A Methodology for Profiling Paraphilic Interest in CSEM Users on Peer-to-peer Networks

Margaret Brennan and Sean Hammond

School of Applied Psychology, University College Cork

Author Note

Margaret Brennan, School of Applied Psychology, University College Cork, Ireland, E: m.brennan@ucc.ie T: +353214904551;

Sean Hammond, School of Applied Psychology, University College Cork, Ireland, E: s.hammond@ucc.ie T: ++353214904551

This work was partly supported by the European Commission's Safer Internet (Plus) programme under Grant [SI-2010-TP-2601002], iCOP: Identifying and Catching Originators in Peer-to-Peer Networks and Grant [SIP-2006-PP-221003], MAPAP: Measurement and Analysis of P2P Activity Against Paedophile Content Project. The authors would like to extend their sincere thanks to Prof. Awais Rashid of Lancaster University and all of our collaborators on the iCOP and MAPAP projects, as well as Interpol and the international law enforcement agencies that supported this work.

Correspondence concerning this article should be addressed to Margaret Brennan, School of Applied Psychology, UCC Enterprise Centre, North Mall, Cork, Ireland. E: m.brennan@ucc.ie T: +353214904551

Abstract

This paper describes the development of a novel methodology for profiling paraphilic interest in the search behaviour of users of online “Peer-to-peer” networks - a major vector for the exchange of Child Sexual Exploitation Material. The profiling methodology focuses on problematic paraphilic interests, involving illegal or non-consensual activities associated with the sexual victimisation of children. This work extends an earlier typal analysis carried out by Hammond, Quayle, Kirakowski, O’Halloran and Wynne (2009) in which a distinct problematic paraphilic typology was uncovered in the search behaviours of Peer-to-peer users. The methodology described focuses on the subsequent development of a Latent Class Model that underpins the operation of the profiling application. The composite profiling process is described. Finally, we discuss the prospective applications of this profiling process and the implications of our methodological design. We identify a series of recommendations for future research and for the design of profiling and risk appraisal processes with application to online CSEM offending behaviour.

Keywords: Child Sexual Exploitation Material; Internet Investigation; Profiling; Paraphilia; Risk Assessment; Peer-to-peer; Latent Class Analysis

Word Count: 6,235

Introduction

Terminology

Throughout this article, the authors use the term "Child Sexual Exploitation Material" (CSEM) in accordance with the definition and associated notation offered by Merdian (2012). We use the acronym "P2P" to denote to "Peer-to-peer" file sharing networks, the online forum for CSEM offending considered in this study.

The Role of P2P Networks in CSEM Offending

Recent years have seen growing concerns about the role of online P2P facilities in the sexual exploitation of children, particularly in terms of their capacity to facilitate expedient, large-scale CSEM access (Choo, 2008; Hughes et al., 2006). Parallel concerns have been expressed around the incidence of children's exposures to illicit and illegal sexual media in these environments (Dombroski, Gischlar & Durst, 2007; Greenfield, 2004; Quayle & Latapy, 2008). To a large extent, these concerns have their impetus in a small but established body of empirical evidence that points to increasing volumes of CSEM and attendant offending activity on P2P networks. While estimates of the extent of CSEM exchange on P2P networks vary considerably, it is now widely conceded that P2P serves as a major vector for the distribution of illegal CSEM (Hughes et al., 2008;

Taylor, Haggerty, Gresty & Fergus, 2010; US Department of Justice, 2010, Wolak, Liberatore & Levine, 2014).

International empirical evidence points to the problem of P2P-facilitated CSEM offending as endemic. A pioneering study by Hughes et al. (2006) found that 1.6% of searches and 2.4% of responses on the "Gnutella" P2P network related to illegal sexual content such as rape, bestiality and child abuse. Hughes et al. (2008) revisited this data to determine how much of this traffic was specifically CSEM-related and determined that approximately 1% of queries (searches) and 1.5% of query hits (returned filenames) were associated with this material. Given the system's scale, and the fact that the study only covered a portion of the Gnutella network, the authors suggested that on the Gnutella network alone, hundreds of searches for CSEM occur each second. Similarly, in an analysis of a larger sample of Gnutella network traffic, Steel (2009) found that a significant proportion of these exchanges were CSEM-related. Here, approximately 1% of all observed queries and 1.45% of query hits on the network were CSEM-related. In a similar vein, substantial volumes of paedophilic queries have been observed on the "eDonkey" P2P network (Latapy, Magnien & Fournier, 2009). Here, the authors identified two keyword-based searches per thousand (0.2%) as CSEM-related with a similar proportion of eDonkey users engaging in such searches. With tens of P2P networks in use, the scale of CSEM-related activity on P2P can be estimated to be in the tens, if not hundreds of thousands of exchanges per day.

The decentralised, private nature of P2P file sharing, its open-access policies and rate of growth make its content extremely difficult to control and enable persons with deviant sexual interests to access and exchange CSEM on these networks with relative ease and anonymity (Nielssen et al., 2011; Westlake Bouchard & Frank, 2011). In networks such as eDonkey and Gnutella, there is no central server that can be traced and held accountable for shared illegal content, making dissemination of CSEM across the network almost impossible to prevent. These features of P2P networks naturally limit law enforcement capacity for intervention and have resulted in a situation where deviant sub-communities have flourished within these forums (Hughes et al., 2006; Steel, 2009).

Paraphilic Activity on P2P

It has long since been suggested that individuals with distinct paraphilic interests use P2P networks to engage with CSEM and other problematic materials. For example, in their analysis the distribution of pornographic material on the Gnutella network, Mehta, Best and Poon (2002) monitored Gnutellameter, a website that captures data exchanged in Gnutella and generates summaries of keywords most commonly entered by its users. The authors identified pornographic imagery as one of the most commonly sought materials on the network with user searches displaying a strong emphasis on paedophile and hebephile content. Similarly, in the Steel (2009) study, the majority (76%) of those who engaged in age-specific searches for CSEM sought imagery featuring children between 11 and 16 years of age, indicative of a prevailing hebephilic disposition in those searching for CSEM on the network. Steel also observed significant correspondence between the use of bestiality and CSEM-related search terms in the search behaviours of P2P users - his analysis of CSEM-related queries established that the search term “Zoofila” was most commonly coincident with CSEM-related terms in relevant user queries on the network. Hammond et al. (2009) determined that the largest proportion of problematic paraphilic searches on the eDonkey network related to hebephilic content. Their typal analysis offered forcible evidence of discrete paraphilic sexual interests in P2P user search behaviours.

