Adaptive algorithms for efficient content management in social networks

Claudia Canali
Michele Colajanni
Riccardo Lancellotti

University of Modena
and Reggio Emilia

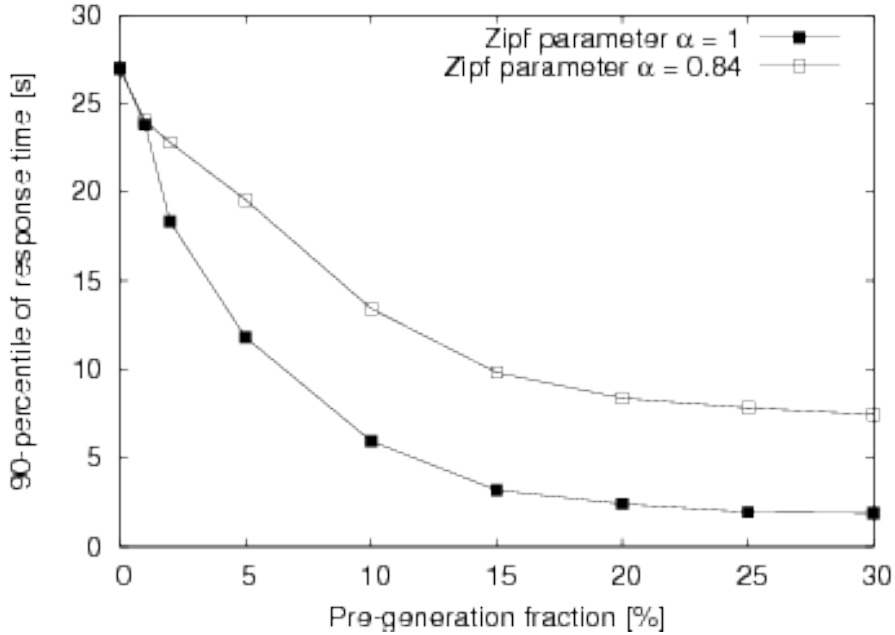
Future Web Scenarios

- **Community-based services**
  - Social networking: support for user interaction be the killer of future Web
  - Rich-media content
  - Presence of Mobile User access

- **Workload evolution in the next four years**
  - Computational demand will grow faster than CPU power (Moore's Law)

Motivations for content management

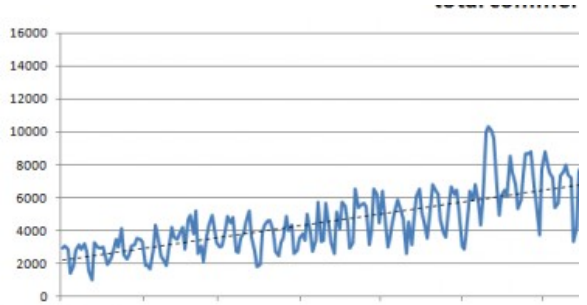
- **Content management**
  - Content replication
  - Caching
  - CDN delivery
  - Resource pre-generation



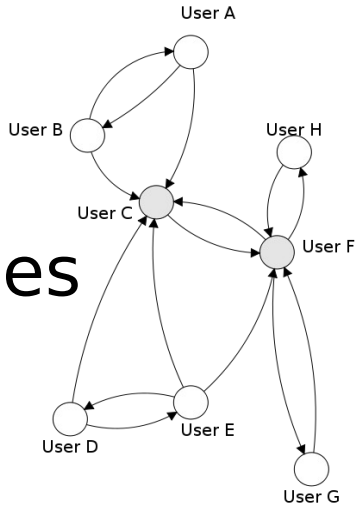
- **→ Need to identify the Hot set of popular resources**
  - Variability in workload characteristics
  - Rapid variations in access patterns
  - Workload dynamics related to social interactions
- **→ Need for algorithms providing early and fast detection of popular resources.**
- **→ Stable performance are not an optional**

Hot set identification

- **The algorithm must identify the most popular resources (Hot set)**
  - Hot set is evaluated periodically with interval Δt
  - Hot set resources will receive the highest number of accesses in the interval [t, t+Δt]
- **Predictive-based algorithm**
  - Evaluates past access patterns and uses a simple predictor to forecast future accesses



- **Social-based algorithm**
  - Evaluates number of incoming social links
  - High connection degree → popular resources
- **Combination of approaches**
  - → must merge heterogeneous information

