produce a great proportion of this type of website in the network, the number of blogs in the blog-seed networks was not ascertained, nor were the number of typical websites in the site-seed network. Undoubtedly, the networks contained both types of websites. Therefore, it may not necessarily be helpful to separate network types in this manner, as websites routinely link to a variety of other website types. Making other network type distinctions might be more useful, such as whether a network is girl- or boy-centered or image-, video-, or text-centered, as these might instead influence network structure.

Finally, Sparrow (1991) notes that while networks are dynamic and ever-changing, network analysis is static. This concern is particularly relevant in the context of the Internet, where sites are constantly appearing, changing, and disappearing. The networks in this paper captured only a snapshot of constantly evolving online networks. Some of the websites in the network may have become broken links and new ones may have appeared, potentially changing the network structure. In addition, it is possible for new websites to quickly gain popularity (Adamic & Huberman, 2000), thereby becoming important hubs. To determine whether there are trends in network growth, a longitudinal research approach might be beneficial. Furthermore, it would be important to track networks pre- and post-disruption. Once a network has been

attacked, examining how the network compensates is important. Certain attacks may be more likely than others to cause a network collapse, whereas others may be easier for a network to recover from. The manner in which a network adapts to an attack may also provide more information on the effects of various intervention strategies.

The dynamic nature of the internet also poses problems for law enforcement. Officers need to be constantly alert for the appearance of new websites. Furthermore, there is no guarantee that a recently closed website will remain that way; the host may simply move the location of the site to a different domain (or move from hosting a site to a blog). Moreover, when one website is removed, another may appear in its place, possibly fulfilling the previous website's function as a bridge or hub (which also emphasizes the importance of a post-network intervention study). Nonetheless, law enforcement interventions are certainly not futile, and to some extent, they may be aided through the use of strategies developed from static network analyses.

6.3. Future Research

In several ways, this project has extended current knowledge on online networks and effective network disruption strategies. Nonetheless, there remain many gaps in knowledge in terms of network content, network evolution, and other online resources used by those interested in child pornography. For instance, this project focused on the links between websites and their positions in the network without accounting for the content of these websites. Yet, knowing about such content may have important law enforcement implications. Websites may differ in terms of the number of images and videos posted, the frequency with which they are posted, and

the severity of the posted materials. To some extent, Westlake *et al.* (2011) did this by introducing the concept of network capital. However, it would be important to extend this work by comparing more closely the websites that hold key network positions and those with the most severe, frequent materials, those websites most frequently visited, those visited for the longest period of time, those from which the most materials are downloaded, and so on. Determining the extent of overlap with respect to websites with prominent network positions and those with the latter features might re-emphasize the important of network-guided attacks or suggest ways in which it might be modified to capture those websites that are important to users in other ways.

Future research might also explore the factors that influence network structure. For instance, it was found that the starting website in this project had no clear influence. An investigation into what qualities might affect network structure would make it easier to assign appropriate intervention strategies, as some network structures are more vulnerable to certain attack strategies than others. For instance, research might examine girl-centered and boy-centered networks. Closer examination of whether this impacts network structure might show that girl-centered networks favour one attack strategy, while boy-centered networks favour another. Network structure might also vary on other measures; some networks may contain mainly softcore or hardcore material, mostly images, videos or text, or other content. Such an investigation will extend knowledge on online child pornography networks, their structure, and the appropriate intervention strategies.

As previously noted, longitudinal research is also valuable to conduct. It is helpful not only for examining how online child pornography networks evolve over time, but also for determining how a network adapts once it has been attacked. Some network structure may be

able to adapt better than others; if this is the case, then more extensive attacks, or a second round of attacks, may be necessary. Such research may provide useful information with respect to how large an attack might have to be in order to prevent a network from easily recovering. In addition, certain attack strategies may be easier to recover from; the loss of a hub might be more important than the loss of a bridge for a network. In these ways, being able to predict how a network will recover from an attack will assist in refining attack methods against networks.

