



13th International Conference on Extending Database Technology (EDBT 2010)  
March 22-26, 2010 Lausanne Switzerland

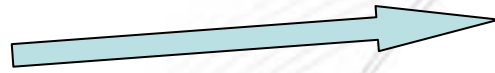
# Gossiping Personalized Queries

*Xiao Bai, Marin Bertier, Vincent Leroy, INSA-Rennes, France*  
*Rachid Guerraoui, EPFL, Switzerland*  
*Anne-Marie Kermarrec, INRIA, Rennes*

# Personalized top-k

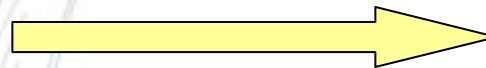


**Lausanne  
tourism**



Museum collection de l'art brut  
The Elysée Museum  
...

Non-personalized



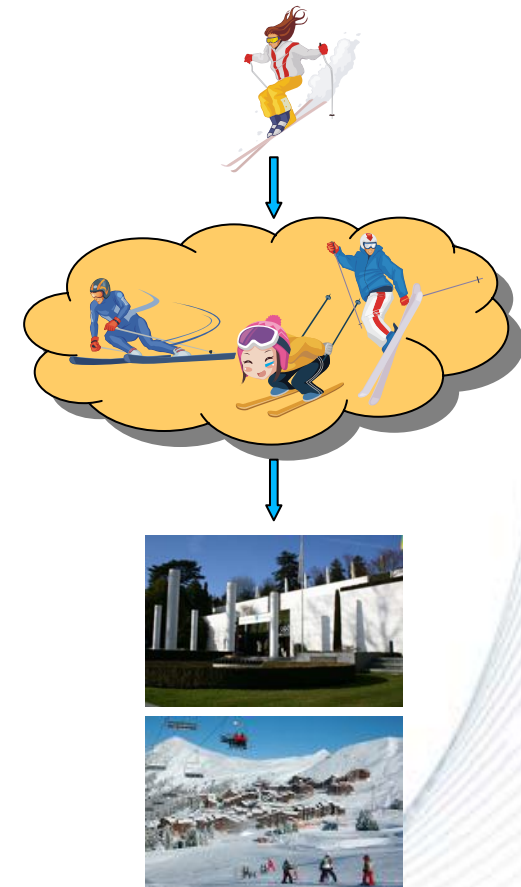
Lake of Geneva  
Lausanne Cathedral  
Olympics museum  
...



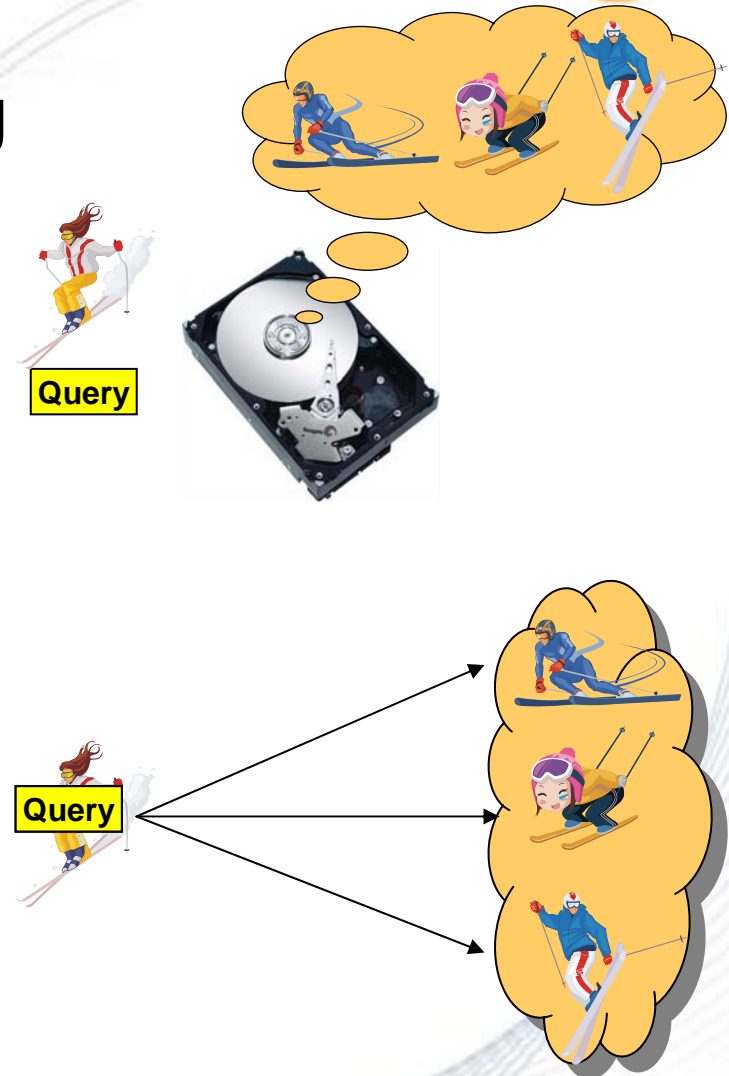
Olympics museum  
Ski station  
...