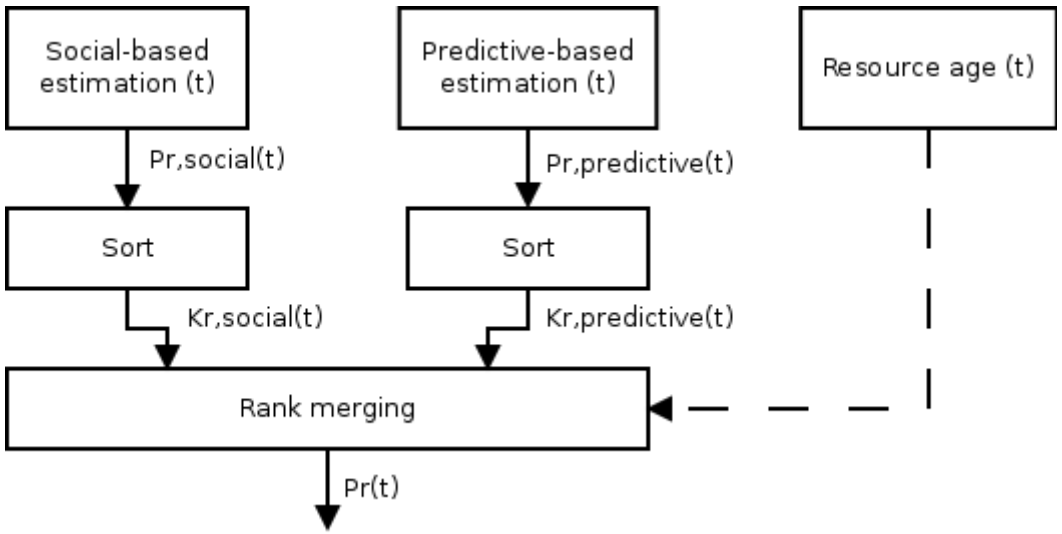


Proposed algorithms

- **Proposal: novel algorithms that merge access pattern prediction and social information**
  - Rank-age
  - Linear-adaptive
  - Rank-adaptive
- **Use of adaptive techniques that takes into account workload characteristics**
- **Comparison with existing solutions**
  - social- and predictive-based

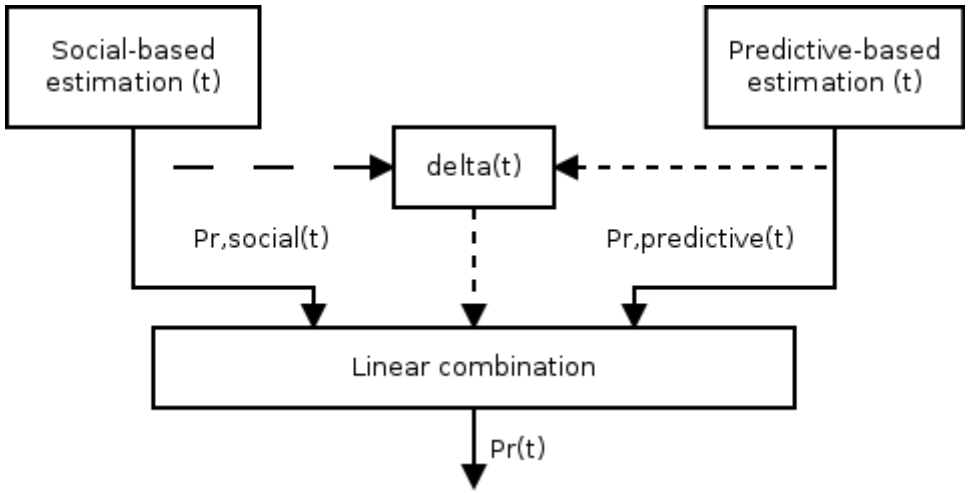
Rank-Age algorithm

- **Social- and predictive-based information have different probability distribution**
  → Use of rank merging
- **Weighting different information:**
  - Predictive information are more reliable for older resources
  - Social-based information may be used otherwise
- **Resource age is used to determine the weight in rank-merging**



