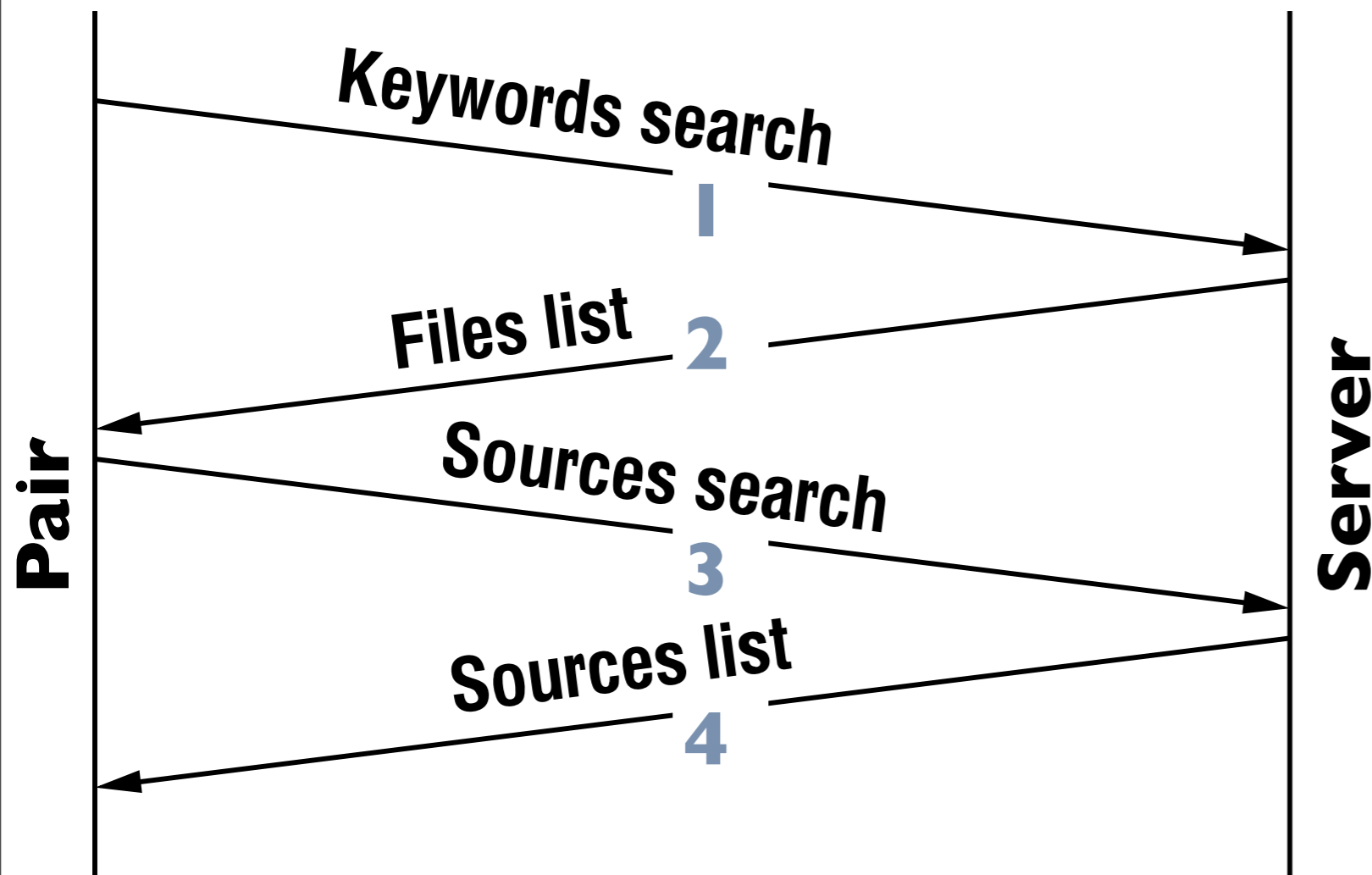


Echange pair-serveur: la recherche de fichier



1. "music mp3"

2. nom:ACDC.mp3

ID:101050

nom.mp3

ID:B16V100

3. ID:B16V100

4. 192.168.0.2

1.2.3.4

Echange pair-serveur: la recherche de fichier

1. "music mp3"

