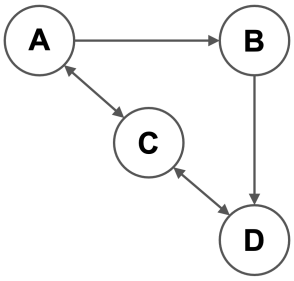
1. Assume documents d_1 and d_2 are two documents where d_1 is the string S and d_2 is the string S+" "+S. Given a query Q, which document will be ranked higher by a cosine-distance-based TF-IDF retrieval model?

a. d_1
b. d_2
c. They will have the same rank (CORRECT)
d. It depends on the query Q

Explanation: the two document will have exactly the same direction, therefore given any vector representation for Q, they will always have the same rank (as it's a cosine-distance-based retrieval)

2. Run the PageRank algorithm on the following graph. What would be the node with the highest PageRank after the 2nd iteration?
Note: In Iteration 0 we assign uniform weights to all nodes. You need to run for the next two iterations: Iteration 1 and Iteration 2.



a. A
b. B
c. C (CORRECT)
d. D

Explanation:

	Iteration 0	Iteration 1	Iteration 2	Ranking
A	¼	1/12	1.5/12	4
B	¼	2.5/12	2/12	3
C	**¼**	**4.5/12**	**4.5/12**	**1**
D	¼	4/12	4/12	2

3. You have a chain of pages where each page links to the next. Additionally, every page in the chain links back to the first page. How will the PageRank probability of the first page behave, using basic PageRank without random jumps, as the chain growths?

a. It will converge to 0
b. It will converge to ½ (CORRECT)
c. It will converge to 1
d. It will converge to infinity

4. What is TRUE regarding Item-based Collaborative Filtering?

a. It does leverage item description
b. It can recommend niche or new items (CORRECT)
c. It recommends items by finding similar users
d. None of the above

5. Using Matrix Factorization we have ended up with two matrices representing the user preferences (for 4 users: A, B, C, D) and item preferences (For 5 items: 1, 2, 3, 4, 5) as shown below.

$$U = \begin{bmatrix} 3 & 0 \\ 2 & 2 \\ 4 & 4 \\ 0 & 4 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 3 \\ 0 & 4 \\ 4 & 3 \\ 1 & 2 \\ 4 & 0 \end{bmatrix}$$

Which top 2 users, the item 3 should be recommended to?

a. A, B
b. B, C (CORRECT)
c. C, D
d. A, D

Explanation:
The UxI matrix will be the following:

	I1	I2	**I3**	I4	I5
UA	3	0	12	3	12
UB	8	8	**14**	6	8
UC	16	16	**28**	12	16
UD	12	16	12	8	0

And the normalized version of it:

	I1	I2	**I3**	I4	I5
UA	0	0	1.5	0	1.5
UB	1	1	**2**	1	1
UC	2	2	**3.5**	3	2
UD	1.5	2	1.5	1	0

6. You have the following sentence: “The **dollar** index dropped today around 0.5% in New York Stock Exchange” and you want to do Entity Linking for the word “**dollar**”.
You retrieve in the KG different nodes that can be related to the mention of “dollar” and end up in an entity graph with the following properties:

KG Node	**Out-degree**	**In-degree**
United States dollar	2	4
Canadian dollar	7	0
Australian dollar	0	6

Using Personalized Pagerank ranking, starting in the entity graph with a node related to the mention “New York Stock Exchange”, which one of the following is always TRUE:

a. P(“Canadian dollar”) <= P(“Australian dollar”) (CORRECT)
b. P(“United States dollar”) < P(“Australian dollar”)
c. P(“United States dollar”) <= P(“Canadian dollar”)
d. None of the above

Explanation: Since the in-degree of the Canadian dollar is zero, in PPR, we never reach this node, for the prob will be zero.