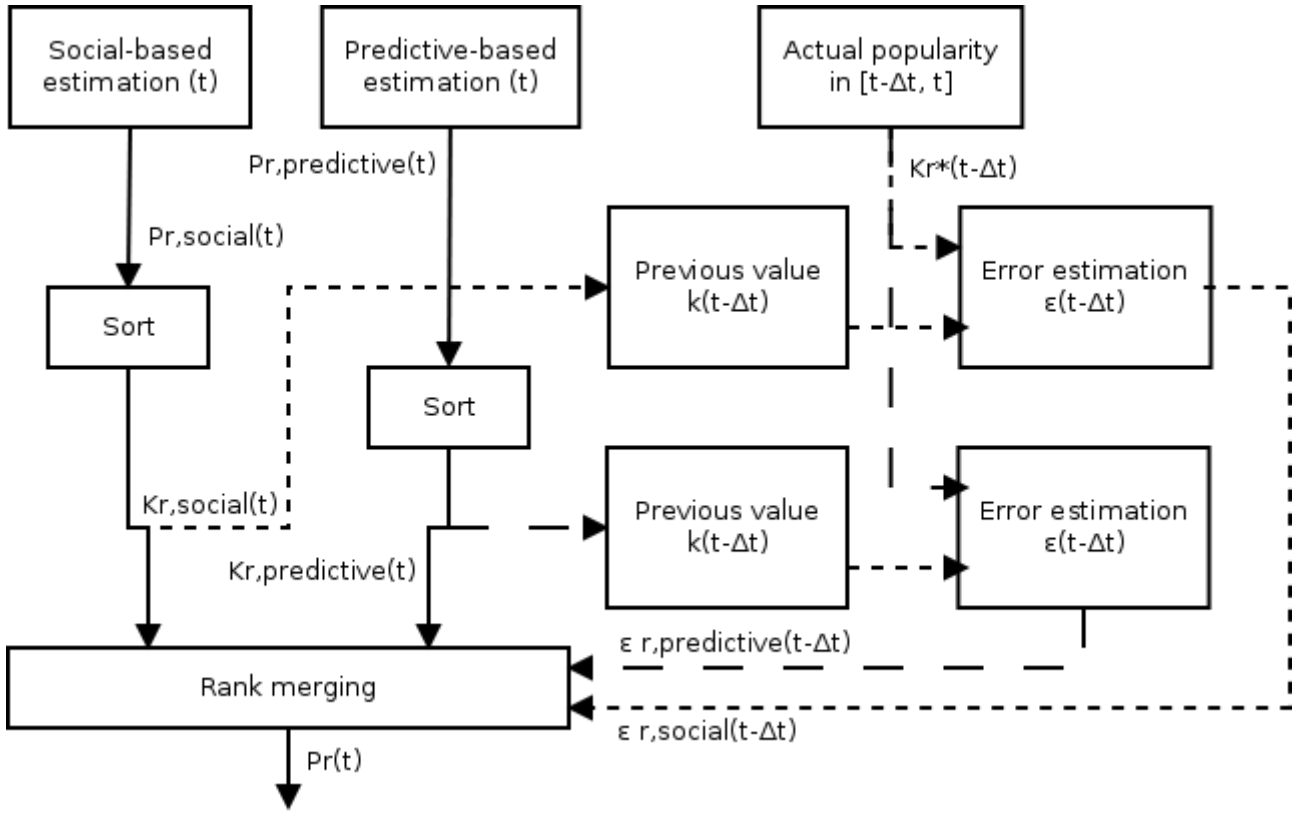
Linear-Adaptive algorithm

- **Social-based and predictive based information have different probability distribution**
  → use of adaptive technique to estimate the weight of each information
  → need to normalize different values
- **The weighting function takes into account *median* and *quartile* information about social information and predicted accesses for the whole working set**



Rank-adaptive algorithm

- **Use of rank merging → handles different probability distribution**
- **Use of a feedback on the popularity estimation errors in previous interval to compute the weight used in rank merging**



