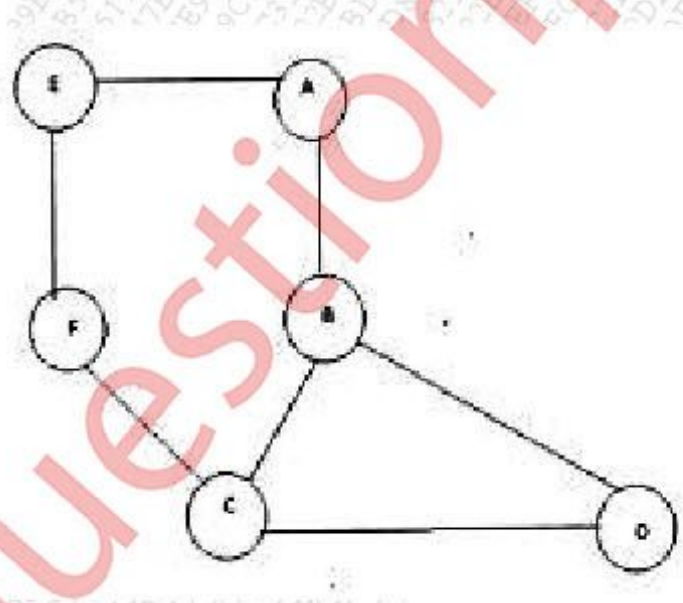
QP Code : 725401

(3 Hours) **[Total Marks :80**

N.B. : (1) Question no. **1** is **compulsory**
(2) Attempt any three from the remaining.
(3) Assume suitable data.

1. (a) What is Big Data? What is Hadoop? How Big Data and Hadoop are linked? **5**
   (b) Explain Page Rank with Example. Can a Website's Page rank Ever Increase? What are its chances of Decreasing? **5**
   (c) Explain Hubs and Authorities with neat diagram. **5**
   (d) With respect to data stream querying, give example of **5**
      (a) One Time queries
      (b) Continuous Queries
      (c) Pre-defined queries
      (d) Ad-hoc queries

2. (a) Explain Hadoop Ecosystem with core components, Explain its Physical architecture. State Limitations of Hadoop. **10**



   (b) What is MapReduce ? Explain How Map and Reduce Work? What is Shuffling in MapReduce? **10**

3 (a) For the Graph given below use betweenness factor and find all communities. **10**
   (b) How would you get the features of the document in a content -based system? Explain document similarity. **5**
   (c) What is triangular matrix? How it is used for main memory counting? **5**

[TURN OVER]

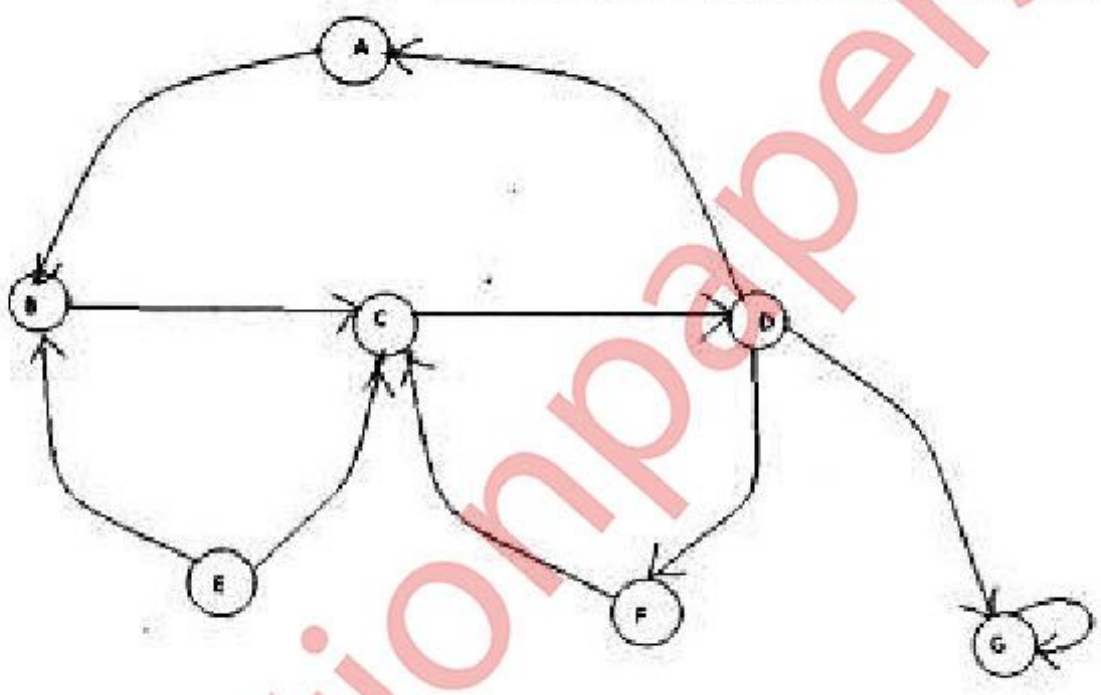
4 (a) Explain Collaborative Filtering based recommendation System. How it is different from content based recommendation systems ? **10**

(b) What are Combiners? When Should one use combiner in mapreduce job? **5**

(c) How to count distinct elements in a stream? Explain Flajolet-Martin Algorithm. **5**

5 (a) Given a l Dim Dataset {1,5,8,10,2} Use the agglomerative clustering algorithm with Euclidean distance to establish hierarchical grouping relationship. Draw the dendrogram . **10**

(b) Consider a Portion of Web Graph Shown below: **10**



(a) Compute the hub and authority scores for all the nodes.

(b) Does this graph contains spider traps? Dead ends? If so, which nodes?

(c) Compute the page Rank of the nodes with teleportation β =0.8? (Show two iterations only)

6. (a) What is NoSQL? What are the business drivers for NoSQL? Discuss any two architectural patterns ofNoSQL. **10**

(b) What is a Data Stream Management System? Explain with Block Diagram **10**
