

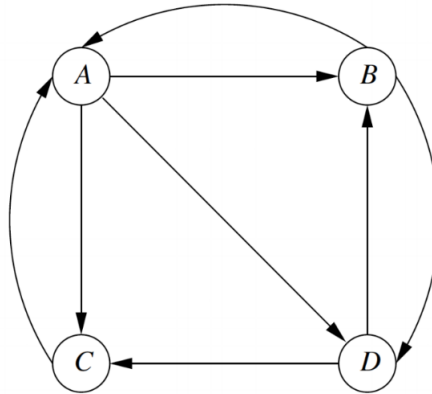
Assignment 3

COSI 120A Topics in Computer Systems 2020 Fall

Due: 12:00 am, 10/23/2020

1 [20 marks]

Given the following graph.



- a) Write down the transition matrix for the graph [10 marks]
- b) Compute the PageRank of each page (describe the process and the results) [10 marks]

2 [30 marks]

Write a hadoop program to complete the PageRank calculation on Berkeley-Stanford web graph. Set the damping factor to 0.85 and run the algorithm for a total of 10 iterations. Report the NodeId of Top 100 pages and the corresponding PageRank scores.

Here is the [link](#) for Berkeley-Stanford web graph. The dataset is stored in a TXT file, where each line denotes an edge. The following screenshot shows the first several lines of this file.

```
# Directed graph (each unordered pair of nodes is saved once): web-BerkStan.txt
# Berkely-Stanford web graph from 2002
# Nodes: 685230 Edges: 7600595
# FromNodeId ToNodeId
1 2
1 5
1 7
1 8
1 9
1 11
1 17
1 254913
1 438238
254913 255378
254913 255379
254913 255383
254913 255384
254913 255392
```