

Vers un nouvel Internet

Anne-Marie Kermarec
INRIA, Rennes



La révolution du Web

Contenu généré par les utilisateurs

Internet social



Openstreetmap.fr



facebook

twitter

citeulike 

LinkedIn®

 delicious
social bookmarking

Le Web devient une gigantesque mine d'information

Que Google ne permet pas d'exploiter...



Exemple: Chercher une baby-sitter anglophone

Alice



« Babysitter anglophone Rennes »

1- AMERICAN GIRL, NATIVE ENGLISH SPEAKING BABYSITTER IN LILLE.

2- Cherche *babysitter* anglophone à Antony - Antony - Babysitters ...

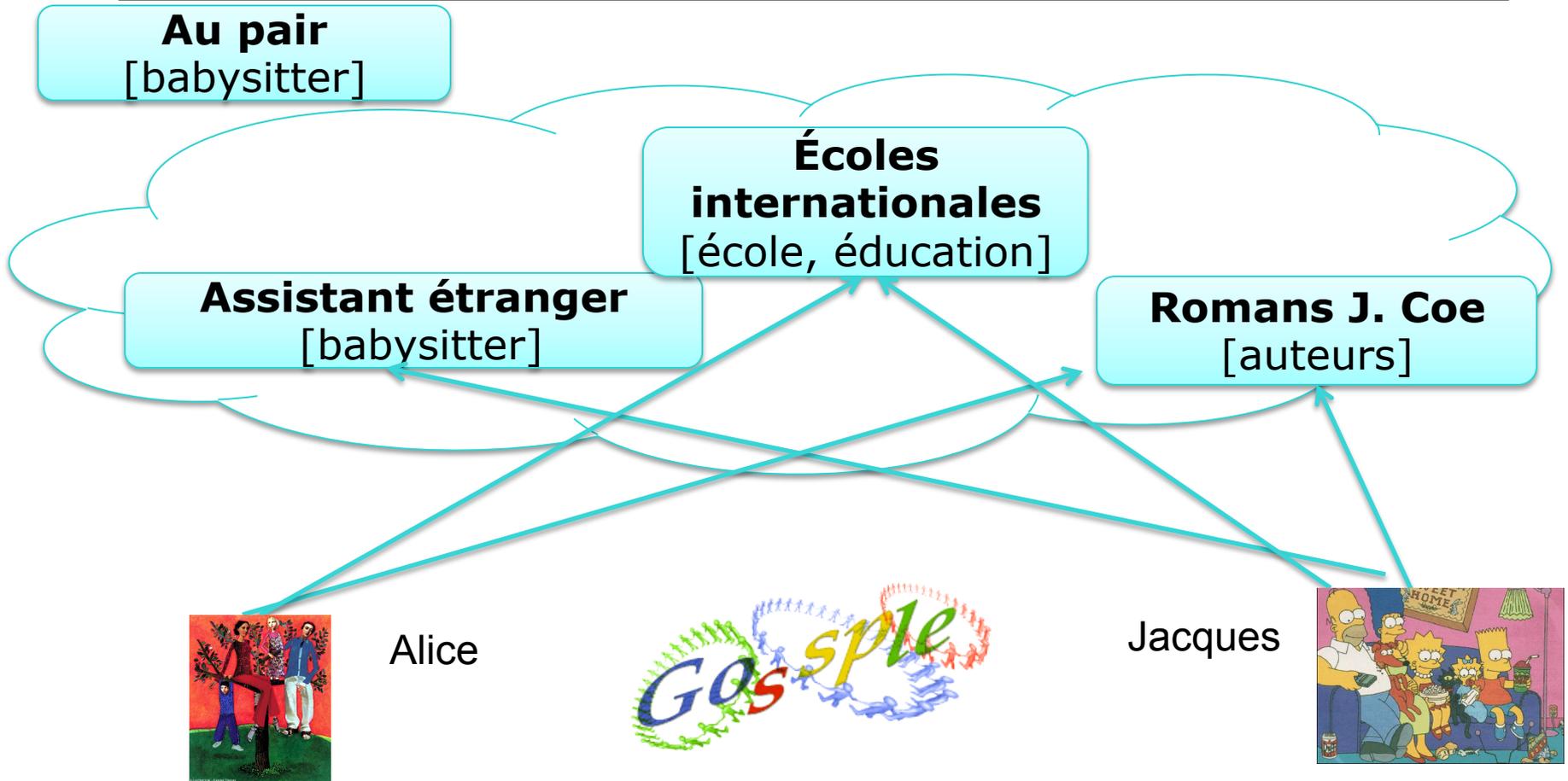
3- [PDF] GOSSPLE: personalized and decentralized queries [...]

