Technical Details

finding email correspondents in online social networks

Yu Chen
Two Assumptions or More…

Profile-based [1]: Match the email accounts and social network accounts according to the profile information. We can regard it as a string similarity problem.

Graph-based [1]: If two persons communicate well by email, they may have a good chance to be connected in a social network. We can regard it as a graph matching problem.

Temporal effects: If two persons communicate well by email, they may have a good chance to be connected in a social network in the near time.

Traffic-based: Assign different weights to network links based on network communication traffic.
Challenges

... 

Incorporating more information

Local Graph Match

Data Quality

Data Access
Some Existing Solutions

Profile-based [1]: Cosine similarity, Jaro-Winkler similarity…
   — PM algorithm

Graph-based [1]: Fuzzy Jaccard similarity…
   — GM algorithm

Hybrid algorithms [1,2]: Profile + Graph
   — MPPGM & PPGM
Experiments
Experiments (cont’d)
References
