



Laboratoire d'Informatique de Paris 6
University Pierre & Marie Curie
104, avenue du Président Kennedy 75016 Paris

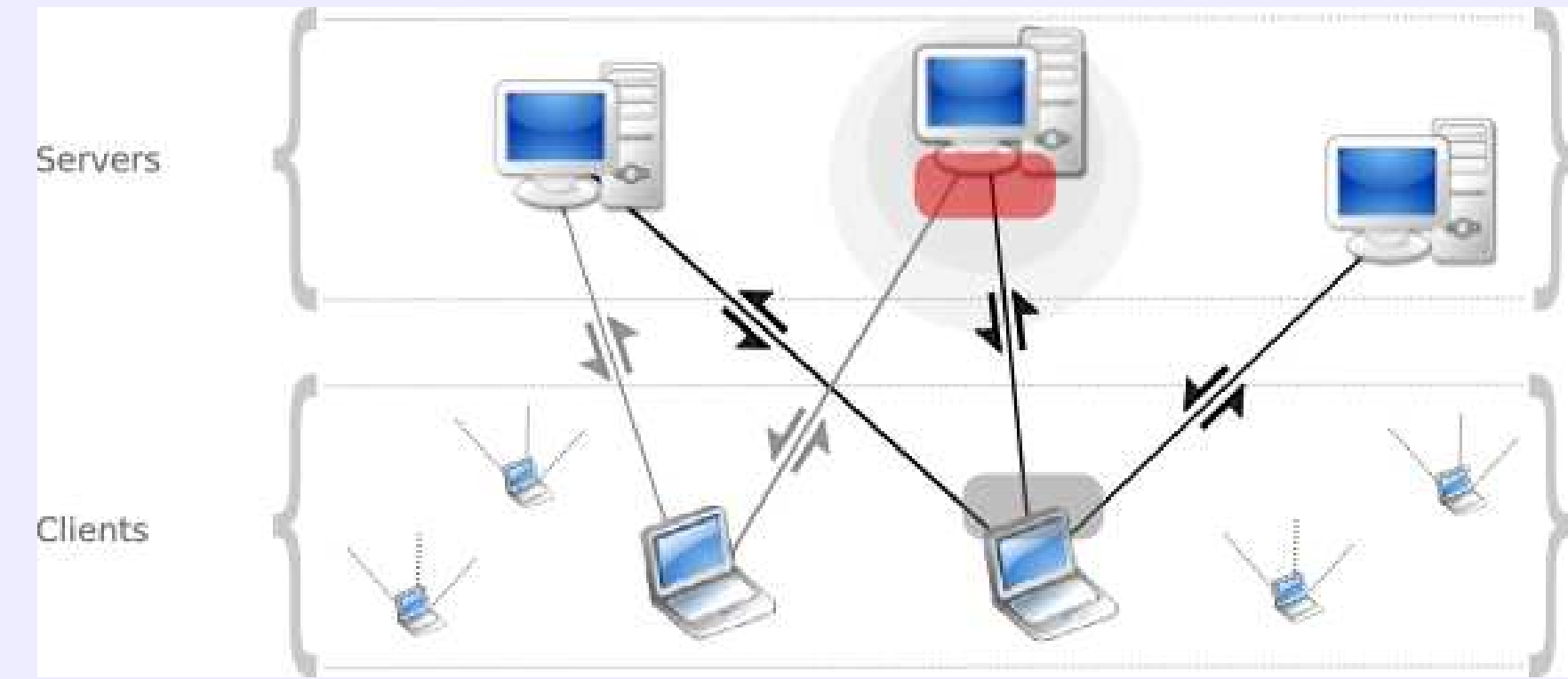
Tracing paedophile eDonkey users through keyword-based queries