P2P-facilitated CSEM Offending - Challenges to Law Enforcement

The prevalence of paraphilic and CSEM-related activity on P2P systems has incited substantial responses from international law enforcement, witnessable in increasing arrest rates for P2P-facilitated CSEM offences and an intensification of investigative activity across these networks. For example, offenders who used P2P networks to access CSEM featured in 4% of US arrests for CSEM possession and distribution offences in 2000. By 2009, P2P-accessed CSEM featured in 61% of all such arrests (Wolak, Finkelhor, & Mitchell, 2012). Similarly, there has been a significant concentration of investigative resources in the development of monitoring solutions such as “Child Protection System” that support the apprehension of CSEM offenders on P2P systems. Notwithstanding these developments, pervasive challenges remain in terms of the current scale of P2P-facilitated CSEM offending.

In their recent investigation of one year of CSEM exchange by US computers on the Gnutella network, Wolak, Liberatore and Levine (2014) illustrated the scale of this challenge to law enforcement, with hundreds of thousands of US computers implicated in P2P-facilitated CSEM exchange over the study period. In view of the scale of this offending activity, these authors highlighted the importance of adopting a more discriminating approach to the investigation of CSEM offences on P2P, whereby problematic P2P offenders are identified and prioritised by law enforcement for urgent intervention. More specifically, having identified that a very small proportion (less than 1%) of computers on the network made comparatively high yearly contributions of 100 or more files to the number of known CSEM files available on the system, these authors suggested that investigating law enforcement should use existing investigative software tools such as “RoundUp” or “ICAC Cops” to prioritise the users of these high-contribution computers for arrest and take their files offline, thereby substantially reducing the number of known CSEM files in circulation.

Liberatore, Erdely, Kerle, Levine and Shields (2010) observed that a primary goal of P2P investigations should be to apprehend child abusers and to help children that are being sexually victimised, rather than simply detecting and confiscating CSEM in the context of possession and distribution offences. These authors advocated the development of strategies to support the identification of those more likely to be directly involved in the sexual victimisation of children. Indeed, this objective is shared by investigating law enforcement internationally (Eke, Seto & Williams, 2011) and to date a small number of profiling strategies have been developed for this purpose. For example, Long, Alison and McManus (2013) developed the Kent Internet Offender Risk Assessment Tool (KIRAT) which discriminates CSEM offenders at risk of contact offending on the basis of a range of factors, including the number and type of collected CSEM files and access to children.

However noble, this objective is particularly difficult to achieve in the investigation of abstract online exchanges on P2P, where little personal and behavioural information is available to profile prospective contact offenders. It is difficult for investigating law enforcement to infer offence motivation or outcome from the limited set of behaviours (e.g. file sharing and downloading) that may be observed on P2P. Indeed, Peersman, Rashid, Schulze, Brennan and Fisher (2014) reported that actualising reliable strategies for the identification of perpetrators of child sexual abuse during P2P investigations persists as a primary operational challenge for law enforcement. Furthermore, P2P offender identities are frequently unknown to online investigators, a situation that precludes the possibility of KIRAT-type assessment (where for example, information regarding the subject's access to children is required to inform the profile). Evidently, a strategy for profiling problematic CSEM offenders on P2P is required, which can accommodate the paucity of personal and behavioural data that is available for profiling purposes on P2P and respond to the reality of online investigations, where identity of the offender is often unknown.

The Psychological Profile

All psychological profiles are, explicitly or implicitly, built around the notion of a taxonomy (Horgan, O'Sullivan & Hammond, 2003). Observed behaviours are then said to flow from the "type" of person we are dealing with. In the early years of offender profiling developments, the taxonomy utilised was derived from an a-priori theoretical model, often with some psychodynamic basis (Groth, Burgess & Holmstrom, 1977; Ressler, Burgess & Douglas, 1988). It quickly became apparent that systems that rigidly adhere to one or another psychological model are somewhat limited in scope and that a more behavioural approach was needed (Canter & Heritage, 1989). From this perspective profiles must be built upon large bodies of empirical data to generate the taxonomic models.

This led to a reliance on data analytic techniques whose primary advantage was that the taxonomies were based upon actual contextualised behaviour (Canter, Hughes & Kirby, 1998). Despite the greater ecological validity of these approaches, a major caveat remains. Data analytic procedures inevitably utilise large samples of offences and offenders to build the behavioural patterns that inform the profile. This entails a degree of aggregation across offenders and the best that can be achieved is a guesstimate of where a particular individual falls within the behavioural pattern. Unique individuals will always confound the profile to some extent. For this reason, this study has adopted a transparent model based upon a simple, empirically derived categorical taxonomy that provides a probability profile for each individual describing possible membership of each category.

Profiling P2P Behaviours

Problematic P2P "types" are still largely unknown, although there is a likely continuity with general sexual offenders. This position assumes that the use of online P2P facilities is simply an extension of a person's normal activities and interests. Thus, searching for and accessing CSEM on P2P networks simply reflects a person's paedophilic predilections or other problematic paraphilic interests (Quayle, Hammond & Wynne, 2007). We know that such an assumption should not be automatic and the relationship that offenders have with technology is often more complex than this suggests (Calder, 2004; Carr, 2006; Quayle & Taylor, 2003). However, P2P user behaviours are difficult to operationalise for profiling or assessment purposes because there is only one source of interface accessible. With the paucity of behaviours available to us in online P2P systems, largely based around file searching and downloading behaviours, the assumption of behavioural continuity was necessary in the early stages of the development of the profiling methodology.

