CSCI5070 Advanced Topics in Social Computing

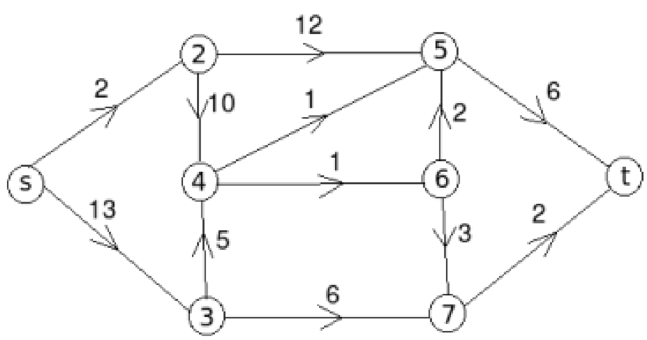
Assignment 2

Deadline: 15:59:59, November 6 (Tuesday), 2012

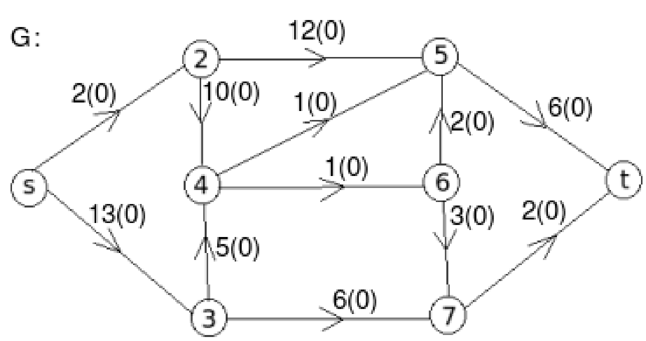
1. Ford-Fulkerson Algorithm.

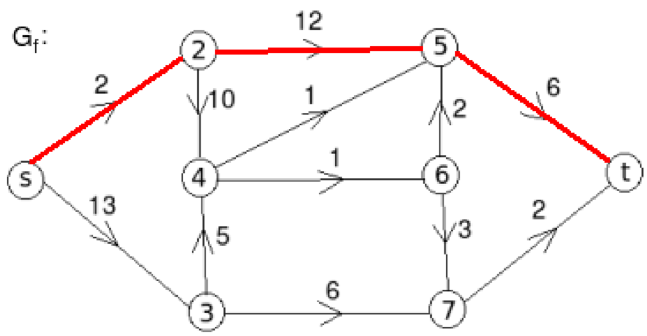
Consider the network flow problem with the following edge capacities, $c(u, v)$ for edge (u, v): $c(s, 2) = 2$, $c(s, 3) = 13$, $c(2, 5) = 12$, $c(2, 4) = 10$, $c(3, 4) = 5$, $c(3, 7) = 6$, $c(4, 5) = 1$, $c(4, 6) = 1$, $c(6, 5) = 2$, $c(6, 7) = 3$, $c(5, t) = 6$, $c(7, t) = 2$.

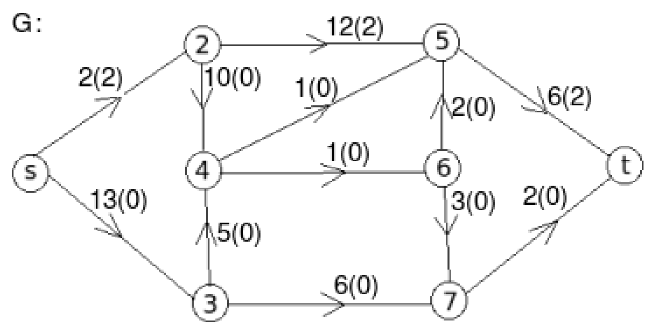
(1) Draw the network.