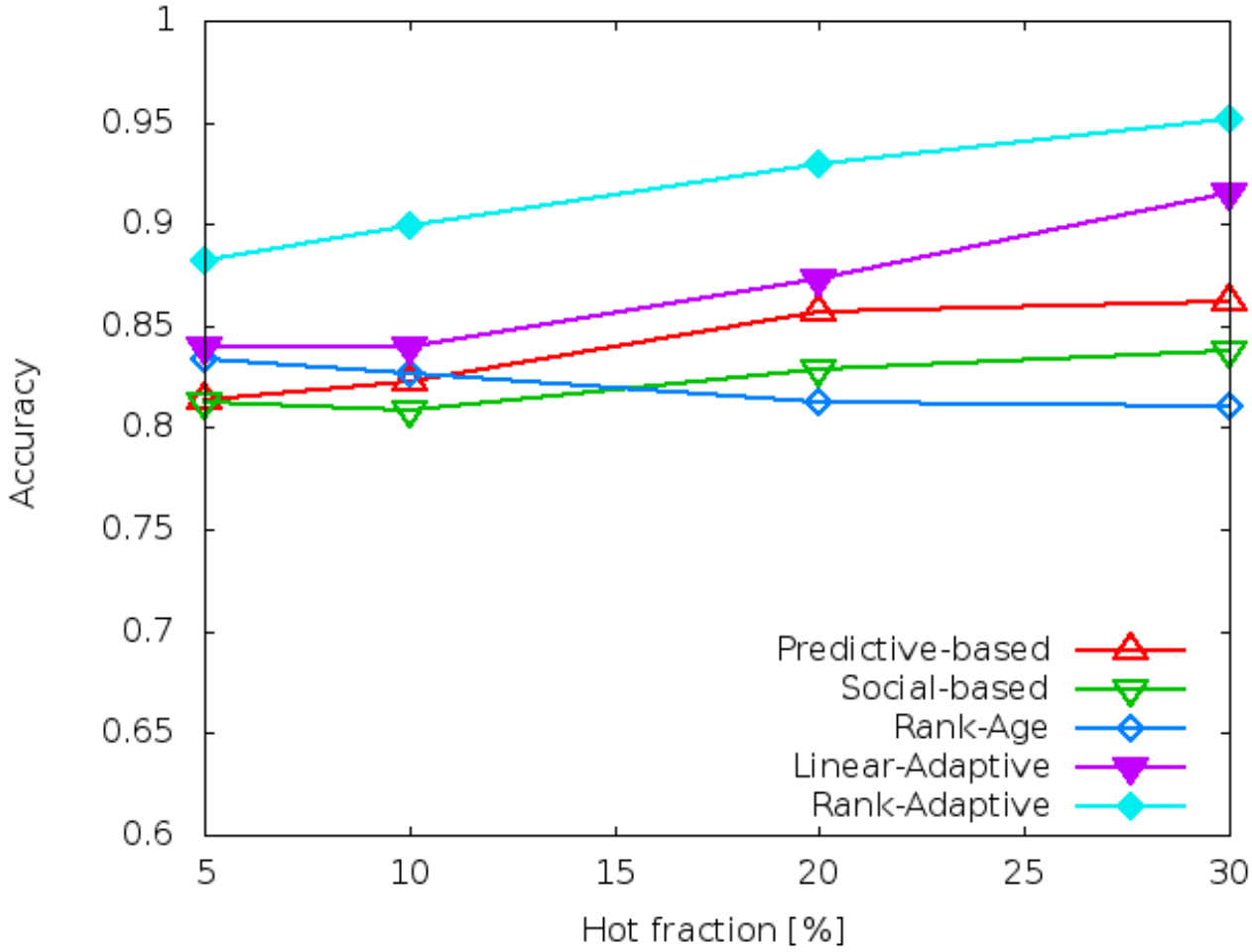
Experimental setup

- **Simulation based on Omnet++ framework**
  - User population up to 20000 units
  - Average of 100 requests/sec
  - 12 hours of simulated time
  - Δt=20minutes
  - Main metric: accuracy=|HS(t) ∩ HS*(t)|/|HS*(t)|

Parameter	Range	Default
Hot fraction [%]	5%-30%	20%
Upload percentage [%]	1%-20%	5%
User/resource popularity correlation	0.6-0.8	0.7

