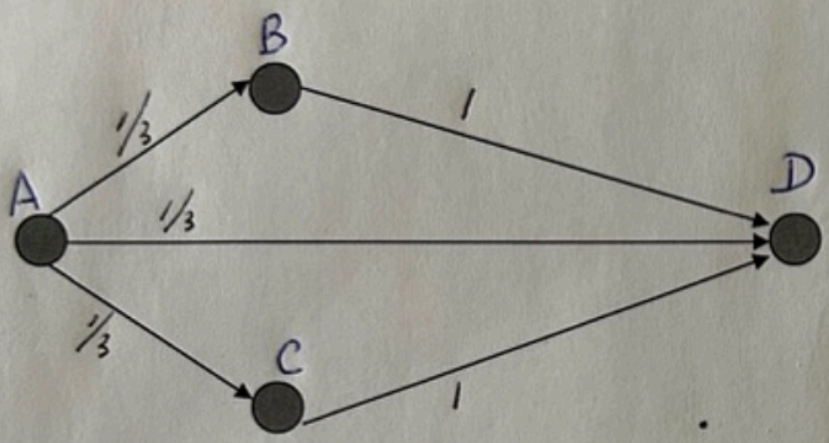
CIS4517/5517 Quiz 4 Name:

1. (12 points) Given the following graph where each node represents a webpage, and each line represents a hyperlink. Assume the initial PageRank of each webpage is 1, calculate the PageRank of each page after 2 iterations using the simple PageRank algorithm (C=1). Show your work.



Initial

PR(A) = 1

PR(B) = 1

PR(C) = 1

PR(D) = 1

1st round:

PR(A) = 0

PR(B) = $\frac{1}{3}$ PR(A) = $\frac{1}{3}$

PR(C) = $\frac{1}{3}$ PR(A) = $\frac{1}{3}$

PR(D) = $\frac{1}{3}$ PR(A) + PR(B) + PR(C) = $\frac{1}{3}$ + 1 + 1 = $2\frac{1}{3}$

2nd round:

PR(A) = 0

PR(B) = $\frac{1}{3}$ PR(A) = 0

PR(C) = $\frac{1}{3}$ PR(A) = 0

PR(D) = $\frac{1}{3}$ PR(A) + PR(B) + PR(C) = $\frac{1}{3}$ + $\frac{1}{3}$ = $\frac{2}{3}$

2. (12 points) For the above graph in Problem 1, assume the initial PageRank of each webpage is 1, calculate the PageRank of each page after 2 iterations using the modified PageRank algorithm (d=0.8). Show your work.

1st round:

$PR(A) = 0.2 + 0.8 \times 0 = 0.2$

$PR(B) = 0.2 + 0.8 \times \frac{1}{3} \cdot PR(A) = 0.2 + \frac{0.8}{3} = 0.467$

$PR(C) = 0.2 + 0.8 \times \frac{1}{3} PR(A) = 0.467$

$PR(D) = 0.2 + 0.8 \times (\frac{1}{3} PR(A) + PR(B) + PR(C))$

$= 0.2 + 0.8 \times (\frac{1}{3} + 1 + 1) = 2.067$

2nd round:

$PR(A) = 0.2 + 0.8 \times 0 = 0.2$

$PR(B) = 0.2 + 0.8 \times \frac{1}{3} \cdot PR(A) = 0.2 + \frac{0.8 \times 0.2}{3} = 0.253$

$PR(C) = 0.2 + 0.8 \times \frac{1}{3} PR(A) = 0.2 + 0.8 \times \frac{1}{3} \times \frac{0.2}{3} = 0.253$

$PR(D) = 0.2 + 0.8 \times (\frac{1}{3} \times 0.2 + 0.467 + 0.467)$

$= 1.0$

3. (3 points) The features of big data are typically characterized as 5 Vs, which one (or more) of the following don't belong to the 5 Vs?

a. Value
b. Visibility
c. Velocity
d. Variety
e. Volume
f. Veracity
g. Virtualization

B, G

4. (3 points) True or false: Judge whether the following statement is true or false, provide a brief explanation if it's false.

In big data, audio and image files are structured data because they have specified formats.

False, they are unstructured data.