...

Emploi.org - Vous cherchez un petit job

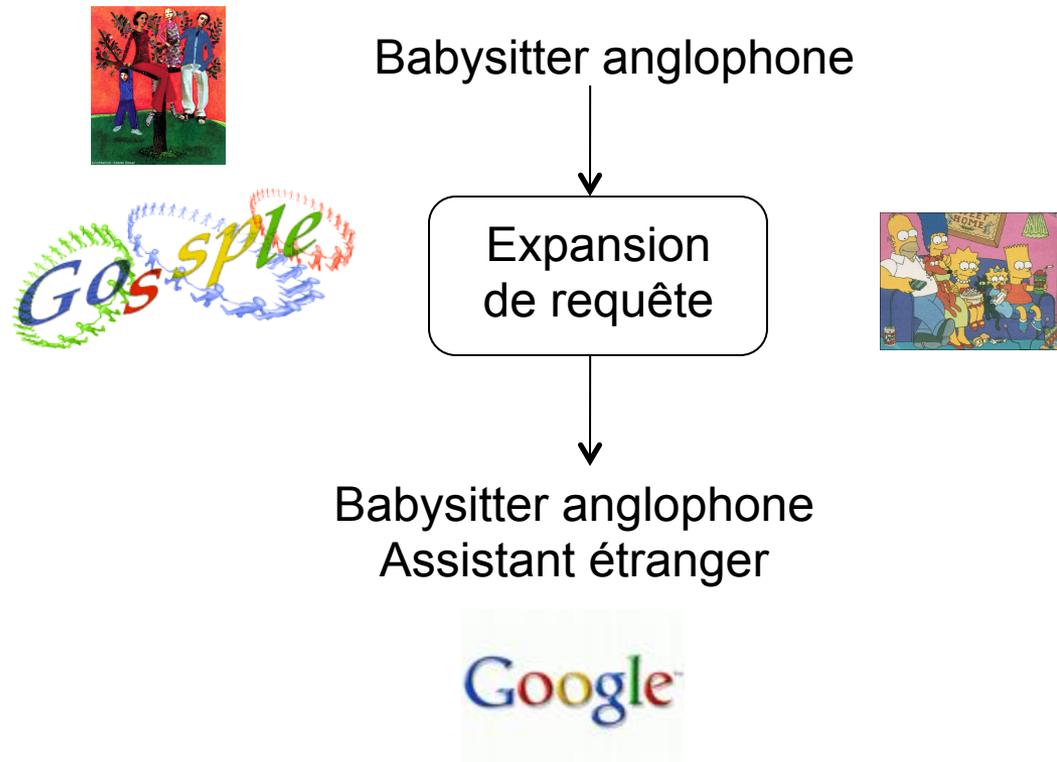
**Et si quelqu'un avait trouvé la
bonne formulation?**

Requête personnalisée



Exploiter la personnalisation

Expansion de requête



Verrous scientifiques

- Comment décider qui sont les voisins pertinents?
- Comment les trouver?
- Comment les exploiter?

Pourquoi est-ce difficile?

- Échelle
- Distribution
- Dynamicité
- Terminaison
- Subjectivité

Quels sont les voisins pertinents?

Similarité entre utilisateurs

Similarité dans le Web 2.0

- Utilisateurs caractérisés par un profil
- Modèle
 - $U(\text{sers}) \times I(\text{tems}) \times T(\text{ags})$
 - $\text{Tagged}_u(i, t)$: u annote item I avec tag t
 - $\text{Profile}(u) = \{\text{Tagged}_u(i, t)\}$

Similarité à base de cosinus

Intersection normalisée

$Items(\{\vec{u}_i\})$: vecteur d'objets de u_{i1}

$$ItemCos(\vec{u}_1, \vec{u}_2) = \frac{|Items(\{\vec{u}_1\}) \cap Items(\{\vec{u}_2\})|}{\sqrt{|Items(\{\vec{u}_1\})| \cdot |Items(\{\vec{u}_2\})|}}$$

Verrous scientifiques

- Comment décider qui sont les voisins pertinents?
- **Comment les trouver?**
- Comment les exploiter?

