

TFC – Seguretat informàtica

Estudi dels mètodes d'estimació de la tipologia de la xarxa en un protocol UUP basat en xarxes socials.

José Cañellas Frau
Enginyeria Informàtica
Universitat Oberta de Catalunya

Consultor: Jordi Castellà-Roca
03/06/2012

Índex de continguts

1. Introducció
2. Objectius del projecte
3. Fases del projecte
4. Protocol
5. Estructura del simulador
6. Simulador gràfic
7. Simulador consola
8. Estudi de les dades: xarxa Wiki-Vote, xarxa Son Epinions, xarxa CA-Hep.Th i xarxa CA-Astro.Ph
9. Nova variable: temps mitjà rebre.
10. Mètode 5: temps mitjà aproximat.
11. Mètode 6: ordenació per temps mitjà
12. Treballs de futur
13. Conclusions

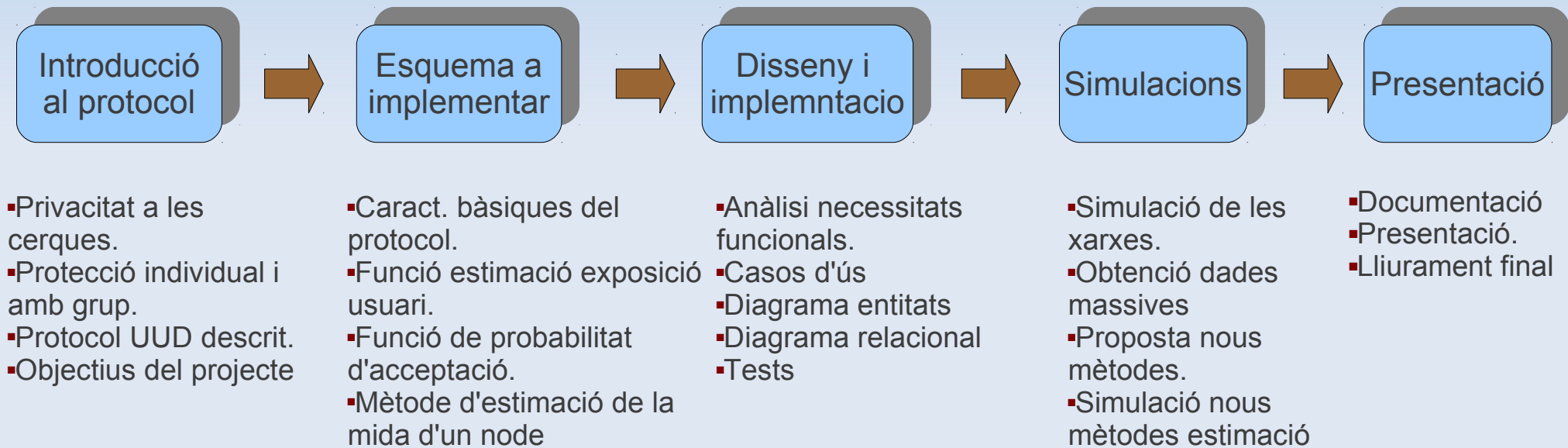
Introducció

- Importància del cercadors d'informació d'Internet (**Web Search Engines**).
- **Perfils d'usuari** → Obtenir informació més precisa.
- Escàndol AOL → Privacitat dels usuaris en relació a les consultes.
- Mecanismes senzills (proxy, IP dinàmiques) → No resolen la privacitat
- Protecció individual i amb grup.
- **Using social networks to distort users' profiles generated by web search engines**

Objectius del projecte

- Implementació d'un **simulador** que es comporti segons el protocol de "Using social networks to distort users' profiles generated by web search engines" Alexandre Viejo i Jordi Castellà-Roca
- Analitzar les dades obtingudes referent al % d'encerts dels mètodes d'estimació de la tipologia de la xarxa.
- Proposar nous mètodes que millorin aquests resultats.
- Contrastar les dades resultants d'aquests mètodes nous

Fases del projecte



Protocol

- Intenta **distribuir** les consultes a enviar al **WSE** amb usuaris d'una xarxa social.
- Pas 1: Decidir cap a quin veïnat enviar o re-enviar la consulta → funció que estima el **nivell d'exposició** de l'usuari candidat.
 - Mètodes d'estimació del nombre de veïnats d'un determinat node
- Pas 2: Acceptar o rebutjar una consulta que arriba d'un veïnat → funció estimació del **nivell d'egoisme**.
- Pas 3: Un cop enviada al WSE reenviam la resposta a través del camí creat.

Estructura del simulador

- S'ha intentat seguir una estructura en 3 capes
 - **Domini:** entitats i classes pròpies del negoci (User, Xarxa, UtilQuery,...).
 - **Persistència:** Accés a dades i conversió estructura relacional – entitats
 - Hibernate
 - DAO
 - **Presentació:**
 - Aplicació escriptori en Swing
 - Aplicació consola: simulacions massives

Simulador gràfic

Options

/home/joan/pfc/files/Wiki-Vote.txt Fitxer Carrega

Maxim usuaris: 30 Factor: 0.02 Nivell: 0.80 Metode: Metode 3 **Simula**

Consultes: 100 Tots els nodes Node

A

- Secció de paràmetres
- Seleccionar la xarxa a carregar.
- Establir: màxim usuaris, factor lambda, probabilitat, mètode d'estimació, consultes generades.

B

D

```
*** Inici simulació (Sat Jun 02 11:48:05 CEST 2012) ***
Fitxer: /home/joan/pfc/files/Wiki-Vote.txt
Mètode: 3
Usuaris: 30
Factor: 0.02
Llindar: 0.8
Consultes: 100
Relacions: 88
%Encerts: 5.8266783
```

Node	Veïnats	Id V.	Temps	Encerts	Intents	Reals	Enviades	Acceptades
30	7	3	173,2	0	229	24	57	19
30	7	5543	10,168539	52	222	2	56	4
30	7	3352	10,179775	207	298	2	47	8
30	7	5254	9,05	113	113	1	56	2