- Network-aware search
  - S. Amer-Yahia et al., *Efficient network-aware search in collaborative tagging sites* [VLDB'08]
  - Interest-based personal network to process query
  - Scalability problem
    - Storage
    - Dynamics

⇒ **Decentralization**



- Distributed local processing
  - Space intensive
  - Top-k freshness threatened
  
- Storage-free solution
  - Many messages
  - No resilience to churn

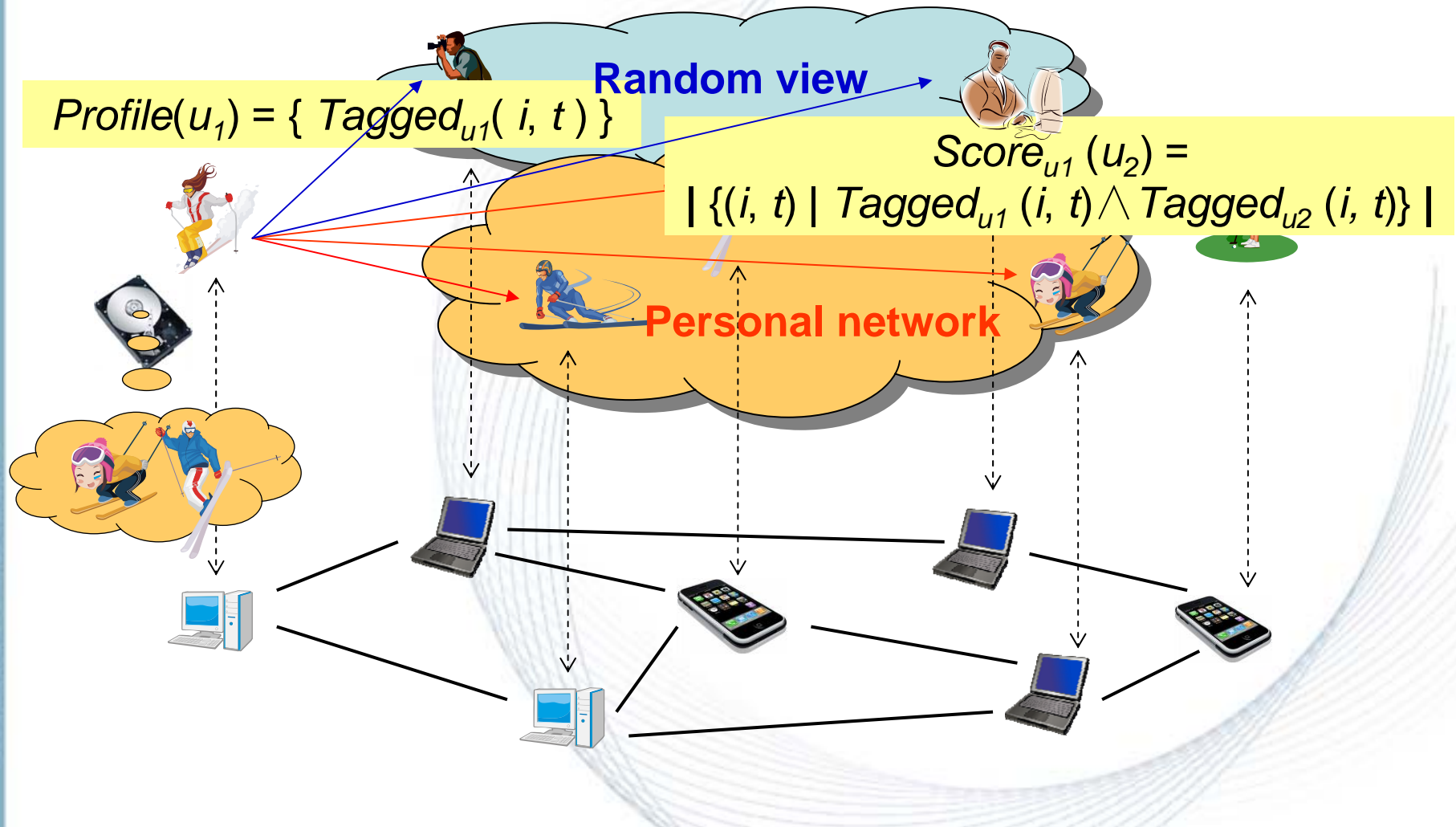


⇒ Gossiping the query



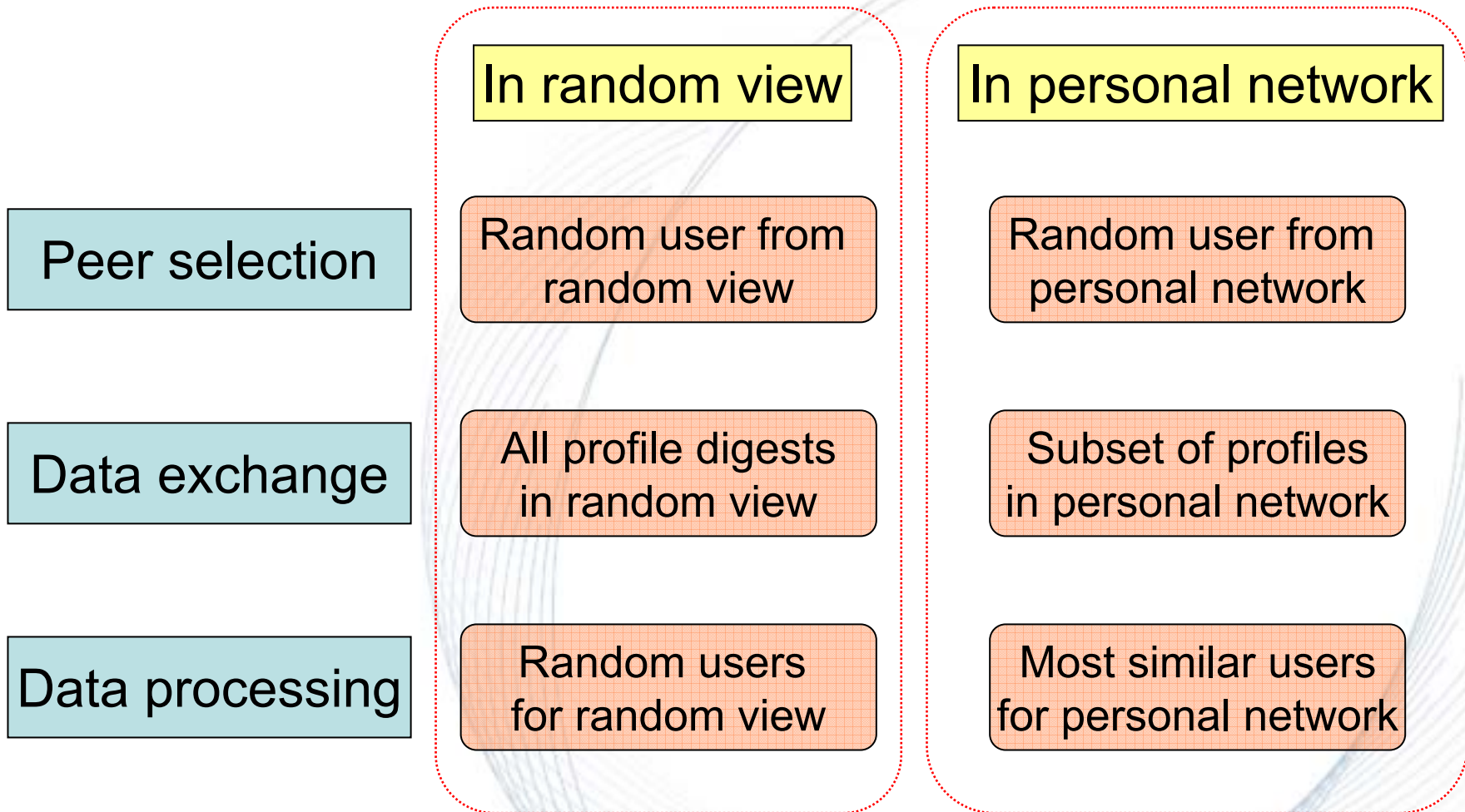
- Goals
  - Efficient query processing
  - Storage at will
  - Low bandwidth consumption
  - Resilient to dynamic users
- Bimodal gossip-based protocol
  - Identifying right users for personal network
  - Propagating the query
  - Gathering the information for query processing

