Project Proposal - Group 51

Jonas Blum, Sanjatul Islam, Abdulfattah Morad, Riccardo Rebecchi, Xin Tian

Research Question:

Can we identify distinct social circles within the Friendfeed social network and characterize them based on their connections, shared interests, and user behaviours?

Data:

To answer the research question, we will analyse the Friendfeed Version 2 dataset which consists of several data tables that are connected to each other:
- "users.csv": Contains user information
- "following.csv": User connectivity information based on their follower or following status
- "subscriptions.csv": Contains user associativity information with external service platforms
- "entries.csv": Contains information about user's posts
- "likes.csv": Contains information about users' reaction on posts
- "comments.csv": Contains information about comments of user on the different posts

Social Circle Detection: To find the distinct social circles we are planning to group users based on their connections and following activity. For this we will use the following tables: users.csv and following.csv.

Shared Interests Detection: Once we detect the communities, we will use the external service platform information to find the interests of the individual communities based on the "subscriptions.csv" table. Here, we expect the different communities to have different interests such as "professional" (e.g. Linkedin, Twitter, Blog etc), "social" (e.g. Facebook, Instagram etc) or "leisure" (e.g. Youtube, Amazon etc).

User Behaviour and Anomaly Detection: Furthermore, we want to analyse the user behaviours of the individual communities and find out if the communities exist mostly of bots, spammers or influencers. Additionally, we can identify if the users are rather introverted or extroverted. For this we will use the following tables: users.csv, following.csv, entires.csv, likes.csv and comments.csv.

Importance:

- By identifying the different social circles, we can recommend new friends to users.
- By analyzing the external service platforms used and patterns of external service usage, we can gain further insights into the interests of the different social circles.
- By identifying the number of extreme users of the different social services, we may be able to better remove bots and spammers from the platform or gain insights whether certain social circles are made up of mostly influential people.

Data Mining Techniques:

- To identify the different social circles, we will use clustering based on the network graph.
- For detecting connections between the different external service platforms, we will use association rule mining.
- For detecting bots, spammers and influencers we will use anomaly detection.