It should be borne in mind that P2P offences largely operate at a distance from the victim and are secondary in nature (Wolak, Liberatore & Levine, 2014), unless the offender generates CSEM shared on the network. Clearly, this does not mitigate the offense, but it does have psychological implications because it is not possible to assume that a P2P offender is automatically a contact offender. However, work with CSEM offenders does suggest that when paedophilic interest is the prime motive the risk of contact offense is high (e.g. Quayle & Taylor, 2003; Sheldon & Howitt, 2008). A considerable proportion of CSEM offenders are likely to be paedophilic or hebephilic and therefore present a direct risk to children (Eke, Seto & Williams, 2011; Seto, Cantor & Blanchard, 2006). In addition, exposure to CSEM and related materials may intensify interest through masturbatory conditioning and greater intensity of interest may drive a motive for contact offence (Sullivan, 2002). A further complication is that multiple paraphilias are common and it is likely that certain combinations are particularly high risk (Abel, Becker, Cunningham-Rathner, Mittelman & Rouleau, 1988). Thus, a paedophile with an interest in coercive and/or sadistic sex presents a dangerous combination (MacCulloch, Snowden, Wood & Mills, 1983). It is also important to note to that paraphilic sexual offenders frequently manifest concomitant impulse control problems (e.g. Dunsieth et al., 2004) and are more likely to recidivate over time (Mann, Hanson & Thornton, 2010).

With this in mind, a typology with implications for risk may be reasonably built around the taxonomic notion of paraphilia (Abel & Osborne, 1992). A paraphilia is an abiding interest and, at the extreme, urge to engage in sexual activity of a deviant or problematic kind. Notable is the class known as paedophilia where the target of sexual interest is the prepubescent child. The fact that there may be subclasses of paedophilia that may have implications for risk, treatment and disposal is not explored here, although it may emerge as the use of the profiling methodology develops (see for example Greenberg, Bradford & Curry, 1995). Another important paraphilia in the context of child abuse is hebephilia. This is a slightly contentious paraphilia where the target of interest is the pubescent or recently post-pubescent child (Blanchard et al., 2009). Legally, the distinction between hebephile and paedophile activity is minor as both involve sex with children; however, psychologically, there is a very strong distinction as there is some evidence to suggest that hebephiles, like fully functioning people, are triggered by secondary sexual characteristics, while paedophiles are not (Griffin, 2010; Hammond et al., 2009).

There are a huge number of paraphilias and most of them have no legal ramifications. However, Hammond et al. (2009) demonstrated that P2P behaviour based upon search terms used could be mapped onto a small subset of paraphilias which the authors termed “problematic paraphilias”, involving illegal or non-consensual activities associated with the sexual victimisation of children. This is the taxonomy that we wish to develop in order to establish the profiling methodology.

The Present Study

The primary aim of this study is to describe a profiling methodology for specific offender cases that may offer a basis for the categorisation of problematic profiles of sexual interest on P2P networks. The intended application of this profiling method is to the online investigation process, where generated profiles may be used to inform assessments of offence severity, candidate risk and related decisions to prioritise specific P2P cases for further investigation. The outcome of this study was designed to serve as a complementary, modular resource in the iCOP system (Peersman et al.,

2014) that was developed at Lancaster University under the aegis of the iCOP Project. Originally, this profiling methodology was envisaged as a post-apprehension aid that would be accessed once law enforcement agencies had identified a perpetrator. In this conception, background information on the offender would be entered into the system to provide a profile to aid disposal decisions, interview strategies, and other investigative decisions. However, in view of the above-identified needs of investigating law enforcement, it became apparent that a screening role for the profiling methodology was envisaged at a much earlier stage in the investigative process, during the online investigation phase.

In the following sections, we outline the development of the paraphilic profiling methodology. First, we describe the methodological considerations and premises that informed the development of the profiling process, as well as any associated limitations. We then present the background to the current study, describing earlier work undertaken by Quayle, Hammond and Wynne (2007) and Hammond et al. (2009) to empirically identify the "problematic paraphilic" categorisation upon which this profiling methodology is based. Next, we describe the development of the Latent Class Analysis Model, which underpins the psychological profiling methodology, and present a composite model of the profiling process. Finally, we discuss the prospective applications of this profiling process and the implications of our methodological design. We identify a series of recommendations for future research and for the design of profiling and risk appraisal processes with application to online CSEM offending behaviour.

Methodological Considerations

Designing a profiling process with application at the online investigation stage presents some challenges in terms of methodological design because this screening role necessitates use of less informative data from an individual offender perspective. In many cases, the only behavioural data available at the online investigation stage is the P2P activity, which is largely limited to the use of search terms and the nature of the files being downloaded or shared. No demographics, criminal or psychiatric history information is available and this imposes severe limitations on the profiling methodology.

In order to accommodate this limitation it was decided that the profiling methodology should be built around a probabilistic model. It is vital to emphasise that the profiling process will not provide deterministic output, and in order to ascertain this we aimed to develop a model with a probabilistic output rather than a clearly defined, deterministic profile. The utility of such an approach may at first seem limited, but it is envisaged that this approach will provide more transparency to law enforcement decision makers and reduce the likelihood of false positive judgments in applied settings.

At the heart of the profiling methodology is the recognition that while empirical models are derived normatively, their application is nearly always idiothetic. This means that the best the generated profiles can be is suggestive. There is always a great concern when developing offender profiles that the profile does not become reified. There are ample cases where profiles have so distorted investigations that cases have collapsed with great collateral damage all round (e.g. Wilson & Soothill, 1996). Therefore, we offer the profiling methodology as a *decision support tool* for law enforcement, with this caveat firmly to the fore. It is intended that the psychological profiles generated by this profiling method would be used as a complementary resource in investigative settings, and would be used as a supplementary point of information to inform case prioritisation and other relevant decisions.