# System model



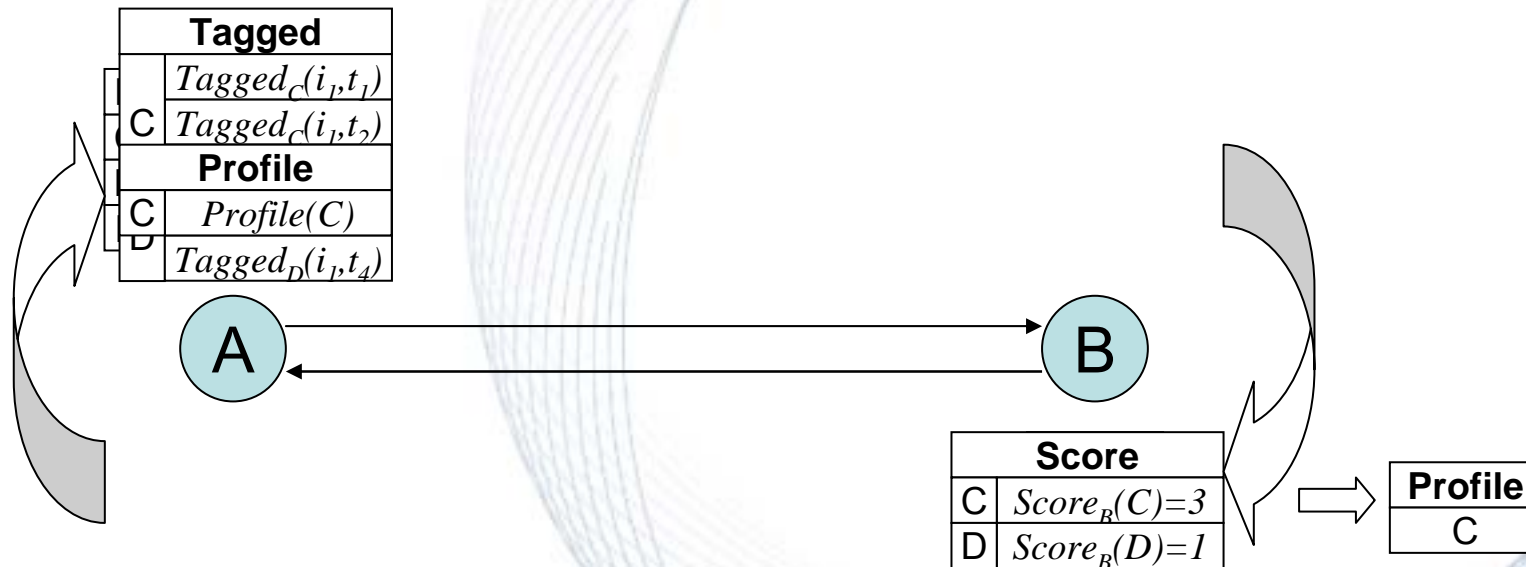


- Lazy mode
  - Measure similarity between users
  - Maintain personal network
  - Periodically at low frequency
- Eager mode
  - Gather information in personal network
  - Process the query
  - Maintain personal network
  - On demand at high frequency

