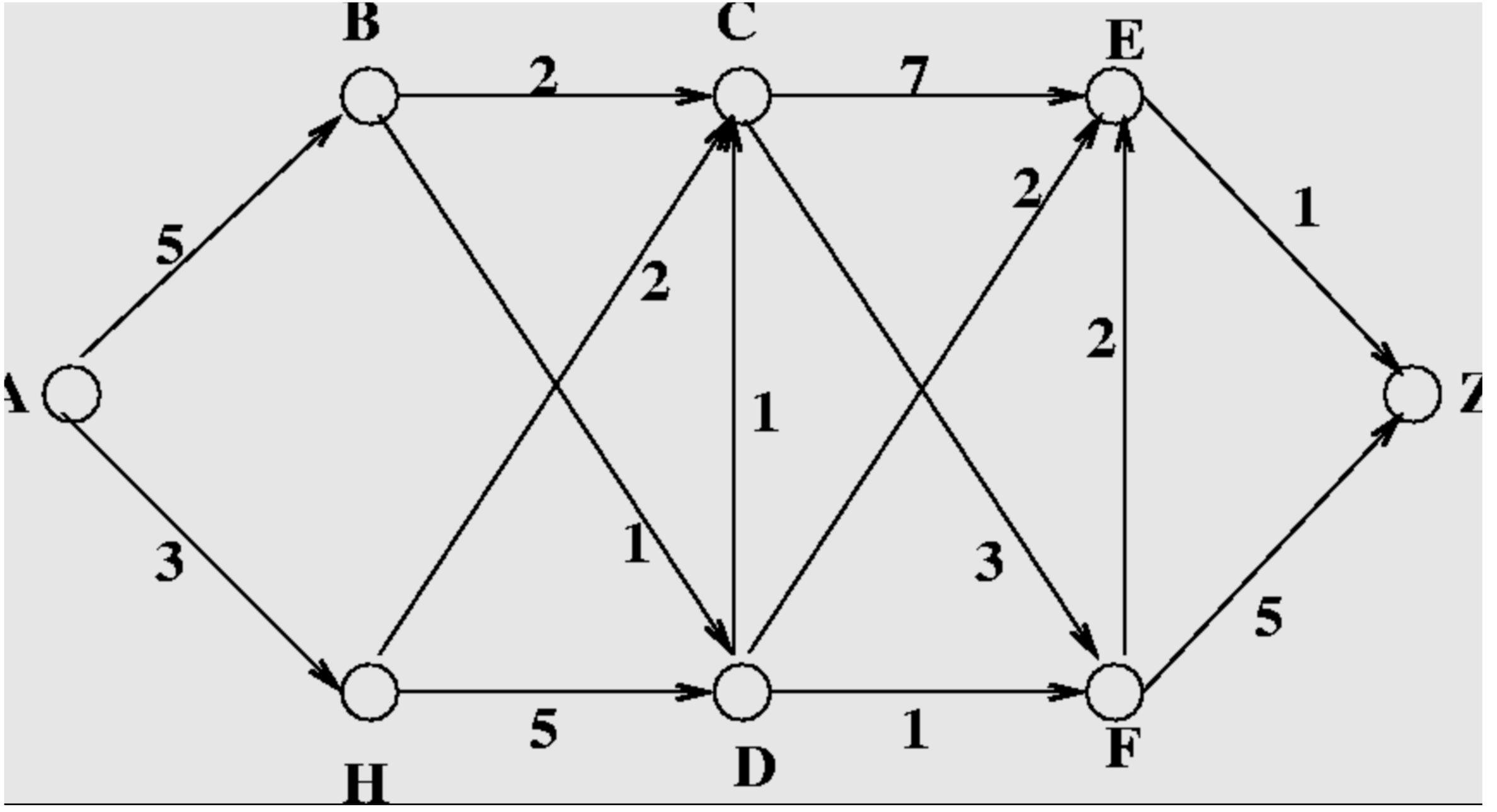

Find the shortest path from A to Z using dynamic programming. Show your work.

Consider an MDP with two states, A and B. In A, there are two actions you can take. Action 1 keeps you in state A, with a reward of one. Action 2 moves you to B, with a reward of zero. In state B, there is only one action to take, which keeps you in B with a reward of 2.

(i) What is the smallest value of the discount factor for which, starting in A, moving to B is optimal?

(ii) Consider the policy which takes a random action. Perform the first two iterations of policy evaluation starting from [16,16] with a discount factor of 1/2. Do this by hand (i.e., do not write code).

(iii) Perform (again by hand) the first two iterations of value iteration for this problem, also starting from [16,16] with discount of 1/2.

- Suppose you add a constant to all costs in an MDP. Is it always true that the optimal policy remains the same?
  Give an answer to this question in the cases when:
  (i) the MDP is continuing.
  (ii) the MDP is terminal.
  You can assume that, in both cases, the discount factor is strictly less than one.
- Suppose someone gives you a randomized policy π in a dynamic programming problem and claims it is optimal. How would you construct a deterministic policy π which is optimal?
  Justify your answer.