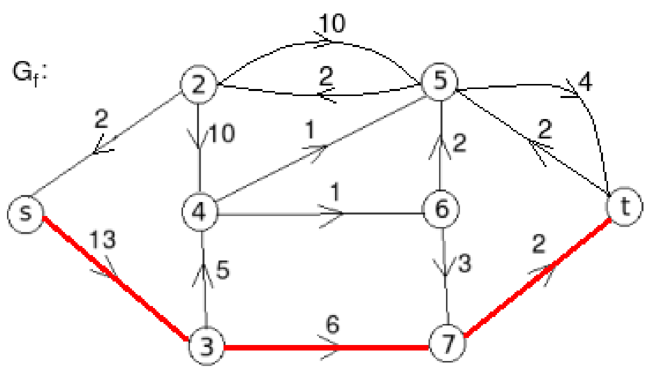


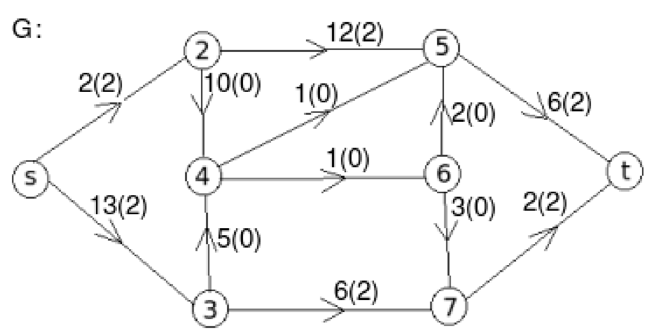
(2) Run the Ford-Fulkerson algorithm to find the maximum flow. Show each residual graph.

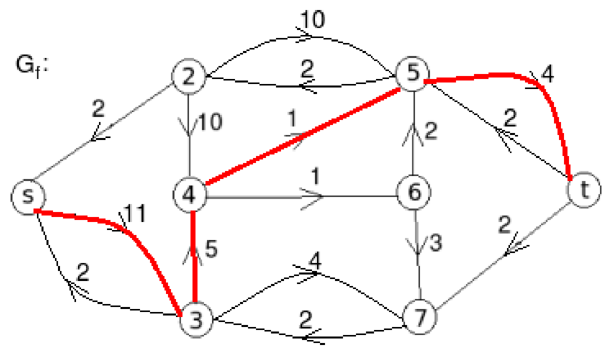


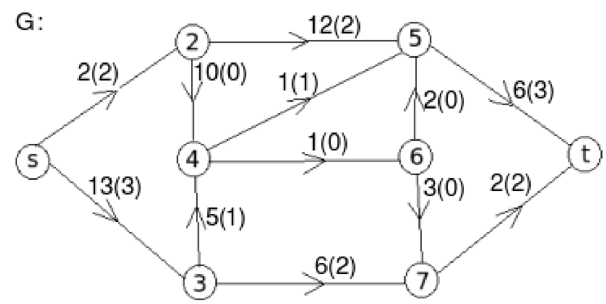

G_f:
s
2
3
4
5
6
7
t
12
2
10
1
2
6
1
13
5
3
2
6

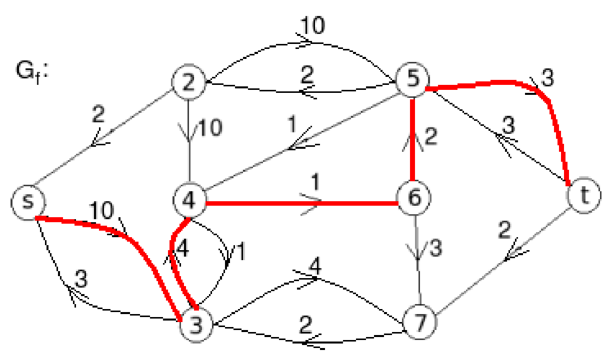

G:
s
2
3
4
5
6
7
t
12(2)
2(2)
10(0)
1(0)
2(0)
6(2)
1(0)
13(0)
3(0)
2(0)
5(0)
6(0)