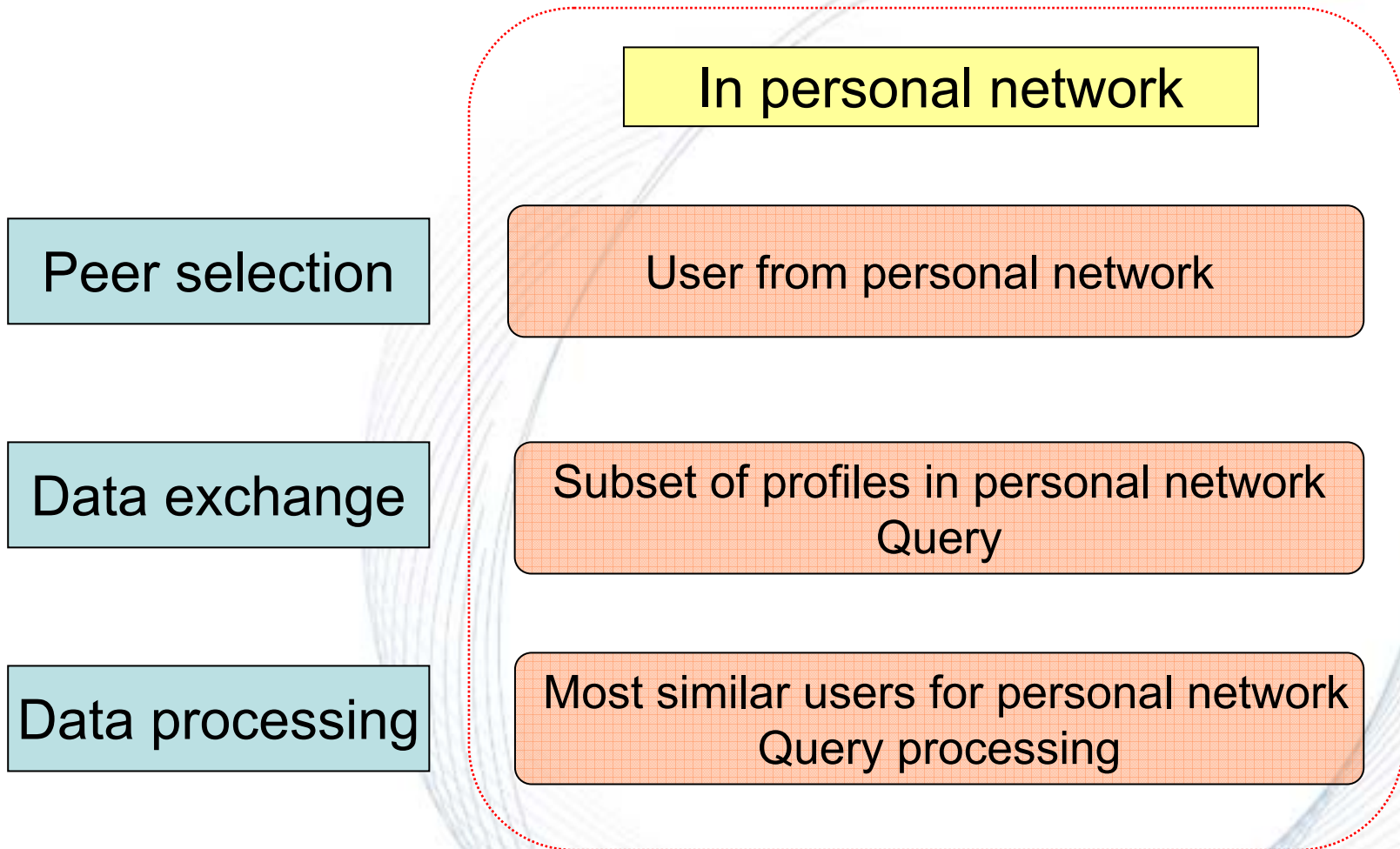




- 3-step data exchange
  - Similarity estimation: profile digest in Bloom filter
  - Similarity measure:  $Tagged_u(i, t)$  of common items
  - Local storage update: profiles