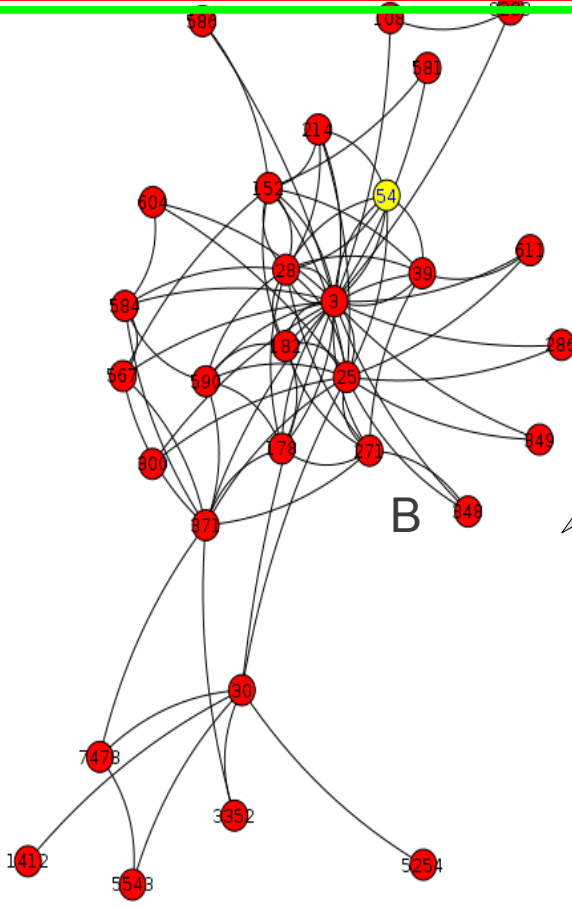
Simulador gràfic

Opcions

/home/joan/pfc/files/Wiki-Vote.txt Fitxer Carrega

Maxim usuaris: 30 Factor: 0.02 Nivell: 0.80 Metode: Metode 3

Consultes: 100 Tots els nodes Node: 8283



B

Factor: 0.02
Llindar: 0.8
Consultes: 100
Relacions: 88
%Encerts: 5.8266783

D

Node	Veïnats	Id V.	Temps	Encerts	Intents	Reals	Enviades	Acceptades
30	7	3	173,2	0	229	24	57	19
30	7	5543	10,168539	52	222	2	56	4
30	7	3352	10,179775	207	298	2	47	8
30	7	5254	9,05	113	113	1	56	2

- Secció de visualització de la xarxa.
- Panell interactiu.
- Selecció d'una node → visualitzar informació.
-

Simulador gràfic

- **Secció de resultats del node**

- Resultats del node seleccionat a B.

- Es mostra:

- Veïnats
- Consultes acceptades del veïnat
- Consultes rebudes
- Consultes acceptades per part el veïnat
- Consultes enviades al veïnat
- Encerts i intents.

A

Simula

Resultats

Seleccionat: 54

envi.	Con. reb.	Con. ac...	WSE	A	I
19	3	384	0	133	
	31	40	0	184	
	7	257	0	160	
	87	87	0	176	
	165	165	0	126	

C

*** Inici simulació (Sat Jun 02 11:48:05 CEST 2012) ***

Fitxer: /home/joan/pfc/files/Wiki-Vote.txt

Mètode: 3

Usuaris: 30

Factor: 0.02

Llindar: 0.8

Consultes: 100

Relacions: 88

%Encerts: 5.8266783

D

Node	Veïnats	Id V.	Temps	Encerts	Intents	Reals	Enviades	Acceptades
30	7	3	173,2	0	229	24	57	19
30	7	5543	10,168539	52	222	2	56	4
30	7	3352	10,179775	207	298	2	47	8
30	7	5254	9,05	113	113	1	56	2

Simulador gràfic

Opcions

/home/joan/pfc/files/Wiki-Vote.txt Fitxer Carrega

Maxim usuaris: 30 Factor: 0.02 Nivell: 0.80 Metode: Metode 3

Consultes: 100 Tots els nodes Node: 8283

A **Simula**

Resultats

Usuari seleccionat: 54

V	Con. acc...	Con. envi.	Con. reb.	Con. ac...	WSE	A	I
3	17	57	19	3	384	0	133
214	16	16	35	31	40	0	184
25	25	40	7	7	257	0	160
39	24	28	10	10	87	0	176
28	18	43	4	4	165	0	126

B

C

D

*** Inici simulació (Sat Jun 02 11:48:05 CEST 2012) ***

Fitxer: /home/joan/pfc/files/Wiki-Vote.txt
Mètode: 3
Usuaris: 30
Factor: 0.02
Llindar: 0.8
Consultes: 100
Relacions: 88
%Encerts: 5.8266783

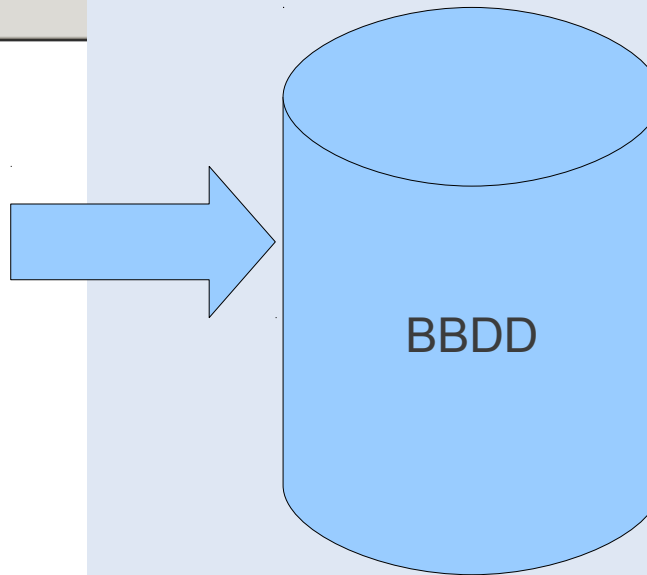
Id V.	Temps	Encerts	Intents	Reals	Enviades	Acceptades
3	173,2	0	229	24	57	19
5543	10,168539	52	222	2	56	4
3352	10,179775	207	298	2	47	8
5254	9,05	113	113	1	56	2

- Secció de resultats globals
- Apareix la recta de regressió.

Simulador consola

- Permet realitzar simulacions massives.
- Permet ser programada (mitjançant crons)
- Pot ser utilitzada en vàries màquines utilitzant el mateix magatzem de dades

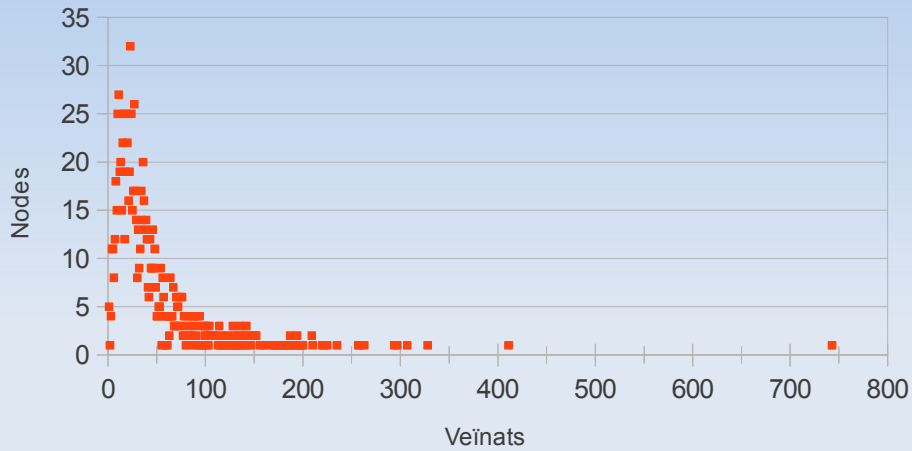
```
xisca@inspiron: ~/Escriptori/Pep/uoc/2012/pfc
File Edit View Search Terminal Help
test:
  [input] metode? [3]
6
  [input] minim usuaris? [100]
400
  [input] maxim usuaris? [500]
500
  [input] interval? [10]
  [input] factor lambda? [0.02]
  [input] nivell prob.? [0.80]
  [input] min. consultes? [100]
  [input] max. consultes? [100]
  [input] interval consultes? [10]
  [input] guardar? [false]
```



