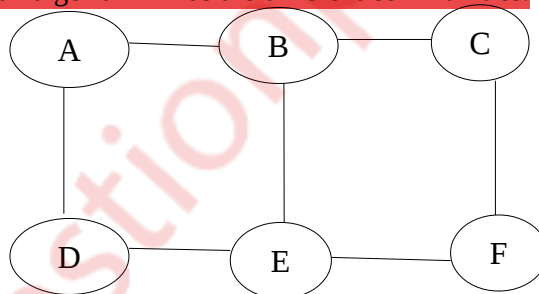


(3 Hours)

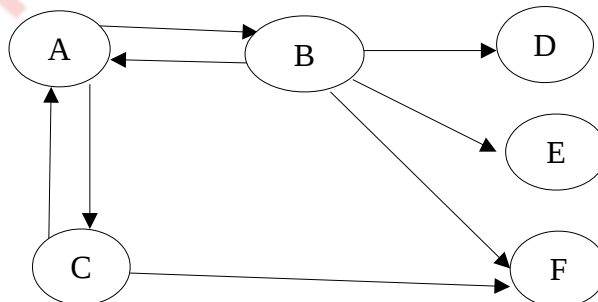
[Total Marks 80]

- i. Q.1 is compulsory
- ii. Attempt any three from the remaining
- iii. Assume suitable data

- Q.1 (a) Explain Edit distance measure with an example. (5)
- (b) When it comes to big data how NoSQL scores over RDBMS. (5)
- (c) Give difference between Traditional data management and analytics approach Versus Big data Approach (5)
- (d) Give Applications of Social Network Mining (5)
- Q.2 (a) What is Hadoop? Describe HDFS architecture with diagram. (10)
- (b) Explain with block diagram architecture of Data stream Management System. (10)
- Q.3 (a) What is the use of Recommender System. How is classification algorithm used in recommendation system. (10)
- (b) Explain the following terms with diagram (10)
- 1) Hubs and Authorities
 - 2) Structure of the Web
- Q.4 (a) What do you mean by Counting Distinct Elements in a stream. Illustrate with an example working of an Flajolet – Martin Algorithm used to count number of distinct elements. (10)
- (b) Explain different ways by which big data problems are handled by NoSQL. (10)
- Q.5 (a) Describe Girvan – Newman Algorithm. For the following graph show how the Girvan Newman algorithm finds the different communities. (10)



- (b) What is the role of JobTracker and TaskTracker in MapReduce. Illustrate Map Reduce execution pipeline with Word count example. (10)
- Q.6 (a) Compute the page rank of each page after running the PageRank algorithm for two iterations with teleportation factor Beta (β) value = 0.8 (10)



- (b) What are the challenges in clustering of Data streams. Explain stream clustering algorithm in detail. (10)