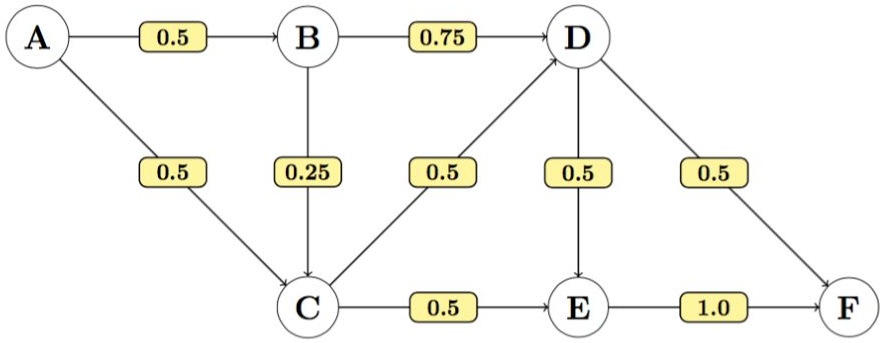
PROBLEM 11

Problem 7 The following graph specifies the states and transition probabilities for a Markov Decision Process (MDP). There are only two actions in this MDP: $a = S$ (stay) or $a = M$ (move). If you elect to stay, you remain in the same state with probability one. If you move, you change states according to the probabilities specified below. F is the only terminal state where you don't move even if you select $a = M$.



The rewards in this MDP are associated with state transitions such that

$$R(B \to D) = -1,\ R(D \to F) = +10,\ R(A \to C) = -2,$$

and all the remaining rewards are zero. The discount factor is $\gamma = 0.5$.

(7.1) **(4 points)** Suppose we initialize the values as

$$V_0(A) = -1,\ V_0(B) = 2,\ V_0(C) = 1,\ V_0(D) = 1,\ V_0(E) = 0,\ V_0(F) = 3$$

What would be the resulting action to take in states A and C?

(7.2) **(3 points)** Calculate $V_1(C)$ after one value iteration.

(7.3) **(2 points)** Suppose we perform value iteration until convergence obtaining $V^*(s)$, $s = A, B, C, D, E, F$. What is the resulting $V^*(F)$? **( )**

(7.4) **(3 points)** Are we guaranteed to get the cumulative discounted reward equal to $V^*(D)$ if we begin in state D and act optimally according to the converged values? (Y/N) **( )**