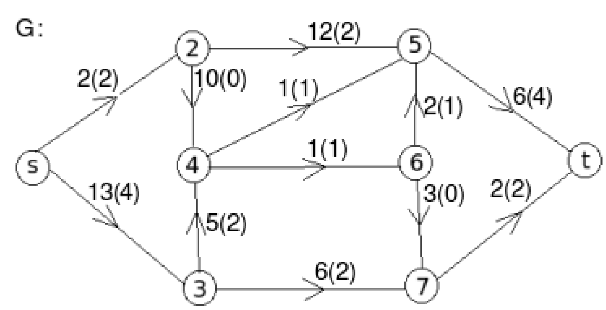

G_f:
s
2
3
4
5
6
7
t
10
2
4
2
10
1
2
2
1
13
5
3
2
6

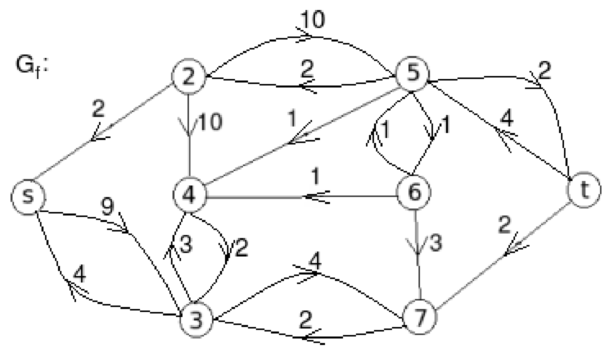

G:
s
2
3
4
5
6
7
t
12(2)
2(2)
10(0)
1(0)
2(0)
6(2)
1(0)
13(2)
3(0)
2(2)
5(0)
6(2)

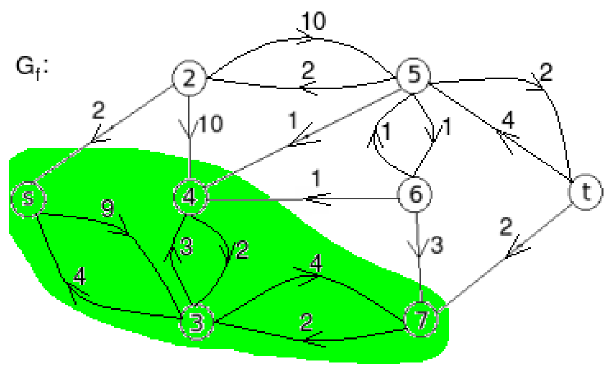

G_f:
s
2
3
4
5
6
7
t
10
2
2
10
1
2
2
4
1
11
5
2
2
3
4
2
2


G:
s
2
3
4
5
6
7
t
12(2)
2(2)
10(0)
1(1)
2(0)
6(3)
1(0)
13(3)
3(0)
2(2)
5(1)
6(2)


G_f:
s
2
3
4
5
6
7
t
10
2
3
2
10
1
2
3
1
10
4
1
3
2
3
4
2

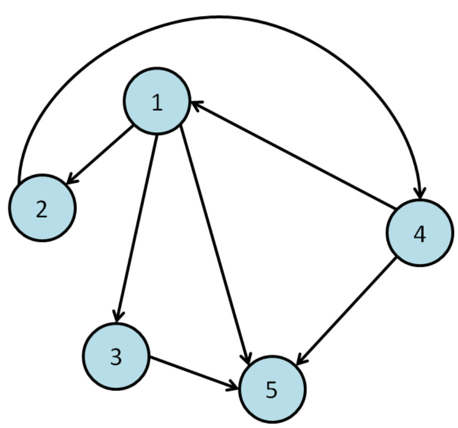

G:
s
2
3
4
5
6
7
t
12(2)
2(2)
10(0)
1(1)
2(1)
6(4)
1(1)
13(4)
3(0)
2(2)
5(2)
6(2)





(3) Show the minimum cut. The minimum cut is highlighted in green. Therefore, the minimum cut has capacity 6 and the maximum flow has flow 6.