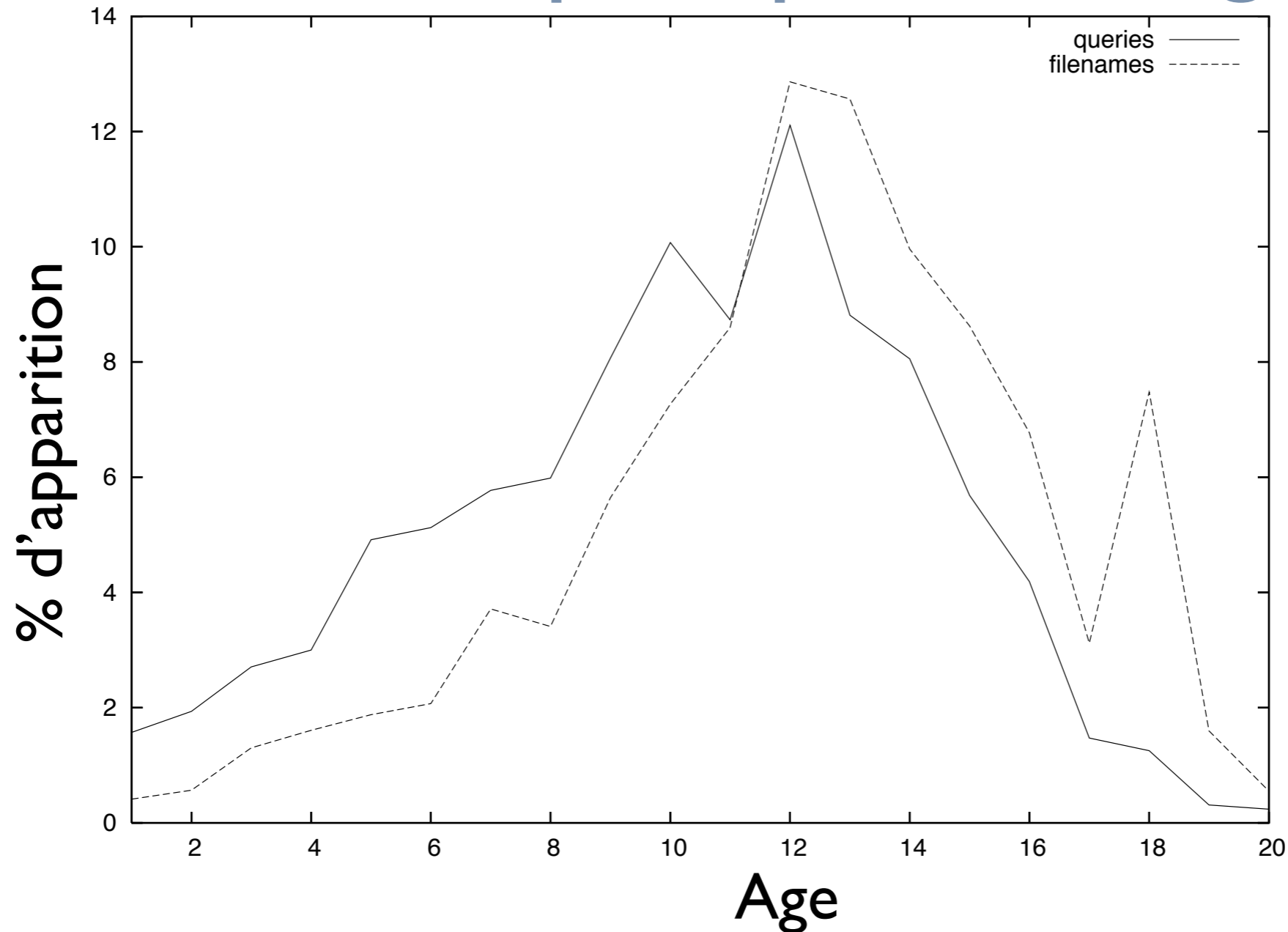
2. nom:ACDC.mp3

Les données

- 10 semaines de mesure
 - ~500 Go de XML compressés
 - ~ 10 milliards de messages
 - ~ 90 millions de pairs
 - ~ 280 millions de fichiers différents
- ➔ adresses IP anonymisées
- ➔ mots anonymisés (< 100 fois dans les données)

1.2.3.4

Contenu pédophile: les ages



- chaînes de caractères contenant des ages
- requêtes ciblent des ages plus jeunes que les fichiers
- plus de demande que d'offre

