WhatsUp: a P2P instant news item recommender

Anne-Marie Kermarrec

INRIA Rennes
The new deal

User centric: Web content is generated by you, me, your friends and millions of others

TIME
You.

You control the Information Age. Welcome to your world.

Inria
Social networking in a 2 billion user world

- YouTube: 1 billion accounts (Oct 2010)
- digg: 20M unique/month
- Facebook: 900M (Spring 2012)
- delicious (social bookmarking): 6M unique/month
- Twitter: 200M registered
- Flickr: 50M registered (Oct 2011)
- Foursquare: 10M (Oct 2011)
- Wikipedia: 15M registered, 140,000 contributors on 30 days
The new deal

A user-centric Web

A small world: everybody sits close to everybody else
The new deal

User-centric

Small world

**Navigation shift:** Notification is taking over
Notification is taking over
En librairie le 12 janvier 2012

Les nouvelles classes moyennes

par Dominique Goux & Eric Maurin

Mots-clés : classes sociales

Les classes moyennes sont souvent considérées comme le noyau stable de la société. En réalité, elles sont le lieu où s'expriment les aspirations les plus intenses à l'ascension sociale et les craintes les plus aiguës face au déclassement. Aiguillonnées par la peur de tomber et le désir de s'élever, elles ont su maintenir leur position tout au long des dernières décennies, au terme d'une compétition sans merci pour les statuts les plus protégés, les quartiers les plus sûrs et les diplômes les plus valorisés. En librairie le 12 janvier 2012.
Livres

Collection dirigée par Pierre Rosanvallon et Ivan Jablonka.

Les nouvelles classes moyennes
par Dominique Goux & Eric Maurin

Les classes moyennes sont souvent considérées comme le noyau stable de la société. En réalité, elles sont le lieu où s'expriment les aspirations les plus intenses à l'ascension sociale et les craintes les plus aiguës face au déclassement. Aiguillonnées (…) 

Refaire société
par Collectif & Pierre Rosanvallon

La crise financière, la montée de la précarité et de la pauvreté, l'accroissement des inégalités mettent en péril la cohésion de notre société. Le délèvement du lien social est aussi aggravé par le repli sur soi et la profonde méfiance des citoyens à l'égard (…)
The new deal

User-centric

Small world

Navigation shift

Personalization around the corner:
recommendation, RSS feeds, social networks, …
A personalized environment
We are not there yet
Why is it so difficult?

• Dynamic interests
• Small portion of the available information is of interest
• Interesting stuff does not come always from friends
• Classical notification systems do not filter enough
What’s wrong with centralized versions?

Privacy: Big Brother is watching you

*If you only knew the power of the Dark Side.*
– Darth Vader

Scaling issues
ACHIEVING PERSONALIZED NOTIFICATION IN A DECENTRALIZED WAY
A call for decentralization

User-centric world
Small world
Navigation shift
Personalization
Privacy-aware
Scalable
WhatsUp: a decentralized recommender
WhatsUp: a decentralized news item recommender

Simple interface: I like it or I don’t

Create a dynamic implicit social network

Disseminate news along this topology

An implicit notification system
WhatsUp in a nutshell

Generate profile

Influence dissemination

Tune dissemination

WUP

BEEP

Influence topology

User Opinion
WhatsUp challenges

Who are my social acquaintances
How to discover them?
How to disseminate news items?
How to preserve users’ privacy

→ Similarity metric
→ Sampling
→ Biased epidemic protocol
Which nodes for the social network?

Model

\( U(\text{sers}) \times I(\text{tems}) \) (news items)

Profile\((u) = \text{vector of liked news items} \)

Cosine similarity metric

\[
\text{Similarity} (n, p) = \frac{n \cdot p}{\|n\| \|p\|}
\]

Minimal information: no tag, no user’s input
WhatsUp challenges

Who are my social acquaintances?

**How to discover them?**

How to disseminate news items?

How to preserve users’ privacy

Through sampling
Gossip Protocols

Probabilistic approach

Goal: system to converge to a desirable outcome

Trade determinism for scalability & robustness
Gossip-based computing

Each node maintains a set of neighbors (k entries)

Parameter Space
- Peer selection
- Data exchange
- Data processing

Random Peer sampling [ACM TOCS 2007] achieves random topologies
Random dynamic graph

Result \rightarrow \text{random graph}

Highly resilient against churn, partition

(1) peer list exchange \hspace{1cm} (2) merge and truncation

Small diameter
Sampling the network with gossip

1. Exchange of neighbors lists

2. Similarity computation

3. Neighborhood optimization
The WhatsUp social network

clustering layer
gossip-based
topology clustering

RPS layer providing
random sampling

Random link  Social link  node
### Data structures

#### Social Network of the closest entries

<table>
<thead>
<tr>
<th>IP:port</th>
<th>132.154.8:2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloom Filter</td>
<td>01011011001</td>
</tr>
<tr>
<td>Profile</td>
<td>I like: N₁, N₂, ...</td>
</tr>
<tr>
<td></td>
<td>I don't: N₁₀, N₁₃,...</td>
</tr>
<tr>
<td>Update time</td>
<td>5</td>
</tr>
</tbody>
</table>

