4/17/2019 P5

```
In [23]:
         import pandas as pd
         import numpy as np
         from tqdm import trange
         import matplotlib.pyplot as plt
         %matplotlib inline
         data = pd.read csv('facebook.txt', header = None, delimiter=' ')
         E = data.values
         V = np.unique(E)
         # function for calculating influence
         def I(E,S):
             N S = S
             for s in S:
                 N_S=np.append(N_S,E[E[:,0]==s,1])
             I S = len(np.unique(N S))
             return I S
         # greed algorithm, we added S as an input so we don't repeat caculations below
         def greedy(E,V,K,S):
             k = len(S)
             while k < K and I(E,S)<len(V):
                 I_S_a_opt = 0
                 a opt = []
                 for a in np.setdiff1d(V,S):
                      I_S_a = I(E,np.append(S,a))-I(E,S)
                      if I_S_a > I_S_a_opt:
                          a opt = a
                          I_S_a_opt = I_S_a
                 S = np.append(S,a_opt)
                 k+=1
             return S, I(E,S)
         # an example of users to seed and total influence for a budge of k=2
         greedy(E,V,2,np.asarray([]))
Out[23]: (array([ 107., 1684.]), 1820)
In [28]: k max = 15
         S = np.asarray([])
         I k = np.zeros(k max)
         for i in trange(k_max):
             S,I_k[i] = greedy(E,V,i,S)
         15/15 [01:40<00:00, 6.71s/it]
```

local host: 8888/nbc onvert/html/Pset 10/P5. ipynb? download=false

4/17/2019 P5

```
In [31]: # plot the influence as a function of K
plt.plot(np.arange(0,k_max),I_k)
plt.ylabel('I(S)')
plt.xlabel('K')
plt.show()
```