Mots clés des fichiers: représentation sous forme d'un graphe

Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

Mots clés des fichiers: représentation sous forme d'un graphe

les betteraves tueur de coiffeur mp3

Dionysos coiffeur d oiseaux mp3

Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

Dionysos

coiffeur

d

oiseaux

mp3

Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

Dionysos

coiffeur

d

oiseaux

mp3

Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

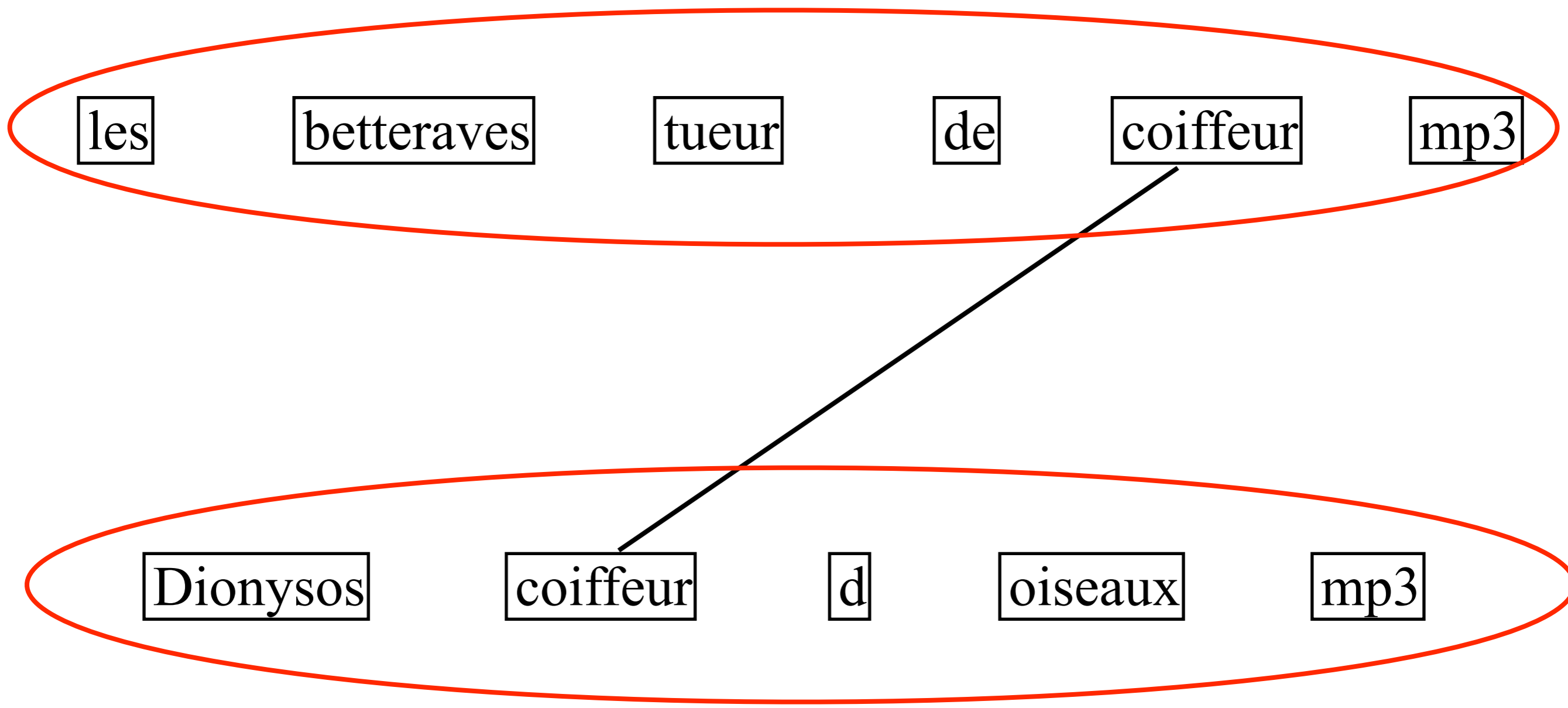
Dionysos

coiffeur

d

oiseaux

mp3



Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

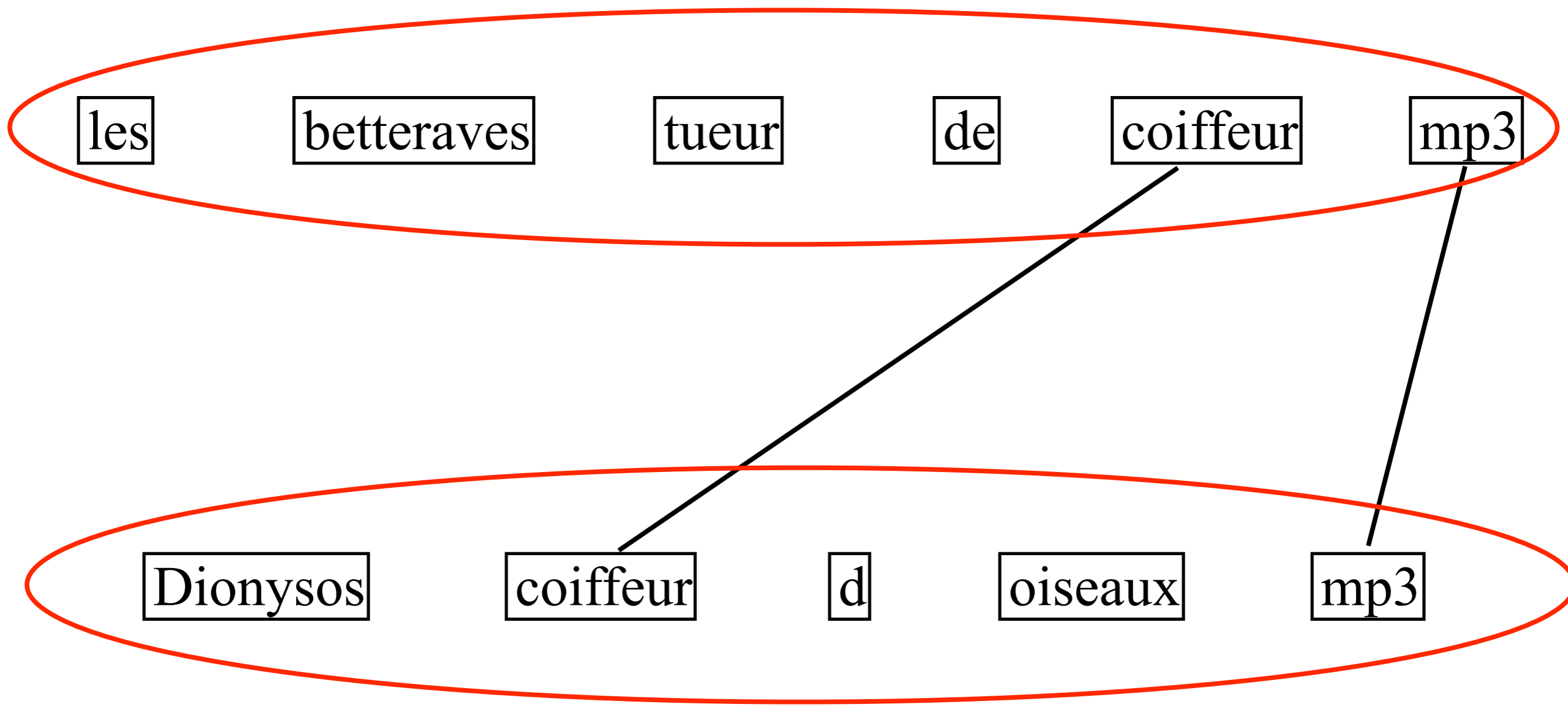
Dionysos

coiffeur

d

oiseaux

mp3



Mots clés des fichiers: représentation sous forme d'un graphe

les

betteraves

tueur

de

coiffeur

mp3

Graphe des mots dans les fichiers

- 1 693 791 de noeuds