#### Uniform (dynamic) sample of k random entries

<table>
<thead>
<tr>
<th>IP:port</th>
<th>102.14.1.1:2110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloom Filter</td>
<td>100100000110</td>
</tr>
<tr>
<td>Update time</td>
<td>30</td>
</tr>
</tbody>
</table>

---

*Exchange of Bloom filters*
Convergence

c current neighbors versus the c closest

Cycles

Random sampling

Biased sampling
WhatsUp challenges

Who are my social acquaintances

How to discover them?

How to disseminate news items?

Biased epidemic protocol (BEEP)
Epidemic dissemination

The $\log(n)$ magic

Each node forwards the message to $f$ nodes chosen uniformly at random

If $f=O(\log(n))$, “atomic” broadcast whp
Result is valid if the fanout for each peer is on average

Fanout: $\log(n) + c$

Random is connected: $p(\text{connect}) = e^{-e^c}$
BEEP: orientation and amplification

Orientation: to whom?
- Forward to friends
- Forward to random users

Amplification: to how many?
- Increase fanout
- Decrease fanout
I like it: Exploit

Orientation: **to whom?**  Amplification: **to how many?**

Forward to friends  Large fanout

\[ F \approx \log(N) \text{ friends in the social network} \]
I don’t like it: Explore

Orientation: to whom?  Amplification: to how many?

Forward to random users  Small fanout

F = 1 node from the RPS, the closest to the news items profile
Evaluation

User Metrics: Recall-Precision

System metrics: Number of messages-Redundancy

Traces
- Real trace from a 480 user survey on 1000 news items
- Delicious and Digg crawls
WhatsUp in action on the survey

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>Redundancy</th>
<th>Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gossip</td>
<td>0.34</td>
<td>0.99</td>
<td>0.85</td>
<td>2.3 M</td>
</tr>
<tr>
<td>Cosine-CF</td>
<td>0.64</td>
<td>0.12</td>
<td>0.27</td>
<td>30k</td>
</tr>
<tr>
<td>Whatsapp</td>
<td>0.53</td>
<td>0.78</td>
<td>0.28</td>
<td>280k</td>
</tr>
</tbody>
</table>
WhatsUp in action

![Graph showing F1-score vs Fanout for different recommendation methods: WhatsUp, Gossip, and Cosine based CF.](image-url)
WhatsUp challenges

Who are my social acquaintances?

How to discover them?

How to disseminate news items?

How to preserve users’ privacy?
A wide range

Blurring the profile

Anonymity: Onion-like solutions

Traditional privacy: threshold protocol

Differential privacy: distributed noise
Blurring the profile

Private User profile

User Profile used locally for similarity computation

I like it

News item profile

Aggregation of profiles of users who liked the news item

Public User profile

User Profile exchanged during gossip
Impact of profile bluring

![Graph showing the impact of profile bluring on F1-Score with different fanout values for Privacy-unaware and Privacy-aware WhatsUp.]
Resilience to attacks

Recall vs. Number of Attackers

- WhatsUp
- Privacy-unaware WhatsUp
Anonymity (sender and receiver)

Dissociates the profile from the user’s identifier

Additional Private RPS

Relay nodes from the RPS

Three types of keys

• Chain keys (RPS)
• Profile keys (Private RPS)
• Symmetric keys for path anonymity

Private RPS provides profile keys and proxys

Onion routing: chain of random nodes, keys provided by the RPS
Chain created of random nodes, keys assigned by the nodes
Anonymous Profile exchange

A encrypts its profile with Profilekey of B (Private RPS)
Then with Chainkeys of R_A and P_A (RPS)
Profile hiding: threshold protocol

The profile might be enough to identify a user

Threshold

Nothing in common | All in common

Only one bit of information revealed

Homomorphic encryption [Golreich 2003] allows operations (multiplication, scaling) to be performed directly on ciphertexts.
Differential Privacy

- Auxiliary knowledge
- Laplacian noise generation [Dwork et al 2006]
- Independent noise generation: Can be applied to cosine-similarity based measures
To take away

Personalization is needed

Decentralization is healthy

Gossip-based computing is one (the) way to go
Thank you

[www.gossple.fr](http://www.gossple.fr)

Joint work with M. Alaggan, A. Boutet, D. Frey, S. Gambs, A. Jégou (INRIA), R. Guerraoui (EPFL)