Estudi dades: Xarxa Wiki-Vote

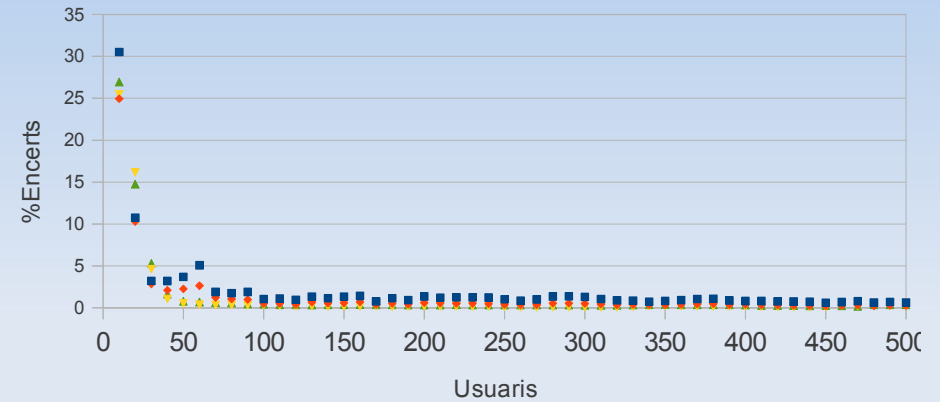
■ Densitat nodes

■ Wiki-Vote



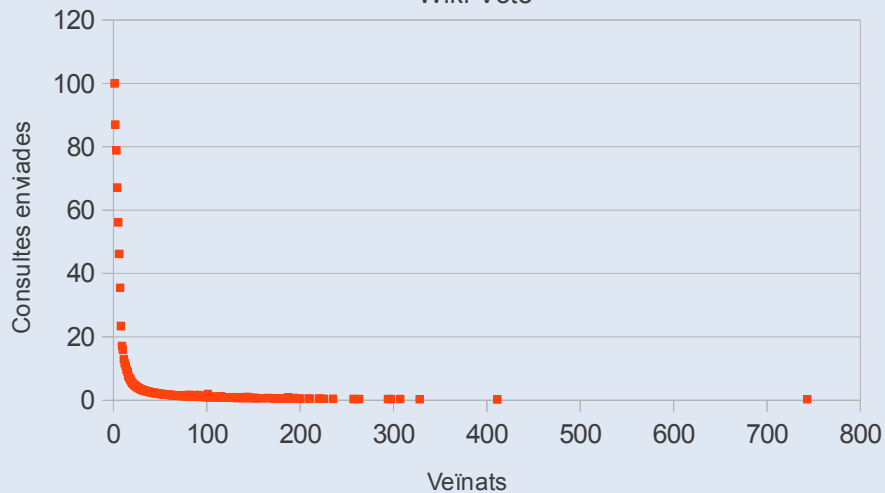
■ %Encerts

■ Mètode 1 ■ Mètode 2 ▼ Mètode 3 ▲ Mètode 4



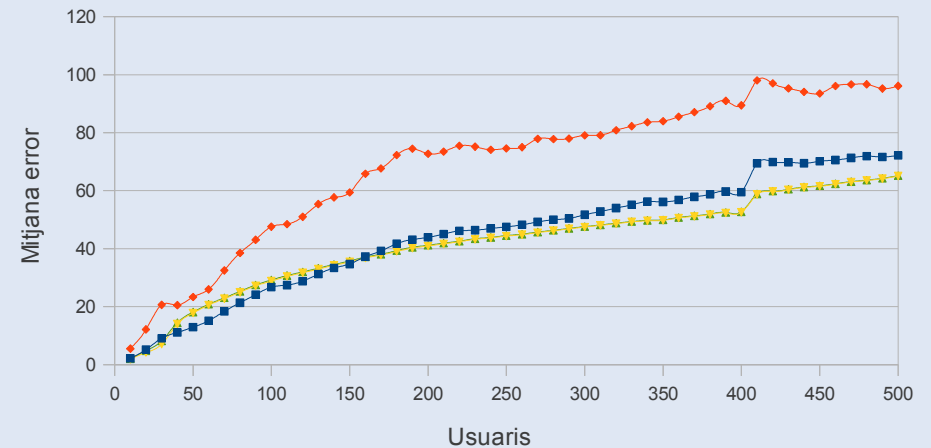
■ Consutes env./num. veïnats

■ Wiki-Vote



■ Mitjana error

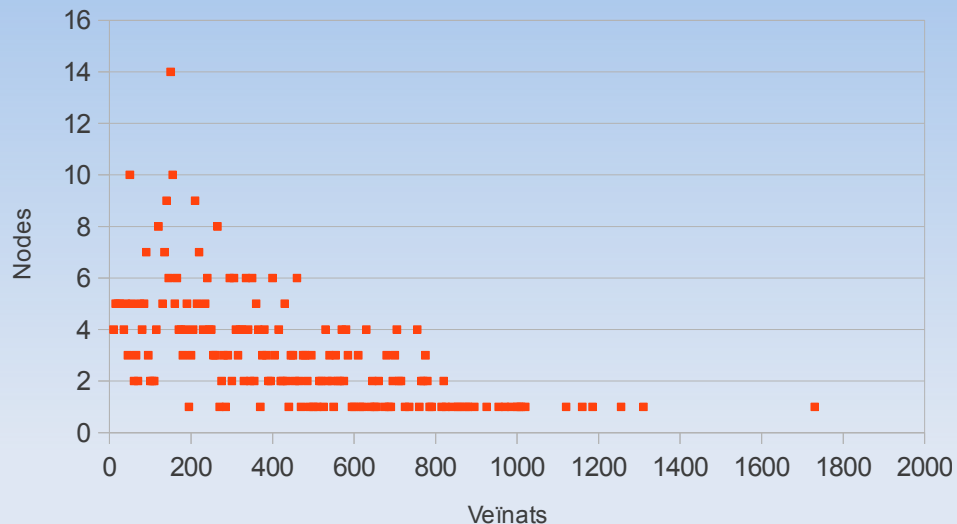
■ Mètode 1 ■ Mètode 2 ▼ Mètode 3 ▲ Mètode 4