2. PageRank and HITS.
The the link structure of five web pages is shown in the following figure.



(1) Suppose $d = 0.7$, please calculate PageRank score of each state in the first and seconde iterations. The initiate score of each state is 0.2.

$$
\begin{aligned}
PR_1(1) &= \frac{1-d}{N} + d \cdot (\frac{PR_0(4)}{L(4)} + \frac{PR_0(5)}{L(5)}) = \frac{0.3}{5} + 0.7 \cdot (\frac{0.2}{2} + \frac{0.2}{5}) = 0.158 \\
PR_1(2) &= \frac{1-d}{N} + d \cdot (\frac{PR_0(1)}{L(1)} + \frac{PR_0(5)}{L(5)}) = \frac{0.3}{5} + 0.7 \cdot (\frac{0.2}{3} + \frac{0.2}{5}) = 0.135 \\
PR_1(3) &= \frac{1-d}{N} + d \cdot (\frac{PR_0(1)}{L(1)} + \frac{PR_0(5)}{L(5)}) = \frac{0.3}{5} + 0.7 \cdot (\frac{0.2}{3} + \frac{0.2}{5}) = 0.135 \\
PR_1(4) &= \frac{1-d}{N} + d \cdot (\frac{PR_0(2)}{L(2)} + \frac{PR_0(5)}{L(5)}) = \frac{0.3}{5} + 0.7 \cdot (\frac{0.2}{1} + \frac{0.2}{5}) = 0.228 \\
PR_1(5) &= \frac{1-d}{N} + d \cdot (\frac{PR_0(1)}{L(1)} + \frac{PR_0(3)}{L(3)} + \frac{PR_0(4)}{L(4)} + \frac{PR_0(5)}{L(5)}) = \frac{0.3}{5} + 0.7 \cdot (\frac{0.2}{3} + \frac{0.2}{1} + \frac{0.2}{2} + \frac{0.2}{5}) = 0.345 \\
PR_2(1) &= \frac{1-d}{N} + d \cdot (\frac{PR_1(4)}{L(4)} + \frac{PR_1(5)}{L(5)}) = 0.188 \\
PR_2(2) &= \frac{1-d}{N} + d \cdot (\frac{PR_1(1)}{L(1)} + \frac{PR_1(5)}{L(5)}) = 0.145 \\
PR_2(3) &= \frac{1-d}{N} + d \cdot (\frac{PR_1(1)}{L(1)} + \frac{PR_1(5)}{L(5)}) = 0.145 \\
PR_2(4) &= \frac{1-d}{N} + d \cdot (\frac{PR_1(2)}{L(2)} + \frac{PR_1(5)}{L(5)}) = 0.203 \\
PR_2(5) &= \frac{1-d}{N} + d \cdot (\frac{PR_1(1)}{L(1)} + \frac{PR_1(3)}{L(3)} + \frac{PR_1(4)}{L(4)} + \frac{PR_1(5)}{L(5)}) = 0.319
\end{aligned}
$$

(2) The initialization of hub score and authority score for each node are both 0.2. Please calculate the hub and authority scores of each state in the first and second iterations.

Let x represent authority score and y represent hub score.

$$\begin{aligned}
x_1(1) &= y_0(4) = 0.2 \\
x_1(2) &= y_0(1) = 0.2 \\
x_1(3) &= y_0(1) = 0.2 \\
x_1(4) &= y_0(2) = 0.2 \\
x_1(5) &= y_0(1) + y_0(3) + y_0(4) + = 0.6 \\
\\
y_1(1) &= x_1(2) + x_1(3) + x_1(5) = 1.0 \\
y_1(2) &= x_1(4) = 0.2 \\
y_1(3) &= x_1(5) = 0.6 \\
y_1(4) &= x_1(1) + x_1(5) = 0.8 \\
y_1(5) &= 0 \\
\\
x_2(1) &= y_1(4) = 0.8 \\
x_2(2) &= y_1(1) = 1.0 \\
x_2(3) &= y_1(1) = 1.0 \\
x_2(4) &= y_1(2) = 0.2 \\
x_2(5) &= y_1(1) + y_1(3) + y_1(4) + = 2.4 \\
\\
y_2(1) &= x_2(2) + x_2(3) + x_2(5) = 4.4 \\
y_2(2) &= x_2(4) = 0.2 \\
y_2(3) &= x_2(5) = 2.4 \\
y_2(4) &= x_2(1) + x_2(5) = 3.2 \\
y_2(5) &= 0
\end{aligned}$$

(3) Please try to find relevant materials to prove the convergence of PageRank algorithm (Hints: PageRank is a special case of Markov Process, where relevant theorems can be found.).

