

Homework 2

*Handed Out: Sep 27, 2018**Due: Oct 10, 2018 11:59 pm*

1 General Instructions

- This assignment is due at 11:59 PM on the due date. Contact TAs if you face technical difficulties in submitting the assignment. We will NOT accept any late submission, no exceptions.
- There is no written part in this assignment, your entire grade will be based on your performance in the coding contest. A part of the grade is based on correctness (test cases passed) and the other on efficiency.
- The programming part of this assignment will be hosted on hackerrank (<https://www.hackerrank.com/>) as a programming contest. To participate in this contest, please create a hackerrank account with your illinois.edu email id. The contest framework will allow you to verify the correctness of your submission based on a set of sample test cases. We will use additional test cases to grade your submission (hidden test cases will be significantly larger than the samples). Please check the course piazza page for the contest URL. You need to register in the contest to solve the programming question. Remember to use the hackerrank account associated with your Illinois email id.
- It is OK to discuss the problems with the TAs and your classmates, however, it is NOT OK to work together or share code. Plagiarism is an academic violation to copy, to include text from other sources, including online sources, without proper citation. To get a better idea of what constitutes plagiarism, consult the CS Honor code (<http://cs.illinois.edu/academics/honor-code>) on academic integrity violations, including examples, and recommended penalties. There is a zero tolerance policy on academic integrity violations. Any student found to be violating this code will be subject to disciplinary action.
- Please use Piazza if you have questions about the homework. Also feel free to send TAs emails and come to office hours.
- The questions are based on the following reference paper¹.

¹<https://arxiv.org/pdf/0712.0499.pdf>

2 Programming Question (20 points)

The reference paper first describes conventional SimRank, a method for computing object similarities in domains with object-object links. Specifically we will be considering Bipartite Simrank with 2 types of objects, users and textual advertisements. Unlike the HITS algorithm, SimRank computes the most similar entities to a given entity (where the entities could belong to either type). This enables advertisers to push the most relevant content to a given user, depending on his/her history and the global click graph obtained from all other users (which is the bipartite graph we are working with).

- **Implement SimRank** : Your first task is to implement conventional SimRank (Eqns 4.1 and 4.2 in the reference paper). You are required to use **Partial Sum Sharing** in your implementation (refer to this presentation² and the SimRank wiki³).
- **SimRank with the evidence metric** : Use two different evidence metrics (Eqn 7.3, 7.4) to alter the similarity scores obtained with SimRank.
- **Weighted SimRank** : The previous 2 parts only required using the link structure between the two types of objects. Now you will use the link weights as well, and compute Weighted Simrank (which works well when the edges have weights assigned to them quantifying how similar a pair of objects are, rather than just a binary label). Refer to Section 8 in the reference paper for this part.

²<http://www.doc.ic.ac.uk/~wyu1/ppt/oral/icde13.pdf>

³<https://en.wikipedia.org/wiki/SimRank>