

CMPE 256 Large Scale Analytics Individual Project

Name : Deepika Yannamani

Student Id : 012443886

Course Instructor : Shih Yu Chang

Facebook Friend/Page recommender system

Facebook is a social networking website with an open platform that enables developers to extract and utilize user information and relationships. My goal is to develop a recommendation system which provides facebook friend/page suggestions for a given user using collaborative filtering method. To make these recommendations, I have gathered user profile information such friends, groups, interests and activities from facebook dataset.

Objective :

In this project, I am implementing a collaborative filtering recommendation system for suggesting friends on Facebook. I implemented two mechanisms for recommending a new friend in a social network. For user X, listed some non-friends in order, starting with the best friend recommendation and ending with the worst. A non-friend is a user who is not X and is not a friend of X. Depending on the recommendation algorithm, the list may include all non-friends or some of them. Further, the recommendations might not be symmetric. I implemented a code in python that produces friend recommendations for U, in order from best to worst. This is done by assigning each potential friend a number called a score, where higher scores indicate a better match. Then the list is returned according to the score. Given user X, if two people Y and Z would be equally good as new friends for X (they have the same score), then they are listed in alphabetical order (for names) or numerical order (for numerical user IDs).

Data Collection:

Dataset is taken from SNAP Source. This dataset consists of 'circles' from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks. Facebook data has been anonymized by replacing the Facebook-internal ids for each user with a new value.

Analysis :

Methodology:

The methodology of this project consists of two core methods of recommendation. The first one is the `recommend_by_common_friends` and `recommend_by_influence`.

The following algorithm is performed 100 times on the dataset and average is calculated of all average values returned to evaluate the results

1. Two nodes are chosen at random.
2. Their friendship is removed from the graph.
3. Friend recommendations for F1 and F2 are computed.
4. Rank of F1 in F2's list of recommended friends is calculated. Rank of F2 in F1's list of recommended friends is calculated. Average of both rank is computed.
5. Friendship is put back to the graph.

Partial results :

I am able to calculate by first method(recommend_by_common_friends) and find its rank. I am yet to work on method 2(recommend_by_influence) and I am planning to compare results.

Recommendations by Method 1(common friends):

[(1376, 97), (1833, 91), (1746, 88), (993, 86), (1390, 86), (1391, 83), (1714, 83), (1059, 81), (1516, 81), (1612, 81)]

Average Rank of Method 1: 4.0