Personnalisation => décentralisation

Passage à l'échelle
Syndrome "big brother"

Modèle de communication pair-à-pair

Comment découvrir ces voisins pertinents?

Exploiter les topologies "Petit monde"
Exploiter la puissance de l'aléatoire

Topologie “petit monde”

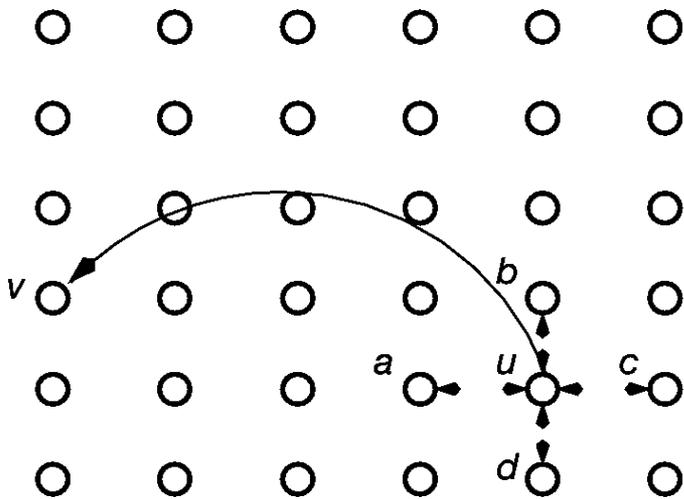
“les six degrés de séparation”

Choix des voisins et routage

- Sélection aléatoire

Performance du routage
glouton

$$O(n^{1/3})$$



- Sélection harmonique

Probabilité dépendante de la
distance

$$\delta(B) = \frac{1}{d(A,B)^2}$$

B choisi par A

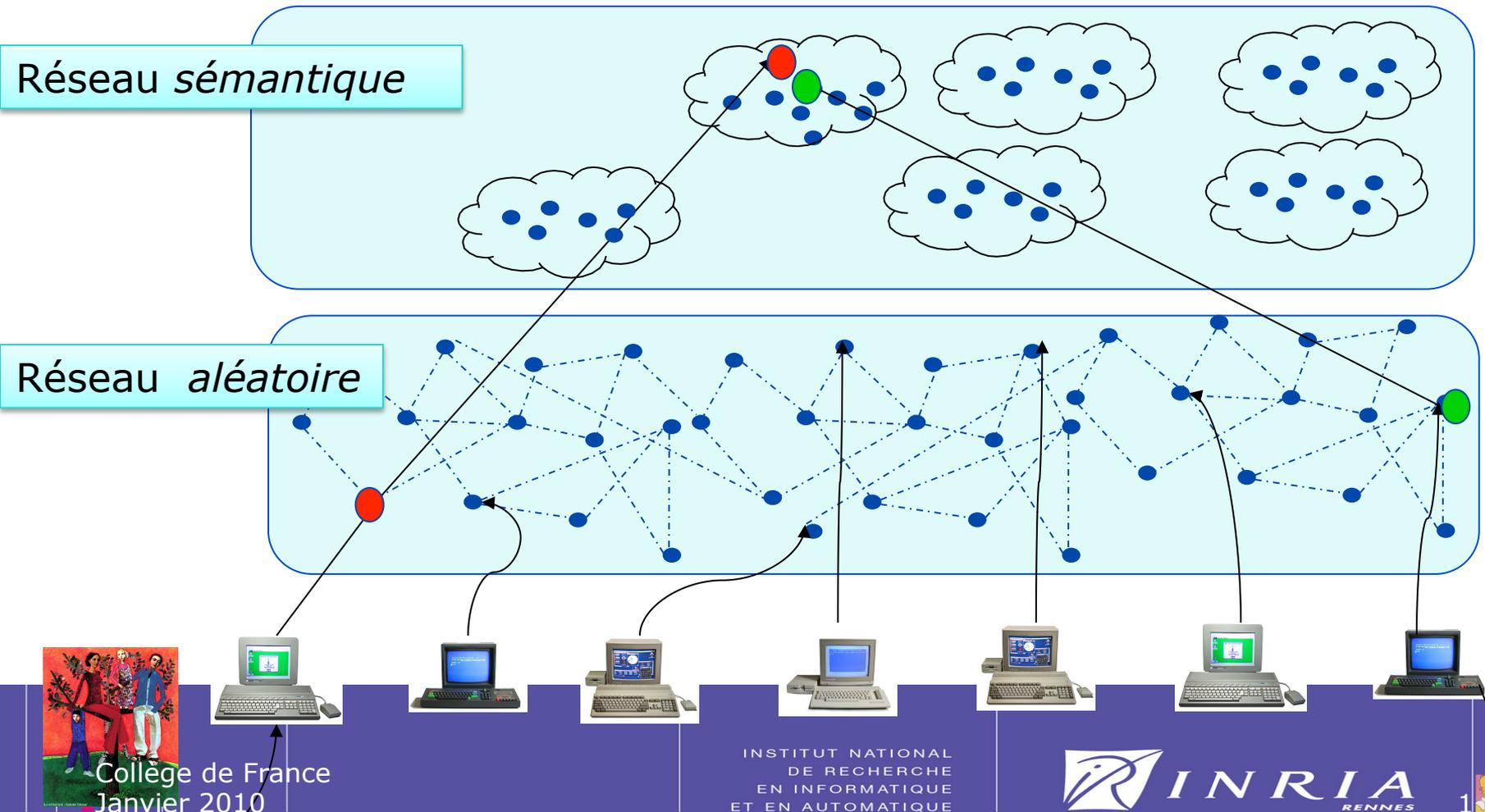
$$P = \frac{\delta(B)}{\sum_{B' \in S} \delta(B')}$$

