Graduation project

Similarity between Tribler peers

The P2P software Tribler is founded on the concept of user similarity. Based on download statistics, Tribler discovers peers that are similar to the user of the client. Peers with a similar interest can aid the user in the discovery of new interesting content by enabling remote search or giving autonomous recommendations. Because searching an entire distributed network is unfeasible the discovery of peers with similar content interest is a vital system component.

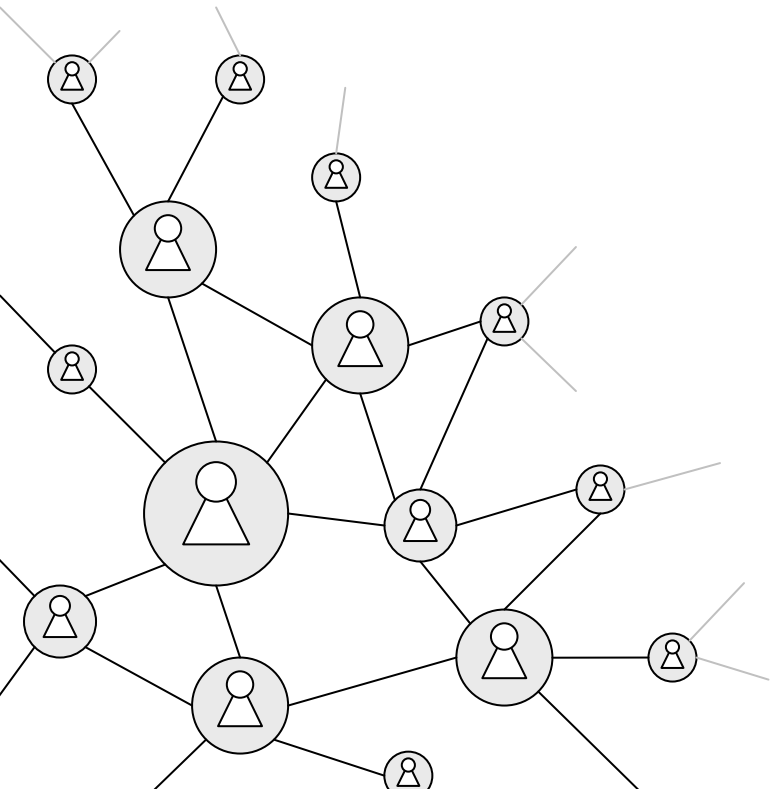
The current similarity function in Tribler is computationally intensive and performs badly because of data sparsity. If two users have no commonalities in their download history the current system is unable to estimate a reliable user similarity.

Research questions

- To derive the similarity between Tribler peers all the available data sources should be exploited. The function should combine download statistics and query logs into a single preference based distance measure between network peers. It should find a useful balance between computational efficiently and accurate user similarity.

- Exchanging all available preference information with the encountered peers is unfeasible. Therefore a strategy needs to be developed to determine which information is most representative of a user's taste. Effectively selecting the most informative data will reduce the necessary bandwidth to create the semantic overlay.

- The BitTorrent network contains many duplicate files. The same content can exist with different compression factors and metadata. Indentifying torrent replicas based on content analysis can reduce the content space and therefore facilitates better content retrieval.



Master variants

Media and Knowledge Engineering
Computer Engineering

Contact

Ir. Maarten Clements
m.clements@tudelft.nl

Dr. J. A. Pouwelse
j.a.pouwelse@tudelft.nl

Prof.dr.ir. Marcel Reinders
m.j.t.reinders@tudelft.nl