library (B)

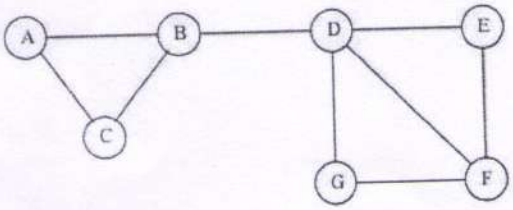
K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Subject Code: 1PCEC202 Subject Name: Big Data Analytics Date: 07/07/2022

April – May 2022
M.Tech(AI) Program: Computer Engineering
Examination: FY Semester: II
Course Code: **1PCEC202** and Course Name: **Big Data Analytics**

Duration: 03 Hours Max. Marks: 60

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3) Assume suitable data if required, and state it clearly.

		Max. Marks	CO	BT Level
Q-1	Solve any **Six** questions out of **Eight.**	**12**		
i)	State 3 Vs of Big Data.	**2**	CO3	Remember
ii)	Explain in short Decision Support System (DSS).	**2**	CO2	Understand
iii)	List different business drivers for Big Data.	**2**	CO1	Remember
iv)	State CPM algorithm for community detection	**2**	CO5	Remember
v)	List any 4 components of hadoop ecosystem.	**2**	CO4	Remember
vi)	Give importance of recommendation systems.	**2**	CO3	Understand
vii)	State steps of FM algorithm.	**2**	CO6	Remember
viii)	Recall Hoeffding trees.	**2**	CO6	Remember
Q-2	Solve any **Four** questions out of **Six.**	**16**		
i)	Describe big data opportunities and challenges.	**4**	CO1	Understand
ii)	Apply different measures of similarity/ distance to find a) Jaccard distance between {a, b, e, d} & {b, d, f, g} b) Cosines of the angles between (−1, 2, 3) and (3, 1, -2). c) Hamming Distance between 10011010 & 10101010	**1** **2** **1**	CO3	Apply
iii)	Explain the significance of Privacy and Ethics in Application of Data Science.	**4**	CO5	Understand
iv)	Find communities in the following graph using GN algorithm 	**4**	CO2	Apply
v)	Explain Bloom's filter for stream data mining.	**4**	CO6	Understand
vi)	Explain Hive and HiveQL with example.	**4**	CO4	Understand

Q-3	Solve any **Two** questions out of **Three.**	**16**		
i)	Explain HDFS architecture with diagram, features and limitations in detail.	**8**	CO3	Understand
ii)	Apply MapReduce technique for two step matrix multiplication for the following two matrices. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ $B = \begin{bmatrix} 6 & 3 \\ 5 & 2 \\ 4 & 1 \end{bmatrix}$	**8**	CO4	Apply
iii)	Discuss Techniques and problems with anonymization Techniques with example.	**8**	CO5	Understand
Q-4	Solve any **Two** questions out of **Three.**	**16**		
i)	a) Describe Business Pressures–Responses–Support Model in detail. b) Explain advantages and limitations of distributed file system	**4** **4**	CO2 CO1	Understand
ii)	What is Data Privacy? Explain different threats and privacy control techniques.	**8**	CO5	Understand
iii)	Investigate number of ones in a given data stream using a sliding window based on DGIM algorithm. Also Estimate the number of 1's in last k bits where k =8 and 18. Comment on in each case, how far off the correct value is your estimate? 1 0 1 1 0 0 1 0 1 1 0 1 0 1 0 1 0 0 1 0 1 0 1 0 1 1 0 1 1	**8**	CO6	Apply

****