Premises of the Profiling Methodology

There are three basic premises to our approach in developing the profiling methodology. Firstly, our psychological profiles should be based upon behaviours rather than assumptions. This means that the methods for obtaining the output profile are entirely transparent and are not based on individualised interpretation. It is true that the formulae used in the probabilistic model are based upon underlying statistical assumptions, but these are kept to a minimum by utilising a nominal level of analysis.

Secondly, the accessible P2P behaviours largely centre around the searches since it is with this volitional behaviour that the user betrays their interests and motives. Where search terms are not available it is possible to use the file sharing or downloading behaviour in terms of the nature of files being downloaded. However, in the context of this study, this strategy was considered problematic. Media files shared or downloaded on a P2P network do not maintain that same volitional quality that characterises P2P search behavior. There are many possible reasons for this, for example, the filenames on downloaded media may have little relation to file content, or P2P users may download CSEM and other media in bulk on P2P networks, when the primary sexual interest may be in only one or two files in a large downloaded set. There are no other volitional features of P2P behaviour that can be reliably utilised for psychological profiling.

Finally, the model we apply is based upon a latent structure in which the latency is viewed as categorical. This enabled us to be consistent with the diagnostic model of paraphilia. Further research may expand this model by opening up multiple latencies or even allowing latent variables to be continuous. However, these extensions may require tighter assumptions that may only be justified by experience of the profiling methodology.

The Profiling Methodology

Background to the Study

The development of the profiling methodology extended work carried out in an earlier European Commission funded study, the Measurement and Analysis of P2P Activity Against Paedophile Content (MAPAP) Project (e.g. Quayle, Hammond & Wynne, 2007).

Quayle, Hammond and Wynne (2007) carried out an analysis in which a list of 119,869 P2P search terms was trawled for words with a sexual connotation. An exhaustive search of the list was carried out to identify terms that indicated sexually related material. To aid this process, a computer program was written in order to isolate words or part-words according to a given theme. The result was the identification of 25 specific themes or categories defined by their sexual and fetishistic content. Specific terms and words associated with each of the 25 categories of sexual interest were identified.

Identification of Individual Sexual Interest Profiles

A computer program was then written to scan, in a serial fashion, over 3,000,000 P2P submissions collected from the eDonkey facility in 2009. For each case, a record containing variables representing the 25 themes was created. Each variable was initially set to zero. If a sought after word occurred in the data set for that case, then the variable representing the theme in which it is placed was incremented by one.

If, after scanning, a case had no occurrence of the critical words it was jettisoned and the program moved onto the next case. If, on the other hand, the case did contain critical words, the record of 25 themes was retained. In this way the program identified 62,940 cases where a P2P user had made one or more sexually related submission. In each case the 25 variables contained the frequency with which terms are submitted within each of the 25 thematic categories. In order to control for the fact that each theme is built of differing numbers of terms the data was represented in binary form thus:

If $y_i>0$ then $x_i=1$

If $y_i=0$ then $x_i=0$

Where y_i is the observed frequency of words in thematic category i and x_i is the binary value.

Each case, then, was recorded as a profile of 25 binary variables in which at least one variable is recorded as 1. Recording the data in this way provided a tractable data set to address the exploration of the relationships between sexual themes and a typal analysis of deviant sexual interest.

Arriving at the Latent Model

A number of multivariate analyses were performed on these profiles and an underlying pattern of paraphilic interest was uncovered. In the first instance a multidimensional scaling revealed regional hypotheses conforming to the paraphilia model. Most of the 25 themes identified were not associated with illegal or non-consensual activity and so Hammond et al. (2009) identified 7 that constituted potential problematic behaviour. Their subsequent Configural Frequency Analysis (Krauth, 1985; Lienert, 1988; von Eye, 1990) demonstrated the discrete nature of these 7 themes. The result of these analyses was to demonstrate seven identifiable paraphilic categories, namely:-

Paedophilic, Hebephilic, Gerontophilic, Sadistic, Coercive, Zoophilic, Incest

This "problematic paraphilic" typology, identifiable in P2P search behaviours, formed the empirical basis of the subsequent Latent Class Modelling approach adopted by the authors in the development of the profiling process.

The Psychological Profiling Model

Building upon these earlier findings it was necessary to fit a model that would be able to infer the positioning of individual cases. To this end, there are several modelling options available, ranging from deterministic models such as clustering through fuzzy set or latent trait modelling. On the basis of the premises noted above, a Latent Class model was selected (Lazarsfeld, 1950; Magidson & Vermunt, 2004). This is a probabilistic approach as opposed to the more deterministic Configural Frequency Analysis. Latent Class modelling maintains an advantage over descriptive analytical approaches like Configural Frequency Analysis as it allows us to describe the probability of membership of multiple categories (problematic paraphilias) for given individual. A further advantage of the Latent Class modelling approach is that it enables meaningful summarisation of very large behavioural datasets, such as that which was utilised in this study.

For simplicity's sake, the particular form of the model is one with one latent variable (Paraphilia) made up of 7-categories.

Latent Class Modelling may be defined in the following formula where, for brevity, 4 indicators (A, B, C and D) are shown.

$$\pi_{ijklt} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}$$

Where π^t_x is the probability of being in class t on latent variable X while $\pi_{it}^{A|X}$ is the conditional probability of i^{th} response on indicator A when a member of class t and π_{ijklt} is the probability of someone presenting with the profile ijkl in latent class t.

Latent Class Modelling requires customised software although modules may be accessed via systems such as R. The most direct way of fitting the model and deriving the conditional probability parameters is to use an iterative approach beginning with rough model estimates in order to generate maximum likelihood estimates (Goodman, 1974). However, there is no guarantee of reaching a local function minimisation so it is advisable to carry out the analysis a number of times using different starting points. This means that the analysis may be time consuming. However, such an analysis is not required for each profiling session as in order to generate the profile the most recent parameters are used in a relatively simple form. To classify individuals, a probability of class membership to each of the t classes is identified based upon the behaviour profile thus:-

$$p_t = \frac{\pi_{ijklt}}{\sum_t \pi_{ijklt}}$$

The modal p_t indicates the most salient class or paraphilia for that person. The relative nature of the classification allows for complex membership profiles such that individuals with multiple paraphilias may be observed.

