Social Network Friend Prediction

Group members: Evan Willenson, Eliot Paisley, Hugh Crockford

Goal: To create a social network friend recommendation system.

Data: http://odysseas.calit2.uci.edu/doku.php/public:online_social_networks#facebook_social_graph

The above link contains two datasets, both roughly sampling 1 million users.One taken from 28 Metropolis-Hastings random walks, the other from a uniform rejection random sample taken from the systems 32-bit ID space.This topic is broad and can be applied to any social network (Facebook, Twitter, Instagram).

Process: From the dataset we will randomly select 2 people. The response variable Y will be binary, 1 if the two users are friends 0 if they are not. We will form a series of covariates based on the overlap of friends for each user. For example, X_1 = # of mutual people between the friends of subject 1 and friends of subject 2 divided by the average number of friends between both subjects. X_2 = # of mutual friends between friends of subject 1 and friends of friends of subject 2 divided by the average. X_3 = # of mutual friends between friends of friends of subject 1 and friends of subject 2 divided by the average. This process can be expanded for K-nearest neighbors, where the nearest neighbors can be thought of as nearest dimension of friends.

We will accomplish the application of our recommendation system in three main phases-

- Look at dataset to make sure friend network is “rich” enough to create the prediction model. In other words, are there enough mutual friends, mutual friends of friends etc. in the dataset.
- Create the prediction model
- Apply the prediction model to two people who aren’t friends. If our model would predict that the two people are friends and they aren’t then we would recommend that they should be friends.

Possible Relevant Methods:

Supervised Learning: *Classifiers*: Support Vector Machine, Logistic Regression, Decision Trees/Random forests, ensemble methods (Bagging, Boosting, Stacking)

Training on known pairs of friends/non-friends with observations of relationships that exist between them as explanatory variables (direct friends, friends of friends, common groups etc).

Unsupervised Learning: *Clustering*: K-Nearest Neighbors, K-means, Recommendation Systems - Collaborative filtering, Neural Networks (Self organizing map), Silhouette.

Grouping people based on similarities (such as common friends and networks), then using these similarities to predict which group a test person may belong to.