- 73 422 135 de liens

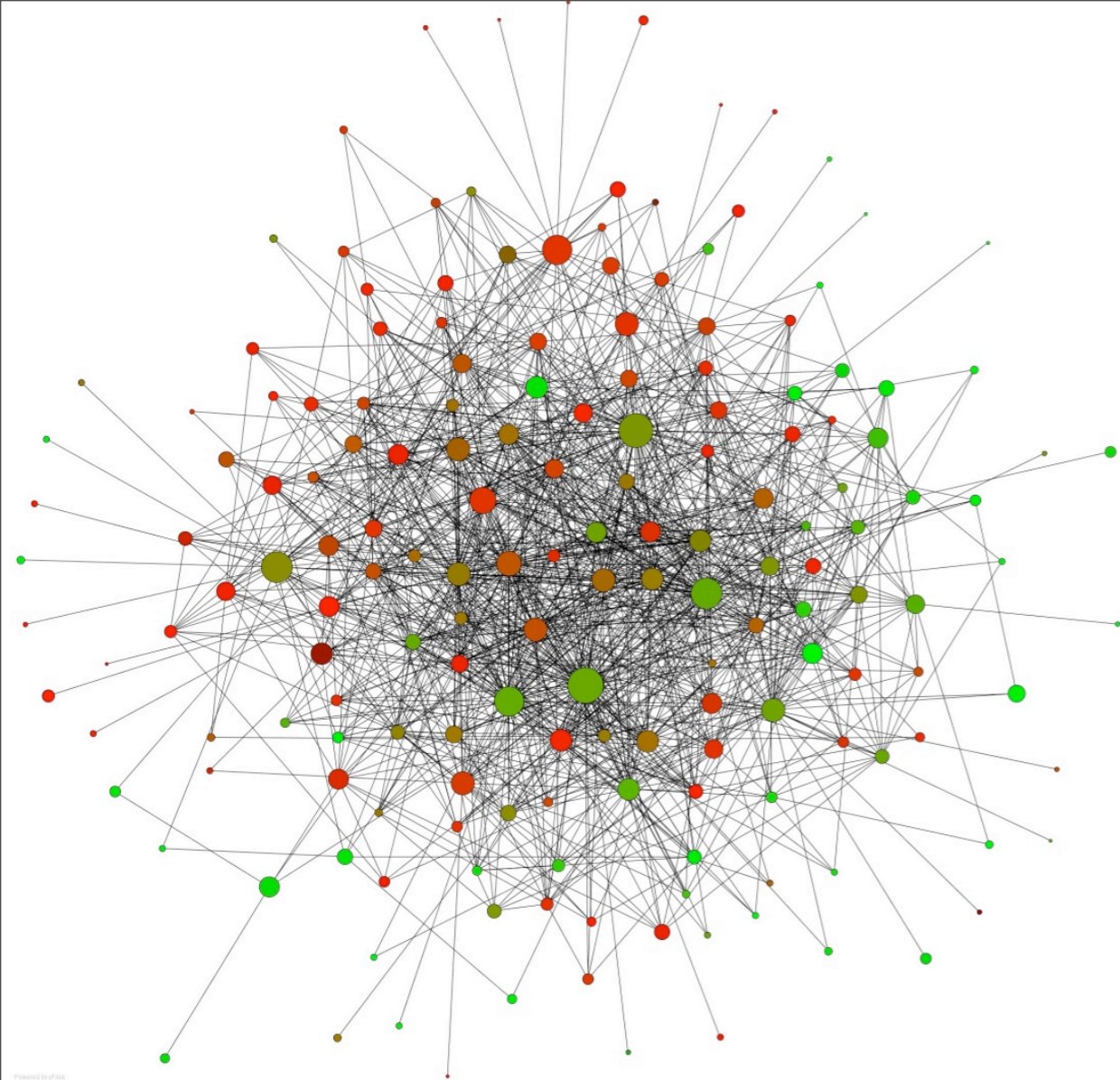
Dionysos

coiffeur

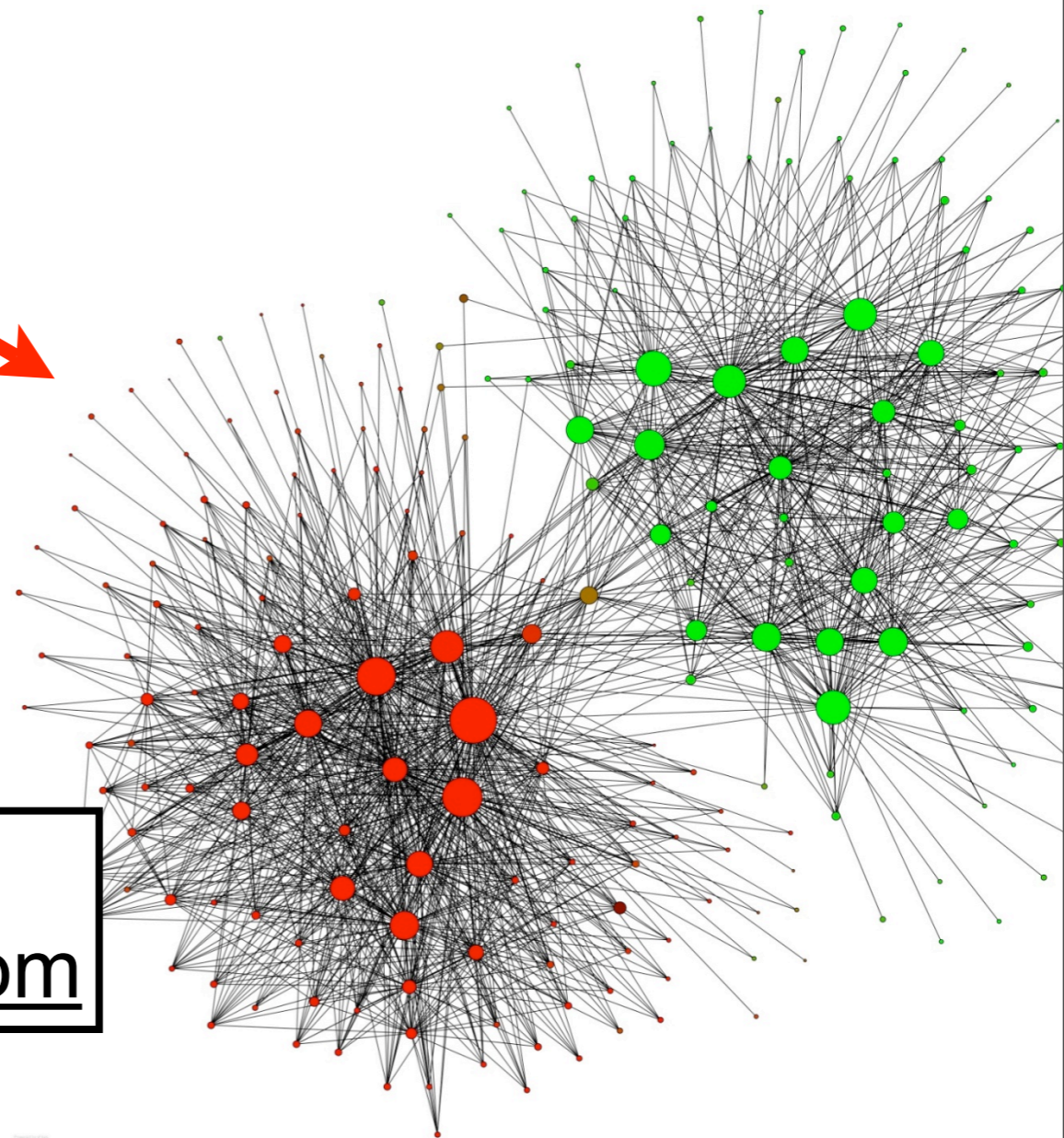
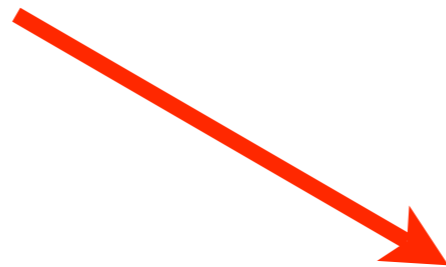
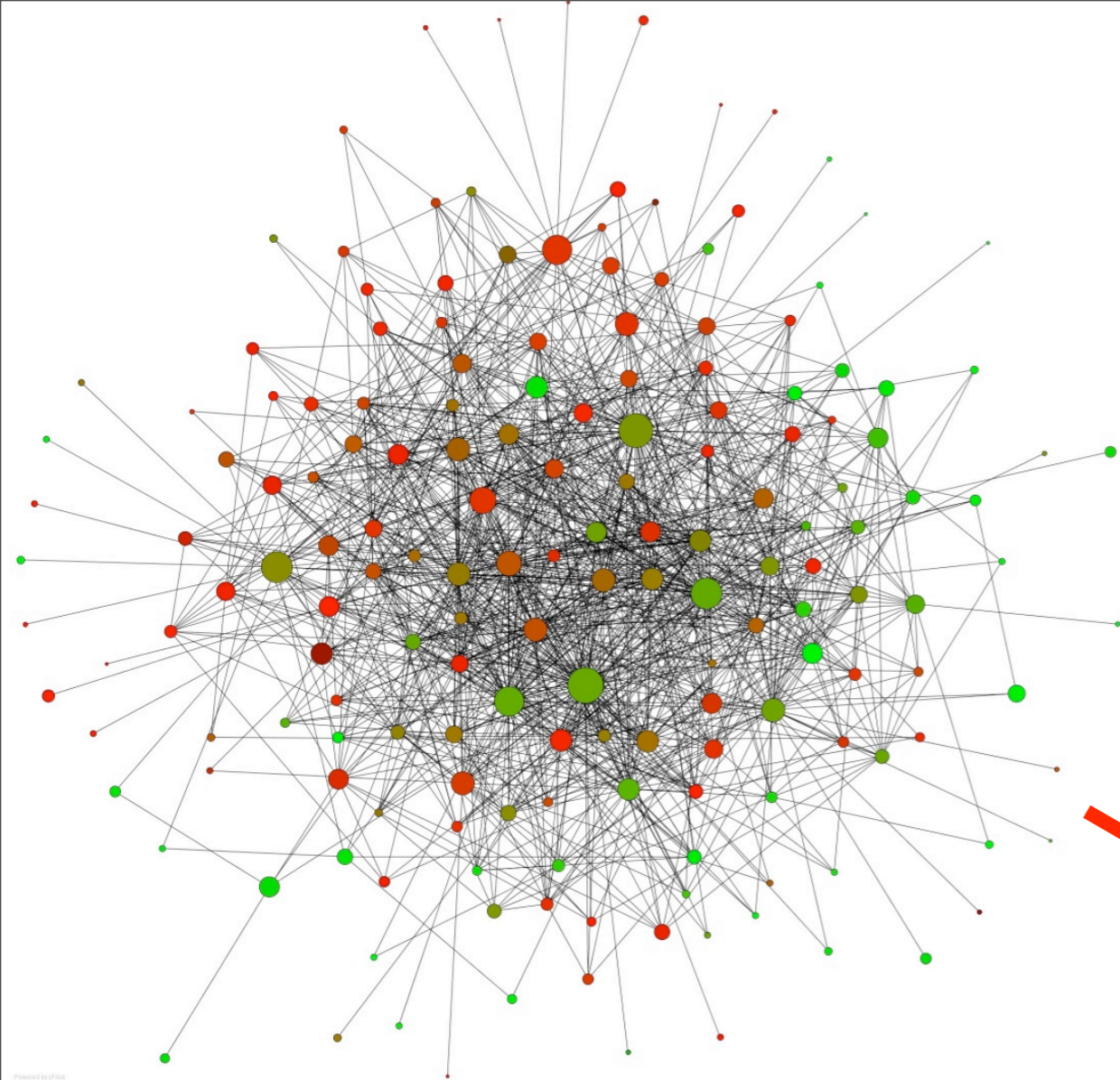
d

oiseaux

mp3



Détection de communautés
<http://findcommunities.googlepages.com>



Détection de communautés
<http://findcommunities.googlepages.com>



Graphe des mots dans les fichiers

Après 5 niveaux de *découpage*:

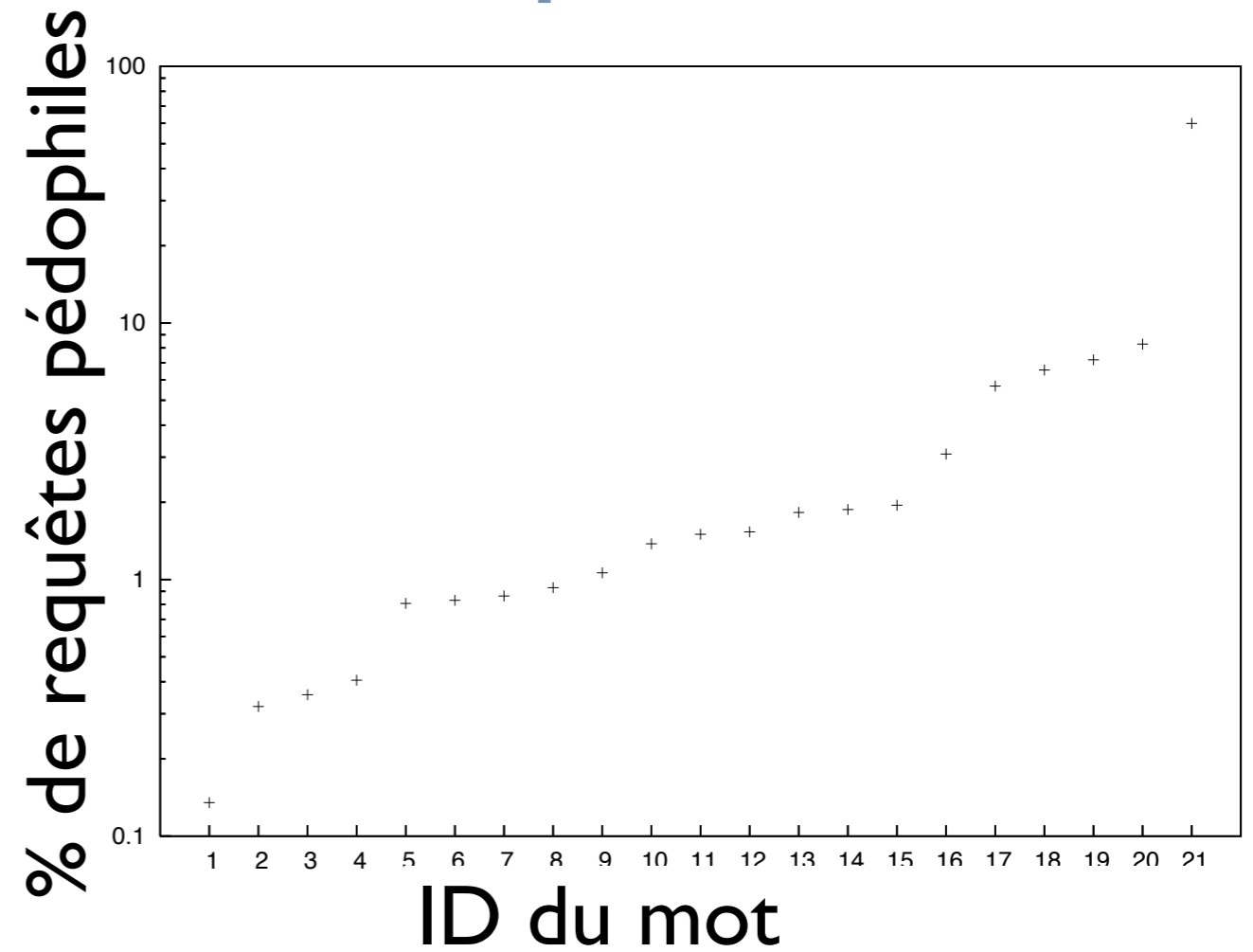
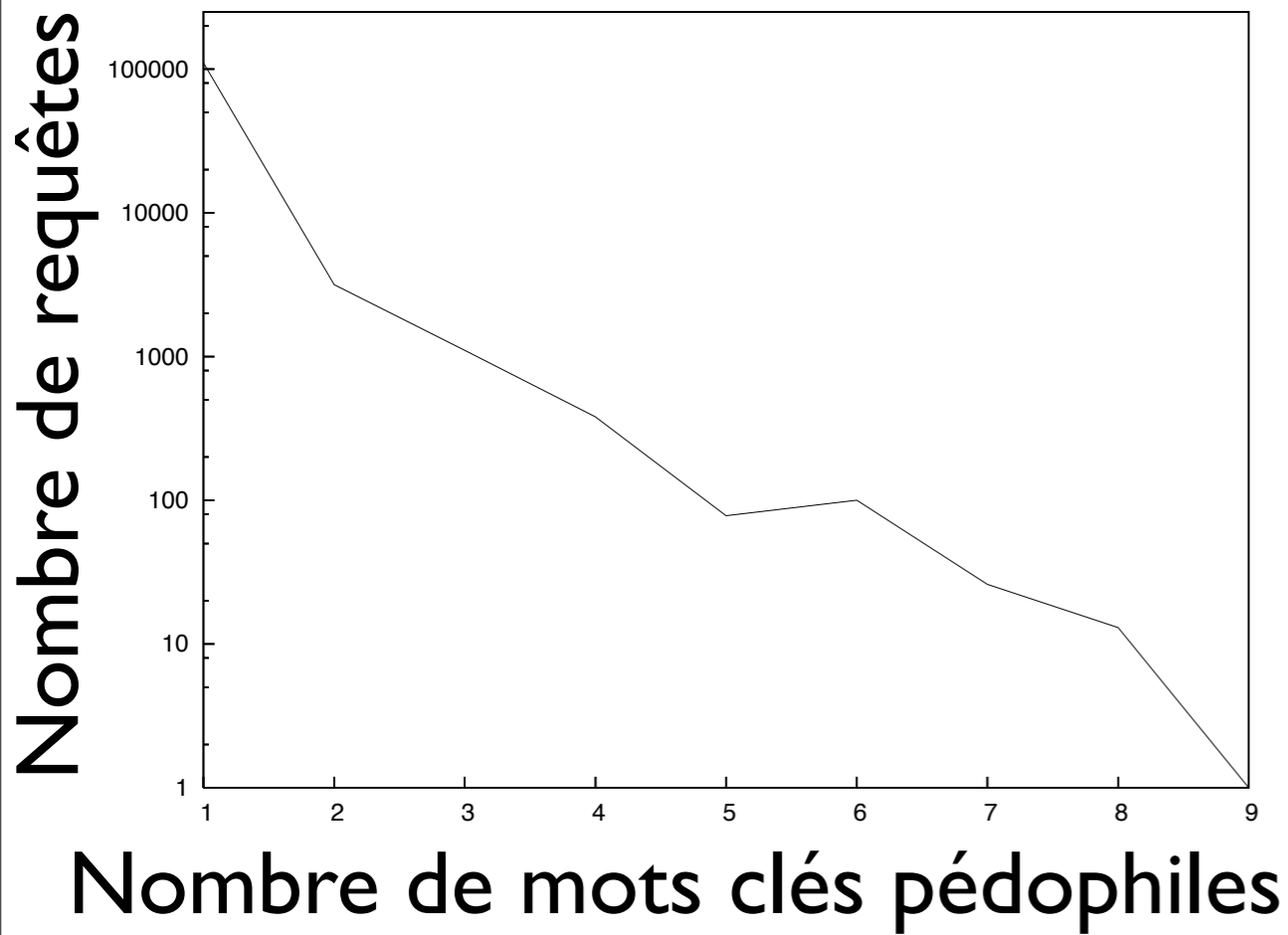
- communauté de 189 mots (**II connus**)



