

## Problem Identification

- Problem Statement
  - As a social network company like Facebook, I want to give an accurate network of potential future connections for users so that I can retain users on the platform since they are able to find and connect with people that they are more likely to form a connection with.
- Context
  - Link prediction can be useful in many contexts such as 'friend suggestions', and 'suggested content creators' in social network applications like Facebook, LinkedIn, etc.
- Criteria for success
  - If the model can with an above 90% accuracy predict past friends in the test data.
- Scope of solution space
  - I will use the facebook user data set with nodes (users) and edges (connections based on different features) to create a model that predicts whether pairs of users are a connection - to understand model accuracy - as well as predict on unconnected users and pick the k most likely potential future connections and display those users as friend suggestions.
- Constraints
  - A potential constraint is the potential lack of a large number of features (edges) connecting many users that might lead to a low probability when checking link prediction between users to give friend suggestions.
- Stakeholders
  - Facebook executives and investors.
- Data sources
  - Facebook user nodes with edges connecting users and their respective features:  
<https://snap.stanford.edu/data/ego-Facebook.html>