CS525: Homework 2
Due Date: 7th October

Make reasonable and well-stated assumptions as needed.

1. In the Reduce phase of MapReduce, there are p mappers and k reducers. Assume that the k reducers are chosen from among the p mappers (you may choose them to maximize performance). Assuming that the reduce operation happens only at source (i.e. the mappers) or the sink (reducers), i.e., there is no reduction in the network, derive the fastest algorithm for this operation. What is the runtime of your operation? (Assume a network with O(p) bisection bandwidth). (You can assume a simple summation map-reduce operation)

2. Repeat Question 1. In this case, assume that you are allowed to do in-network reduction. Derive the fastest algorithm for this operation.

3. In the all-to-all broadcast algorithm developed in class, the goal was to develop the fastest algorithm. Consider an alternate case in which the goal is to minimize the number of bits communicated between any pair of processors (i.e., minimum total communication cost). What is the total communication cost of the optimal all-to-all broadcast method presented in the class? Develop a method that reduces total communication cost. What is the total communication cost of your algorithm? Assume a network with O(p) bisection.

4. Repeat Question 3 for an all-to-all personalized communication operation.

5. Consider the case where you have a k-to-all broadcast. k processors initially have messages. At the end of the operation, all of the processors must have all of the messages from the k sources. Derive the best algorithm for this operation. Assume a network with O(p) bisection.

6. Develop an algorithm for minimizing total communication cost of the k-to-all broadcast operation. What is the total communication

cost of your algorithm? Assume a network with O(p) bisection.

7. Consider the optimal one-to-all broadcast algorithm (implemented as one-to-all personalized followed by all-to-all broadcast). What is the total communication cost of this algorithm? How would you establish a lower bound for this cost? Is the time-optimal one-to-all broadcast operation also total communication optimal? Assume a network with O(p) bisection.