S = ensemble de pairs
non voisins de A

Performance du routage

glouton $O(\log^2(n))$

Navigation personnalisée



Créer un réseau *petit monde* sans serveur

c liens courts

- $c \ll N$ voisins les plus similaires
- mesure de similarité entre utilisateurs
Exemple: Cosinus

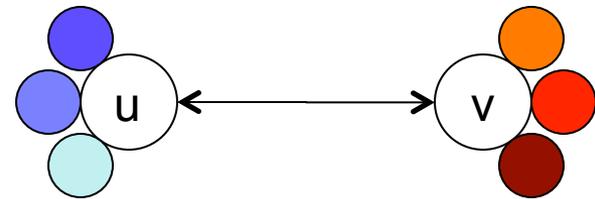
k liens longs

- échantillonnage aléatoire (ou harmonique)

Algorithme épidémique

Chaque nœud connaît c autres nœuds

Échange périodique



Espace des paramètres:

- Sélection de la cible de l'échange
- Information échangée
- Traitement de l'information

Topologie: graphe aléatoire et uniforme

Réseau petit monde pour le Web 2.0

c liens courts

@IP:port	132.154.8.5:2020	
Profil	www.inria.fr : inria, computer www.assistants.fr : baby-sitter, english ...	

c entrées

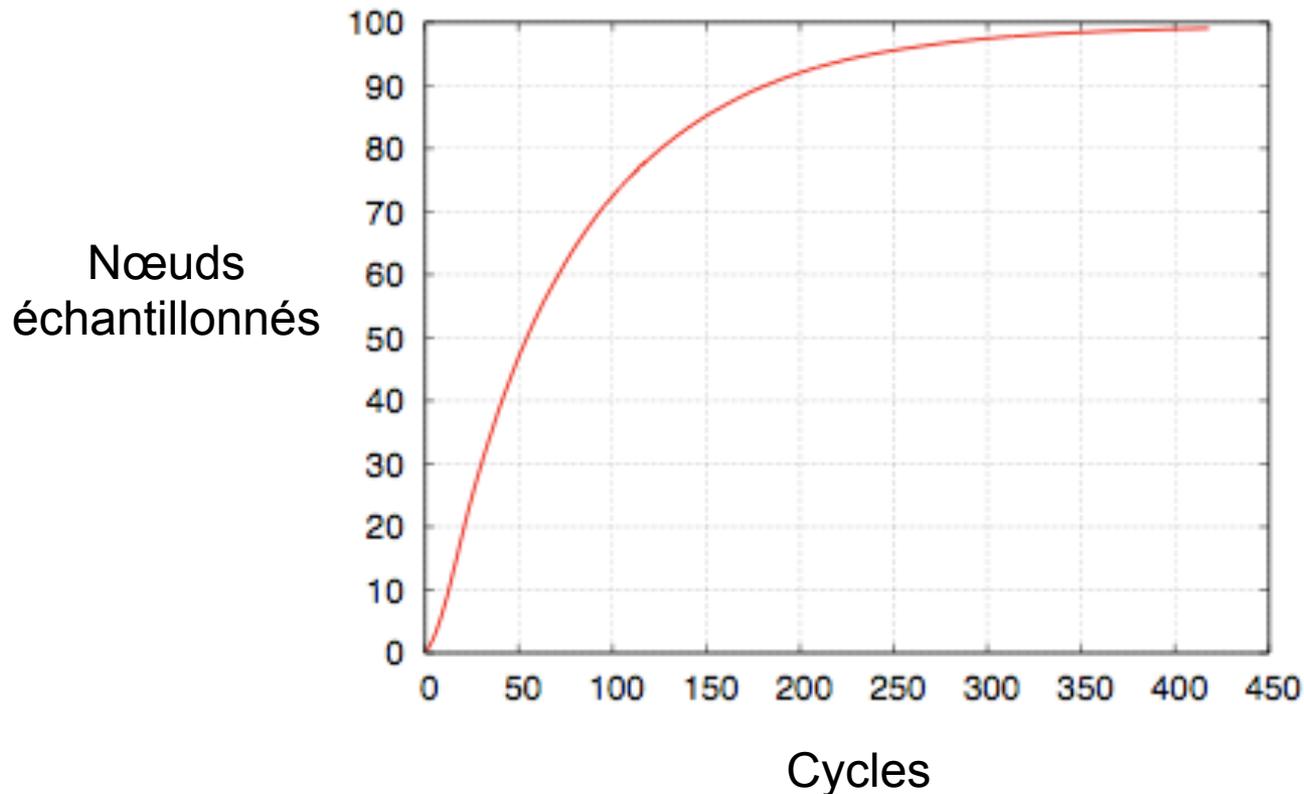
k liens longs

@IP: port	102.14.18.1:2110	
...		

k entrées

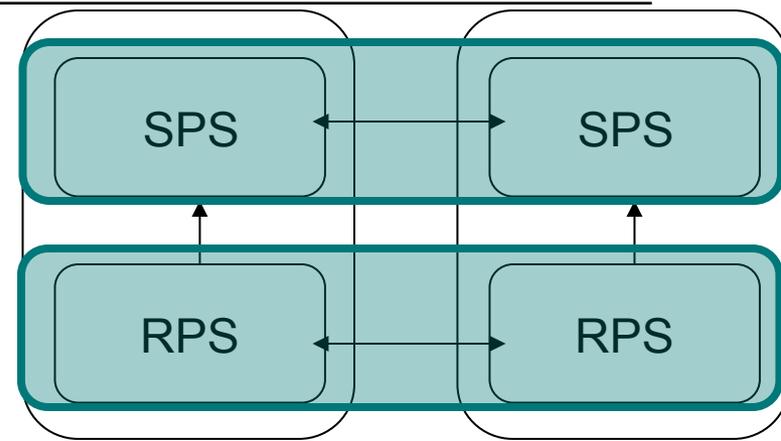
Échantillonnage aléatoire

- $O(n/k \log n)$ itérations.



