

Article

Comparing paedophile activity in different P2P systems

Raphaël Fournier^{1,*}, Thibault Cholez², Matthieu Latapy^{3,4}, Isabelle Chrisment², Clémence Magnien^{3,4}, Olivier Festor², and Ivan Daniloff

¹ L2TI/Institut Galilée, Université Paris-Nord, 99 avenue JB Clément, 93430 Villetaneuse, France

² LORIA/INRIA Nancy-Grand Est, 615 Rue du Jardin Botanique, 54600 Villers-lès-Nancy, France

³ Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France

⁴ CNRS, UMR 7606, LIP6, F-75005, Paris, France

* Author to whom correspondence should be addressed; raphael.fournier@univ-paris13.fr

Version May 14, 2014 submitted to SocialSciences. Typeset by \LaTeX using class file *mdpi.cls*

Abstract: Peer-to-peer (P2P) systems are widely used to exchange content over the Internet. Knowledge on paedophile activity in such networks remains limited while it has important social consequences. Moreover, though there are different P2P systems in use, previous academic works on this topic focused on one system at a time and their results are not directly comparable.

We design a methodology for comparing *KAD* and *eDonkey*, two P2P systems among the most prominent ones and with different anonymity levels. We monitor two *eDonkey* servers and the *KAD* network during several days and record hundreds of thousands of keyword-based queries. We detect paedophile-related queries with a previously validated tool and we propose, for the first time, a large-scale comparison of paedophile activity in two different P2P systems. We conclude that there are significantly fewer paedophile queries in *KAD* than in *eDonkey* (approximately 0.09% vs 0.25%).

Keywords: P2P networks; eDonkey; paedophile activity

1. Introduction

Paedophile activity is a crucial social issue and is often claimed to be prevalent in peer-to-peer (P2P) file-sharing systems [1,2]. However, current knowledge on paedophile activity in these networks remains limited.

18 Recently, research works have been conducted to improve this situation by quantifying paedophile
19 activity in *Gnutella* and *eDonkey*, two of the main P2P systems currently deployed [3,4]. They
20 respectively conclude that 1.6% and 0.25% of queries are of paedophile nature, but these numbers are
21 not directly comparable as the authors use very different definitions and methods. Such comparisons are
22 of high interest though, since differences in features of P2P systems, such as the level of anonymity they
23 provide, may influence their appeal for paedophile users.

24 In this paper, we perform for the first time such a comparison. We focus on the *KAD* and *eDonkey*
25 P2P systems, which are the names given to the two underlying P2P networks used by the popular eMule
26 file-sharing application. They are both widely used, accounting together for almost 10% of the global
27 Internet traffic in Europe in 2012 [5], but they differ significantly in their architecture: while *eDonkey*
28 relies on a few servers, *KAD* is fully distributed. This lack of centralisation may lead users to assume that
29 *KAD* provides a much higher level of anonymity than *eDonkey*. Comparing the two systems sheds light
30 on the influence of a distributed architecture on paedophile behavior and increases general knowledge
31 on paedophile activity in P2P systems.

32 The term paedophilia is popularly used to denote adult sexual engagement with children, both
33 prepubescent and pubescent. The definition of paedophilia we use in this article thus encompasses both
34 the medical definition of paedophilia (sexual interest in prepubescent children) and hebephilia (sexual
35 interest in pubescent children not sexually mature).

36 We discuss related work in Section 2, to give an overview of the state-of-the-art on online paedophile
37 activity detection and analysis. Section 3 presents a short introduction to P2P systems, before our
38 description of our datasets and how we collected them (Section 4). We then present the details of our
39 comparison of the amount of paedophile queries in *KAD* and *eDonkey* in Section 5. Section 6 focuses on
40 an important feature of paedophile activity: ages entered in queries. Finally, in Section 7, we introduce
41 a methodology to estimate the fraction of paedophile queries in *KAD* from the one in *eDonkey*.

42 2. Related work

43 Collecting P2P traces is an active topic for years, but it is mostly aimed at analysing peer behaviour
44 to help with future P2P protocol design. In 2006, authors of [6] and [7] explored the social and technical
45 issues related to online child pornography and opened the way to the research in the field. The first
46 detailed quantitative study focusing on a P2P system was proposed in [3], using an active methodology
47 (sending specific queries and analysing the answers provided by the search engine). Since then, several
48 approaches have been proposed to gauge the extent of the phenomenon. Among them, [8] presented
49 filename categorization tool, while [9,10] proposed to label suspicious chat conversations. [11] especially
50 analysed aged-related queries.