# Eager mode



# Gossiping queries

Network(A)	
C	Profile(C)
D	Profile(D)
L	Profile(E)
B	
F	
G	
H	
I	
J	
K	
L	

Partial Result List A

Local Partial Top-k

Network(B)	
A	Profile(A)
F	Profile(F)
G	Profile(G)
C	
D	
I	
K	
L	
M	
N	
O	

Partial Result List B



Partial Top-k Cycle 1

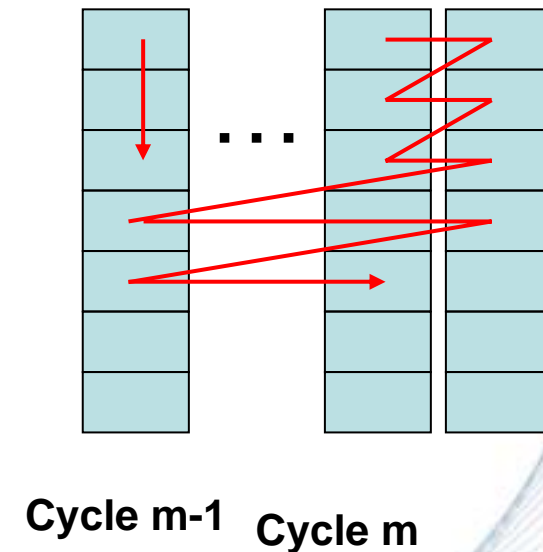
$L_o(A)$
H
I
J

$L_o(B)$
K
L

# Top-k processing

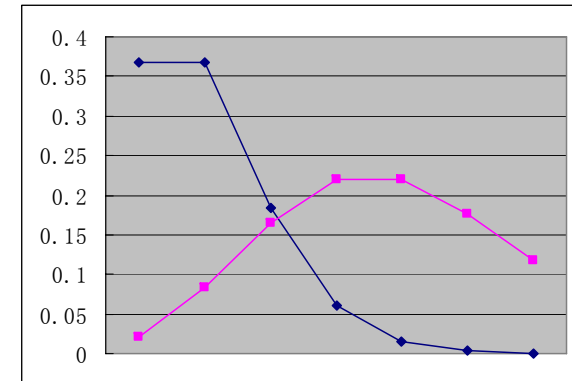
- Gossiped user
  - Partial result list
- Querier
  - Extended NRA at each cycle
  - Stop conditions
    - All profiles in personal network are used
    - Satisfied with partial top-k

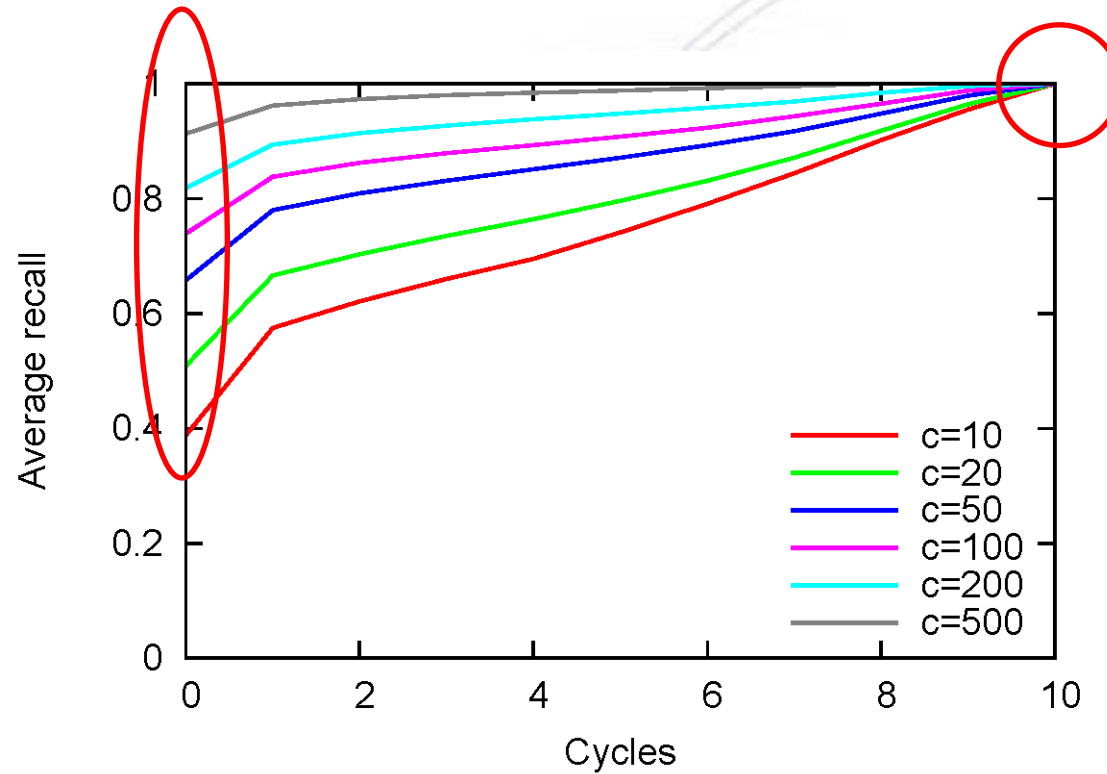
<i>Item</i>	<i>PartialScore<sub>Q,u</sub>(i)</i>
$i_1$	10
$i_2$	5