A set of routines written by the authors for a Windows platform (Pascal Code) was used to perform the Latent Class Analysis and the subsequent classification.

The data (comprising 62,940 cases, as described above) was fitted to a number of unrestricted latent class models ranging from 2 to 8 underlying classes.

Results

The 7-class model, fitted to the data from 62,940 cases, was found to be the best fitting using the log-likelihood statistic and the Bayesian Information Criterion (BIC). The log likelihood emerged at 59.12 with 107 degrees of freedom and the BIC at 93.90. The Disimilarity Index for this model was 0.004, showing an excellent fit of the model and the data. The conditional probabilities are presented in Table 1. The existence of a set of relatively "pure" types corresponding with each of the 7 paraphilias is evident.

Themes	Classes……						
	1	2	3	4	5	6	7
Gerontophilic	0.000	0.998	0.000	0.000	0.000	0.000	0.000
Bestiality	0.008	0.028	0.026	0.000	1.000	0.019	0.054
Paedophilic	0.000	0.000	0.001	0.000	0.046	0.004	1.000
Hebephilic	0.000	0.035	0.008	1.000	0.010	0.020	0.003
Sadistic	0.000	0.031	0.001	0.000	0.000	1.000	0.000
Rape	0.000	0.000	1.000	0.001	0.000	0.000	0.001
Incest	0.997	0.000	0.000	0.000	0.000	0.000	0.000
Class Probabilities	0.019	0.072	0.143	0.279	0.230	0.101	0.154

Diagnostic Statistics

Log Likelihood	59.12
2LL	118.23
Pearson χ^2	194.60
Pearson χ^2 under independence	8281.38
Dissimilarity Index	0.004
BIC	93.90

Table 1. Latent Class Analysis of Paraphilic Themes: The 7-Class Solution

Discussion

Application of the Latent Class Model: P2P Investigation

State of the art P2P investigation systems maintain a range of tools that allow investigators to prioritise persons of interest based on the number or the type of CSEM files they engage on these networks. However, the available literature suggests that a persistent challenge is to develop a solution that discriminates those who may present enhanced risk for contact sexual offending and recidivism, such as those presenting with profiles of problematic paraphilic interest. Given that no personal information about CSEM users themselves is gathered by existing monitoring systems (Wolak et al., 2014), this requirement is not currently supported, as no personal information is accessible to investigators that could readily support such decisions.

Therefore, the authors have designed the methodology to be implemented as a complementary, elective function that would be integrated into existing monitoring systems. In its current conception, this function may be invoked at the discretion of the investigator to support their decisions to prioritise detected CSEM offenders for investigation based on a psychological profile of their problematic paraphilic interest. For example, where resources are limited, the paraphilic profiling function may be invoked to help investigators to discriminate amongst individuals with comparatively high contributions of CSEM files to P2P systems, such that priority cases may be identified for further investigation and prosecution.

Importantly, this profiling methodology supports an initial psychological profiling of cases where no background information on the individual is known, providing psychological indicators that are not available in other systems at pre-arrest level. While comparatively simple strategies such as paraphilic keyword or high intensity search detection may be feasible, the latent class model and associated profiling system provides the possibility to identify paraphilic profiles in a more sophisticated fashion. The presented methodology organises search-related information in a way that flags up potentially high-risk combinations of paraphilic interest and provides the user with a probabilistic profile, comprising a series of indices denoting that individual's likelihood of membership of each problematic paraphilic category. These features enable investigators to make more discriminating judgements to between CSEM users on P2P for prioritisation purposes. Moreover, the profiling methodology described here does not assume the presence of particular paraphilic types in P2P systems in a way that could increase the likelihood of false positive judgements, or otherwise misinform investigative decisions to prioritise certain cases. Rather the methodology relies on a series of empirically identified problematic paraphilic categories, discriminated in the preceding typal analysis of paraphilic interest on P2P (Hammond et al., 2009).

The Composite Profiling Methodology

The composite profiling methodology is illustrated at Figure 1, below. In the context of the profiling methodology, the Latent Class Analysis operates as a separate and static module and is not run every time there is a query on the profiling process. Rather, it is run once to generate the model parameters that are then stored. The classification module simply draws down the predefined parameters to produce the paraphilia profile, as illustrated (in blue) in the following process:

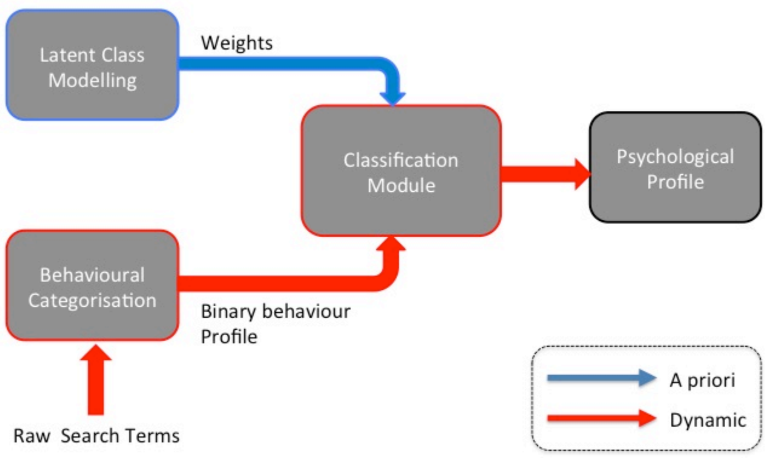

Figure 1. The Composite Profiling Methodology

The Behavioural Categorisation process indicated in Figure 1 is based upon Quayle, Hammond and Wynne's (2007) a-priori thematic analysis carried out for the MAPAP Project described above. Where search terms input for a given individual, this categorisation process involves a simple "look-up" process in order to generate the binary indicators required for profiling. It should be noted that the words and search terms utilised by this behavioural categorisation process may need to be updated and extended in accordance with the emergence of new domain terminology, as it is identified by future law enforcement and related initiatives in the online child protection domain.

