JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

M.Tech-II Semester (Data Science)

COURSE CODE(CREDITS): 22M111CI213   MAX. MARKS: 35

COURSE NAME: BIG DATA ANALYTICS

COURSE INSTRUCTORS: Er. Nitika   MAX. TIME: 2 Hours

Note: (a)All questions are compulsory.

(b)Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q1. Consider the given social network graph and find the communities within the graph using Girvan Newman algorithm. [6]
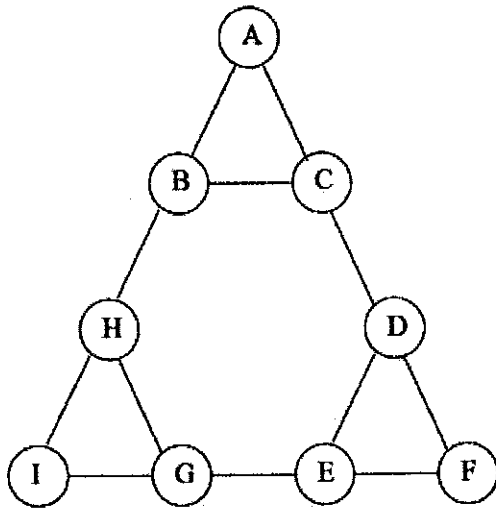
Q2. What is hierarchical Clustering? How do you represent a cluster with more than one point? How will you choose which two clusters to merge? When will stop combining clusters? [5]

Q3. Solve the given numerical using bloom filtering. Insert elements 10 and 7 in the bloom filter of size 5.

Consider these two hash functions:

$H1(x) = x \bmod 5$

H2(x) = (2x+6) mod 5

Find the presence of element 14 and 15. [5]

Q4. Describe SON algorithm with its principle and how SON algorithm helps in finding frequent itemsets in large datasets. [5]

Q4. Give the explanation of overlapping communities in social networks. Why is it important to identify overlapping communities, and what are some methods used to detect them? Provide an example to illustrate your points. [4]

Q5. Explain the difference between collaborative filtering and content-based recommendations? [4]

Q6. Assume you have a content-based recommendation system that suggests movies to users based on their genre preferences. Consider a user who has rated the following movies:

Movie A: Comedy (rating: 4)

Movie B: Action (rating: 3)

Movie C: Comedy (rating: 5)

Additionally, the system has the following movies available for recommendation:

Movie X: Comedy

Movie Y: Action

Movie Z: Drama

Using content-based recommendation, calculate the recommendation score for each of the movies X, Y, and Z for the given user. Assume the recommendation score is calculated based on genre similarity and user ratings. [3]

Q7. Define the concept of "Counting Distinct Elements in a Stream" with reference to data stream mining. [3]