51 A first large-scale study of P2P search-engine queries was presented in [12]. Their study focused on
52 “onset”, the first deliberate viewing of child pornography. They gathered the Top 300 queries submitted
53 to the popular *Isohunt* tracker (part of the BitTorrent network) and published on the website Isohunt.com.
54 Their study lasts for 3 months, a scale similar to ours, but they resort to manual classification of the
55 queries. Their dataset is particular, as it only gives a relative popularity order for some queries, and
56 may not provide any indication on the extent of child pornography in the network. Plus, with only 300

57 queries collected daily, they get very few pedophile queries (only 3), which leads to results with a limited
58 statistical significance. However, their discussion is truly interesting, including comments on whether
59 “regular paedophile users” are likely to submit several times the same query (to “build a collection”),
60 while first-time users may not (they do not progress to downloading material once they have discovered
61 the meaning of intriguing pedophile sequences such as *pthc*). This bias may lead query-based studies
62 like ours to slightly overestimating the demand for child pornography, and would impact estimations on
63 the number of paedophile users, but additional filtering based on the IP-address or the client ID could
64 limit this issue.

65 In [4], the authors developed and assessed a dedicated tool for search engine query classification, and
66 collected large-scale datasets on *eDonkey* (up to 28 weeks of uninterrupted experiment). We use here
67 their tool and one of their datasets. Part of their work was later reused by another team to study another
68 P2P network, BitTorrent [13]. The European Commission has set up a “Safer Internet” program [1],
69 which funded some large research projects such as MAPAP¹ and iCOP².

70 In parallel, authors of [14,15] provided an extensive study (one-year long) on child pornography on
71 Gnutella and eMule, partnering with law enforcement to develop software platforms and collect data
72 on child pornography trafficking. They made a precious contribution to understand the “supply”: how
73 many users are involved in the distribution of files, what are their importance in the network, etc. In [14],
74 they evaluated different strategies to best fight paedophile activity given the limited resources of law
75 enforcement and proposed an efficient metric to target the most prominent peers.

76 While having a smaller scale, our study is the first to provide a methodology to gain new knowledge
77 from the proper comparison of data collected from two P2P networks which architecture and monitoring
78 capacity are totally different. Moreover, if the general user behavior in the *KAD* network was detailed
79 in [16], our article is the first to study whether its decentralized architecture is prone to favor criminal
80 activity.

81 3. P2P systems

82 P2P systems are computer networks in which every user may share content with others members.
83 They have become popular because they gather large amount of digital contents (books, movies,
84 music) which can be obtained for free. Copyrighted material is available (however not authorized)
85 and pornography is widespread. Accessing a P2P network is generally easy: a user only needs to
86 download and install on his computer a single application, which will handle the connection process
87 to the network. Then, he can search for files with some keywords, and gets a list of corresponding
88 available files. The application sends messages to the network to find providers of the selected files, and
89 then users interconnect directly to exchange them.

90 P2P networks are easy to access for both providers and consumers. Contents are obtained free of
91 charge, and rather anonymously (no personal details are required). These features make such networks
92 appealing for illegal activities such as paedophile material trafficking.

¹ <http://antipaedo.lip6.fr>

² <http://scc-sentinel.lanacs.ac.uk/icop/>

93 P2P networks account for approximately a fifth of the global bandwidth use on the Internet. Bittorrent
94 is the most prominent P2P network nowadays, preceding *eDonkey* and *KAD* (the usage of which decline
95 in Europe). For instance, an important *eDonkey* server received on average 8.8 million queries per week
96 between 2009 and 2012 [4].

97 4. Experimental setup and datasets

98 In order to compare paedophile activity in two different P2P systems, we first need appropriate
99 datasets, the collection of which is a challenge in itself. In *KAD* and *eDonkey*, different kinds of
100 measurements are possible, depending on the details of the network's architecture.

101 In *eDonkey*, servers index files and providers for these files, and users submit keyword-based queries
102 to servers to seek files of interest to them [17]. By monitoring such a server, one may collect all those
103 queries [18]. Here, we record all queries received by two of the largest *eDonkey* servers during a
104 three-month period in 2010. The servers are located in different countries (France and Ukraine) and
105 have different filtering policies: the French server indexes only non-copyrighted material, while the
106 Ukrainian server openly indexes all submitted files. Monitoring two such different servers will allow us
107 to compare them in order to know if server policy impacts our results.