The binary indicator profile is passed to the classification process where the probabilistic profile is generated using formula 2 (above). In the end user scenario as currently envisaged, this output is relayed to the law enforcement user for interpretation. A potential form of the output relayed to the end user is demonstrated in the relative profile illustrated at Figure 2, below.

P	**H**	**G**	**S**	**C**	**I**	**Z**
0.194	0.055	0.305	0.037	0.287	0.009	0.111

(Gerontophile with coercive interests)

P = Paedophilic; **H** = Hebephilic; **G** = Gerontophilic; **S** = Sadist; **C** = Coercive; **I** = Incest; **Z** = Zoophilic

Figure 2. Sample Profile Generated by the Profiling Process

No natural language formulation of this profile has yet been constructed because of the caveats mentioned above about reification of profiles. It is more important that detailed training is provided on the interpretation of these profile outputs, to reinforce the practical importance of using these profiles as a means of decision support rather than "decision making."

General Discussion

The profiling methodology developed at this study demonstrates a clear paraphilic typology in the search behaviours of P2P users that may maintain some salience in assessing and managing this population. The primary intended application of this profiling method is to online investigation processes, where generated profiles may be used to inform assessments of offence severity, candidate risk and related decisions to prioritise specific P2P cases for further investigation. Evidently, these profiles may have a secondary application in investigative settings in the sense that they may inform assessments of seized CSEM, or help to determine the nature or sequence of topics that could be explored in suspect interviewing strategies.

It should be borne in mind that the results presented above come from the use of P2P search behaviour (search terms and submissions). The utility of the profiling system is contingent upon access to forms of behavioural data that are volitional, and that allow us to meaningfully identify paraphilic profiles (or other indicators of sexual risk) in P2P contexts. Arguably, all psychological profiling should be based upon volitional behaviours. Examining search terms appears to allow us to profile in this way, however other behavioural features of P2P use (e.g. downloading and sharing behaviours) do not appear to maintain that same, volitional quality that characterises P2P search behavior and may not hold same value for profiling purposes. State of the art investigative monitoring systems for P2P such as iCOP (Peersman et al., 2014) and Child Protection System tend to prioritise shared or downloaded files as the principal unit of analysis. From a policing perspective, an approach that prioritises file content is perfectly justifiable as it offers direct avenues to victim identification and the identification of sexual victimisation. However, for the purposes of a psychological profile search behaviours may be preferred, in view of their highly volitional character.

The model developed for the purposes of the profiling methodology shows good promise in terms of its content validity. However, further work is required to establish the predictive validity of the profiling system using live search data on P2P systems; to further determine the salience of the paraphilic profiles developed with this methodology to investigative and disposal decisions. This suggests a requirement for an end-user supported trial in live P2P investigation settings. Such a trial would establish apposite mechanisms for positioning and integrating the profiling methodology with investigative workflows and processes, and help identify what policies and training needs are required to support its deployment in live investigative settings.

The psychological profiles of paraphilic interest generated by this methodology may also maintain some value when applied to more traditional, clinical-forensic risk assessment processes. Notwithstanding the substantive mediating influence of the online environment in the commission of child sexual offences within P2P and other online forums, as well as the salience of contextual "crime scene" information that may be drawn from these environments to forensic risk assessment processes (e.g. West & Greenall, 2011), it is common that clinical-forensic professionals do not have access to such data when formulating assessments of online child sexual offenders. This situation is largely attributable to the illegal character of CSEM and the inaccessibility of materials and online forums implicated these offences to non-law enforcement professionals. As aforementioned, this situation can be problematic given the apparent salience of this information to the formulation of comprehensive risk assessments, disposal and other management decisions. Therefore, the paraphilic profile generated by the law enforcement user may serve as a useful, supplementary point of information for the clinical-forensic practitioner regarding the character of the P2P offender's sexual deviance; such information may not otherwise be accessible for assessment purposes. Furthermore, the profiling methodology delivers this information in a way that does not require the assessor to directly engage with CSEM and related online behaviours and environments. This feature of the process may also be beneficial given the distressing and potentially corrosive nature of exposures to CSEM and related offending practices (e.g. Powell, Cassematis, Benson, Smallbone & Wortley, 2014).

A further, final advantage of the profiling methodology presented in this study is that it is sensitive to the situational factors that impact upon the commission of CSEM offences on P2P. In the field of offender profiling, a broad range of studies have demonstrated the variability of offending behaviour in response to contextual and situational factors; particularly in relation to sexual offences (Alison, Goodwill, Almond, van den Heuvel & Winter, 2010). Moreover, sexual offences committed through P2P systems are almost entirely mediated by the online environment; in this way the enactment of the offence behaviour and offending outcome may be shaped by the functionalities the online interface, and other dynamic contextual factors. Notwithstanding the formative influence of online offending contexts on criminal outcomes, current systems offered for the purposes of investigative risk appraisal and case prioritisation (of which there are very few) do not appear to attend to the mediating role of these contextual influences on offending behavior. Specifically, these systems do not appear to account for the possibility that extraneous features of the online environment may enable or constrain the expression of volitional components of offence-related behaviours, so colouring attendant conclusions that may be drawn about the nature of the individual's sexual interest or prospective risk. As we have seen in the context of P2P, these specific contextual influences may substantially compromise the volitional character of certain P2P behaviours (e.g. number/type of filenames downloaded), and, by extension, their utility for risk assessment purposes.