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)},$$

where $p_1, p_2, ..$ are the pages under consideration, $M(p_i)$ is the set of pages that link to p_i, $L(p_j)$ is the number of outbound links on page p_j, and N is the total number of pages. Let $\pi = [PR(p_1), PR(p_1), ..., PR(p_N)]$, we have $\pi_s G = \pi_{s+1}$, where

$$G = d \cdot \begin{pmatrix} \frac{l(p_1,p_1)}{L(p_1)} & \frac{l(p_1,p_2)}{L(p_1)} & \cdots & \frac{l(p_1,p_n)}{L(p_1)} \\ \frac{l(p_2,p_1)}{L(p_2)} & \frac{l(p_2,p_2)}{L(p_2)} & \cdots & \frac{l(p_2,p_n)}{L(p_2)} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \frac{l(p_n,p_1)}{L(p_n)} & \frac{l(p_n,p_2)}{L(p_n)} & \cdots & \frac{l(p_n,p_n)}{L(p_n)} \end{pmatrix} + \frac{(1-d)}{N} \cdot ee^T$$

, and $l(p_i, p_j) = 1$ if p_i links to p_j, 0 otherwise.
Thus, G is a stochastic matrix (G is also called "Google matrix") for a Markov process. A stationary probability vector π is defined over G: $\pi G = \pi$. That is exactly the pagerank value for each webpage. Therefore, the PageRank algorithm converges.

3. Memory-based Collaborative Filtering.

	i_1	i_2	i_3	i_4	i_5	i_6
u_1	0	2	5	3	1	0
u_2	3	3	4	3	0	2
u_3	3	0	0	5	2	2
u_4	5	0	4	4	5	3
u_5	2	3	3	0	2	2

The above table shows the ratings of 5 users on 6 items (The value 0 means the user has not rated the item). Please utilize Pearson Correlation Coefficient (PCC) similarity and Memory-based CF algorithms introduced in the lecture notes to

(1) find top 2 most similar users of u_3 and estimate u_3's rating on i_2 using user-based CF.

Given X and Y, $PCC(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$. Thus,

$$\begin{aligned} PCC(u_3,u_1) &= 1 \\ PCC(u_3,u_2) &= 0.756 \\ PCC(u_3,u_3) &= 1 \\ PCC(u_3,u_4) &= 0 \\ PCC(u_3,u_5) &= 0 \end{aligned}$$

, from which we know u_1 and u_2 are the top 2 most similar users of u_3.
$r_{32} = \bar{r}_3 + \frac{PCC(u_3,u_2)*(r_{22}-\bar{r}_2)+PCC(u_3,u_1)*(r_{12}-\bar{r}_1)}{PCC(u_3,u_2)+PCC(u_3,u_1)} = 2.573.$

(2) find top 2 most similar items of i_5 and estimate u_2's rating on i_5 using item-based CF.

$$\begin{aligned} PCC(i_5, i_1) &= 0.945 \\ PCC(i_5, i_2) &= 1 \\ PCC(i_5, i_3) &= -0.240 \\ PCC(i_5, i_4) &= 0.240 \\ PCC(i_5, i_5) &= 1 \\ PCC(i_5, i_6) &= 1 \end{aligned}$$

, from which we know i_6 and i_2 are the top 2 most similar items of i_5. $r_{25} = \frac{PCC(i_5,i_6)*r_{26}+PCC(i_5,i_2)*r_{22}}{PCC(i_5,i_6)+PCC(i_5,i_2)} = 2.500$.