108 To collect *KAD* data, we use the HAMACK monitoring architecture [19], which makes it possible
109 to record the queries related to a given keyword by inserting distributed probes close to the keyword
110 ID onto the *KAD* distributed hash table. We supervise 72 keywords, which we choose to span well
111 the variety of search requests entered in the system, with a focus on paedophile activity: a set of 19
112 *paedophile* keywords (*babyj*, *babyshivid*, *childlover*, *childporn*, *hussyfan*, *kidzilla*, *kingpass*, *mafiasex*,
113 *pedo*, *pedofilia*, *pedofilo*, *pedoland*, *pedophile*, *pthc*, *ptsc*, *qqaazz*, *raygold*, *yamad*, *youngvideomodels*),
114 which are known to be directly and unambiguously related to paedophile activity in P2P networks; a
115 set of 23 *mixed* keywords (*1yo*, *2yo*, *3yo*, *4yo*, *5yo*, *6yo*, *7yo*, *8yo*, *9yo*, *10yo*, *11yo*, *12yo*, *13yo*, *14yo*,
116 *15yo*, *16yo*, *boy*, *girl*, *mom*, *preteen*, *rape*, *sex*, *webcam*) frequently used in paedophile queries but also in
117 other contexts (for instance, *Nyo* stands for *N years old* and is used by both paedophile users and parents
118 seeking games for children of this age); and a set of 30 *not paedophile* keywords (*avi*, *black*, *christina*,
119 *christmas*, *day*, *doing*, *dvdrip*, *early*, *flowers*, *grosse*, *hot*, *house*, *housewives*, *live*, *love*, *madonna*, *man*,
120 *new*, *nokia*, *pokemon*, *rar*, *remix*, *rock*, *saison*, *smallville*, *soundtrack*, *virtual*, *vista*, *windows*, *world*) used
121 as a test group and *a priori* rarely used in paedophile queries. The sets of keywords were established
122 using the work on paedophile query detection presented in [4]. Notice that our set of keywords contains
123 mainly common English words (*love*, *early*, *flowers*), but some are in other languages (*saison*, *pedofilia*),
124 and some are also brand names (*pokemon*, *nokia*).

125 Because of the differences in architectures of the two networks and of the measurement
126 methodologies, we obtained very different datasets, which are not directly comparable: in *eDonkey*,
127 we observe all queries from a subset of users whereas in *KAD* we only observe queries related to a given
128 keyword, but from all users. In addition, based on various versions of *KAD* clients, the measurement
129 tool only records the queries containing a monitored keyword placed in first position or being the longest
130 in the query. As a consequence, with a short keyword such as *avi*, a name extension for video files, we
131 almost only record queries in which it is the unique keyword, because otherwise it most likely is neither

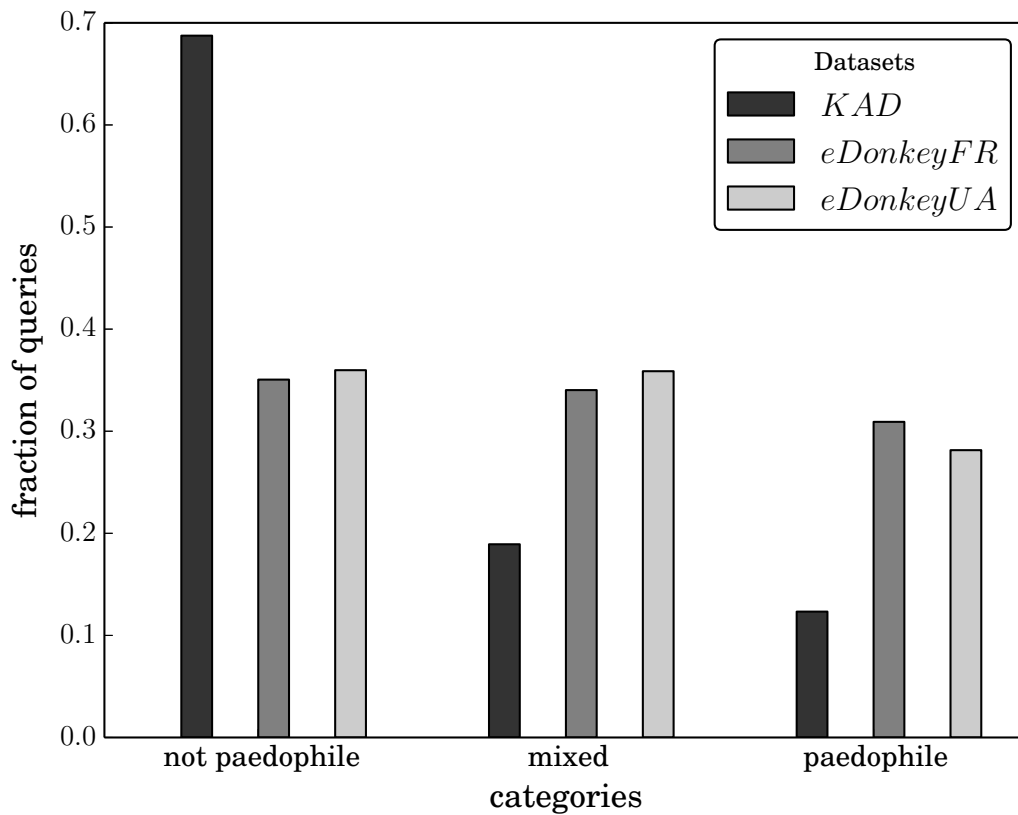
132 the longest nor the first word in any query. In order to obtain comparable datasets, we therefore limit our
 133 study to a subset of our datasets: the queries composed of exactly one word among the 72 keywords we
 134 monitor.

