Project Proposal Group 51

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

# Dataset

For our project we will use the Friendfeed Version 2 dataset.

# Research Questions

**Combined Research Question (based on the research questions below):**  
Can we find different social circles based on their connections (using clustering)? Do the different social circles use different external services and are there different patterns of external service usage between the different social circles (using association rule mining)? Do the number of extreme users (bots, spammers, influencers) differ in the different social circles (anomaly detection).

**Importance**: By identifying the different social circles, we can recommend new friends to existing users based on the identified social circles. By analyzing the external services used and the different 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.

* Clustering:
  + Can we find circles of friends based on their conntections?  
    Importance: If new users register, we could recommend them posts they might enjoy based on their identified circle of friends.
  + What are the most common topics of the users’ posts? Can we determine similar groups of users based on the description of their posts?  
    Importance: We could recommend posts and other users to existing users based on the topics they both find interesting.
* Association Rule Mining:
  + Are there any patterns in the external services the users have connected to the Friendfeed platform? For example can the following connection of social media platforms be made: Twitter, Linkedin -> Facebook?  
    Importance: This helps us understand connections between different platforms and let’s us recommend users to link the most likely platform.
  + Are there any patterns of the topics of posts that the users like? For example can the following connection of topics of liked user posts be made: Technology -> Business?  
    Importance: This helps us understand the different topics that are being discussed on the social network and their connections.
* Anomaly Detection:
  + Can we identify extreme users (bots, spammers, influencers) by analyzing their behavior patterns such as number and frequency of posts, likes, comments.   
    Importance: This helps us remove unwanted users (bots and spammers) or recommend influential users to other users.
  + Can we detect posts that receive an unusual amount of likes and comments compared to the other posts of a user?  
    Importance: This lets us recommend certain posts with lots of likes and comments to other users.