Performance evaluation

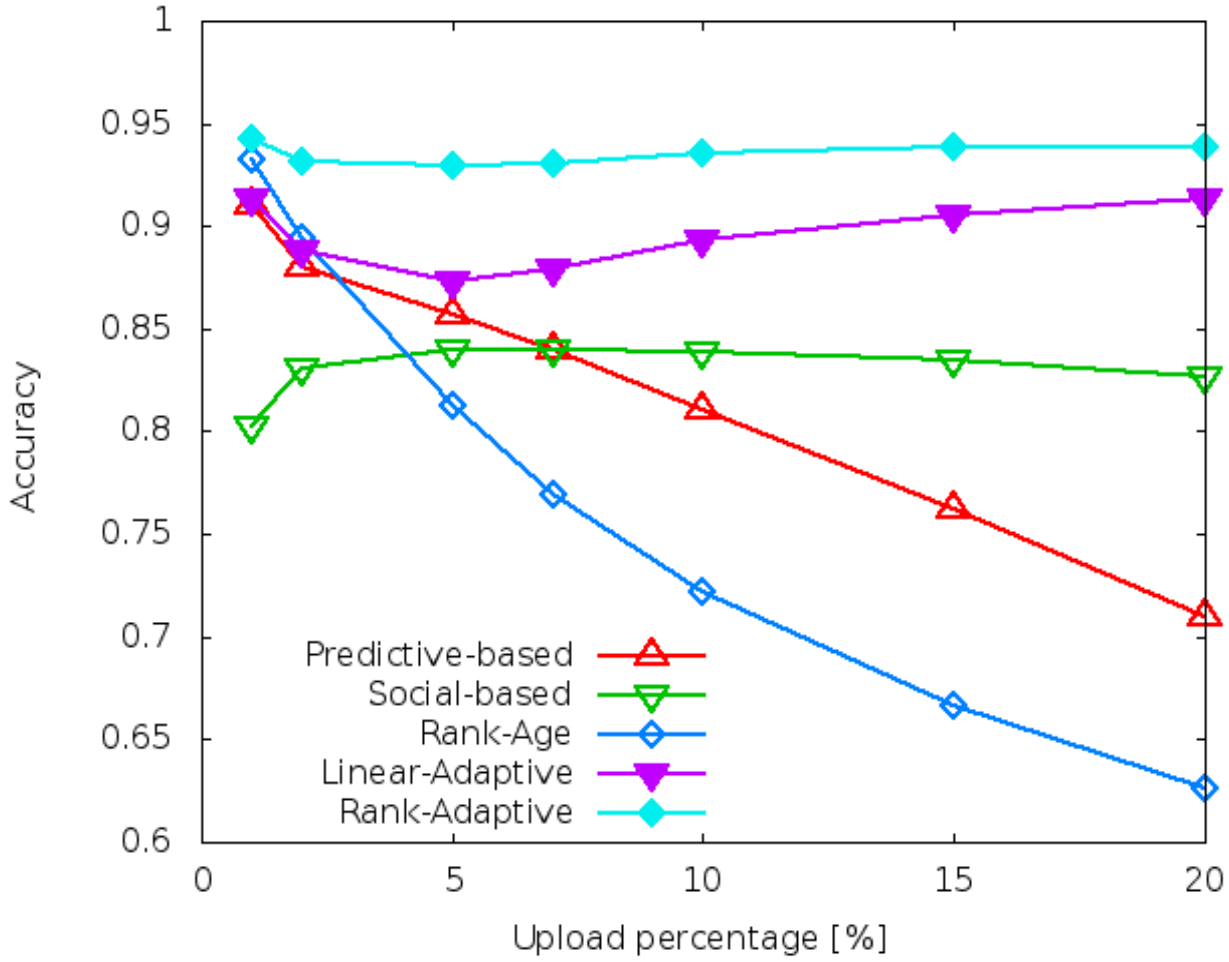
- Predictive and social-aware algorithms can be improved
- Adaptive algorithms outperforms other solutions
- Rank-age algorithm provides poor performance because it tends to prefer younger resources even when they are not popular