- Data
  - Real trace from delicious
  - 10,000 random users' profiles
- System setting
  - Size of personal network: 1000
  - Profiles in personal network:
    - Homogenous systems: 10, 20, 50, 100, 200, 500, 1000
    - Heterogeneous systems: Poisson Distribution  $\lambda = 1$ ,  $\lambda = 4$
  - Size of random view: 10





The more profiles stored, the better the local top-k  
 Regardless of storage, queries are satisfied within 10 cycles

# Bandwidth consumption

- Personal network maintenance

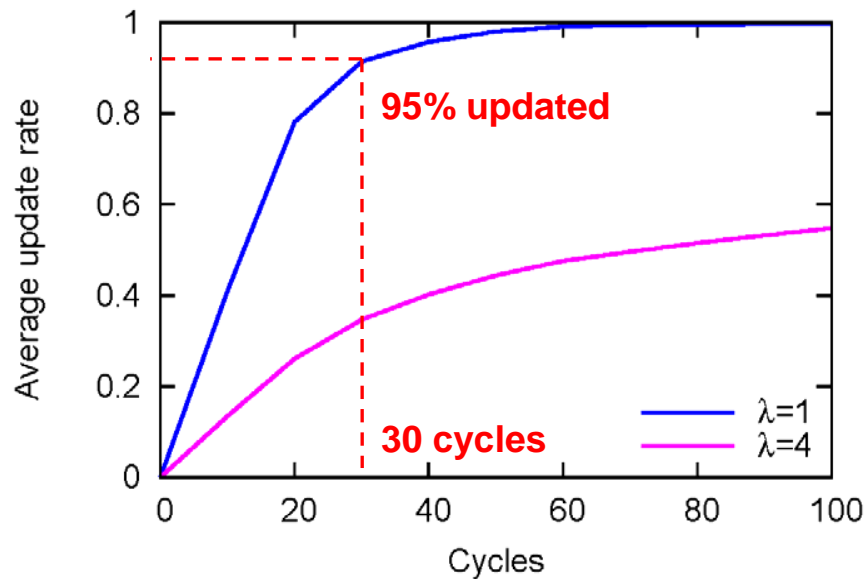
Storage	% users		bytes/cycle and user	
	2-step	3-step	2-step	3-step
$\lambda=1$	87.6%	4.1%	15.8K	503K
$\lambda=4$	88.0%	12.9%	23.6K	545K

