CMSC5733 Social Computing

Exercise 2

Deadline: 23:59:59, October 21(Monday), 2013

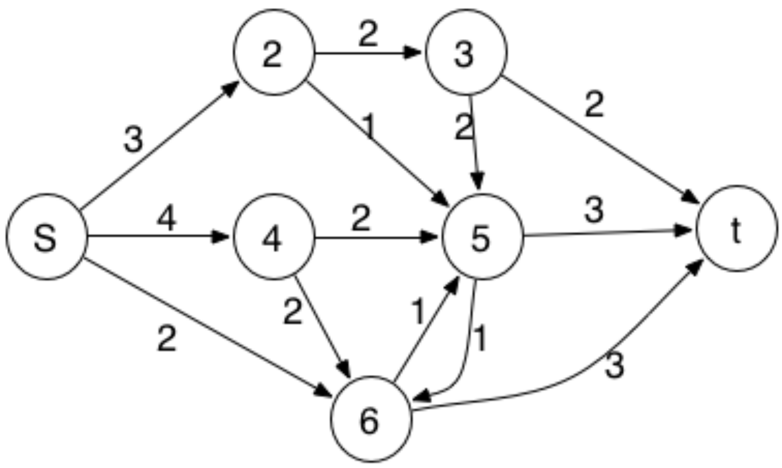
Late submission will lead to marks deduction. Days of 1, 2, 3, and 4 or above will cause 10%, 30%, 60% and 100% marks deduction respectively.

Submission Guidelines: Please send the PDF file to email address cmsc5733@gmail.com with your name and student ID.

1. Ford-Fulkerson Algorithm.

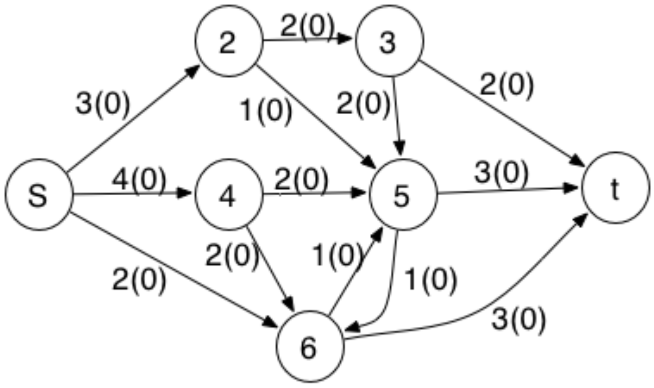
Consider the network flow problem with the following edge capacities, c(u, v) for edge (u, v): c(s, 2) = 3, c(s, 4) = 4, c(s, 6) = 2, c(2, 3) = 2, c(2, 5) = 1, c(4, 6) = 2, c(4, 5) = 2, c(6, 5) = 1, c(6, t) = 3, c(5, 6) = 1, c(5, t) = 3, c(3, 5) = 2, c(3, t) = 2.

(1) Draw the network.



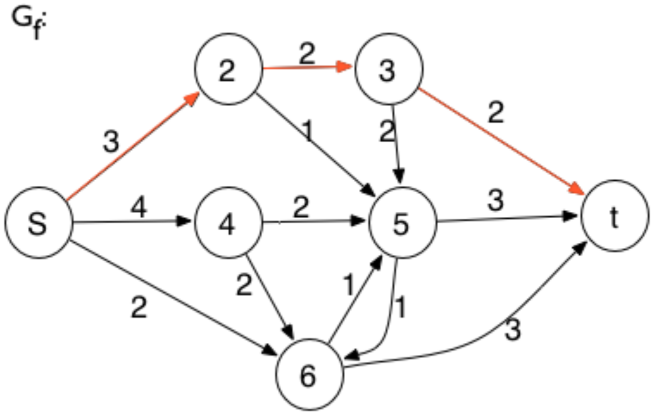
(2) Run the Ford-Fulkerson algorithm to find the maximum flow. Show each residual graph.

