Agnel Charities'

Fr. C. Rodrigues Institute of Technology, Vashi, Navi-Mumbai

Department of Computer Engineering

Question Bank for AT-2 and IA-2

Course Code and Name: **CSC702 Big Data Analytics** Academic Year:**2023-2024**

Name of the Faculty: **Mrs.Dakshayani R** Semester:**VII(A)**

Module No:4 – Mining Data Streams

1.Summerize the work flow of Data stream Management System with a neat diagram.
2.With respect to data stream queries ,give example of (i)Adhoc query (ii) Standing Query
(iii) Continuous query (iv)one time query (iv)Predefined query.
3.What are the challenges of querying large data streams
4.Compare and contrast DBMS and DSMS
5.Give FM algorithm to count distinct elements in a stream,
Solve (i) x=1,2,3,4,5,6,4,2,5,9,1,6,3,7,1,2,2,4,2,1 given h(x) =1x+6 mod 32
6.Summerize the concept of Blloms filter with suitable example
Solve (i)m=5,h1(x)=x mod 5,h2(x)=(2x+3)mod 5 ,insert 9,11 and query (check) 15,16
(ii) h1(x)=(s.length)mod10 ,h2(x)=(2*s.length)mod 10 ,h3(x)=(no of t's in s)mod 10
Insert 'data','cat','ram' and check 'bata','rita'
7.Estimate the number of buckets using DGIM
(i) 1001011011101
(ii)10110110111011011 (estimate buckets if current stream with 011 enters)

Module no: -5: - Real-Time Big Data Models

1.Compare and contrast Content Based filtering and collaborative filtering.
2.Given user-item matrix determine the missing rating apply both user-user similarity measure and item-item similarity measure to find missing rating

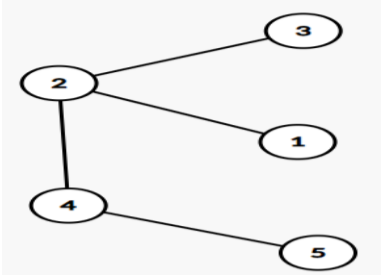
(i)

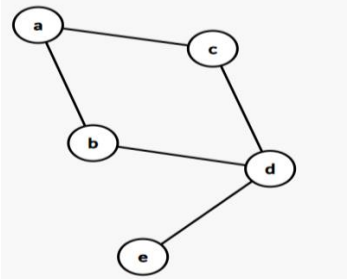
	I1	I2	I3	I4	I5	I6
U1	3	7	4	9	9	7
U2	7		5	3	8	8
U3	7	5	5	??	8	4
U4	5	6	8	5	9	8
U5	5	6	8	8	10	9
U6	7	7		4	7	8

(ii)

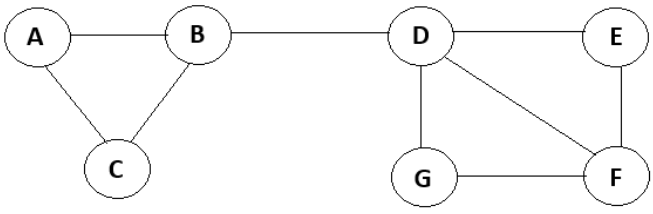
	I1	I2	I3
U1	3	7	??
U2	5	7	5
U3	3	6	7

3.Calculate the centrality measure of the following graph



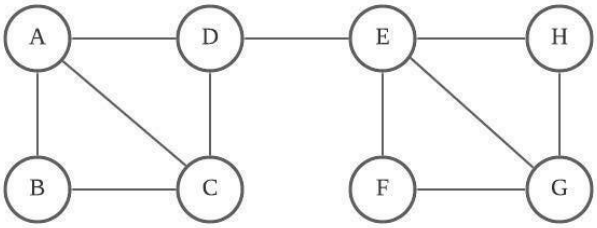


4.Apply Girvan Newman for identify communities in the following graph
(i)



Sample Graph

(ii)



5.Apply Clique Percolation method (k=3) to identify communities in the above graphs

Module No: -6 Data Analytics with R

1.Apply merge(),append(),c(),rbind(),cbind() on the following dataset A=c(1,2,3,4) and B=c(5,6,7,8) and represent the output of each function.
2.Given the vector x=c(9,4,3,6,7,8,3,4) apply function to sort the numbers in ascending order and descending order and also find the sum and product of numbers >5
3.Given a list x=list("apple","orange","banana") perform following operations ,
(i)insert "pineapple",
(ii)insert "strawbrerry" after "apple"
(iii)delete "orange"
(iv)update "banana" to "custardapple"
(v)given y=list("blue berry","dragon fruit","mango") merge both list x and y.
4.create student data frame as given in the figure

Roll Number	Name	CGPA
001	Vaibhav	9.1
002	Neha	9.5
003	Harsh	8.5
004	Shreya	9.3

Perform following operation on the data frame
(i) Vibhav and Neha are 14 years old; harsh is 15, Shreya is 16 years old
Append these data as a new numeric column variable in data frame called age.
(ii)Add a row for Klaus with Roll_Number=005, CGPA=8.5
(iii) Retrieve 1st ,3rd rows and 2nd ,4th columns.
(iv) Find student with cgpa>9
(v) Find Students with CGPI>9 and age<16
(vi) Find Students with CGPI>9 and age<16 display only name .
5.with suitable example write about
(i)List (ii)Vectors (iii)data frame