135 As a result of this construction process, we obtain three datasets, which we call *eDonkeyFR*,
 136 *eDonkeyUA* and *KAD*. They contain 241,152, 166,154 and 250,000 queries respectively, all consisting
 137 of a unique keyword from our list of 72 monitored keywords, which ensures that they are comparable.
 138 The server corresponding to the *eDonkeyFR* dataset is located in France, while the one corresponding to
 139 *eDonkeyUA* is in Ukraine. Their large sizes make us confident in the reliability of our statistical results
 140 presented hereafter.

141 5. Amount of paedophile queries in *eDonkey* versus *KAD*

142 The most straightforward way to compare the paedophile activity in different systems certainly is to
 143 compare the fraction of paedophile queries in each system. Figure 1 presents the fraction of queries for
 144 each category of keywords. This plot clearly shows that there are very distinct search behaviors in the
 145 two networks, since values obtained for the *paedophile* and *not paedophile* categories significantly differ
 146 between *KAD* and the two *eDonkey* datasets. More surprisingly, the fraction of paedophile queries is
 147 significantly lower in *KAD* than in *eDonkey* which is in sharp contradiction with previous intuition, as
 148 *KAD* is assumed to provide a higher level of anonymity. The plot also shows that values obtained for the
 149 two *eDonkey* servers are similar, which indicates that very different filtering policies have no significant
 150 influence on the amount of paedophile queries.

151 In order to gain a more detailed insight on this phenomenon, we study the frequencies of each keyword
 152 separately in the three datasets. As we want to explore possible correlations between the paedophile
 153 nature of a keyword and its frequency, we need a way to quantify the paedophile nature of a keyword.
 154 To do so, we use the 28-week dataset and the paedophile query detection tool from [4], which divides a
 155 dataset between *paedophile* and *not paedophile* queries (with a precision above 98% and a recall above
 156 75%). We denote by Q the whole dataset of queries, and by $Q(k)$ the set of queries containing a given
 157 keyword k . For each keyword k , we obtain $Q(k) = N(k) + P(k)$, where $N(k)$ and $P(k)$ are the subset
 158 of queries containing keyword k and tagged as *not paedophile* or *paedophile*, respectively. We then
 159 define the *paedophile coefficient* $\pi(k)$ of keyword k as: $\pi(k) = \frac{|P(k)|}{|Q(k)|}$. If all the queries with keyword
 160 k are paedophile queries, $\pi(k) = 1$, and if none of them are, $\pi(k) = 0$. All keywords in the *not*
 161 *paedophile* category have a *paedophile coefficient* below 0.006. For keywords in the *mixed* category,
 162 the *paedophile coefficient* is above 0.01 and below 0.4. All *paedophile* keywords have a *paedophile*
 163 *coefficient* above 0.885. Finally, we plot in Figure 2 the ratios $\frac{f_{eDonkeyFR}(k)}{f_{kad}(k)}$ and $\frac{f_{eDonkeyUA}(k)}{f_{kad}(k)}$, where
 164 $f_s(k)$ denotes the frequency of queries composed of keyword k in the dataset s , for each of our 72
 165 keywords. We rank keywords on the horizontal axis in increasing order of *paedophile coefficient*. The
 166 horizontal line represents $y = 1$, which enables a visual comparison of the values: if the point is below
 167 the line, then the keyword is more frequent in *KAD*, otherwise it is more frequent in the *eDonkey* dataset.
 168 This plot gives a clear evidence for a correlation between the paedophile nature of a keyword and its
 169 higher presence in *eDonkey* than in *KAD*. In addition, the frequencies in both *eDonkey* datasets are very
 170 similar for the vast majority of keywords.

Figure 1. Fraction of queries of each kind in our three datasets.

