



Register Number: _____
Date: _____

ST. JOSEPH'S UNIVERSITY, BENGALURU-27
M.Sc.(BIG DATA ANALYTICS) —II SEMESTER
SEMESTER EXAMINATION —APRIL 2023
(Examination conducted in May 2023)
BDA2121:Foundations of Data Science
(For current batch students only)

TIME: 2 hrs

MAXIMUM MARKS: 50

This paper contains ONE printed page and THREE parts.

Part A

Answer ALL questions.

(2 × 5 = 10)

1. Briefly motivate the frequency moments of data streams problems.
2. What is a Markov chain? Give an example.
3. Give two use cases of a bipartite graph.
4. Define the $G(n, p)$ model.
5. What is SVDs? State an application of it.

Part B

Answer ANY FIVE questions.

(4 × 5 = 20)

6. Stat Jonhson-Lindenstrauss Lemma. Explain any one of it's application.
7. Prove that an acyclic graph has a topological sorting. Hence write an algorithm to find a topological sorting of the graph.
8. Explain the power method for computing SVDs.
9. Explain the majority algorithm in the context of frequency moments of data streams.
10. Explain what a phase transition is in a large random graph. Give two examples.
11. Briefly compare and contrast streaming, sketching, and sampling.

Part C

Answer ANY TWO questions.

(2 × 10 = 20)

12. Explain the following as applications of SVDs:
 - A. Principal Component Analysis (PCA).
 - B. Clustering a mixture of sperical gaussians.
13. Compare and contrast $G(n, p)$ and $G(n, m)$ models of graphs. Give two aapplications of each.
14. Explain the existence of large components and emergence of cycles as phase transition properties of a large random graph.