CENE can also be changed in ways that allow it to explore other Internet domains or other networks with illicit content. This project focused on a particular area of the Internet: typical websites or blogs. However, the Web also provides access to instant messaging systems, forums and newsgroups, and peer-to-peer networks. Carr (2004) found that while 42% of online sex offenders used the Web to access images, 39% used newsgroups, and 78% used Internet Chat Relay (IRC). Peer-to-peer networks are also popular for trading child pornography images (Steel, 2009). Some research has examined the network properties of peer-to-peer networks and forums (Adamic, *et al.*, 2003; Aiello *et al.*, 2000; De Laat, 2002; Ellison *et al.*, 2007; Iamnitchi *et al.*, 2002; Kaafar *et al.*, 2006; Preece & Krichmar, 2003), but no such research has been extended to the context of child pornography. As such, modifying the web-crawler to explore different Web domains would allow researchers to ascertain their properties and accordingly, develop the appropriate intervention strategies. In addition, specific tools that involve automatic scanning of IRC channels and image scanning could be generated.

Future research can also adapt CENE to extract other networks that deal with illegal or illicit content. For instance, it is currently being re-designed to extract extremist or terrorist networks. Extracting such networks can provide an idea of how large or small some of these

networks are, examine which extremist groups have ties to other groups, and determine whether such networks are vulnerable to those attack strategies identified for online child pornography networks. An adapted web-crawler might also provide opportunities to delve qualitatively into the content of such websites and, for instance, examine the manner in which such websites recruit others, seek donations, incite others to action, and provide training. Chen (2012) and his colleagues have begun research in this area by developing a semi-automatic web-crawler designed to extract networks of extremist websites.

The question remains with respect to how this project's findings might extend to other types of networks, particularly networks of individuals. Previous research has published networks of drug groups, gangs, and terrorist networks (e.g., Krebs, 2002; Morselli, 2009, 2010; Natarajan, 2006). These networks all show small-world characteristics (an average of less than 6 paths between any individual) and while they may not strictly follow a power-law distribution, all of them contain hubs. In criminal networks, those individuals who form hubs are often considered to be more susceptible to law enforcement intervention, as their network positions render them more visible (Morselli, 2009). For instance, Morselli (2010) found that members of the Hell's Angels group with high centrality were more likely to be arrested. Using a drug-smuggling and trafficking network, Morselli (2009) also found that those with higher centrality, despite placing themselves in a vulnerable position, fulfilled a "hands-on" position that assured them "strategic control of the resources [such as money] exchanged within the network at hand" (p. 60). Furthermore, those who scored high in betweenness in Morselli's (2010) Hell's Angel network were often not part of the actual group. In addition, in Krebs' (2002) 9/11 terrorist network, hubs denoted the "ring leader" of the terrorist group and the pilots crucial for the

mission. Given that these networks share some of the small-world and scale-free characteristics of online networks, some of the network disruption strategies identified in this project may remain effective. For instance, the removal of hubs found in criminal networks may be important for both reducing network density and clustering, and in some cases, eliminating important players in the network (although, in some networks, the leader may purposefully isolate themselves to avoid prosecution). However, the strategies suggested are effective only to the extent that criminal networks share the small-world and scale-free properties of the online networks examined in this study. Street gangs who form cohesive clusters without a notable hub may require other intervention strategies. In addition, these network strategies ignore the role individuals might play in the network. While centrality may indicate leadership in one network, isolation may suggest leadership in another. Future research would assist in determining how useful the attack strategies introduced in this project would be for criminal networks of individuals.

In essence, this project makes concrete suggestions for the manner in which online child pornography should be approached by introducing a strategy that accounts for, and takes advantage of, the Web's networked nature. Such an approach seeks to attack a network in ways that prevent users from traveling through it and impede their access to information. Attack strategies were derived from the literature on social network analysis, and it was found that hub attacks were particularly useful at reducing network density and clustering, while fragmentation and betweenness attacks were effective at reducing network reachability. While there are some limitations to this study, several of these can be addressed through future research (e.g., by performing longitudinal studies, examining different network types, accounting for more content measures, and perfecting CENE).

REFERENCE LIST

Adamic, L.A., & Huberman, B. A. (2000). Power-law distribution of the World Wide Web. *Science, 287,* 2115a.