Détection de communautés

<http://findcommunities.googlepages.com>

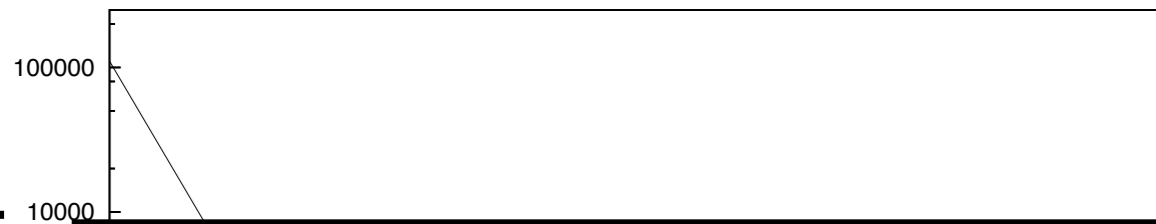
Mots clés dans les requêtes



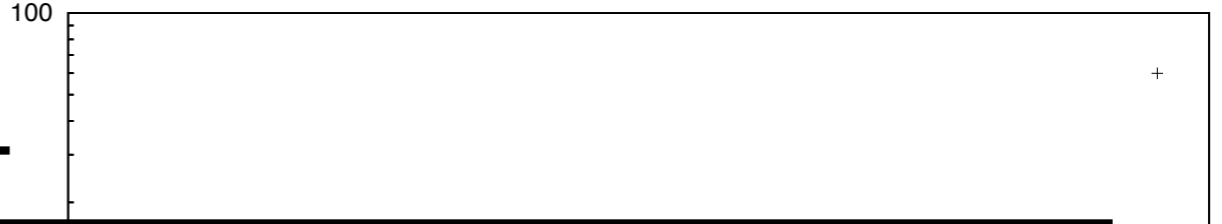
- 21 mots clés pédophiles
- requête pédophile si elle contient au moins 1 mot clé
 - 94% contiennent seulement un mot clé pedophile
 - le mot clé 21 est utilisé dans 60% des requêtes