Estudi dades: Xarxa Soc Epinions

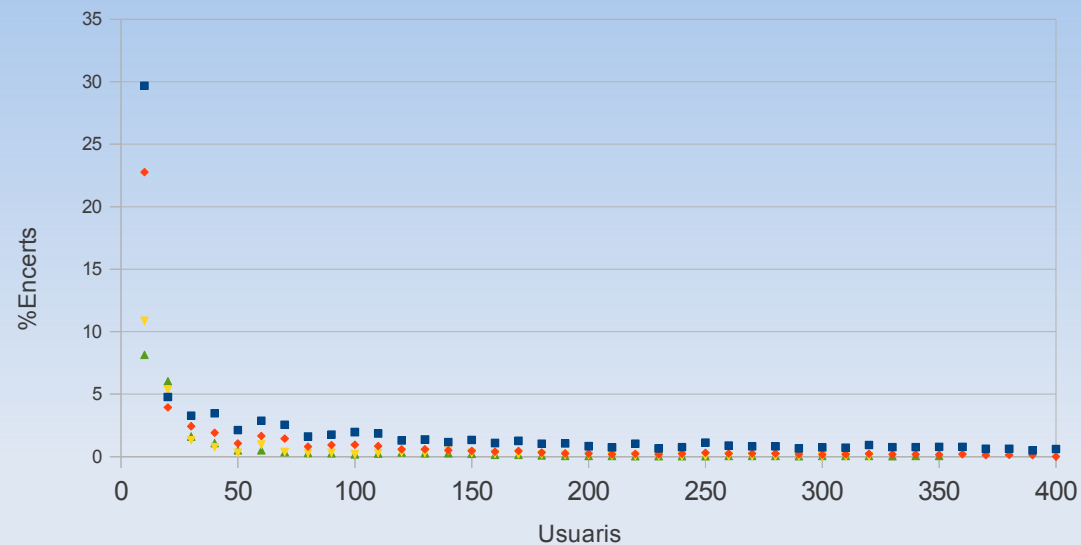
■ Densitat nodes

■ Soc-Epinions



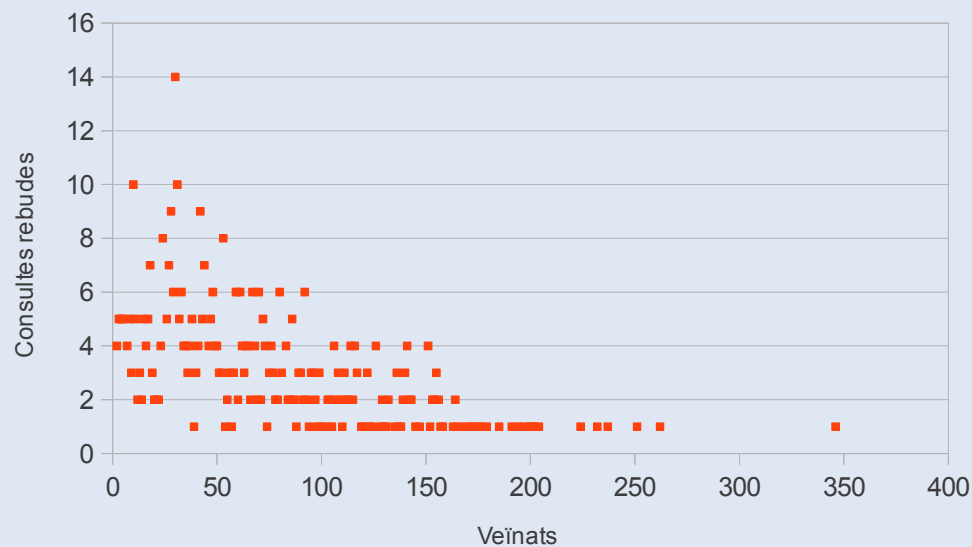
■ %Encerts

■ Mètode 1 ■ Mètode 2 ▼ Mètode 3 ▲ Mètode 4



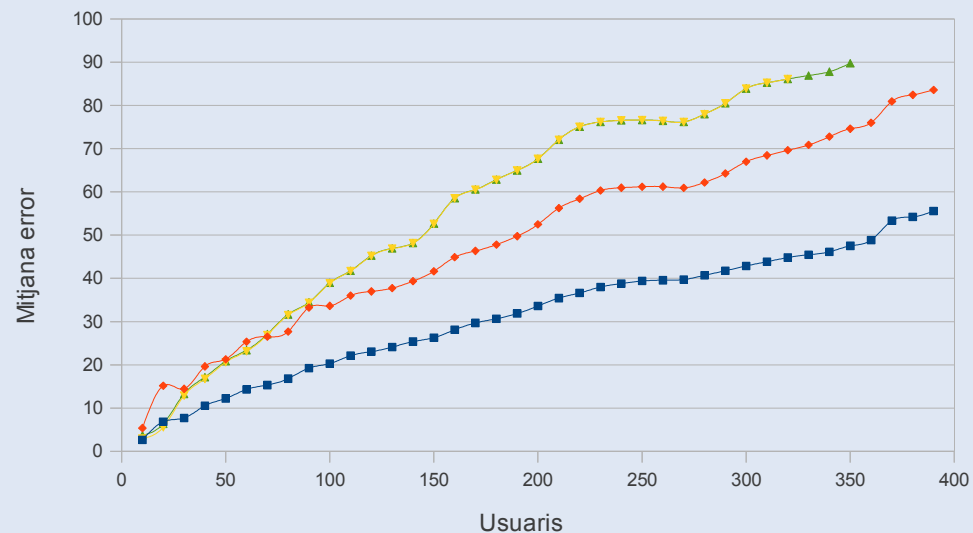
■ Consutes env./num. veïnats

■ Soc-Epinions



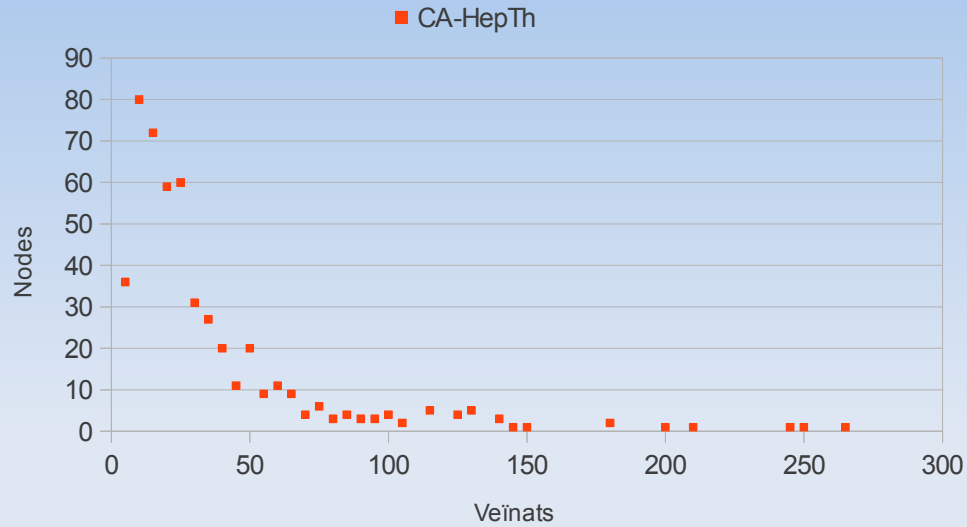
■ Mitjana error

■ Mètode 1 ■ Mètode 2 ▼ Mètode 3 ▲ Mètode 4

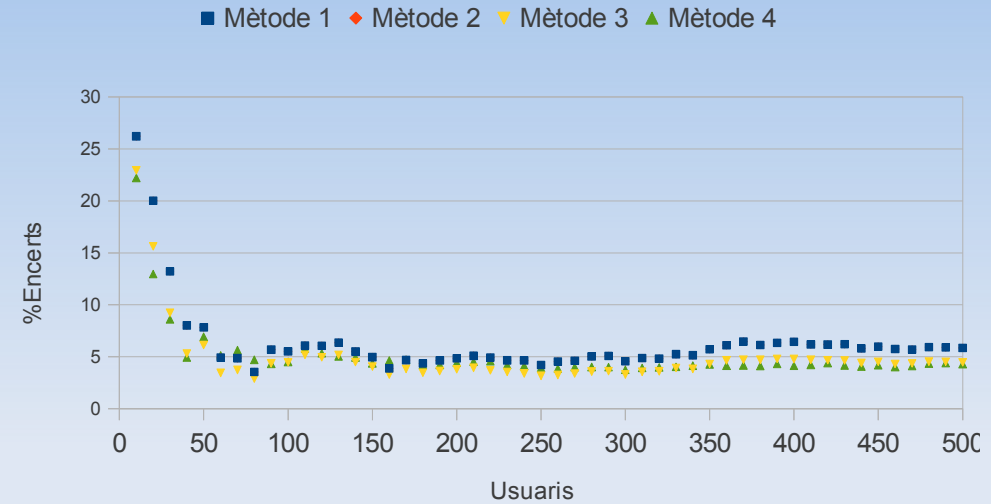


Estudi dades: Xarxa CA-Hep.Th

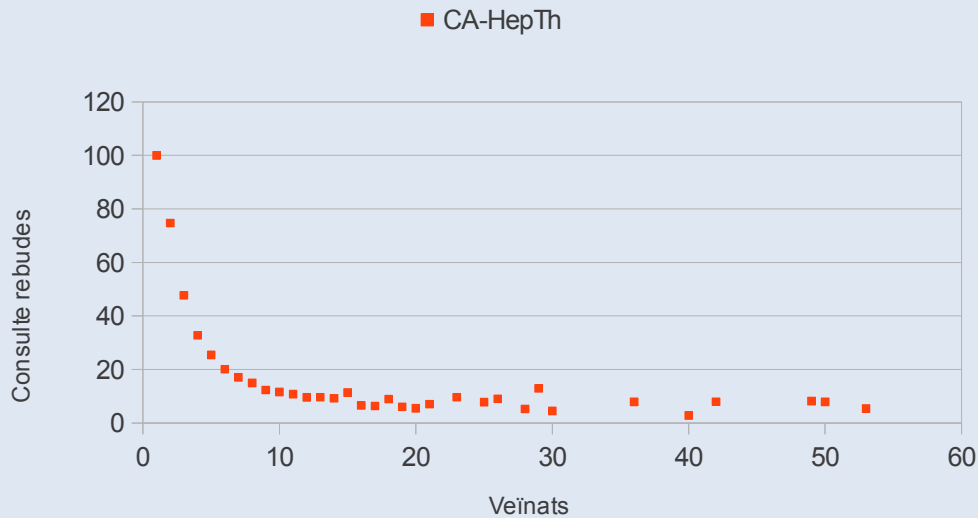
■ Densitat nodes



