

Notes. Please work on this with your group-mate(s); just submit *one* writeup per group. Consulting other sources (including the Web) is not allowed. Write your solutions neatly; if you are able to make partial progress by making some additional assumptions, then state these assumptions clearly and submit your partial solution.

Problem 1 (the second problem) involves simulation; it is suggested that you start on it at the earliest, to give you some time to do and interpret the simulations.

0. Download and read the following two papers of Jon Kleinberg, which are essentially different versions of the same paper, but have somewhat differing content and style: *Navigation in a Small World* (Nature, 2000) and *The small-world phenomenon: an algorithmic perspective* (Proc. ACM STOC, 2000).

This important work of Kleinberg has led to many interesting applications in peer-to-peer networks and sociology. I will assume for the final exam, that you have read and understood these papers.

1. Consider the directed cycle C on $(n + 1)$ nodes: we have nodes $0, 1, \dots, n$, with a directed edge from node i to node $(i + 1) \bmod (n + 1)$, for each i . Let $d(i, j)$ be the length (i.e., number of directed edges) in the unique path from node i to node j in C . (Note that $d(i, j)$ and $d(j, i)$ are in general different.) We augment C with additional directed edges as follows. Each node i independently chooses a long-range contact $j \neq i$ (i.e., node i constructs a *directed* edge from itself to j), by choosing node j from the following distribution:

$$\Pr[i \text{ chooses } j] = \frac{1/d(i, j)}{1/1 + 1/2 + 1/3 + \dots + 1/n}.$$

In this augmented graph, we want to route a message from node 0 to node n using a distributed algorithm. It has been shown that such routing can be accomplished in $O(\log^2 n)$ expected number of steps. As usual, a single step taken by the message can only be from node u to a node v , if there is a *directed* edge from u to v .

Do a simulation of this process for many “large” values of n , and use the simulation runs to (approximately) explain why the expected time should be upper-bounded by something like $O(\log^2 n)$; even giving a sound explanation for any upper bound of the form $O(\log^c n)$ for some constant c , is fine. You can choose any language, platform etc. for the simulation. Use your creativity to come up with as good an experimental setup and interpretation of the experimental runs, as you can!