Raphaël Fournier, Guillaume Valadon, Clémence Magnien and Matthieu Latapy

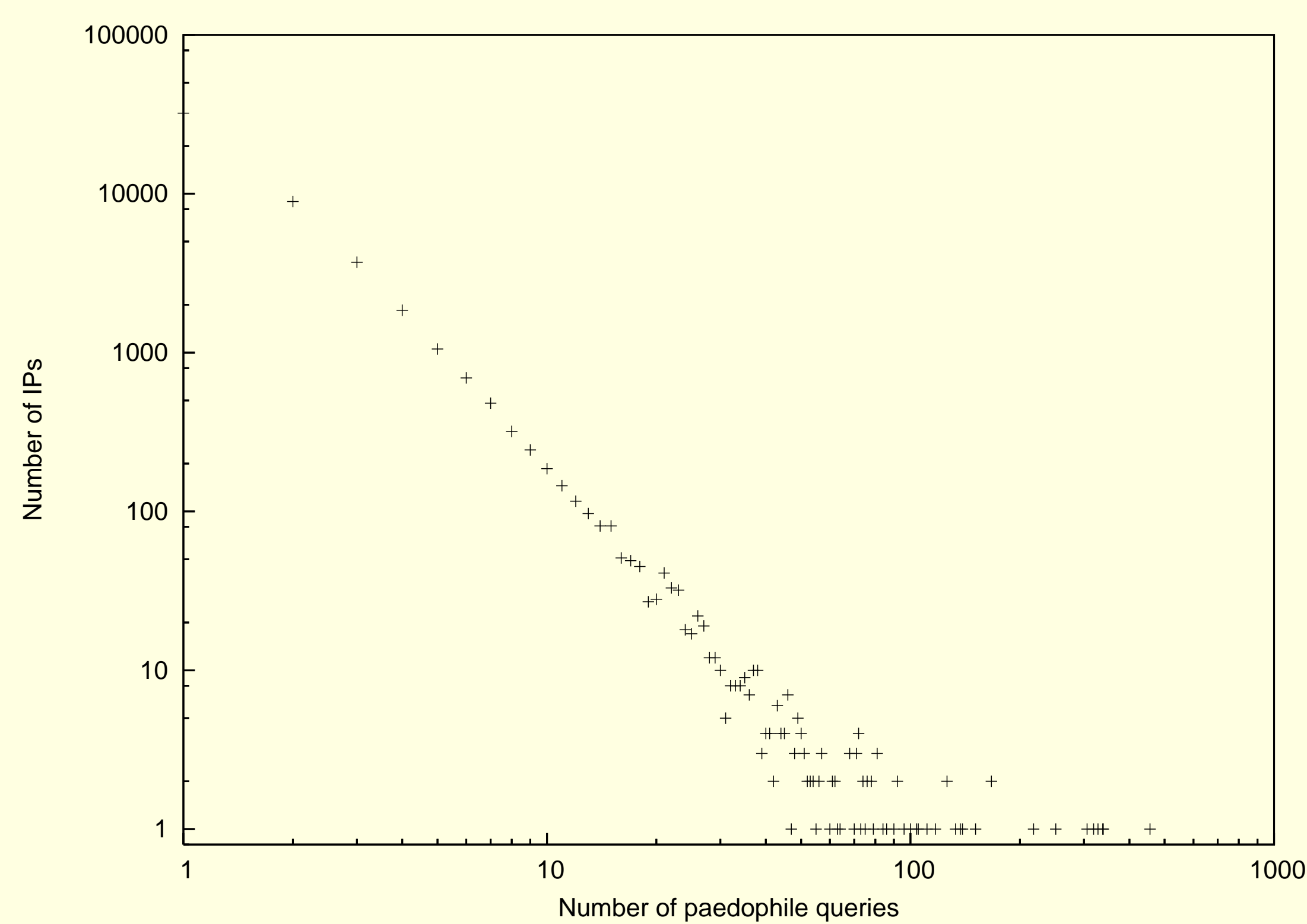


{prenom.nom}@lip6.fr
<http://www.complexnetworks.fr>

- data: keyword-based content queries submitted to an eDonkey server
- parameters: 21 paedophile keywords, 10 weeks 24/7 recording
- model: paedophile query = at least one paedophile keyword
paedophile user = at least one paedophile query
- goal: accurate facts about paedophile users on an eDonkey server



