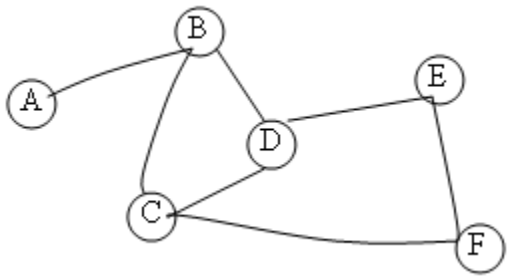
Invatare prin recompensa
Q-learning

Algoritmul Q-learning

initializeaza Q(s, a) pentru toate starile s si actiunile a
pentru fiecare episod
    alege s
    repeta
        alege actiunea a
        executa actiunea a, noua stare devine s', se obtine recompensa r
        $Q(s, a) = r + \delta \max_{a'} Q(s', a')$
        s = s'
    pana la starea de stop



Recompense immediate:

Stare	**Stare urmatoare**					
	A	**B**	**C**	**D**	**E**	**F**
A	-5	0	-	-	-	-
B	0	-5	0	0	-	-
C	-	0	-5	0	-	100
D	-	0	0	-5	0	-
E	-	-	-	0	-5	100
F	-	-	0	-	0	-

Resurse:

1. Curs 3 – Q learning
2. http://people.revoledu.com/kardi/tutorial/ReinforcementLearning/index.html