Long Questions

1. What is the CURE algorithm, and how does it work for clustering?
2. Discuss the advantages of the CURE algorithm over traditional clustering algorithms.
3. Explain the concept of clustering in non-Euclidean spaces and its significance in real-world applications.
4. How do distance measures differ in non-Euclidean spaces compared to Euclidean spaces?
5. Describe the challenges of clustering for streaming data and how they are addressed.
6. What are some popular algorithms used for clustering in streaming data?
7. Explain the concept of parallelism in the context of clustering algorithms.
8. How can parallel processing improve the scalability and efficiency of clustering algorithms?
9. Discuss the role of distributed computing frameworks like Apache Spark in parallel clustering.
10. What are some key considerations when implementing parallel clustering algorithms?
11. Describe the MapReduce paradigm and its relevance to parallel clustering.
12. How does the K-means algorithm perform in a distributed computing environment?
13. Explain how partitioning-based clustering algorithms leverage parallelism.
14. Discuss the challenges of maintaining data consistency and synchronization in parallel clustering.
15. How can fault tolerance be achieved in parallel clustering systems?
16. Write a Python function to implement the PageRank algorithm for a given graph representation.
17. Develop a Python script to detect and filter out link spam in a web graph dataset using PageRank scores.
18. Write a SQL query to perform limited-pass algorithms for counting frequent items in a database table.
19. Develop a Python script to perform clustering in non-Euclidean spaces using techniques like DBSCAN or OPTICS.

20. How would you use Python libraries like Scikit-learn or PySpark to implement clustering algorithms for streaming data?
21. What are the primary challenges faced in online advertising, particularly concerning issues like ad fraud and ad blockers?
22. Explain the concept of online algorithms and how they are applied in the context of online advertising.
23. What is the matching problem in online advertising, and how is it typically addressed by ad networks?
24. Discuss the Adwords problem and its significance in the field of online advertising.
25. Can you outline the process involved in implementing Adwords campaigns for advertising purposes?
26. Describe the basic model used in recommendation systems and its components.
27. How do content-based recommendation systems generate recommendations for users?
28. Explain the collaborative filtering approach used in recommendation systems and its variants.
29. What role does dimensionality reduction play in recommendation systems, and how is it achieved?
30. What was the Netflix Challenge, and why was it significant in the field of recommendation systems?
31. How does the emergence of ad blockers impact online advertising strategies and revenue models?
32. Discuss the importance of real-time bidding (RTB) in online advertising auctions.
33. Can you explain the role of machine learning in optimizing ad targeting and personalization?
34. What ethical considerations are involved in targeted advertising and personalized recommendations?
35. How do recommendation systems handle the cold start problem, particularly in new or sparse datasets?
36. How do advertisers optimize their ad campaigns using performance metrics like click-through rate (CTR) and conversion rate?

37. Discuss the concept of ad relevance and its impact on the effectiveness of online advertising.
38. Can you explain the difference between first-price and second-price auction mechanisms in online advertising?
39. Describe the role of data mining techniques in extracting insights from user behavior for targeted advertising.
40. What are some common strategies used to combat click fraud in online advertising?
41. Explain the concept of contextual advertising and provide examples of its implementation.
42. How do ad networks determine the quality score of ads, and how does it affect ad placement?
43. Discuss the challenges associated with ad viewability and methods used to improve it.
44. Can you outline the process of retargeting in online advertising and its benefits?
45. What is the role of programmatic advertising in automating ad buying and selling processes?
46. Describe the use of recommender systems in e-commerce platforms for product recommendations.
47. How does collaborative filtering handle the sparsity problem in recommendation systems?
48. Discuss the trade-offs between content-based and collaborative filtering approaches in recommendation systems.
49. What are some common evaluation metrics used to assess the performance of recommendation systems?
50. Can you explain the concept of matrix factorization and its application in recommendation systems?
51. How would you implement an online algorithm for ad bidding in Python, considering factors like bid price and quality score?
52. Develop a Python program to simulate the AdWords problem and determine the optimal bidding strategy using dynamic programming.
53. Develop a Python script to download and preprocess the Netflix Prize dataset for collaborative filtering experiments.

54. Write Python code to deploy a recommendation system model as a RESTful API using Flask or FastAPI for real-time inference.
55. Develop a Python program to implement the Content-Delivery Network (CDN) algorithm for efficient ad serving based on user location.
56. What are the key characteristics of social networks represented as graphs?
57. How do social network graphs facilitate the analysis of relationships between individuals or entities?
58. Can you explain the process of representing social networks mathematically as graphs?
59. Discuss the challenges associated with modeling large-scale social networks as graphs.
60. What role do nodes and edges play in social network graph representation?
61. How can graph theory concepts be applied to analyze social network structures effectively?
62. Explain the significance of node centrality measures in understanding social network dynamics.
63. How does the density of edges impact the interpretation of social network graphs?
64. Discuss the importance of community detection algorithms in analyzing social network graphs.
65. What are some common metrics used to evaluate the connectivity of social network graphs?
66. Can you describe the concept of homophily in the context of social network graphs?
67. How do directed and undirected edges differ in social network graph representation?
68. What are the implications of edge weights in social network graphs?
69. How can social network graphs be visualized to gain insights into network structures?
70. Discuss the ethical considerations involved in mining social-network graphs for analysis.
71. What is the significance of clustering in social-network graph analysis?
72. Can you explain the process of community detection in social-network graphs?

73. How do modularity-based algorithms contribute to clustering in social-network graphs?
74. Discuss the challenges associated with partitioning large social-network graphs.
75. Explain the concept of Simrank and its applications in social-network graph analysis.
76. How does Simrank measure the similarity between nodes in a social network?
77. Can you provide examples of real-world scenarios where Simrank is applied?
78. What are the limitations of Simrank in analyzing large-scale social-network graphs?
79. How does counting triangles contribute to understanding the connectivity of social-network graphs?
80. What algorithms are commonly used for counting triangles in social-network graphs?
81. Discuss the role of triangles in identifying cohesive substructures within social networks.
82. Can you explain the concept of transitivity in the context of social-network graphs?
83. How do triadic closure and triadic formation affect the evolution of social networks?
84. What are some strategies for mitigating biases in social-network graph mining?
85. How can findings from social-network graph analysis be leveraged to improve recommendation systems or targeted advertising?
86. Develop a Python function to identify and remove isolated nodes (singletons) from a social network graph.
87. Write Python code to compute the shortest path between two nodes in a social network graph using Dijkstra's algorithm.
88. Write Python code to compute the clustering coefficient of a node in a social network graph.
89. Write Python code to partition a social network graph into balanced subgraphs using the Kernighan–Lin algorithm.

90. Write Python code to calculate the degree centrality of each node in a social network graph.