Number of queries by IP

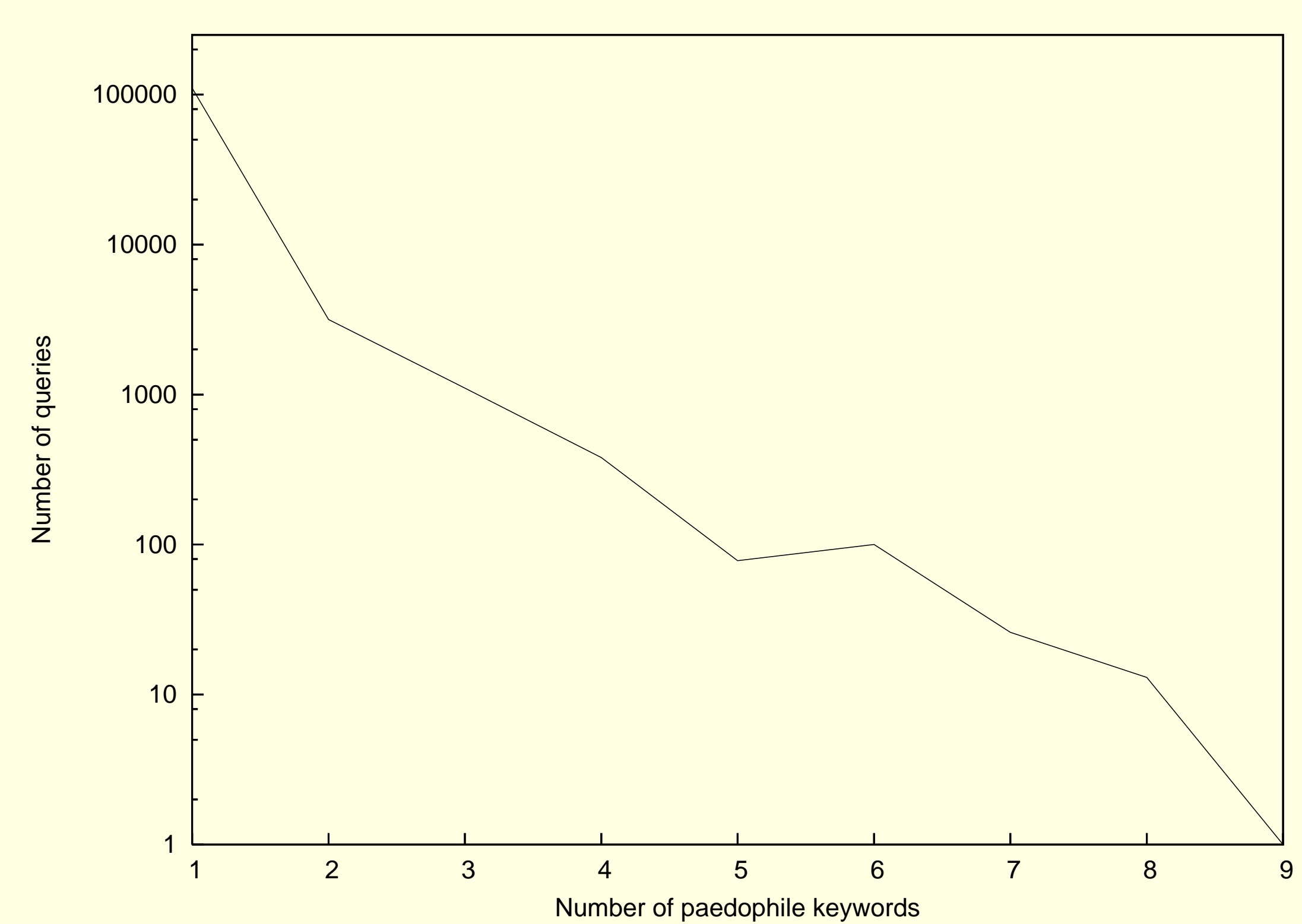


- highly heterogeneous distribution, power-law
- 66% of the paedophile IPs submit only one query

Open issues:

- upper and lower thresholds of queries to be considered paedophile
- automatic re-submission of queries