Mots clés dans les requêtes

Nombre de requêtes



pedophiles



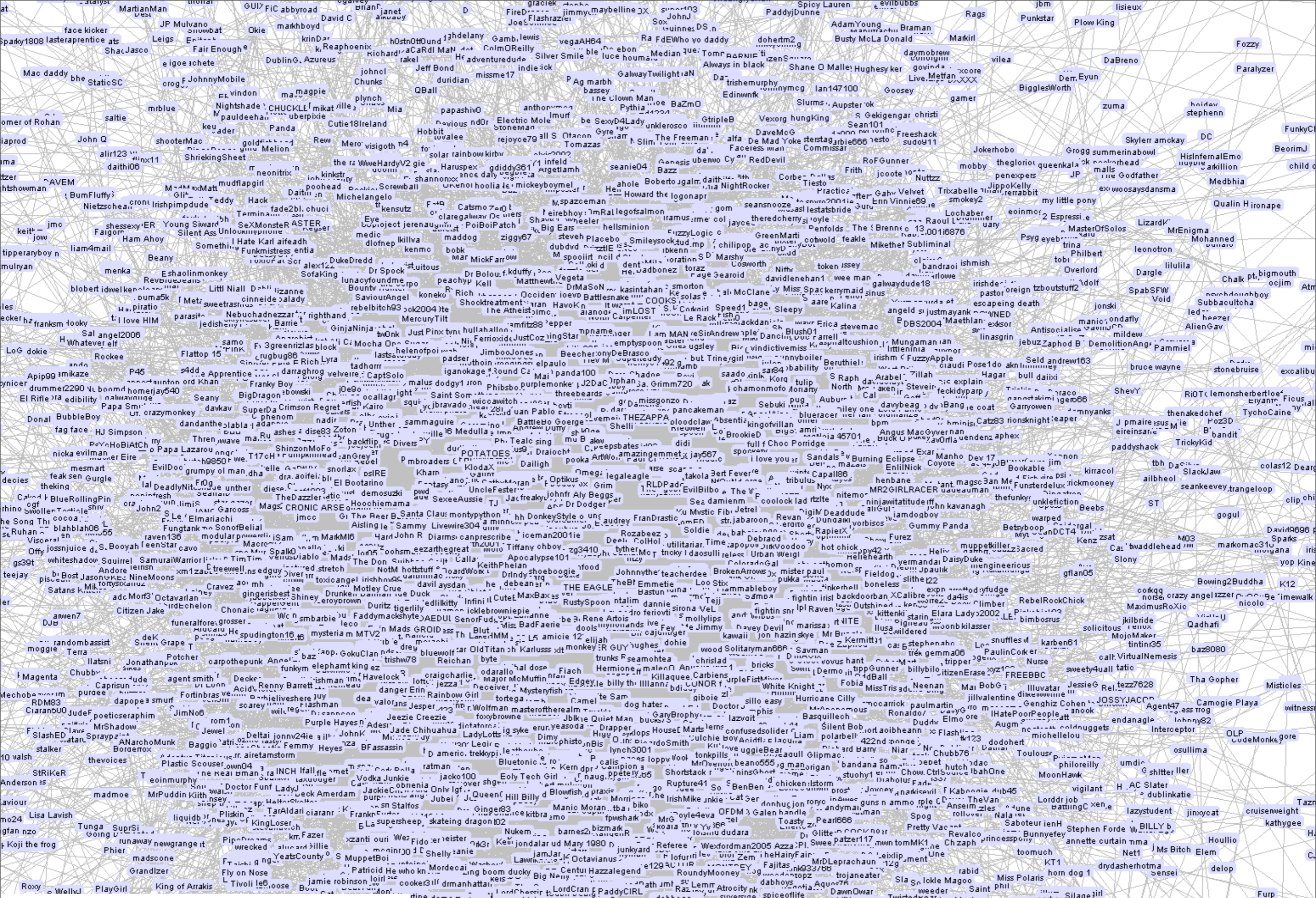
■ Perspectives & ouvertures

- définition d'une IP et d'une requête pédophile
- combien de fichiers et d'utilisateurs pédophiles ?
- identification du contenu via les noms de fichiers
- impact des lois (hadoppi, loppssi, ...)

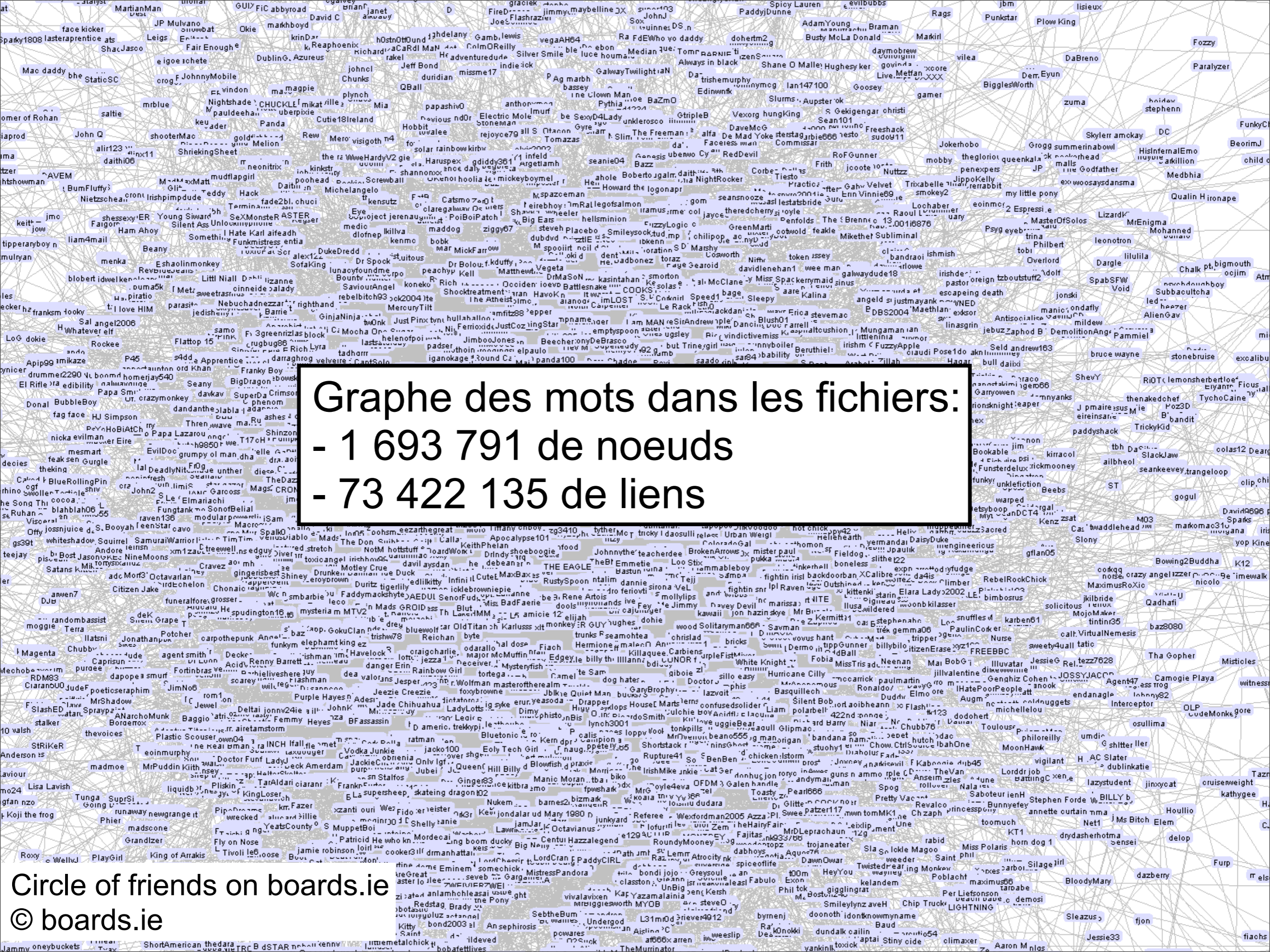
■ **<blink>je cherche du travail</blink>**

- le mot clé 2l est utilisé dans 60% des requêtes

by karl photos zoophilie serpent gratuit
amateur florence -229121 etudiante lyon 1
tel -184778 jeune salope 20 -1730843 wmv



Circle of friends on boards.ie
© boards.ie



Graphe des mots dans les fichiers:
- 1 693 791 de noeuds
- 73 422 135 de liens

Communautés contenant 12yo

	Requêtes	Fichiers
graphe	3 380 213	1 693 791
découpage 1	2 469 134/ 107 688	828 913/ 97 111
découpage 2	785 606/ 23 680	314 665/ 30 367
découpage 3	232 620/ 3 954	13 540/ 1259
découpage 4	51 026/ 836	1310/ 94
découpage 5	1614/ 70	189/ 11

NOIR mots/noeuds dans la communauté

ROUGE mots connus (> 100 apparition dans les données)