

CSCI-4974/6971: Assignment 3 (100 pts)

Take-home Final

Soft Due Date: Wednesday 13 Dec., Hard Deadline: Sunday 17 Dec.

This assignment will just involve some short response questions based on Lectures 1–20.

1 Response Questions

In a single file, submit your responses to the following questions:

1. How does varying the ordering of vertices within a graph might affect the performance of a computation performed on it – e.g., for Pagerank?
2. Describe how the graph partitioning problem is utilized to improve execution time of a distributed graph computation. What are variants of the partitioning problem, such as what constraints and objectives might one consider?
3. Describe two different random graphs and discuss some of their properties (e.g., degree distribution/skew, diameter, clustering, etc.)
4. Discuss a specific application of graph analysis for each of the following graph types: web graphs, social network graphs, road networks, biological networks.
5. You're using label propagation to detect community structure of a changing graph. Assume you have an initial static graph, and then future updates in terms of edge additions/deletions (e.g. “edge between vertex 10 and vertex 55 is added”) are streamed in. Describe an algorithm that processes these updates. Make any assumptions you need to.
6. GraphBLAS is a push towards formulating graph analytics as linear algebraic operations. Describe at least two motivating factors.
7. Describe two methods we looked at for graph sampling. Give at least one pro and one con for each.
8. What is a 2D distributed graph layout? What is its primary benefit relative to a 1D layout?
9. Give at least two challenges involved with processing graphs on a manycore such as a GPU relative to processing on a CPU.

This take-home final will be evaluated on your answers to these questions (10 pts each).