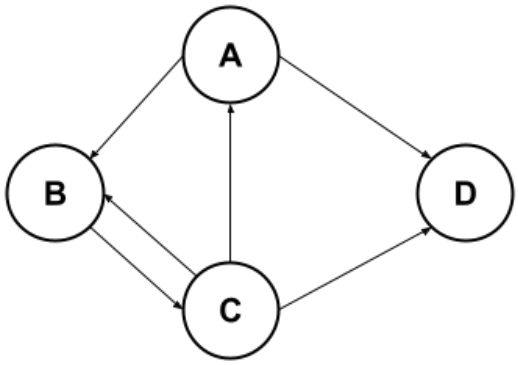
Question 4 - GraphProcessing [20 Points]



a) Given the above graph, write down the adjacency matrix used to compute the PageRank of the graph. (Use the naive formulation without using teleportation) [4 points]

[2 points] Has the right entries in the matrix filled in
[2 points] Entries have the correct weights
Subtract one point for minor errors

b) State the initial condition $\boldsymbol{r}_0$ for power iteration. Perform 3 iterations of power iteration to find $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, and $\boldsymbol{r}_3$. [4 points]

[1 point] Has the right initial condition
[1 point] for each iteration that is correct (still give credit if answer is correct given an incorrect adjacent matrix)
Subtract one point if most entries are correct but there are some minor mistakes

c) Will the power iteration solution for the above graph converge to what we want? Why or why not? If not, explain how to implement a fix. [6 points]

[2 points] No. (take a away a point if they claim it will not converge - it does converge)
[2 points] Recognize there is a dead end, D. (there is no spider trap)
[2 points] Fix is teleportation.

d) Describe at least 2 differences in the MapReduce implementation of PageRank. [6 points]

[3 points] Adjacency list instead of adjacency matrix (or graph must be split across many nodes/can't use matrix/represented in KV pairs, etc)
[3 points] Need a separate step to redistribute rank from spider traps and dead end nodes
[2 points] The computation is split over mappers and reducers to run in parallel.