Adamic, L.A., Lukose, R.M., & Huberman, B.A. (2003). Local search in unstructured networks. In S. Bornholdt and H. G. Schuster (Eds.), Handbook of Graphs and Networks: From the Genome to the Internet (pp. 295-317). Berlin: Wiley-VCH.

Adler, A. (2001). The perverse law of child pornography. *Columbia Law Review, 209,* 265-73.

Aiello, W., Chung, F., Linyuaa, L. (2000). *A random graph model for massive graphs. Proc. of the Thirty-Second Annual ACM Symposium on Theory of Computing,* Portland, Oregon, USA.

Akdeniz, Y. (2001). Governing pornography and child pornography on the Internet: The UK approach. *University of West Los Angeles Law Review, 247-275.* Retrieved from http://www.cyber-rights.org/reports/child.htm.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the World Wide Wed. *Nature, 401,* 130-131.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Attack and error tolerance of complex networks. *Nature, 406,* 378.

Ali-Hasan, N., & Adamic, L. (2007). Expressing social relationships on the blog through links and comments. *ICWSM,* Boulder CO.

Baker, W.E., & Faulkner, R.R. (1993). The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review, 58,* 837-860.

Barabási, A.-L. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. New York: Penguin Group.

Barabási, A.-L. (2001). The physics of the web. *Physics World, 14,* 33-38.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 186,* 509-512.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 186,* 509-512.

Barabási, A.-L., Albert, R., Jeong, H., & Bianconi, G. (2000). Power-law distribution of the World Wide Web. *Science, 287,* 2115a

Beech, A.R., Elliott, I.A., Birgden, A., & Findlater, D. (2008). The Internet and child sexual offending: A criminological review. *Aggression and Violent Behavior, 13,* 216-228.

Bollobás, B. (1985). *Random Graphs.* London: Academic Press.

Bollobás, B., Riordan, O., Spencer, J., & Tusnády, G. (2001). *The Degree Sequence of a Scale-Free Random Graph Process.* New York: John Wiley & Sons.

Borgatti, S. (2003). The key player problem. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers (pp.241-252).* Washington D.C.: National Academy of Science Press.

Borgatti, S. P. & Everett, M.G. (2006). A graph-theoretic perspective on centrality. *Social Networks, 28,* 466-484.

Bornholdt, S., & Ebel, H. World Wide Web scaling exponent from Simon's 1955 model. *Physical Review E, 64,* no. 035104.

Broder, A. Kumar, K., Maghoul, F., Raghavan, P., Rajagopalan, S, Stata, R., Tomkins, A, & Wiener, J. (2000). Graph structure in the web. *Computer Networks, 33,* 309-320.

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

Burt, R. (1992). *Structural holes: The social structure of competition.* Cambridge, MA: Harvard University Press.

Carr, J. (2004). Child abuse, child pornography and the internet. London: NCH.

Chase-Dunn, C, Kawano, Y., & Brewer, B. (2000). Trade globalization since 1795: Waves of integration in the world-System. *American Sociological Review, 65,*77-95.

Chen, H. (2012). Dark Web research overview. *Integrated Series in Information System, 1,* 3-18.

Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S.J., & Willinger, W. (2002). The origin of power laws in Internet topologies revisited. *In Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE Computer Society, Los Alamitos, CA.*

Cohen, R., Erez, K., ben-Avraham, D., & Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical Review Letters, 85,* 4626-4628.

Criminal Code, R.S. c. C-46 (1985). Retrieved from http://laws.justice.gc.ca/PDF/ Readability/C-46.pdf

De Laat, M. (2002) Network and content analysis in an online community discourse. In G. Stahl (Ed.), *Proceedings of Computer Support for Collaborative Learning (CSCL) 2002 Conference*, Jan. 7-11, Boulder, CO. Mahwah, NJ: Lawrence Erlbaum (pp. 625-626).

Dern, D.P. (1994). *The Internet Guide for New Users.* NY: McGraw-Hill, Inc..

Durkin, K.F. (1997). Misuse of the Internet by pedophiles: Implications for law enforcement and probation practice. *Federal Probation, 61,* 14-18.