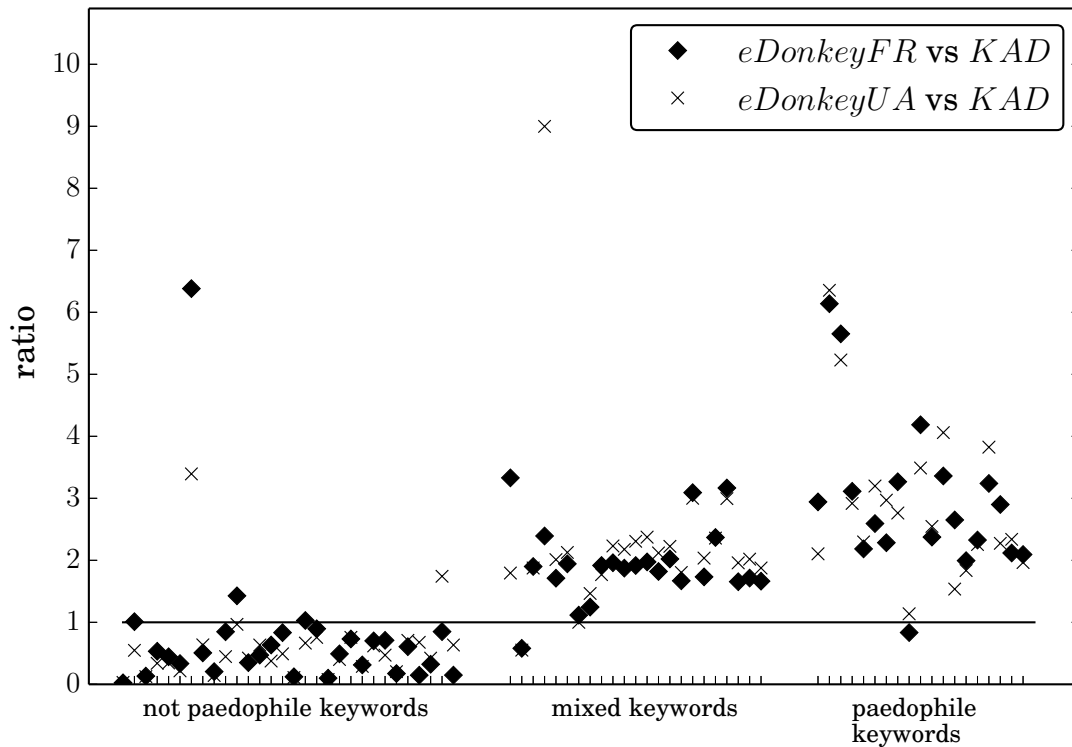
171 We therefore conclude that anonymity is not the prevailing factor when paedophile users choose a
 172 network, since neither the decentralised architecture of *KAD* nor the different filtering policies increase
 173 the frequency of paedophile queries. Instead, the frequency of paedophile queries is even higher in
 174 *eDonkey* than in *KAD*. Finding an explanation for this unexpected phenomenon is still an open question.
 175 The higher technical skills required to use *KAD* may be part of the explanation. Users may also search
 176 content on *eDonkey* while protecting their privacy with other tools, such as Virtual Private Networks
 177 or TOR [20]. The fact that in *KAD* search requests are sent over UDP and cannot benefit from TOR
 178 anonymisation could explain the difference in the network usage.

179 6. Ages indicators in queries

180 A way to gain more insight on observed paedophile activity is to study the distribution of age
 181 indicators in queries [11]. Notice that age indicators are sometimes used in other contexts than
 182 paedophile activity, especially when parents seek content suitable for children of a certain age. However,
 183 one can observe on Figure 2 that ages indicators have similar behavior to those obtained for the
 184 *paedophile* group, and are therefore closely related to the topic.

185 We plot the distribution of age indicators on Figure 3: for each integer n lower than 17, we plot the
 186 number of queries of the form nyo in each dataset (yo stands for *years old*). The three plots have similar
 187 shape, with mostly increasing values from 1 to 10, a little drop at 11, a peak at 12 and a fall from 13
 188 to 16. These values for *KAD* are below the values for the *eDonkey* servers, which is due to the fact that

Figure 2. Ratio of keyword frequencies in *eDonkey* vs *KAD*. Keywords are ranked in increasing order of *paedophile coefficient*. Points above the $y = 1$ horizontal line indicates keywords more frequent in the corresponding *eDonkey* dataset; below the line keywords are more frequent in *KAD*.



189 this dataset is a bit smaller than others and that paedophile queries are rarer in it. The key point here is
 190 that the distributions are very similar in all three datasets. This indicates that, although the *amount* of
 191 paedophile activity varies between systems, its nature is similar, at least regarding ages.

192 7. Quantifying paedophile activity in *KAD*

193 In [4], the authors establish a method to quantify the fraction of paedophile queries in *eDonkey*. It
 194 relies on a tool able to accurately tag queries as paedophile or not, and on an estimate of the error rate
 195 of this tool. Such an approach cannot directly be applied to *KAD* though, as only a small (and biased)
 196 fraction of all queries may be observed in this system. We however show in this section how to derive
 197 the fraction of paedophile queries in *KAD* from the one in *eDonkey*.

198 In a given system, *eDonkey* or *KAD* here, we consider different sets of queries and we denote by Q the
 199 set of all queries, P the subset of paedophile queries in Q , \bar{Q} the subset of queries composed of one word
 200 among the 72 monitored keywords, \bar{P} the subset of paedophile queries with one word, *i.e.* consisting of
 201 one of the 19 monitored paedophile keywords (and so: $\bar{P} = \bar{Q} \cap P$). Figure 4 illustrates our notations.

202 In both our *eDonkey* measurements, $|P|$ and $|Q|$ may be directly estimated, as shown in [4], and one
 203 can then obtain the fraction $\frac{|P|}{|Q|}$ of paedophile queries in the dataset. We give the results for our two
 204 measurements in Table 1. On the contrary, in *KAD*, one may only estimate $|\bar{P}|$ and $|\bar{Q}|$.

Figure 3. Distribution of age indicators in our three datasets.

