(3 Hours)

Q. 1 (a) Explain what characteristics of Social Networks make it Big Data. (5)
(b) What do you mean by Jaccard Similarity? Illustrate with an example. Describe any two applications that can use Jaccard Similarity. (5)
(c) Define concept of a Link Farm using a diagram. How does it lead to Link Spam? (5)
(d) What are the challenges of querying on large Data Streams? (5)

Q. 2 (a) What do you understand by BASE properties in NOSQL Database? (10)
Explain in detail any one NOSQL architecture pattern. Identify two applications that can use this pattern.
(b) Write Map Reduce Pseudocode to multiply two matrices. Illustrate the procedure on the following matrices. Clearly show all the steps.

\[
A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}
\]

Q. 3 (a) For the graph given below show the page ranks of all the nodes after running the PageRank algorithm for two iterations with teleportation factor with Beta (\(\beta\)) value = 0.8 (10)

(b) Give two applications for counting the number of 1’s in a long stream of binary values. Using a stream of binary digits, illustrate how the DGIM algorithm will find the number of 1’s. (10)
Q. 4 (a) What do you mean by the Hadoop Ecosystem? Describe any three components of a typical Hadoop Ecosystem.

(b) Explain the following concepts with respect to world wide web
   A. Topic Specific Page Rank
   B. Bowtie structure of the Web

Q. 5 (a) Explain the design of a recommender system used to recommend movies to users. The recommender system should use Collaborative filtering.

(b) Explain the CURE algorithm for clustering large datasets. Please illustrate the algorithm using appropriate figures.

Q. 6 (a) Explain the SON algorithm for Frequent Pattern mining. Illustrate how Map Reduce can be used for implementing this algorithm

(b) What is a "Community" in a Social Network Graph? For the following graph show how the Girvan Newman algorithm finds the different communities.