Durrett, R. (2007). *Random graph dynamics.* Cambridge: Cambridge University Press.

Easton, S.T., & Karaivanov, A.K. (2009). Understanding optimal criminal networks. *Global Crime, 10,* 41-65.

Ellison, N.B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook ‘‘friends:’’ Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication, 12,* 1143-1168.

Engeler, E. (2009, September 16). UN expert: Child porn on internet increases. The Associated Press. Retrieved from http://www.msnbc.msn.com/id/32880508/ ns/technology_and_science-security

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrencen, 6,* 290-297.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5,* 17-61.

Erdős, P., & Rényi, A. (1961). On the strength of connectedness of a random graph. *Acta Mathematica Hungarica, 12,* 261-267.

Esposito, L.C. (1998). Regulating the Internet: The new battle against child pornography. *Case Western Reserve Journal of International Law 30*¸ 541-564.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computer Communications Review, 29,* 251-262.

Frank, R., Westlake, B., & Bouchard, M. (2010). The structure and content of online child exploitation networks. *Proceedings of the tenth ACM SIGKDD Workshop on Intelligence and Security Informatics.*

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

Frezza, B. (1997). Morality and imagination: Technology challenges both. *COMM. WK*, 31.

Furukawa, T., Ishizuka, M., Matsuo, Y., Ohmukai, I., & Uchiyama, K. (2007). *Analyzing reading behavior by blog mining*. 22nd Annual Conference on Artifical Intelligence (AAAI-07), 1353-1358.

Glaister, D. (1995, July 3). Tap of the Devil. *The Guardian.*

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. In S.I. Feldman, M. Uretsky, M. Najork, and C.E. Wills (Eds.), *Thirteenth International World Wide Web Conference (pp.491-501).* New York, NY: ACM Press.

Gulli, A., & Signorini, A. (2005). The indexable Web is more than 11.5 nillion pages. In *Special Interest Tracks and Posters of The 14th International Conference on World Wide Web, WWW'05*, pages 902–903, Chiba, Japan.

Hanneman, R., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved from http://faculty.ucr.edu/ ~hanneman/nettext/

Haynie, D. (2001). Delinquent peers revisited: does network structure matter? *American Journal of Sociology, 106*, 1013-57.

Healy, M.A. (2004). Child pornography: An international perspective. *Proceedings from the World Congress against Commercial Sexual Exploitation of Children.* Retrieved from http://www.crime-research.org/articles/536/

Iamnitchi, A., Ripeanu, M., & Foster, I. Locating data in (small world?) peer-to-peer scientific collaborations. (2002). In *Proceedings of the First International Workshop on Peer-to-Peer Systems (IPTPS'02)*.

Internet World Stats (2012, June 30). *World Internet usage and population statistics.* Retrieved from: http://www.internetworldstats.com/stats.htm

Janson, S., Luczak, T., & Rucinski, A. (2000). *Random graphs.* Canada: John Wiley & Sons, Inc.

Jeong, H. Mason, S., Barabási, A.-L., & Oltvai, Z.N. (2001) Lethality and centrality in protein networks. *Nature, 411,* 41-42.

Jeong, H, Néda, Z., & Barabási, A.-L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters, 61*, 567-572.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature, 407,* 651-654.

Johnson, B. (2008, June 6). Time taken to shut child abuse sites criticised. *The Guardian*. Retrieved from: http://www.guardian.co.uk/technology/2008/jun/06/internet.child protection

Jones, J.H., & Handcock, M.S. (2003). An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society B: Biological Sciences, 270,* 1123-28.

Kaafar, M.A., Mathy, L., Turletti, T., & Dabbous, W. (2006). Virtual networks under attack: Disrupting Internet coordinate Systems. *Proc. of Second CoNext Conference,* Lisbon, Portugal.

Klein, E.J., Davies, H.J. and Hicks, M.A. (2001) *Child Pornography: The Criminal Justice-System Response.* Alexandria, VA (extended survey by American Bar Association Center on Children and the Law). National Center for Missing and Exploited Children. Retrieved from http://www.missingkids.org

Kleinberg, J..M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). The web as a graph: Measurements, models, and methods. In *Proceedings of International Conference on Combinatorics and Computing*, pp. 1–17.

