SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
Continuous Assessment Test - II, September 208
Fall Semester 2018 -19

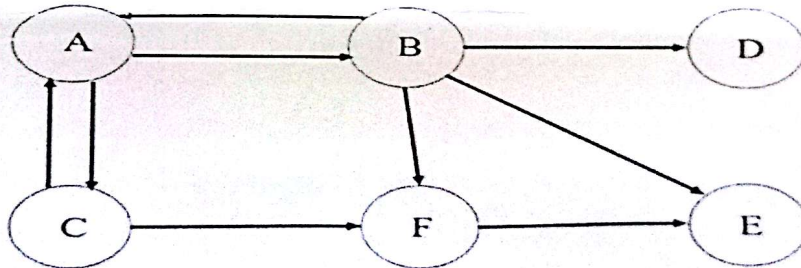|			
Course Code	: CSE 3024	**Duration**	: 90 Minutes
Course Name	: Web Mining	**Max Marks**	: 50
Slot	: F2 + TF2		

Answer ALL the questions

1. Consider the following term-frequency matrix. Rank the document for the given query "Sea Voyage" using Latent Semantic Index. **[15]**

DOC#	Ship	Sea	Voyage
D1	0	2	5
D2	3	1	4

2. Find the page rank of the given below web page linkage graph using power-iteration method. Consider damping factor (d) = 0.8 and number of iteration = 2 **[10]**

3. Considering the following document set. Classify the documents D4, D5 and D6 into class is A or B or C using Naïve Bayes classifier. (use multinomial distribution) **[15]**

Doc #	Content	Class
D1	ant ant bee	A
D2	dog bee dog ant	B
D3	cat dog ant dog	C
D4	dog bee ant bee	?? (B)
D5	dog ant cat ant	?? (A)
D6	ant dog bee	?? (B)

1

Prepare the confusion matrix and evaluate the performance of the model using F-measure.

4. Suppose we have the following set of five documents [10]

D1: Romeo Juliet sad.

D2: happy dagger happy

D3: Romeo died dagger.

D4: Juliet free died free.

D5: Romeo Juliet happy dagger.

Apply the hierarchical clustering on the above document set considering term-frequency matrix and draw the dendrogram using complete linkage method.

****