To conclude, this study demonstrates a clear speciation of pornographic interest, identifiable in the searching behaviours of P2P users. As mentioned above, future analyses to refine the model may reveal other relevant distinctions or sub-classifications in profiles of paraphilic search behaviour, e.g. preferences for boys vs. girls; for specific age groups and developmental stages, etc. Successful demonstration of such subclasses of search behaviour and paraphilic interest may hold significance for risk assessment purposes. For example, it has been well established that sexual offenders who maintain preferences for male children are more resistant to treatment and tend to recidivate at a higher rate (e.g. Hanson & Bussiere, 1998; Harris & Hanson, 2004; Petrunik & Deustchmann, 2008); much as those who sexually offend against both girls and boys maintain a particularly high risk of sexual recidivism (e.g. Langevin et al., 2004). The suggested salience of such sub-classifications to the identification of recidivistic and contact offending potential in CSEM offenders is perhaps all the more pertinent in light of the findings of a recent study of recidivism predictors in a Canadian sample of “child pornography offenders” (Seto & Eke, 2015), which suggests that those whose CSEM collections indicate paedophilic or hebephilic preference for male children are more likely to reoffend across CSEM and contact sexual offence types.

References

Abel, G. G., Becker, J. V., Cunningham-Rathner, J., Mittelman, M., & Rouleau, J. L. (1988). Multiple paraphilic diagnoses among sex offenders. *Journal of the American Academy of Psychiatry and the Law Online*, *16*(2), 153-168.

Abel, G. G., & Osborn, C. (1992). The paraphilias: the extent and nature of sexually deviant and criminal behavior. *Psychiatric Clinics of North America, 15*, 675 – 687.

Alison, L., Goodwill, A., Almond, L., Heuvel, C., & Winter, J. (2010). Pragmatic solutions to offender profiling and behavioural investigative advice. *Legal and Criminological Psychology*, *15*(1), 115-132.

Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T., Dickey, R., & Klassen, P. E. (2009). Pedophilia, hebephilia, and the DSM-V. *Archives of Sexual Behavior*, *38*(3), 335-350.

Calder, M. C. (Ed.). (2004). *Child sexual abuse and the Internet: Tackling the new frontier*. Lyme Regis: Russell House Publishing Ltd.

Canter, D. V., & Heritage, R. (1989). A multivariate model of sexual offence behaviour: Developments in offender profiling. Vol. 1. *Journal of Forensic Psychiatry, 1*, 185-212.

Canter, D. V., Hughes, D., & Kirby, S. (1998). Paedophilia: Pathology, criminality, or both? The development of a multivariate model of offence behaviour in child sexual abuse. *The Journal of Forensic Psychiatry*, *9*(3), 532-555.

Carr, A. (2006). *Internet censorship offending: A preliminary analysis of the social and behavioural patterns of offenders*. (Unpublished doctoral dissertation). Bond University, Australia. Retrieved from http://epublications.bond.edu.au/theses/carr/

Choo, K. K. R. (2008). Organised crime groups in cyberspace: a typology. *Trends in organized crime*, *11*(3), 270-295.

Dombrowski, S. C., Gischlar, K. L., & Durst, T. (2007). Safeguarding young people from cyber pornography and cyber sexual predation: A major dilemma of the Internet. *Child abuse review*, *16*(3), 153-170.

Dunsieth Jr, N. W., Nelson, E. B., Brusman-Lovins, L. A., Holcomb, J. L., Beckman, D., Welge, J. A., ... & McElroy, S. L. (2004). Psychiatric and legal features of 113 men convicted of sexual offenses. *Journal of Clinical Psychiatry*, *65*(3), 293-300.

Eke, A. W., Seto, M. C., & Williams, J. (2011). Examining the Criminal History and Future Offending of Child Pornography Offenders. *Law and Human Behavior*, *35*(6), 466-478.

Goodman, L.A. (1974). Explanatory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Greenberg, D. M., Bradford, J., & Curry, S. (1995). Infantophilia - a new subcategory of pedophilia? A preliminary study. *Journal of the American Academy of Psychiatry and the Law Online*, *23*(1), 63-71.

Greenfield, P. M. (2004). Inadvertent exposure to pornography on the

Internet: Implications of peer-to-peer file-sharing networks for child development and families. *Journal of Applied Developmental Psychology*, *25*(6), 741-750.

Griffin, M. E. (2010). *An exploration into the measurement of hebephilic interest in a forensic context*. (Unpublished doctoral dissertation). University College Cork, Ireland.

Groth, N. A., Burgess, A. W., & Holmstrom, L. L. (1977). Rape: Power, anger and sexuality. *American Journal of Psychiatry*, *134*, 1239-1243.

Hammond, S., Quayle, E., Kirakowski, J., O'Halloran, E., & Wynne, F. (2009). An examination of problematic paraphilic use of peer to peer facilities. In *Advances in the Analysis of Online Paedophile Activity* (pp. 65-73). Paris: MAPAP Project.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348-362.

Harris, A. J. R., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question* (Vol. 3). Ottawa: Public Safety and Emergency Preparedness Canada.

Horgan, J., O'Sullivan, D., & Hammond, S. (2003). Offender profiling: A critical perspective. *The Irish Journal of Psychology*.

Hughes, D., Walkerdine, J., Coulson, G., & Gibson, S. (2006). Peer-to-peer: Is deviant behavior the norm on P2P file-sharing networks? *IEEE Distributed Systems Online*, *7*(2), 1-11.

Hughes, D., Rayson, P., Walkerdine, J., Lee, K., Greenwood, P., Rashid, A., May-Chahal, C., Brennan, M. (2008). Supporting law enforcement in digital communities through natural language analysis. In *Proceedings of the 2nd International Workshop on Computational Forensics*: *Lecture Notes in Computer Science* (pp. 122-134). Heidelberg: Springer-Verlag.

Krauth, J. (1985). 'Typological personality research by Configural Frequency Analysis.' *Personality and Individual Differences. 6*, 161-168.

Langevin, R., Curnoe, S., Fedoroff, P., Bennett, R., Langevin, M., Peever, C., ... & Sandhu, S. (2004). Lifetime sex offender recidivism: A 25-year follow-up study. *Canadian Journal of Criminology and Criminal Justice/La Revue canadienne de criminologie et de justice pénale*, *46*(5), 531-552.