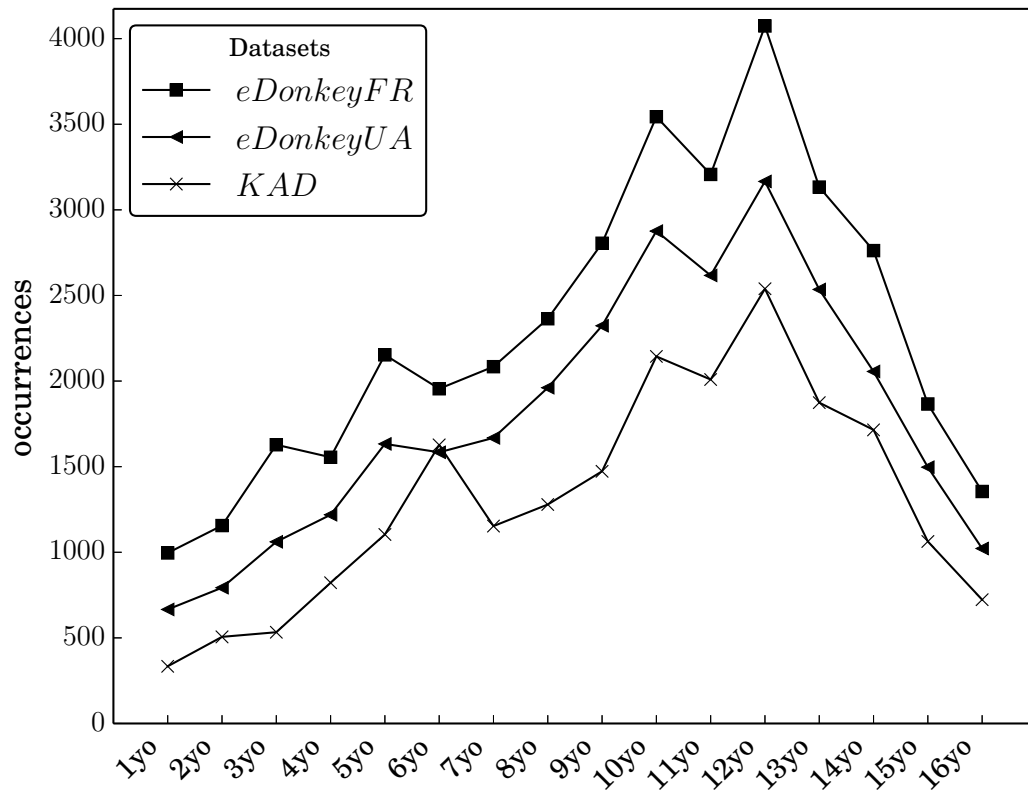


Figure 4. The different sets of queries defined for each dataset.

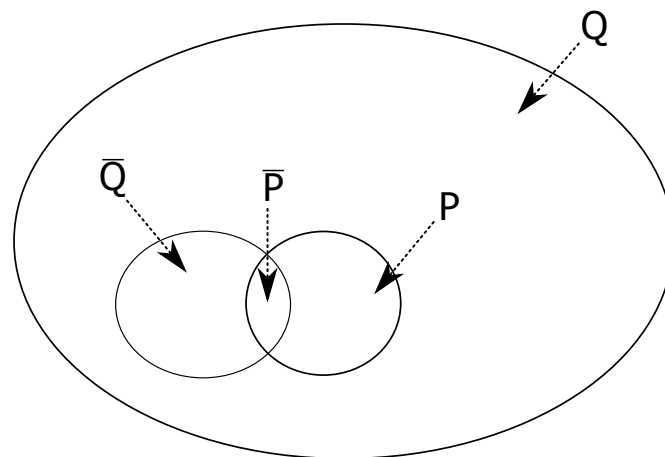


Table 1. Results for our three datasets.

dataset	$\frac{ P }{ Q }$	$ \bar{P} $	$ \bar{Q} $	α	β
<i>edonkeyFR</i>	$2.554 \cdot 10^{-3}$	74,557	241,152	$1.431 \cdot 10^{-3}$	0.2502
<i>edonkeyUA</i>	$2.668 \cdot 10^{-3}$	46,763	166,154	$1.538 \cdot 10^{-3}$	0.2251
<i>KAD</i>	n/a	30,821	250,000	n/a	n/a

205 However, we define $\alpha = \frac{|\bar{Q}| - |\bar{P}|}{|Q| - |P|}$ and $\beta = \frac{|\bar{P}|}{|P|}$, which capture the probability for a non paedophile
 206 query, respectively paedophile, to make a query of one word among one of our monitored keywords.
 207 Given the definition of α and β , there is no *a priori* reason to assume that they have significantly different
 208 values between *eDonkey* and *KAD*. From the definitions of α and β , we have:

$$\begin{aligned} \alpha = \frac{|\bar{Q}| - |\bar{P}|}{|Q| - |P|} &\implies |Q| = \frac{\alpha|P| + |\bar{Q}| - |\bar{P}|}{\alpha} \\ \beta = \frac{|\bar{P}|}{|P|} &\implies |P| = \frac{|\bar{P}|}{\beta} \end{aligned}$$