Number of keywords in queries

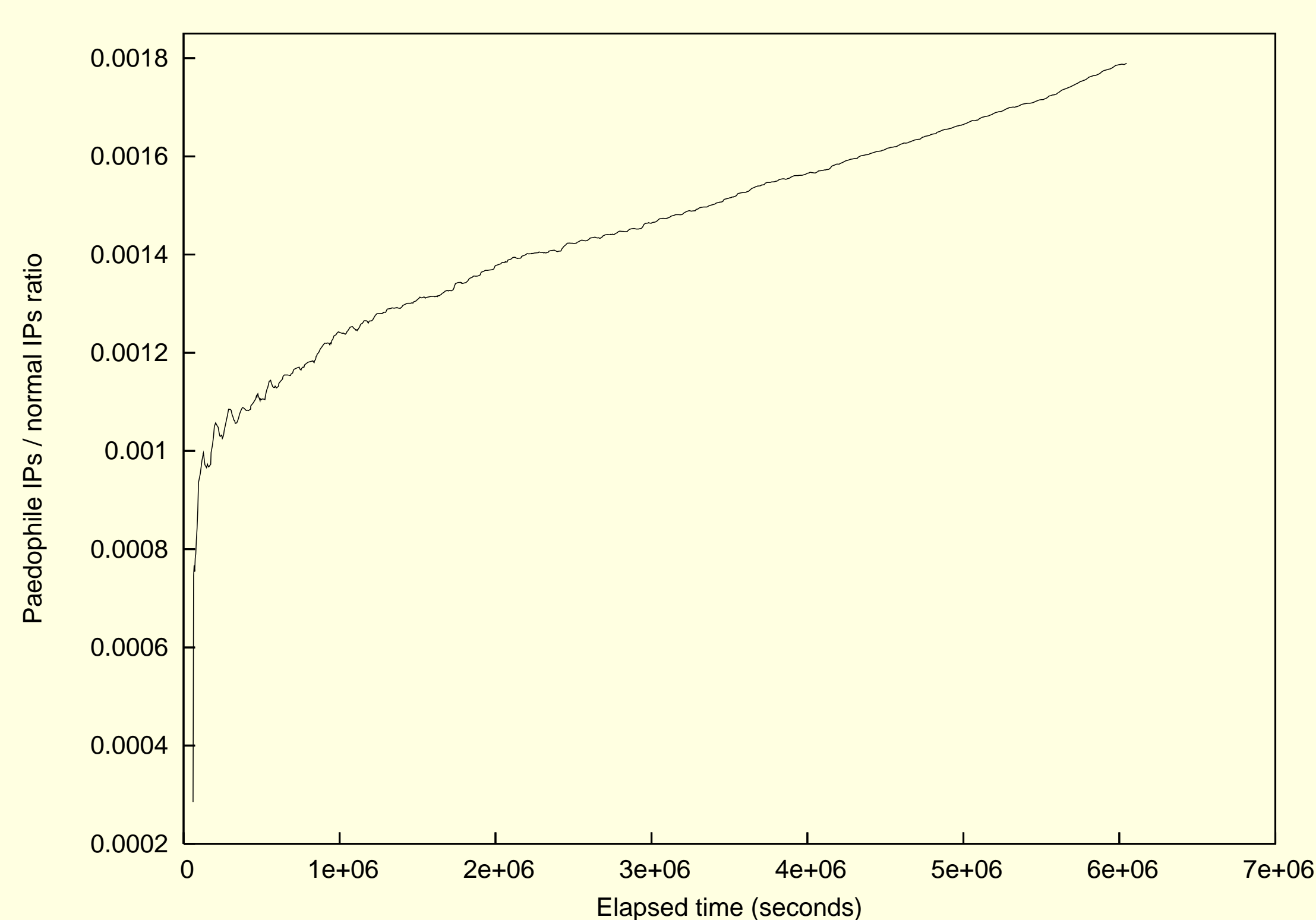


- variety of paedophile keywords contained in queries
- 96% of the paedophile queries contain one keyword

Open issues:

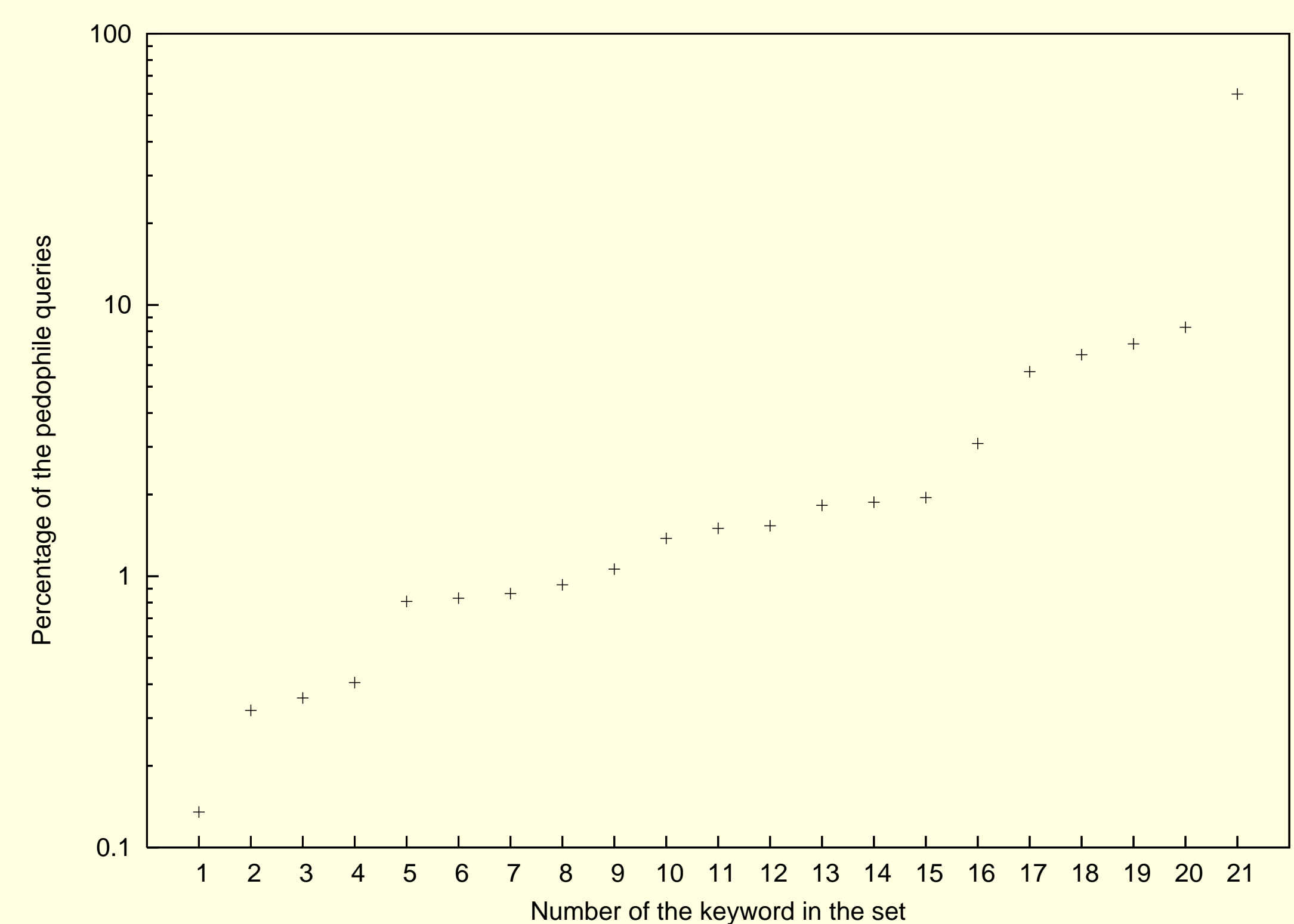
- some queries with up to 9 keywords: is it a human behavior?
- should we eliminate queries with more than x keywords?

Paedophile IPs / normal IPs ratio



- once an IP makes a paedophile query it remains paedophile till the end
- new paedophile and non-paedophile IPs continuously discovered

Total number of queries by keyword



- paedophile keywords, sorted by frequency of appearance

Open issue:

- one keyword in 60% of the paedophile queries: may be not-so-specific

Conclusion

- some keywords may be too common
- eDonkey clients may have unexpected behaviours
- one query over 1,000 is paedophile

Future work

- study combination of keywords
- consider ages as paedophile keywords
- refine the definition of a paedophile query