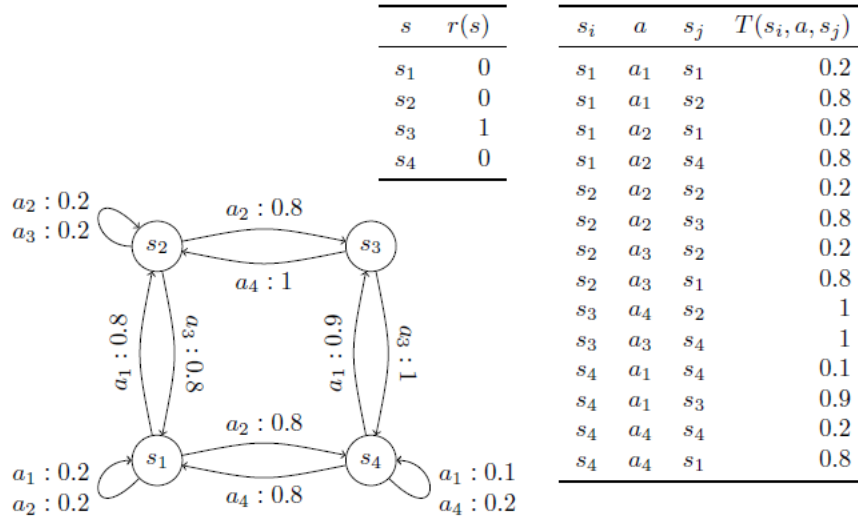
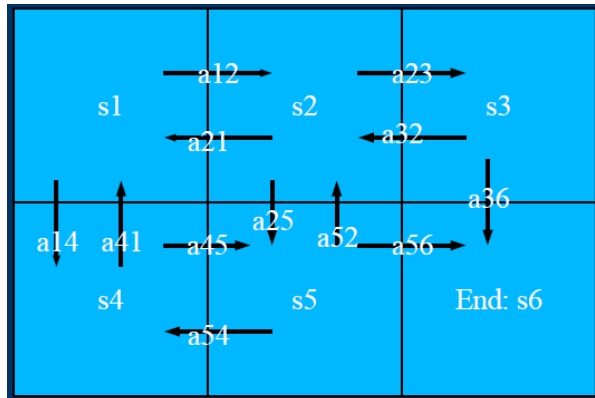


## Homework 2

- 1- Use Value Iteration and Policy Iteration algorithms and solve the following MDP. Stop after 5 time steps.



- 2- Consider the following scenario with  $\gamma = 0.5$  and  $r=100$  if moving into  $S_6$  and 0 otherwise.



Use Q-Learning algorithm and calculate  $(S_k, a_{i,j})$ . You must have at least one nonzero value for each state action pair.