COMP9313 2017s1 Assignment

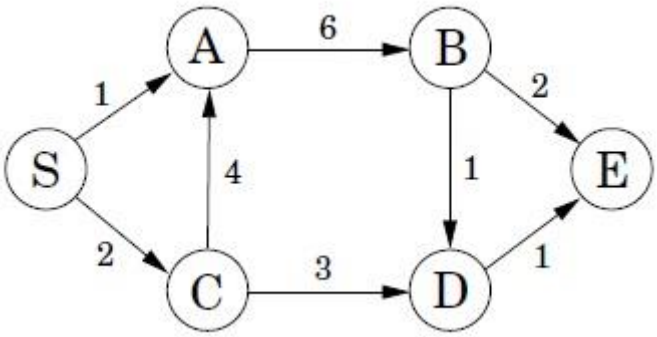
Question 1. MapReduce (4 pts)

The following code of for computing the relative frequency is problematic. Describe how you can fix it.

```
class Mapper
        method Map(docid a, doc d)
                for all term w in doc d do
                        for all term u in Neighbors(w) do
                                Emit(pair (w, u), count 1)
                                Emit(pair (w, *), count 1)
class Reducer
        curMarginal <- 0
        method Reduce(pair p, counts [c1, c2, …])
                s <- 0
                for all count c in counts [c1, c2, …] do
                        s <- s + c
                if(p.contains(*))
                        Emit(p, s/curMarginal)
                Else
                        curMarginal <- s
```

Question 2. Graph Algorithms (6 pts)

Given the following graph, assume that you are using the single shortest path algorithm to compute the shortest path from node S to node E. Show the output of the mapper (sorted results of all mappers) and the reducer (only one reducer used) in each iteration (including both the distances and the paths).



Question 3. Streaming Data Processing (5 pts)

Suppose we are maintaining a count of 1s using the DGIM method. We represent a bucket by (i, t), where i is the number of 1s in the bucket and t is the bucket timestamp (time of the most recent 1).

Consider that the current time is 200, window size is 60, and the current list of buckets is: (16, 148) (8, 162) (8, 177) (4, 183) (2, 192) (1, 197) (1, 200). At the next ten clocks, 201 through 210, the stream has 0101010101. What will the sequence of buckets be at the end of these ten inputs?

Question 4. Recommender Systems (5 pts)

Consider three users u_1, u_2, and u_3, and four movies m_1, m_2, m_3, and m_4. The users rated the movies using a 4-point scale: -1: bad, 1: fair, 2: good, and 3: great. A rating of 0 means that the user did not rate the movie.

The three users' ratings for the four movies are: u_1 = (3, 0, 0, -1), u_2 = (2, -1, 0, 3), u_3 = (3, 0, 3, 1)

(i) (3 pts) Which user has more similar taste to u_1 based on cosine similarity, u_2 or u_3? Show detailed calculation process.

(ii) (2 pts) User u_1 has not yet watched movies m_2 and m_3. Which movie(s) are you going to recommend to user u_1, based on the user-based collaborative filtering approach? Justify your answer.

Submission:

Deadline: Sunday 4th June 09:59:59 PM

Please provide your solutions to these questions in a pdf file named as "answers.pdf". Log in any CSE server (williams or wagner), and use the give command below to submit your solutions:

$ give cs9313 assignment5 answers.pdf

Or you can submit through:
https://cgi.cse.unsw.edu.au/~give/Student/give.php

If you submit your assignment more than once, the last submission will replace the previous one. To prove successful submission, please take a

screenshot as assignment submission instructions show and keep it by yourself.

Late submission penalty

You will receive zero marks for this assignment.

Plagiarism:

The work you submit must be your own work. Submission of work partially or completely derived from any other person or jointly written with any other person is not permitted. The penalties for such an offence may include negative marks, automatic failure of the course and possibly other academic discipline. Assignment submissions will be examined manually.

Relevant scholarship authorities will be informed if students holding scholarships are involved in an incident of plagiarism or other misconduct.

Do not provide or show your assignment work to any other person - apart from the teaching staff of this subject. If you knowingly provide or show your assignment work to another person for any reason, and work derived from it is submitted you may be penalized, even if the work was submitted without your knowledge or consent.