Knoke, D., & Yang, S. (2008). *Social Network Analysis.* Los Angeles: Sage.

Krebs, V.E. (2002). Mapping networks of terrorist cells. *Connections, 24*, 43-52.

Krone, T. (2004). A typology of online child pornography offending. *Trends and Issues in Crime and Criminal Justice, 279,* 1-6.

Kumar, R., Rajalopagan, S., & Tomkins, A. (1990) Extracting large-scale knowledge bases from the web. *Proceedings of the 9th ACM Symposium on Principles of Database Systems 1.*

Labovitz, C., & Ahuja, A. (1999). Experimental study of Internet stability and wide-area backbone failures. *Fault-Tolerant Computing Symposium,* 278 - 285.

Lawrence, S. & Giles, C.L. (1999). Accessibility of information on the web. *Nature, 400,* 107-109.

Lanning, K. 1992. *Child Molesters: A Behavioral Analysis*. Washington, DC: National Center for Missing and Exploited Children.

Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., & Aberg, Y. (2001). The web of human sexual contacts, *Nature, 411,* 907-908.

Malm, A. and Bichler, G. (2011). Networks of collaborating criminals: Assessing the structural vulnerability of drug markets. *Journal of Research in Crime and Delinquency, 00,* 1-25.

Maslov, S., Sneppen, K., Zaliznyak, A. (2004). Detection of toplogical patterns in complex networks: Correlation profile of the Internet. *Physica A: Statistical Mechanics and its Applications, 333,* 529-540.

McGloin, J. (2005). Policy and intervention considerations of a network analysis of street gangs. *Criminology andPublic Policy, 4,* 607-636.

McGrath, C., & Krackhardt, D. (2003). Network conditions for organizational change. *The Journal of Applied Behavioral Sciences, 39,* 324-336.

McLaughlin, J. (2004). Cyber child sex offender typology. Available at: http://www. ci.keen.nh.us/police/typology.html

Medina, R., & Hepner, G. (2008). Geospatial analysis of dynamic terrorist networks. In I. Karawan, W. McCormack and S.E. Reynolds (Eds.), *Values and Violence: Intangible Aspects of Terrorism* (pp.151-167). Berlin, Germany: Springer.

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 60-67.

Mitchell, K.J., Finkelhor, D., & Wolak, J.W. (2008). Are blogs putting youth at risk for online sexual solicitation or harassment? *Child Abuse & Neglect, 32,* 277-294.

Moreno, J.L. (1937). Sociometry in relation to other social sciences. *Sociometry, 1,* 206-219.

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

Morselli, C. (2010). Assessing vulnerable and strategic position in a criminal network. *Journal of Contemporary Criminal Justice, 26*(4), 382-292.

Morselli, C., & Petit, K. (2007). Law-enforcement disruption of a drug importation network, *Global Crime, 8*(2), 109–130.

Morselli, C, & Roy. (2008). Brokerage qualifications in ringing operations. *Criminology, 46,* 71-98.

Morselli, C., & Tremblay, P. (2004). Criminal achievement, offender networks, and the benefits of low self-control. *Criminology, 42,* 773-804.

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

Newman, M.E.J. (2008). The physics of networks. *Physics Today, 61,* 31-33.

Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167-256.

O'Donnell, I., & Milner, C. (2007). *Child Pornography: Crime, Computers and Society.* Michigan: Willan Publishing.

Papachristos, A. (2009). Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology, 115,* 74-128.

Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Dynamical and correlation properties of the Internet. *Physical Review Letters, 87,* 258701.

Preece, J., Maloney-Krichmar, D. & Abras, C. (2003). History of emergence of online communities. In K. Christensen and D. Levinson (Eds.), *Encyclopedia of Community: From the village to the virtual world* (pp. 1023-1027). Thousand Oaks: Sage.

Price, J.D.S. (1965). Networks of scientific papers. *Science, 149*. 510-515.

Price, D.J.S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal American Social Information Science, 27,* 292-306.

