

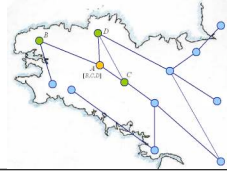
PROXSEM: Interest-based Proximity Measure to Improve Search Efficiency in P2P Systems

Yann Busnel, Anne-Marie Kermarrec

IRISA / ENS Cachan, IRISA / INRIA

Peer-to-Peer File Sharing System

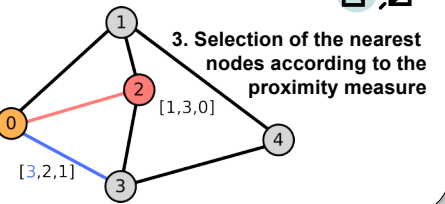
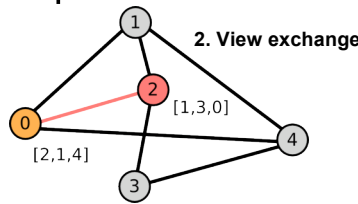
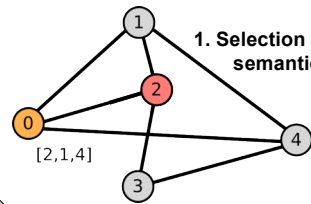
Example of a file sharing peer-to-peer network



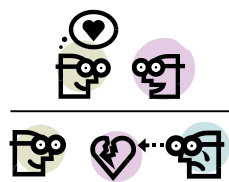
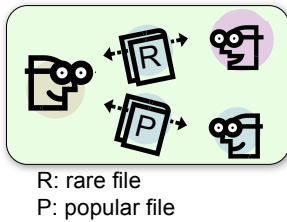
- Each node owns:
- A view: list of neighbours
 - A cache: set of files

- Semantic neighbours:** nodes which have several common interests
- Search in these systems:** send a query to semantic neighbours
- Naïve interest-based proximity measure:** number common files
→ Biased genuine proximity most cases.

Gossip-based Communication

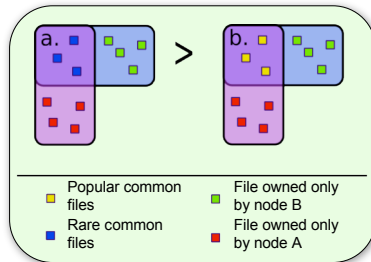


File popularity impact

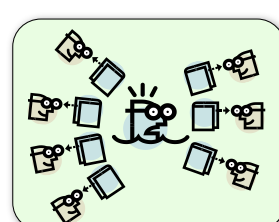


Contribution strategy

Favour nodes which cache overlap contains the largest number of rare files

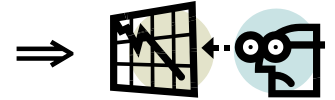


Node generosity impact

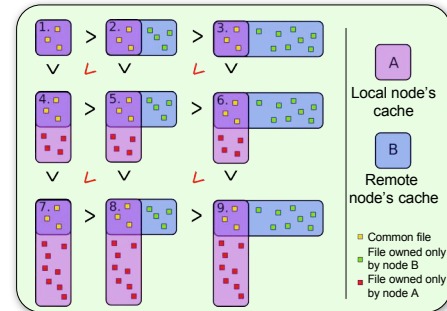


A generous node has a higher probability to be chosen as a semantic neighbour:

Node Overload

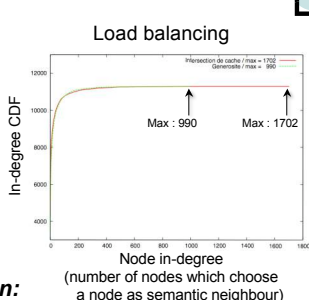
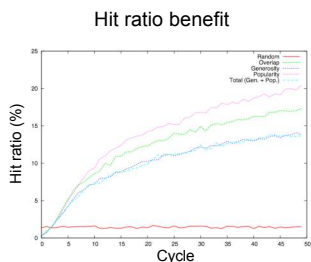


Contribution Strategy



Case classification to minimize this impact

Simulation Results



Semantic neighbours selection:

Random

Overlap: According to the size of cache overlap

Generosity: According to PROXSEM without popularity ($\gamma=0$)

Popularity: According to PROXSEM without generosity ($\alpha=1; \beta=0$)

Total: According to complete PROXSEM

PROXSEM: Semantic Proximity Measure

$$\xi_A(B) = \overbrace{|\sigma_{A,B}|}^{\text{Overlap}} \cdot \overbrace{\left(\frac{\alpha}{|K_A|} + \frac{\beta}{|K_B|} \right)}^{\text{Generosity}} \cdot \overbrace{\left(1 - \frac{\tau}{|K_B|} \right)^{\gamma}}^{\text{Popularity}} \quad \text{with} \quad \begin{cases} \alpha + \beta = 1 \\ \alpha < \beta \\ \gamma \geq 0 \end{cases}$$

Conclusion

Tacking into account node generosity

Improved node load balancing

Tacking into account file popularity

Improved hit ratio for file searching request