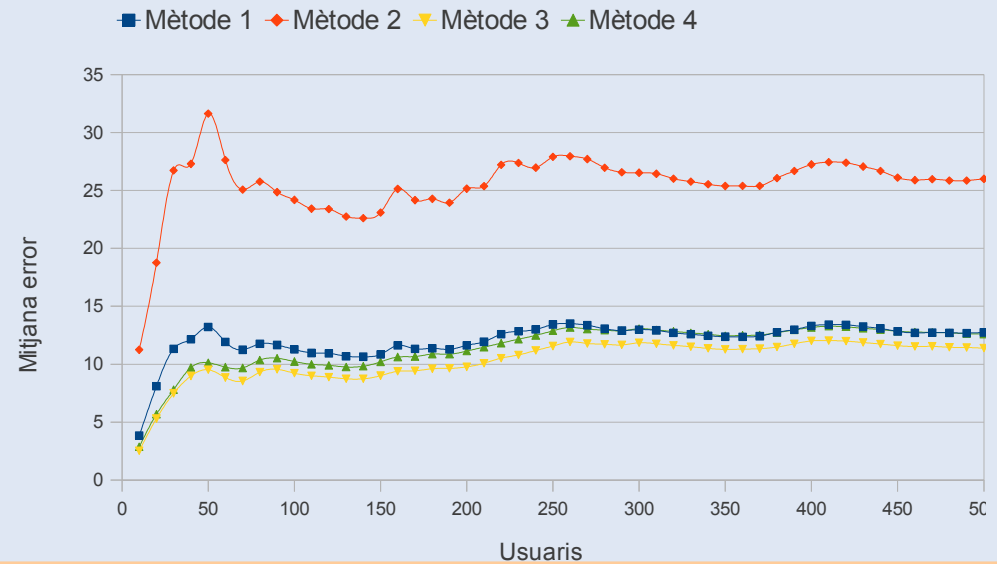
■ %Encerts



■ Consutes env./num. veïnats

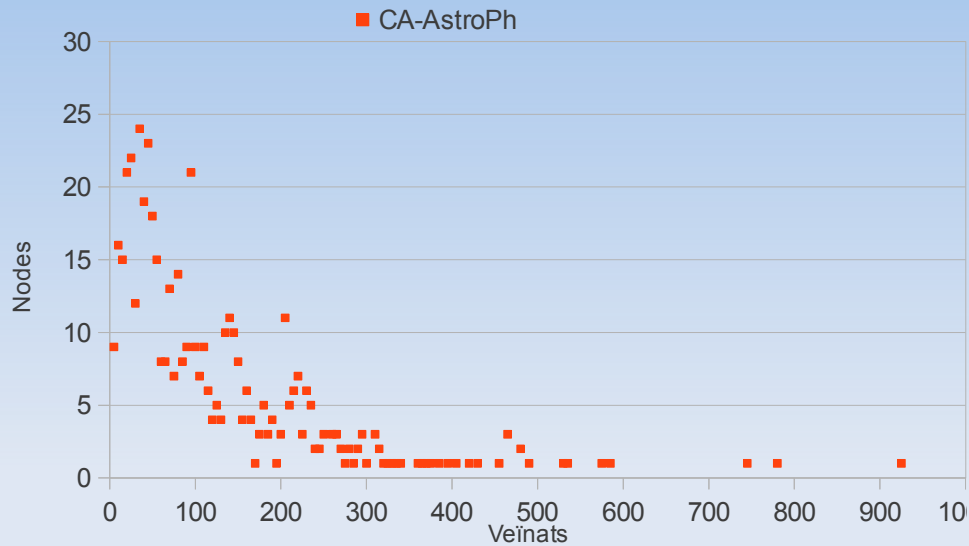


■ Mitjana error

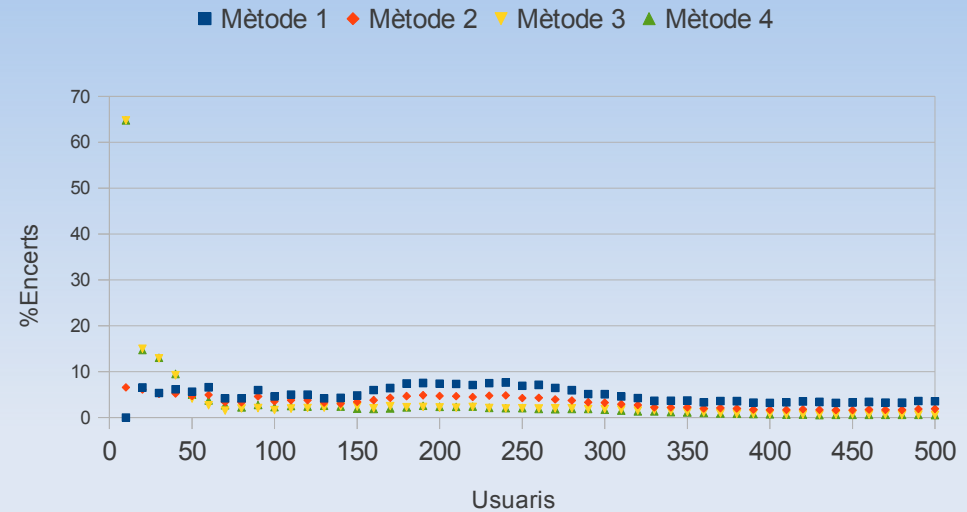


Estudi dades: Xarxa CA-Astro.Ph

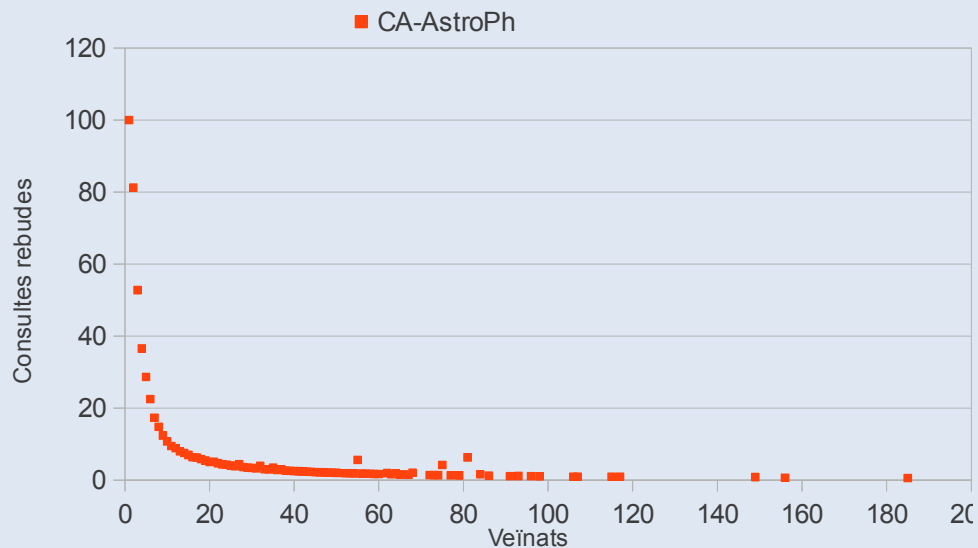
■ Densitat nodes



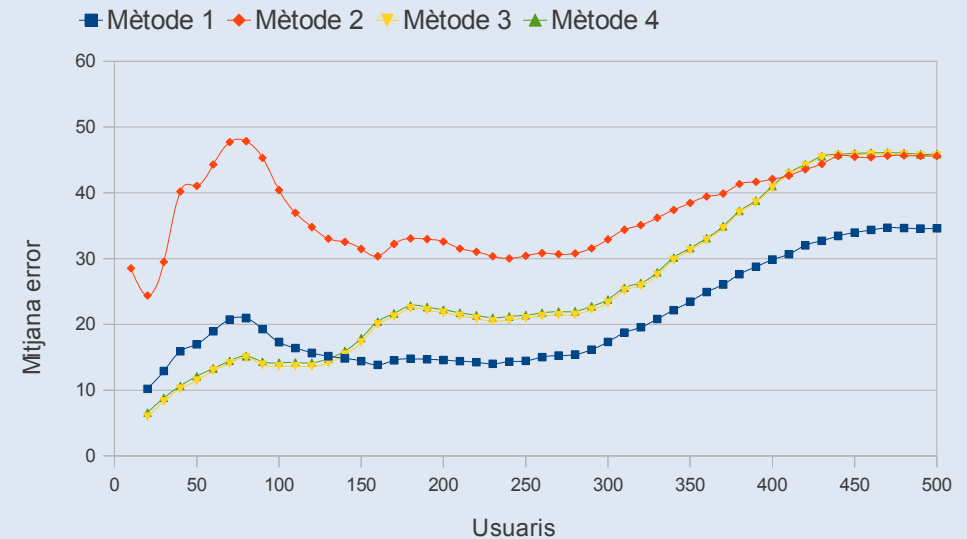
■ %Encerts



■ Consutes env./num. veïnats

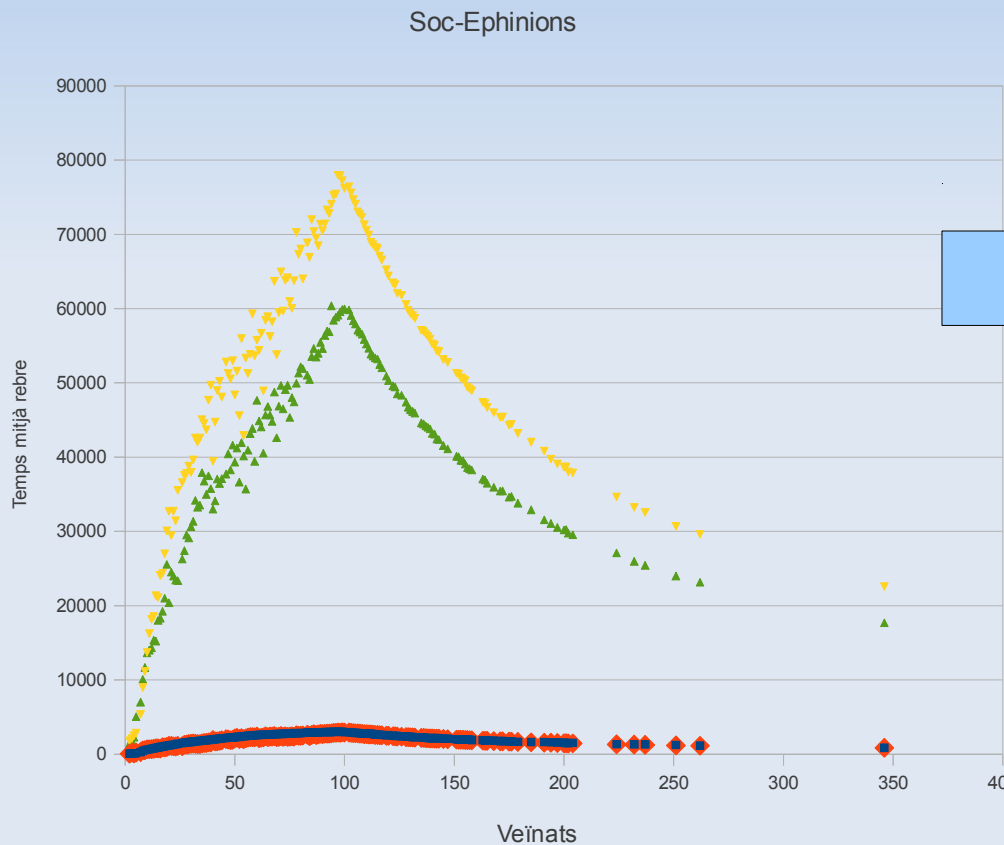


■ Mitjana error

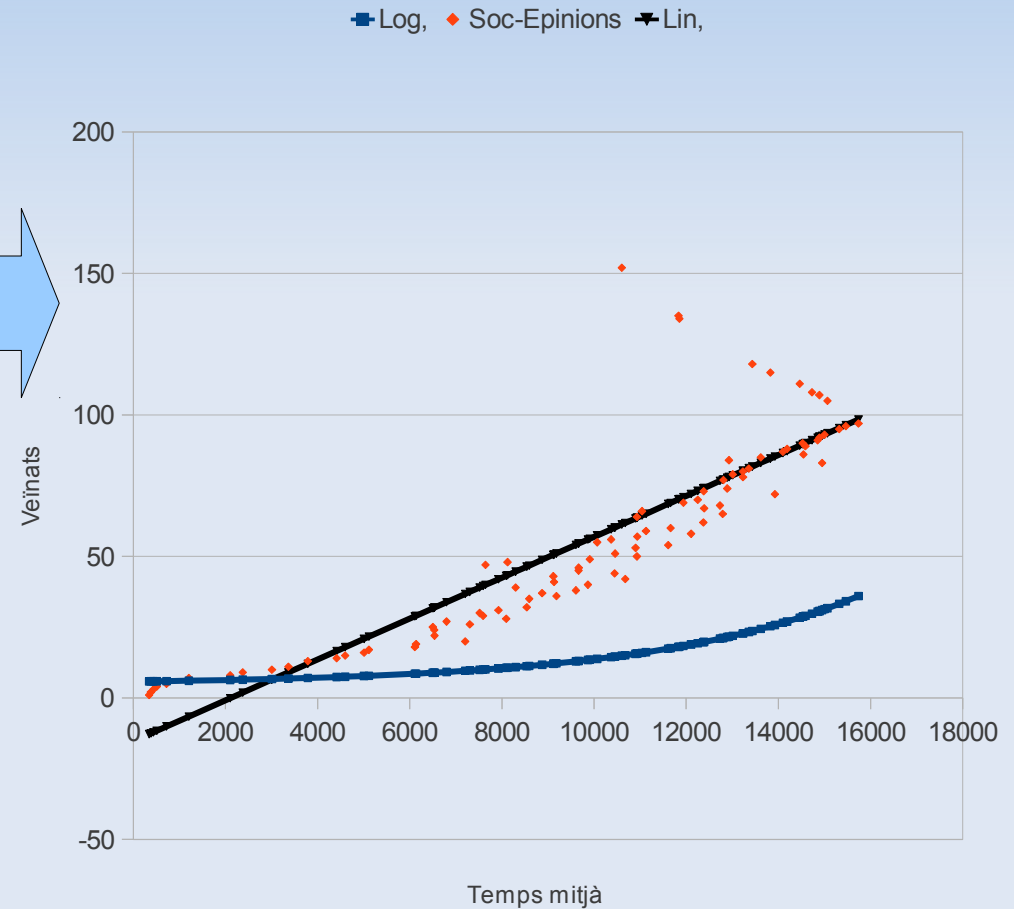


Nova variable: temps mitjà rebre

Temps mitjà en rebre consultes del veïnat

