

Programming Project 1

Due date: Wed, Oct. 16, 2013 at 4pm.

Description: For the first project, you are requested to decide if a pair of nodes in a social graph is connected with an edge or not. We will be using an online platform (Kaggle, <http://inclass.kaggle.com>) to host this competition. The tasks that you need to solve are the following:

Task 1: Identification of co-authorship links:

We provide you with a snapshot from DBLP. This network contains information about researchers and their work. The nodes of the network are people and the edges represent co-authorship. This means that if two researchers have collaborated on at least one paper, then there exists an edge that connects them. As supporting material, we give you, for each researcher, the most important terms from the titles of their papers and their respective frequencies.

Task 2: Identification of social links:

You will be working on a sample from Flickr's network. Each node in the network is a Flickr user and an edge between two nodes indicates that these users are friends. To help you with this task, you will have access to the groups in which each user is a member.

Evaluation: The evaluation method that will be used is the AUC (area under curve). This is a measure that handles unbalanced data (e.g, in case there exist more true negatives than true positives).

Data: For details on the data files check each individual competition, under the page "Data".

Writeup: In addition to submitting your solution online, you need to provide us with a 2-page writeup that describes the algorithm you have implemented and the special tricks you used in order to make it work. Also, describe your strategy for selecting that particular algorithm and how you did your offline evaluation of the method.

Instructions: If you don't have an account for Kaggle, you will need to signup. Go to <http://inclass.kaggle.com> and use your BU email. After that, search for *BU CS 565 : Project 1* and you will find 2 active competitions. **Take part and submit solutions for both of these competitions.** You may upload

your predicted solution as a *.csv* file at “Make a submission”. We suggest that you don’t just upload your final solution, but instead submit all your intermediate ones so that you can check your standing. You can see how well your solution is ranking if you go to the “Leaderboard”. Notice that the deadline on the competition page may not be accurate.

If you encounter any problems, send an email at *cmav@bu.edu*

Privacy and confidentiality: Please write the usernames you used to sign up with on your project report.