How Much do Your Friends Know About You?
Reconstructing Private Information from the Friendship Graph

Norbert Blenn, Christian Doerr, Nasireddin Shadravan and Piet Van Mieghem

TU Delft - Network Architectures and Services (NAS)
Outline

“Birds of a feather flock together”

“Hiding properties in my profile protects my privacy.”

1. Dataset(s)
2. Representative Subset of the Population
3. Reconstructing Profiles
   1. Homophily
   2. Association Rule Learning
4. Conclusion
Hyves.nl

- Largest Online Social Network in the Netherlands
- ca. 11 million user accounts
  (Inhabitants of the Netherlands: 16.8 million)

Meertens Institute

- Research institute of the Royal Netherlands Academy of Arts and Sciences.

CBS (Statistics Netherlands)

- Collects and processes data in order to publish statistics
Privacy Settings in Hyves.nl

Public vs. private Profiles
- ca. 1/3 of all profiles publicly viewable
- Levels of privacy
  - Visible for:
    - Everyone (public),
    - Hyvers (Users registered at Hyves.nl),
    - Friends of Friends,
    - Friends,
    - no one

Scattered plot of the number of friends a user has, to the number of friends having a public (visible for everyone) profile.
Comparison of Online Social Network Data with General Statistics

- Percentage of citizens of Dutch towns towards the location of Hyves.nl users
  (only cities that are found in both datasets)

- Distribution of the age of Hyves.nl users

- Relationship status compared to statistical data of the Netherlands
Reconstructing Profiles

Homophily

- Inferring the Age of a User
  - Using the mode of friends age

- Inferring the Location of a User
  - Finding groups that inherit location information by calculating the probability of being in a group dependent on the probability to live in a specific city.
Reconstructing Profiles

Homophily

- Inferring different tastes in age groups
Reconstructing Profiles

Homophily
- How similar are users

- Similarity based on different topics
Association Rule Learning

Basket analysis

- Support
- Confidence

- Predictability of a user based on association rules:
  - The average confidence of applicable rules
Association Rule Learning

Predictability of a user:

If the user has 3 groups, A, B, C and the following rule set exists:

- A<-B confidence: 80%
- B<-A, C confidence: 80%
- B<-A confidence: 50%

Then 2/3 of the groups are predictable with an average confidence of 80%
I just woke up, what was he saying?

- Users are similar to each other
- Small number of users is enough to reconstruct a profile
  - But: Who are those few?
- Association rules – global rules vs. local applicability
  - Predict more than the user knows / specifies of himself

Attributes of a user are linked with themselves as well as through the friendship graph

Future Work:
- Triadic closure
- Prediction over multiple hops
- More comprehensive model of privacy
Thank you for your attention

Questions
Average Number of Groups vs. Age of a User

How Much do Your Friends Know About You?

Reconstructing Private Information from the Friendship Graph

Norbert Blenn, Christian Doerr, Nasireddin Shadravan and Piet Van Mieghem
Distribution of Surnames
Average distance to friends

![Graph showing the average distance to friends with distance to friend $d$ in km on the x-axis and probability a friend lives $d$ km far away from you on the y-axis.](image.png)
How Much do Your Friends Know About You?
Reconstructing Private Information from the Friendship Graph
Norbert Blenn, Christian Doerr, Nasireddin Shadravan and Piet Van Mieghem

Community detection
based on a weighted graph of private and public profiles

Weights are based on the profile setting:
High weight
  if both profiles are private / public
Low weight
  if profile settings differ.

Community detection based on a modularity maximizing method
(Louvain (la-Neuve) Method)
Age prediction full range

![Graph showing absolute difference to predicted age vs. predicted age for different age ranges.](image-url)