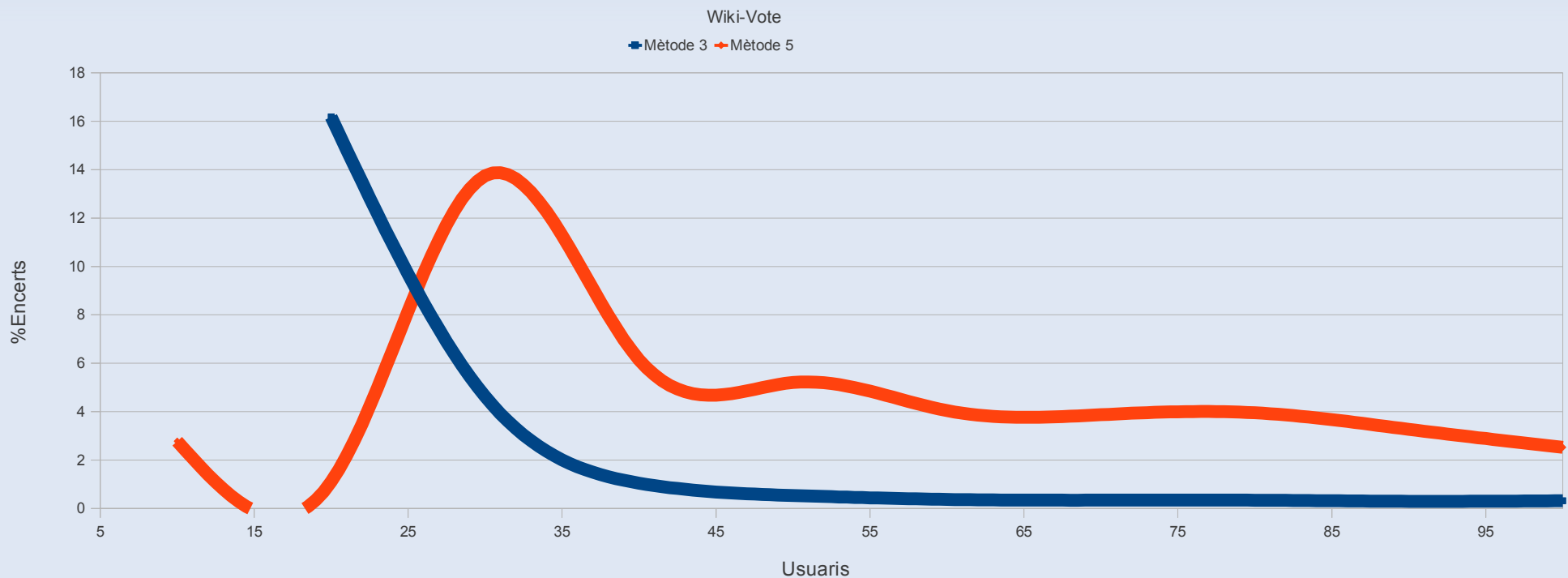


Càlcul de regressions



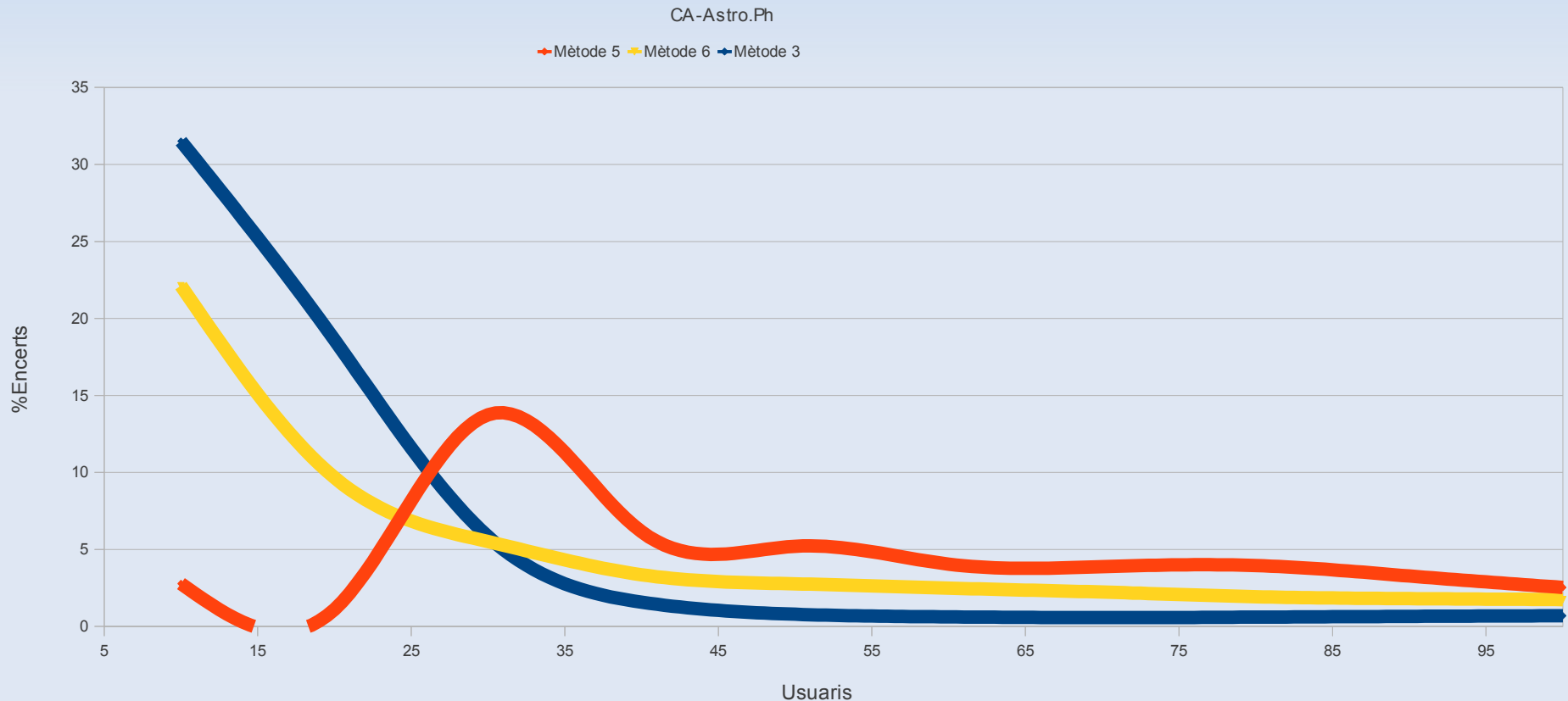
Mètode 5: temps mitjà aprox.

- Es calcula a partir de les regressions per a cada xarxa.
- Millora rengiment usuaris > 50 .
- No implementable món real (necessita simulacions prèvies, massa dependències amb el sistema)



Mètode 6: ordenació temps mitjà

- Llista ordenada segons temps mitjà.
- El veïnat amb menor temps es situa en primera opció.
- Semblant al mètode 3 i implementable món real.



Treballs de futur

- Incorporar al simulador llibreries gràfiques per apoderar representar els resultats segons esculli l'usuari.
- Proporcionar més realisme al simulador → comunicació entre els nodes via sockets o http i distribuir-ho en xarxa.
- Projecte de mineria de dades per analitzar les dades obtingudes i **extreure'n coneixement**.
- Fer el simulador multi-thread de manera que augmenti el seu propi rendiment.
- Incorporar a més de la variable del temps mitjà, la **probabilitat en com evoluciona el mètode**.

Conclusions I

- Privacitat entronc als WSE és un tema important.
- Compromís **privacitat<>informació acurada**.
- Necessitam mètodes d'estimació de la xarxa per aplicar el protocol.
- Simulador per a obtenir dades i comportament protocol.
- Simulador gràfic i simulador bàsic.

Conclusions II

- Mètodes d'estimació poc eficients si hi ha molts d'usuaris i relacions.
- Variable **temps mitjà** per a predir la mida d'un node.
- Mètode 5: millora però no és realista (massa dependència al sistema i simulacions anteriors).
- **Mètode 6: millora segons quins casos i és realista.**