→ Need to evaluate performance stability

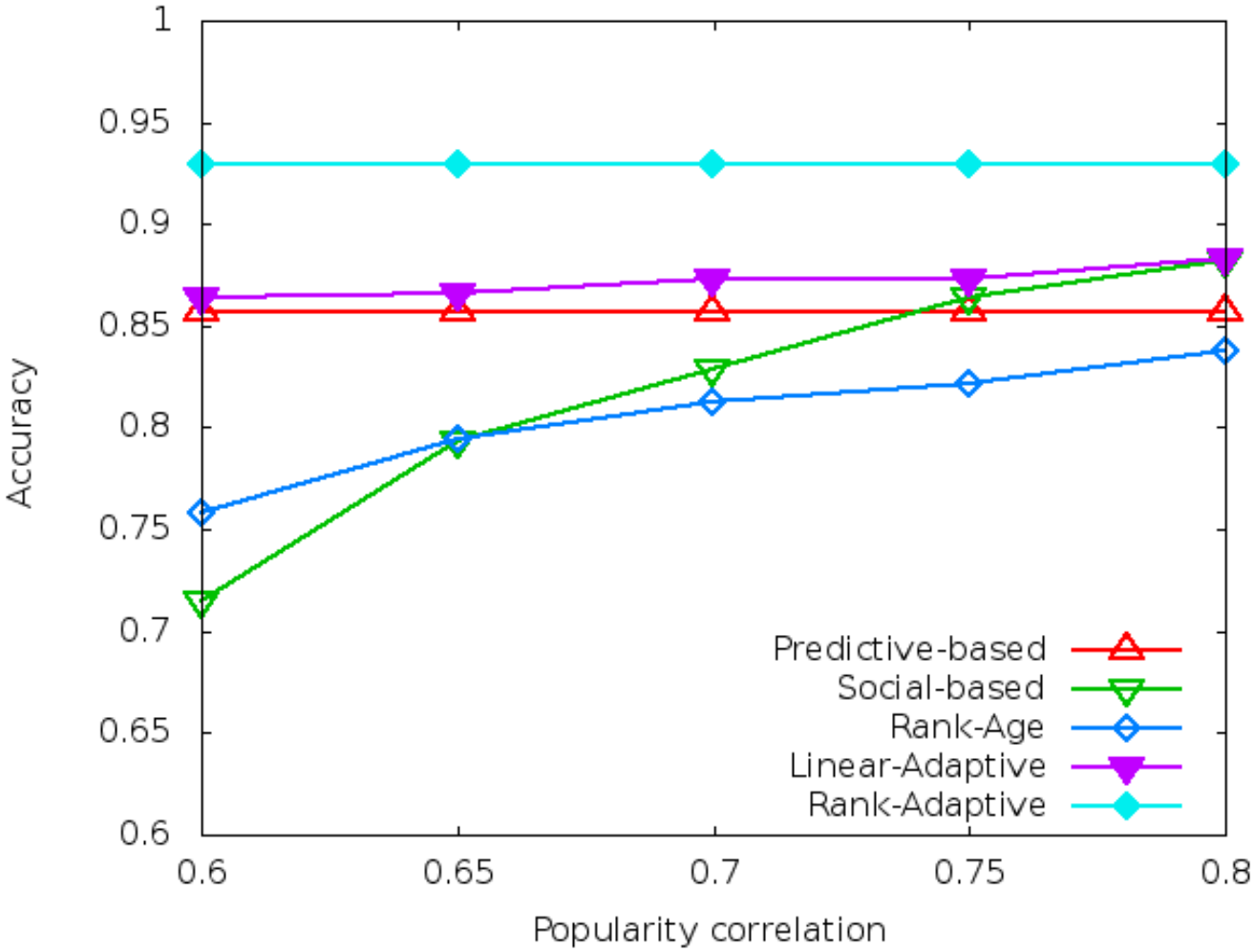
Sensitivity to workload dynamics

- Prediction is highly sensitive to upload percentage
- Social-aware algorithm is not sensitive to workload dynamics
- Rank-age algorithm provides poor performance when many young resources are present
- Adaptive algorithms provide stable performance



Sensitivity to social parameters

- Prediction is not affected by social phenomena
- Social-aware is highly sensitive to the correlation between user and resource popularity
- Rank-age relies on social-aware algorithm and shares its drawback
- Adaptive algorithms provide very stable performance