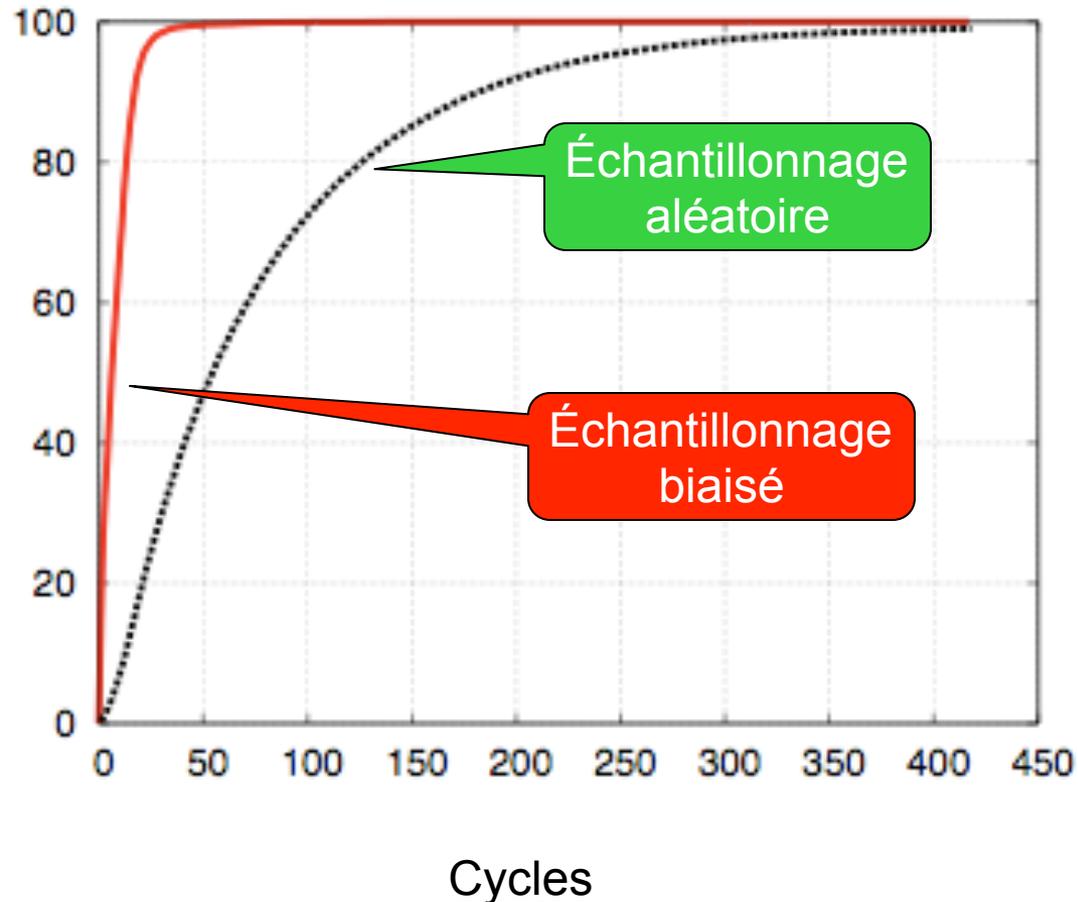
Échantillonnage biaisé

- Deux protocoles épidémiques
 - Similarité (SPS)
 - Aléatoire (*RPS*)
- Échantillonnage biaisé par la similarité
 - Sélection de pair aléatoire dans SPS
- Accélérateur de convergence



Convergence du SPS

Ratio entre
les c voisins
courants et
les c idéaux



Verrous scientifiques

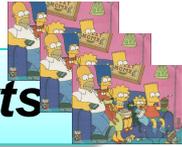
- Comment décider qui sont les voisins pertinents?
- Comment les trouver?
- **Comment les exploiter?**

Expansion de requête



c liens courts

k liens longs



TagMap (n)
Vue personnalisée
des relations entre
mots clés

	babysitter	assistant étranger	au pair
babysitter	1	0.9	0.1
assistant		1	0
au pair			1

Babysitter anglophone

Expansion de requête (1)
(ala PageRank)

Babysitter anglophone: 0.44
Assistant étranger: 0.42



www.assistantinfrance.com

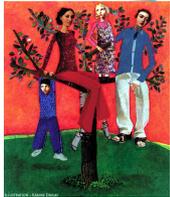
En résumé

Au pair
[babysitter]

Écoles internationales
[école, éducation]

Assistant étranger
[babysitter]

Romans J. Coe
[auteurs]



Les défis

- Efficacité
- Critères
- Validation
- Sécurité
- Anonymat