R. v. Sharpe, 1 S.C.R. 45 (2001). Retrieved from: http://scc.lexum.umontreal.ca/en/2001 /2001scc2/2001scc2.html

Raab, J., & Milward, H.B. (2003). Dark networks as problems. *Journal of Public Administration Research and Theory, 13,* 413-439.

Rapoport, A. (1968). Cycle distribution in random nets. *Bulletin of Mathematical Biophysics, 10,* 145-157.

Rapoport, A., & Horvath, W.J. (1961). A study of a large sociogram. *Behavioral Science, 6,* 279-291.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution, *European Physical Journal B, 4,* 131-134.

Schwartz, D.M., & Rouselle, T. (2009). Using social network analysis to target criminal networks. *Trends in Organized Crime, 12,* 188-207.

Scott, J. (2000). *Social network analysis: A handbook.* London: SAGE Publications Ltd.

Seglen, P.O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43,* 628-638.

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

Simon, A. (1955). On a class of skew distribution functions. *Biometrika, 42*, 425-440.

Solomonoff, R., & Rapoport, A. (1951). Connectivity of random nets. *Bulletin of Mathematical Biophysics, 13*, 107-117.

Sparrow, M.K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, *13*(2), 251-274.

Spink, A., Ozmutlu, H.C., & Lorence, D.P. (2004). Web searching for sexual information: An exploratory study. *Information Processing and Management: An International Journal, 40*, 113-123.

Stanley, J. (2001). Child abuse and the Internet. *Child Abuse Prevention Issue, 15*, 1-20.

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

Taylor, M., & Quayle, E. (2003). *Child Pornography: An Internet Crime*. East Sussex: Brunner-Routledge.

TopTenREVIEWS (2004 February 6). *TopTenREVIEWS Releases Porn Industry Statistics.* Retrieved from: http://www.toptenreviews.com/2-6-04.html

Tyler, R.P., Stone, L.E. (1985). Child pornography: Perpetuating the sexual victimization of children. *Child Abuse & Neglect, 9*(3), 313-318.

U.S. Customs. (2001). *Moscow city police team up against child pornography.* Retrieved from http://www.cbp.gov/xp/CustomsToday/2001/April/custoday_bluorchid .xml

Vázquez, A., Pastor-Satorras, Y., & Vespignani, A. (2002). Large-scale topological and dynamical properties of the Internet. *Physical Review Letters, 65*(6), no. 066130.

Wasserman, S., Faust, K., Iacobucci, D., & Granovetter, M. (1994). *Social Network Analysis: Methods and Applications.* Cambridge, MA.: Cambridge University Press.

Watts, D.J. (1999). *Small Worlds*. Princeton, NJ: Princeton University Press.

Watts, D.J. (2003). *Six Degrees: The Science of a Connected Age*. New York: Norton.

Wells, M., Finkelhor, D., Wolak, J., & Mitchell, K. J. (2007). Defining child pornography: law enforcement dilemmas in investigations of internet child pornography possession. *Police Practice and Research*, 8(3), 269-282.

Westlake, B.G. (2011). *Analyzing online child exploitation networks: An examination of severity and connectivity.* Retrieved from Summit, Simon Fraser University's Institutional Repository.

Westlake, B., Bouchard, M., & Frank, R. 2011. Finding the Key Players in Online Child Exploitation Networks. *Policy & Internet, 3*(2), 104.

Wortley, R., & Smallbone, S. (2006). *Child pornography on the Internet.* Washington, DC: Office of Community Oriented Policing Services.

Xu, J. & Chen, H. (2008). The topology of dark networks. *Communications of the ACM, 51,* 58-65

Yook, S.H., Jeong, H., & Barabási, A.-L. (2001). Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences, USA, 99,* 13382-13386.

Young, K.S., Griffin-Shelley, E., Cooper, A., O'Mara, J., & Buchanan, J. (2000). Online infidelity: A new dimension in couple relationships with implications for evaluation and treatment. In A. Cooper (Ed.), *Cybersex: The dark side of the force* (pp. 59-74). Philadelphia: Brunner Routledge.