209 Then, the following expression holds:

$$\begin{aligned} \frac{|P|}{|Q|} &= \frac{|\bar{P}|}{\beta} \times \frac{\alpha}{\alpha|P| + |\bar{Q}| - |\bar{P}|} \\ &= \frac{\alpha|\bar{P}|}{\beta|\bar{Q}| + (\alpha - \beta)|\bar{P}|} \end{aligned} \quad (1)$$

210 We now use expression (1) to infer the fraction of paedophile queries that were submitted in the *KAD*
 211 P2P network during our experiment. Using the values from Table 1 and the average values of α and β
 212 between our *eDonkey* datasets, we obtain:

$$\frac{|P|}{|Q|} \approx 0.087\% \pm 0.008$$

213 This value is of similar magnitude to the one of *eDonkey* (approx. 0.25%) but close to three times lower.

214 This estimation of $\frac{|P|}{|Q|}$ relies on the value of α . One may wonder whether the choice of keywords
 215 from which we built $\bar{Q} \setminus \bar{P}$ has a significant impact on the estimated value of $\frac{|P|}{|Q|}$ in *KAD*. We check this
 216 as follows: we randomly select 1,000 subsets of 26 keywords out of the 53 keywords which compose
 217 the queries in $\bar{Q} \setminus \bar{P}$. We then compute, for each subset, the number of queries consisting of exactly
 218 one of those keywords and the resulting value of alpha. For *eDonkeyFR*, we obtain an average value of
 219 $\bar{\alpha} = 0.000889$ (minimum: 0.000256, maximum: 0.00153, and 90% of the values in [0.000463;0.00133]).
 220 For *eDonkeyUA*, we obtain an average value of $\bar{\alpha} = 0.00105$ (minimum: 0.000352, maximum: 0.00172,
 221 and 90% of the values in [0.00062;0.00148]). This means that we would obtain very similar results with
 222 26 keywords only and so we may be confident in our estimate obtained with 53 keywords.

223 8. Conclusion

224 We performed a comparative study of two large-scale peer-to-peer networks, *KAD* and *eDonkey*,
 225 with regards to the queries related to child pornography. We designed a methodology to collect and
 226 process datasets allowing to compare them in a relevant manner. We obtained the counter-intuitive result
 227 that paedophile keywords are significantly more present in *eDonkey* than in *KAD*, despite the higher
 228 anonymity level it provides. On the contrary, our study of age indicators in queries showed that the
 229 nature of paedophile queries is similar in these systems. We finally established the first estimate of
 230 the fraction of paedophile queries in *KAD*. We obtained a value close to 0.09%, which is of the same
 231 magnitude but significantly lower than in *eDonkey* (0.25%).

232 Our approach here is similar to the one used in [4]: we focus on search queries, which help to grasp
233 the demand for paedophile material. It differs from [14,15] which focused on the files. In P2P networks
234 such as *eDonkey* and *KAD*, a single file may have several names, most of which describe its content.
235 However, filenames are prone to pollution and often exhibit keywords unrelated to the real content of the
236 file, for instance a paedophile file may have a non-paedophile name [21,22]. Thus, estimations relying
237 on specific filenames are likely to underestimate the true extent of child pornography distribution, while
238 estimations relying on file-based honeypots are likely to overestimate the demand due to false-positive
239 download requests. Query-based estimations using search requests do not suffer from such a bias, but, as
240 mentioned earlier, may be impacted by repetitive queries from regular paedophile users. Nevertheless,
241 both the considered P2P networks (*KAD* and *eDonkey*) should be equally affected, thus making their
242 comparison valid to this regard.

243 Our contributions open various directions for future work. In particular, our methodology may be
244 applied to compare other systems, and our datasets may be used to perform either deeper analyses on
245 paedophile activity or on general search engine behaviors.