Latapy, M., Magnien, C., & Fournier, R. (2009). *Quantification of Paedophile Activity in a Large P2P System.* (MAPAP Project: SIP-2006-PP-221005). Retrieved from http://antipedo.lip6.fr/T24/TR/quantification.pdf

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. In Stouffer, S., et al. (Ed.), *Measurement and Prediction.* Wiley: New York.

Liberatore, M., Erdely, R., Kerle, T., Levine, B. N., & Shields, C. (2010). Forensic investigation of peer-to-peer file sharing networks. *Digital Investigation*, *7*, 95-103.

Lienert, G.A. (1988). *Angewandte Konfigurationsfrequenzanalyse.* Frankfurt: Athenaum.

Long, M. L., Alison, L. A., & McManus, M. A. (2013). Child pornography and likelihood of contact abuse: A comparison between contact child sexual offenders and noncontact offenders. Sexual Abuse: A Journal of Research and Treatment, *25*, 370–395.

MacCulloch, M. J., Snowden, P. R., Wood, T. J. W., & Mills, H. E. (1983). Sadistic fantasy, sadistic behaviour, and offending. *The British Journal of Psychiatry, 143*(1), 20–29.

Magidson, J. & Vermunt, J.K. (2004). Latent Class Analysis. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Thousand Oakes: Sage Publications.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, *22*(2), 191-217.

Mehta, M., Best, D., & Poon, N. (2002). Peer-to-Peer Sharing on the Internet: An analysis of how Gnutella networks are used to distribute pornographic material. *Canadian Journal of Law and Technology*, *1*(1), 30-40.

Merdian, H. L. (2012). *Offenders who use child sexual exploitation material: Development of an integrated model for classification, assessment, and treatment*. (Unpublished doctoral dissertation). University of Waikato, New Zealand. Retrieved from http://researchcommons.waikato.ac.nz/handle/10289/6566

Nielssen, O., O'Dea, J., Sullivan, D., Rodriguez, M., Bourget, D., & Large, M. (2011). Child pornography offenders detected by surveillance of the Internet and by other methods. *Criminal Behaviour and Mental Health*, *21*(3), 215-224.

Peersman, C., Schulze, C., Rashid, A., Brennan, M., & Fischer, C. (2014). iCOP: automatically identifying new child abuse media in P2P networks. In *Security and Privacy Workshops* (pp. 124-131). IEEE.

Petrunik, M. & Deustchmann, L. (2008). "The exclusion inclusion spectrum in state and community response to sex offenders in Anglo-American and European Jurisdictions”. *International Journal of Offender Therapy and Comparative Criminology*, *52*, 499-519.

Powell, M., Cassematis, P., S. Benson, M., Smallbone, S., & Wortley, R. (2014). Police officers’ perceptions of the challenges involved in Internet Child Exploitation investigation. *Policing: an international journal of police strategies & management*, *37*(3), 543-557.

Quayle, E., Hammond, S., & Wynne, F. (2007). *An Empirical Investigation into the Sexual Interest Profiles Manifest in P2P Activity: A Preliminary Report*. (MAPAP Project: SIP-2006-PP-221005).

Quayle, E., & Latapy, M. (2008). *Current situation regarding our knowledge of paedophile activity in P2P networks.* (MAPAP Project: SIP-2006-PP-221005). Retrieved from http://antipaedo.lip6.fr/Current_situation.pdf

Quayle, E., & Taylor, M. (2003). Model of problematic Internet use in people with a sexual interest in children. *CyberPsychology & Behavior*, *6*(1), 93-106.

Ressler, R. K., Burgess, A., & Douglas, J. E. (1988). Sexual homicide: Patterns and motives. Lexington, MA: Lexington Books.

Seto M. C., Cantor J. M. & Blanchard, R. (2006). Child pornography offenses are a valid diagnostic indicator of pedophilia. *Journal of Abnormal Psychology*, *115*, 610-615.

Seto, M. C., & Eke, A. W. (2015). Predicting Recidivism Among Adult Male Child Pornography Offenders: Development of the Child Pornography Offender Risk Tool (CPORT). *Law and Human Behaviour.* Advance online publication. doi: http://dx.doi.org/10.1037/lhb0000128

Sheldon, K., & Howitt, D. (2007). *Sex offenders and the Internet*. Sussex: John Wiley & Sons, Ltd.

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

Sullivan, J. (2002). The spiral of sexual abuse: A conceptual framework for understanding and illustrating the evolution of sexually abusive behaviour. *NOTA News*, *41*, 17-21.

Taylor, M., Haggerty, J., Gresty, D., & Fergus, P. (2010). Forensic investigation of peer-to-peer networks. *Network Security*, 9, 12-15.

US Department of Justice. (2010). *The national strategy for child exploitation prevention and interdiction: A report to congress.* Washington DC: US Department of Justice.

von Eye, A. (1990) *Introduction to Configural Frequency Analysis.* Cambridge: Cambridge University Press.

West, A. G., & Greenall, P. V. (2011). Incorporating index offence analysis into forensic clinical assessment. *Legal and criminological psychology*, *16*(1), 144-159.

Westlake, B. G., Bouchard, M., & Frank, R. (2011). Finding the key players in online child exploitation networks. *Policy & Internet*, *3*(2), 1-32.

Wilson, P., & Soothill, K. (1996). Psychological profiling: Red, green or amber. *Police Journal*, *69*, 12.

Wolak, J., Finkelhor, D., & Mitchell, K. (2012). Child pornography possessors: Trends in offender and case characteristics. *Sexual abuse: a journal of research and treatment*, *23*(1), 22-42.

Wolak, J., Liberatore, M., & Levine, B. N. (2014). Measuring a year of child pornography trafficking by US computers on a peer-to-peer network. *Child abuse & neglect*, *38*(2), 347-356.