Conclusions

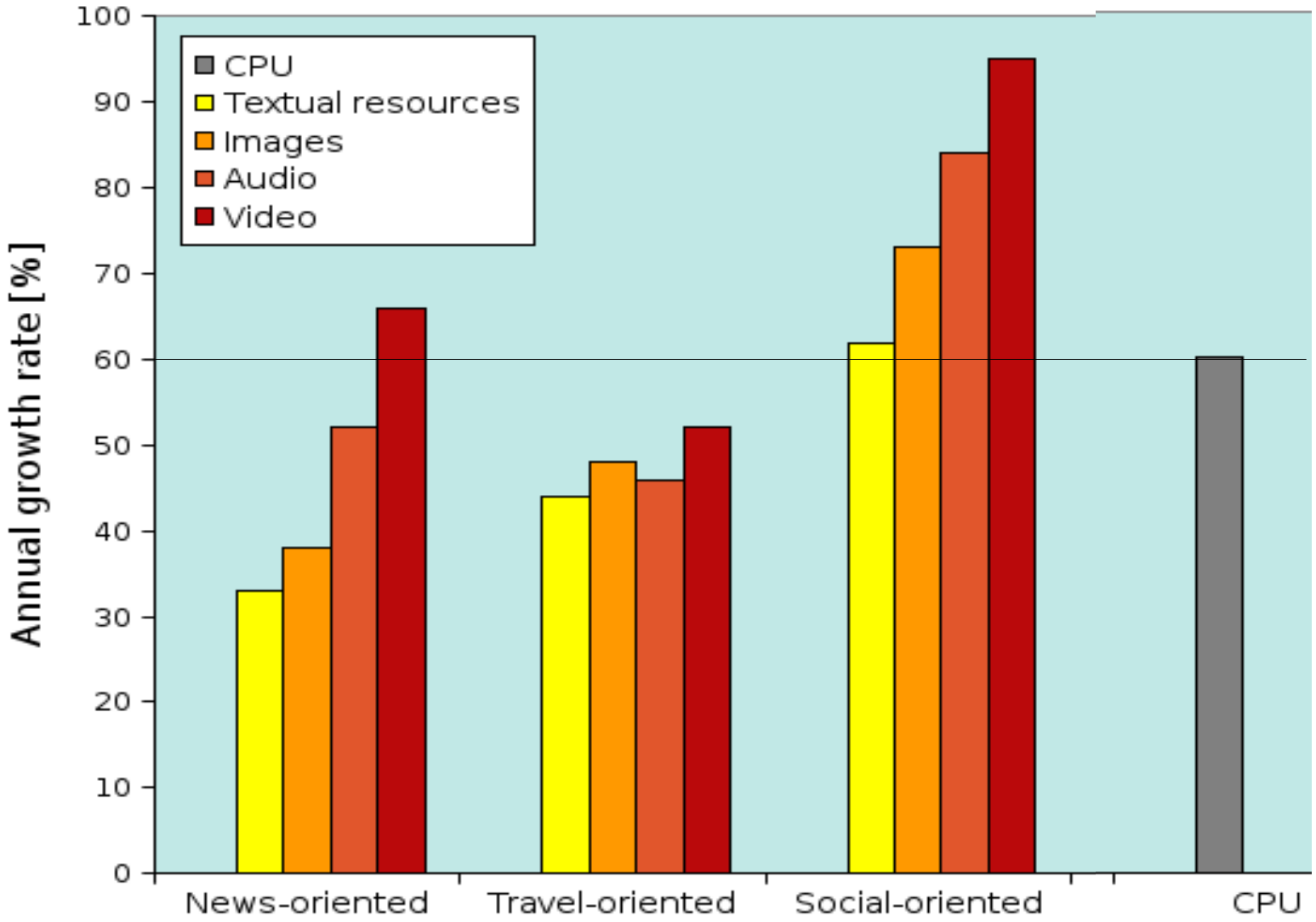
- **Content management will be fundamental for future social network applications**
  - Need to identify the Hot set
  - Must cope with novel challenges (social interaction, short resource lifespan, ...)
  - Need for high accuracy and stable performance
  - Can rely on heterogeneous information, but we must combine them
- **Proposal of different algorithms that combine heterogeneous information**
  - Adaptive techniques allow to exploit the benefits of predictive and social-aware information
  - Non-adaptive approach result in poor and highly variable performance