- Query processing

	bytes/query	messages/query
$\lambda=1$	573K	228
$\lambda=4$	360K	70

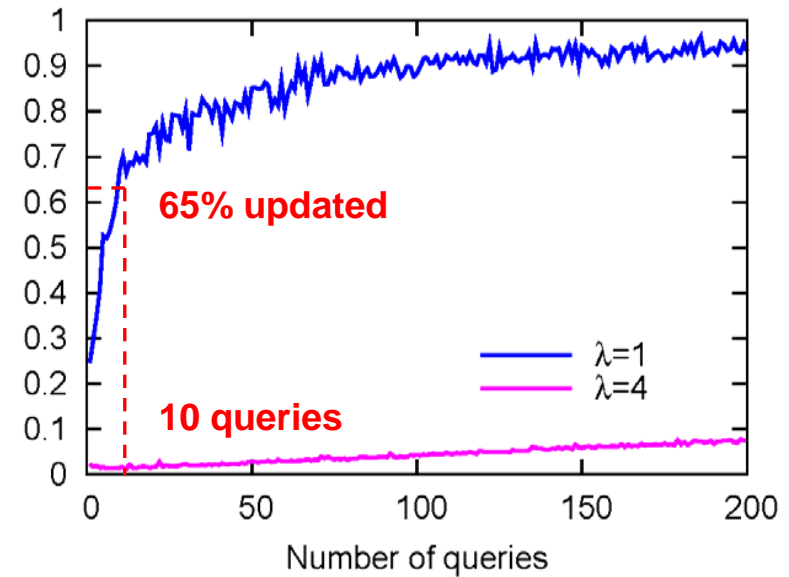
# Profile dynamics

- Lazy mode



Changes quickly captured

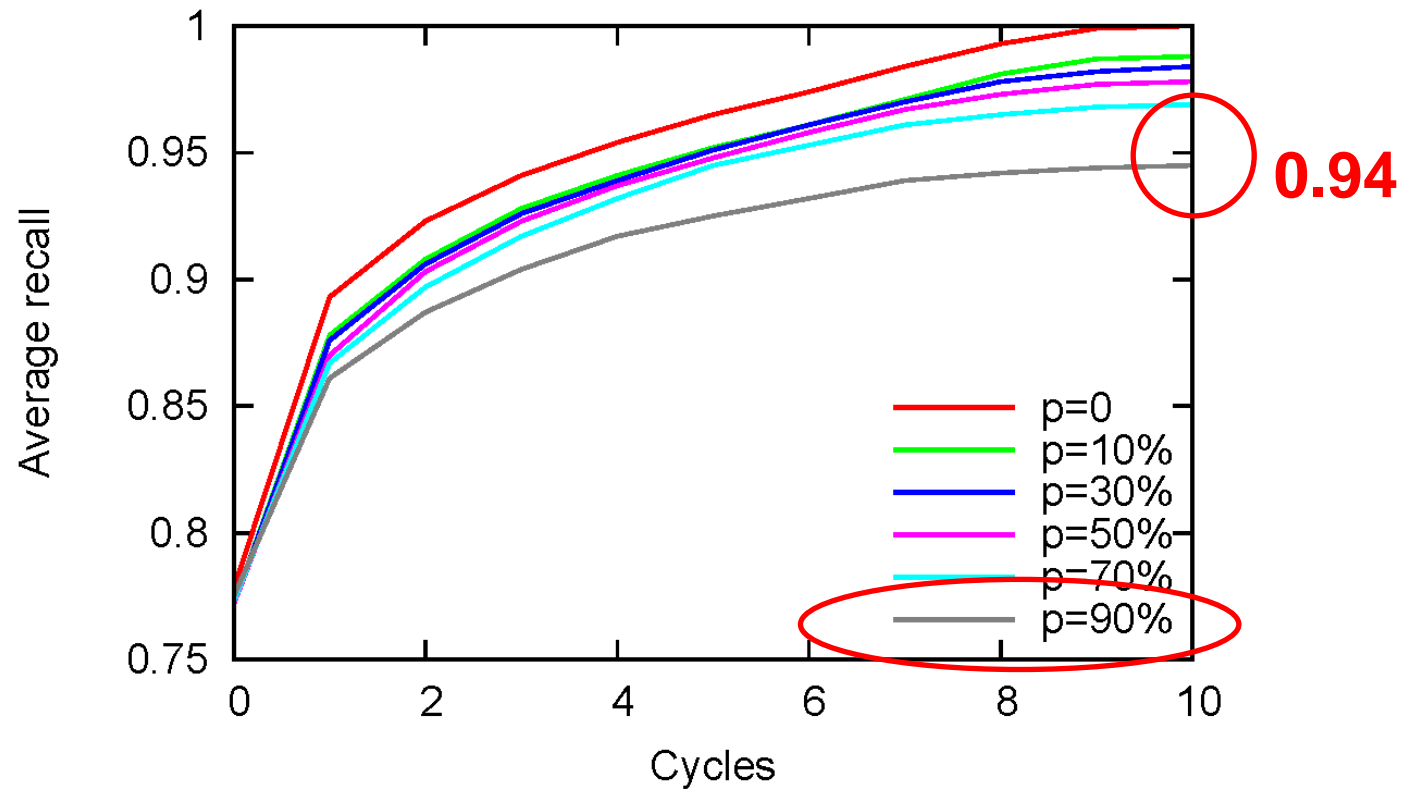
- Eager mode



Fresh wave for gossiped users



# User departure



Top-k quality is insensitive to the user departure

- Personalized query processing in a fully decentralized and collaborative way
- Bandwidth effective
  - Low maintenance cost
  - Forwarding query while maintaining personal network
- Adaptive storage
- Resilient to dynamic users
  - Profile update
  - User departure

Thank you !

13th International Conference on Extending Database Technology (EDBT 2010)  
March 22-26, 2010 Lausanne Switzerland

# Gossiping Personalized Queries

*Xiao Bai*

[xbai@irisa.fr](mailto:xbai@irisa.fr)

*INSA-Rennes, France*