Register Number:……..

KUMARAGURU college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2022
(Regulation 2018)
V Semester
ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
U18AIE5005**:** Mining Big Data

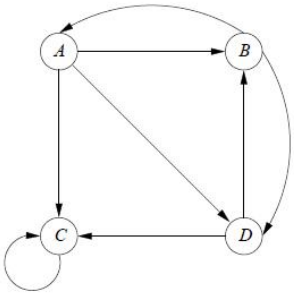
COURSE OUTCOMES

CO1: Choose b tools to carry out exploratory data analysis and produce effective visualization of given data.
CO2: Perform parallel data processing and duplication with Hadoop and Map-Reduce
CO3: Identify suitable data model and algorithms for mining mass data set.
CO4: Apply link analysis & mining social network graphs in real time problem

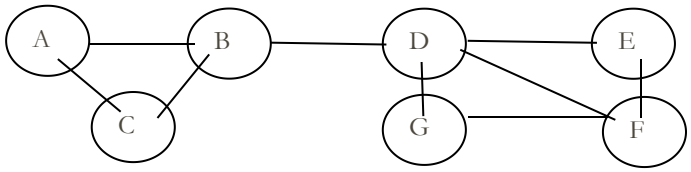
Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1.	In order to illustrate the scenario of clustering, Plot the cholera cases on a map of an Indian City	CO1	[K_3]
2.	Perform a univariate analysis on the employee dataset to mark the difference in the salary component of the employees in a graph. Write a python code to demonstrate the scenario	CO1	[K_3]
3.	Analyze the amount of data involved in tweets and logs in social media on a day-to-day basis	CO2	[K_4]
4.	Apply MapReduce computation to count the number of occurrences for each word in a collection of documents.	CO2	[K_3]
5.	Write a SQL query to report the number of unique users for a web site over the past month	CO3	[K_3]
6.	State the purpose of a bloom filter.	CO3	[K_1]
7.	Compute the page rank fare for the graph given below. Assume $\beta = 0.8$	CO4	[K_3]

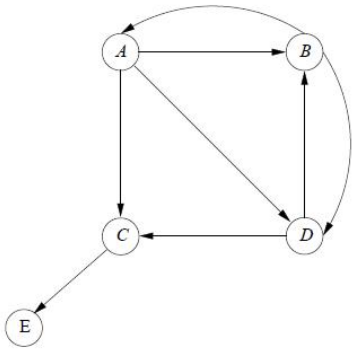
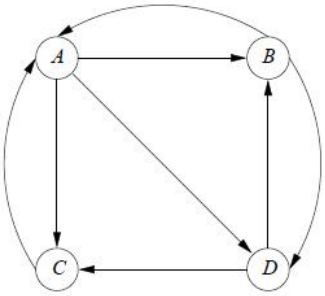


8.	Write about the alternative approaches for eliminating the link spam.	CO4	[K_2]
9.	Draw a tripartite graph with 3 different set of nodes like Users, Tags and Web Pages.	CO4	[K_3]
10.	Draw the BFS for the graph given below starting at node ‘E’	CO4	[K_3]



Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11.	a)	Explain Bonferroni’s Principle with a suitable example	8	CO1	[K_2]
	b)	Explain the impact of the 6 V’s of Big Data trait.	8	CO1	[K_2]
12.	a)	Assume that a social-networking site has a relation of kind Friends(User, Friend). Apply MapReduce methodologies to determine the number of friends each user has.	8	CO2	[K_3]
	b)	Create an empty matplotlib figure with three subplots and write suitable commands to construct a random graph.	8	CO2	[K_4]

13.	a)	Explain in detail about the different types of sampling and data classification with suitable examples.	16	CO3	[K_2]
14.	a)	Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2, 6, 5. Our hash functions will all be of the form h(x) = ax+b mod 32 for some a and b. You should treat the result as a 5-bit binary integer. Determine the tail length for each stream element and the resulting estimate of the number of distinct elements if the hash function is: (a) h(x) = 2x + 1 mod 32. (b) h(x) = 3x + 7 mod 32. (c) h(x) = 4x mod 32.	8	CO3	[K_3]
	b)	Assume we have a stream with a, b, c, b, d, a, c, d, a, b, d, c, a, a, b. The length of the stream is n = 15. Estimate the second moment of the stream using a limited amount of space. Apply Alon-Matias-Szegedy Algorithm for Second Moments	8	CO3	[K_3]
15.	a)	A graph with 2 level of dead ends is given below  Identify and remove the dead ends and compute the PageRank for each node.	16	CO4	[K_3]
16.	a)	Consider the Web graph given below  Compute the transition matrix for the graph and determine a compact representation of this transition matrix .	8	CO4	[K_3]
	b)	Explain about Topic-Sensitive PageRank	8	CO4	[K_2]

****