Adaptive algorithms for efficient content management in social networks

Claudia Canali
Michele Colajanni
Riccardo Lancellotti

University of Modena
and Reggio Emilia

Expected growth of computational demands



Blue

Predictive-Social algorithms

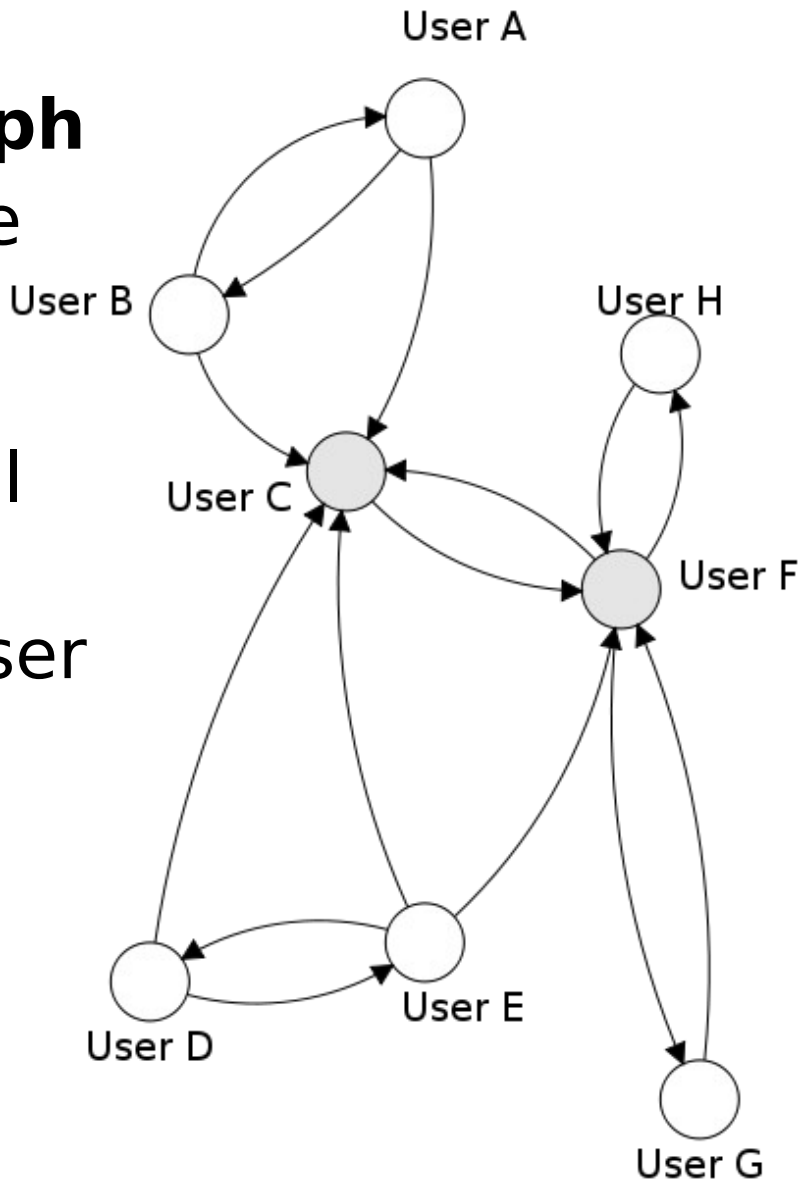
- **Merging social-aware and predictive information**
  - $p_r P(t)$ → predictive
  - $p_r S(t)$ → social
  - $\delta(t)$ → weight
- **That is:**
  - $p_r(t)=\delta(t)\ p_r P(t) + (1-\delta(t))\ p_r S(t)$
  - $\delta(t)=QWM(PS(t))/(QWM(PS(t)) + QWM(PP(t)))$

Predictive algorithms

- **History of past accesses to resource r represented as a time series:**
  - $D_r(t)=\{d_r(t), d_r(t-\Delta t), \ldots, d_r(t-(n-1)\Delta t)\}$
  - $d_r(t)$ is number of accesses to resource r in interval $[t-\Delta t, t]$, $d_r(t-\Delta t)$ refer to $[t-2\Delta t, t-\Delta t]$, ...
- **Use of an EWMA model for prediction:**
  - $d_r^*(t,t+\Delta t)=\gamma d_r^*(t,t+\Delta t)+(1-\gamma)d_r(t)$
  - $\gamma=2/n$, where n is the time series length
- **Other prediction models are possible**

Social-aware algorithms

- **Social network can be represented as a directed graph**
  - Reverse contact represent the popularity of a user within the social network
  - User navigation exploits social links
  - Strong correlation between user popularity and popularity of uploaded resources
  - → Popular users are likely to publish popular content



Predictive-Social algorithms

- **Most innovative class of algorithms**
  - Merges information from two sources:
  - Prediction
  - Social information
- **Need for a reliable way to merge two completely different sets of data**
  - Different value ranges
  - Different probability distributions
- **Use of a robust weighting function**
  - Two-sided quartile weighted median
  - Given distribution P(t):
  - $QWM(P(t))=(Q_{25}(P(t))+2Q_{50}(P(t))+Q_{75}(P(t)))/4$

Red

Green

Black