246 Conflicts of Interest

247 The authors declare no conflict of interest.

248 References

- 249 1. Commission, E. Safer Internet Programme: Empowering and Protecting Children Online, 2010.
250 http://ec.europa.eu/information_society/activities/sip/index_en.htm.
- 251 2. McCullagh, D. RIAA: Child porn rife on P2P networks. *CNet* **2003**. [http://news.cnet.com/
252 RIAA-Child-porn-rife-on-P2P-networks/2100-1028_3-5073817.html](http://news.cnet.com/RIAA-Child-porn-rife-on-P2P-networks/2100-1028_3-5073817.html).
- 253 3. Hughes, D.; Walkerdine, J.; Coulson, G.; Gibson, S. Is Deviant Behavior the Norm on P2P
254 File-Sharing Networks? *IEEE Distributed Systems Online* **2006**, *7*.
- 255 4. Latapy, M.; Magnien, C.; Fournier, R. Quantifying paedophile activity in a large P2P system.
256 *Information Processing & Management* **2012**.
- 257 5. Demographics, S.N. Global Internet Phenomena Report: Spring 2011, 2012.
- 258 6. Chopra, M.; Martin, M.; Rueda, L.; Hung, P.C.K. Toward New Paradigms to Combating Internet
259 Child Pornography. *Electrical and Computer Engineering, 2006. CCECE '06. Canadian
260 Conference on, 2006*, pp. 1012–1015.
- 261 7. Shupo, A.; Martin, M.V.; Rueda, L.; Bulkan, A.; Chen, Y.; Hung, P.C. Toward efficient detection
262 of child pornography in the network infrastructure. *IADIS International Journal on Computer
263 Science and Information Systems* **2006**, *1*.
- 264 8. Panchenko, A.; Beaufort, R.; Naets, H.; Fairon, C. Towards Detection of Child Sexual Abuse
265 Media: Categorization of the Associated Filenames. In *Advances in Information Retrieval*;
266 Serdyukov, P.; Braslavski, P.; Kuznetsov, S.; Kamps, J.; R uger, S.; Agichtein, E.; Segalovich,
267 I.; Yilmaz, E., Eds.; Springer Berlin Heidelberg, 2013; Vol. 7814, *Lecture Notes in Computer
268 Science*, pp. 776–779.

- 269 9. Kontostathis, A.; Garron, A.; Reynolds, K.; West, W.; Edwards, L. Identifying Predators Using
270 ChatCoder 2.0. CLEF (Online Working Notes/Labs/Workshop), 2012.
- 271 10. Pendar, N. Toward Spotting the Pedophile Telling victim from predator in text chats. *Semantic*
272 *Computing*, 2007. ICSC 2007. International Conference on, 2007, pp. 235–241.
- 273 11. Steel, C.M. Child pornography in peer-to-peer networks. *Child Abuse & Neglect* **2009**.
- 274 12. Prichard, J.; Watters, P.A.; Spiranovic, C. Internet subcultures and pathways to the use of child
275 pornography. *Computer Law & Security Review* **2011**, 27, 585–600.
- 276 13. Rutgaizer, M.; Shavitt, Y.; Vertman, O.; Zilberman, N. Detecting Pedophile Activity in BitTorrent
277 Networks. PAM; Taft, N.; Ricciato, F., Eds. Springer, 2012, Vol. 7192, *Lecture Notes in*
278 *Computer Science*, pp. 106–115.
- 279 14. Wolak, J.; Liberatore, M.; Levine, B.N. Measuring a year of child pornography trafficking by US
280 computers on a peer-to-peer network. *Child Abuse & Neglect* **2013**, p. 10.
- 281 15. Hurley, R.; Prusty, S.; Soroush, H.; Walls, R.J.; Albrecht, J.; Cecchet, E.; Levine, B.N.;
282 Liberatore, M.; Lynn, B.; Wolak, J. Measurement and Analysis of Child Pornography Trafficking
283 on P2P Networks (Runner-Up, Best Paper Award). Proc. Intl. World Wide Web Conference
284 (WWW), 2013.
- 285 16. Locher, T.; Mysicka, D.; Schmid, S.; Wattenhofer, R. A Peer Activity Study in eDonkey and Kad.
286 International Workshop on Dynamic Networks: Algorithms and Security (DYNAS), Wroclaw,
287 Poland, 2009.
- 288 17. Heckmann, O.; Bock, A.; Mauthe, A.; Steinmetz, R. The eDonkey File-Sharing Network. GI
289 Jahrestagung (2); Dadam, P.; Reichert, M., Eds. GI, 2004, Vol. 51, *LNI*, pp. 224–228.
- 290 18. Aidouni, F.; Latapy, M.; Magnien, C. Ten weeks in the life of an eDonkey server. *International*
291 *Workshop on Hot Topics in P2P Systems* **2009**.
- 292 19. Cholez, T.; Chrisment, I.; Festor, O. Monitoring and Controlling Content Access in KAD.
293 International Conference on Communications - ICC 2010; IEEE, , 2010.
- 294 20. The Tor Project, Inc. TOR project, 2012. <https://www.torproject.org/>.
- 295 21. Montassier, G.; Cholez, T.; Doyen, G.; Khatoun, R.; Chrisment, I.; Festor, O. Content Pollution
296 Quantification in Large P2P networks : a Measurement Study on KAD. 11th IEEE International
297 Conference on Peer-to-Peer Computing (IEEE P2P'11); IEEE Communications Society, , 2011;
298 pp. 30–33. Projet ANR MAPE et Projet GIS 3SGS ACDAP2P.
- 299 22. Brennan, M.; Hammond, S. Complete critical literature review. Technical report, iCOP:
300 Identifying and Catching Originators in P2P Networks, 2011.