G:



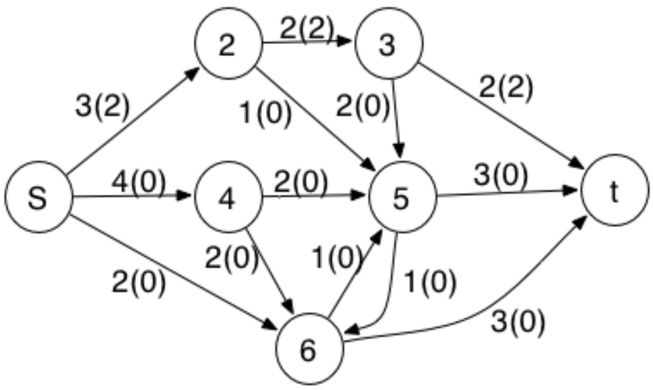
S
2
3
4
5
6
t
3(0)
2(0)
1(0)
2(0)
2(0)
4(0)
2(0)
3(0)
2(0)
1(0)
1(0)
2(0)
3(0)

G_f:



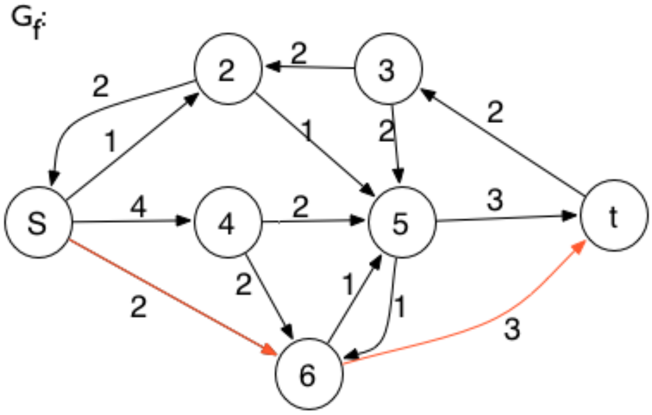
S
2
3
4
5
6
t
3
2
1
2
2
4
2
3
2
1
1
2
3

G:



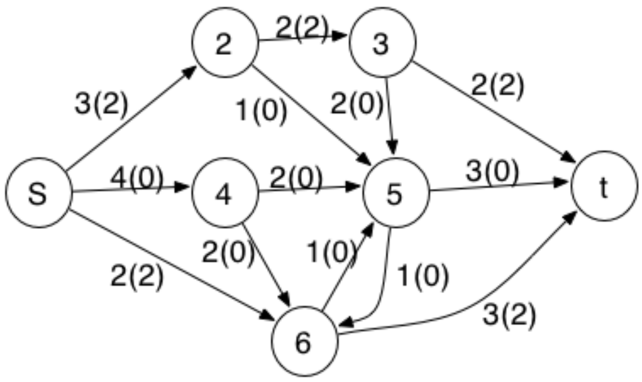
S
2
3
4
5
6
t
3(2)
2(2)
1(0)
2(0)
2(2)
4(0)
2(0)
3(0)
2(0)
1(0)
1(0)
2(0)
3(0)

G_f:



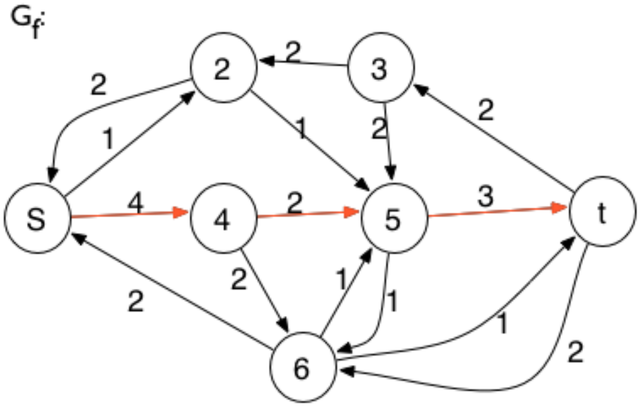
S
2
3
4
5
6
t
2
1
2
1
2
2
4
2
3
2
1
1
2
3

G:



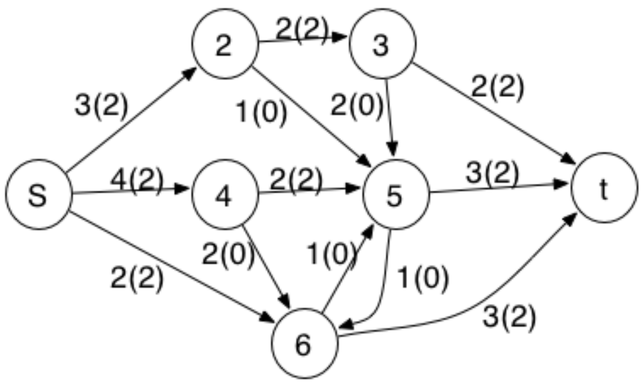
2
3
S
4
5
t
6
2(2)
3(2)
1(0)
2(0)
2(2)
4(0)
2(0)
3(0)
2(0)
1(0)
2(2)
1(0)
3(2)

G_f:



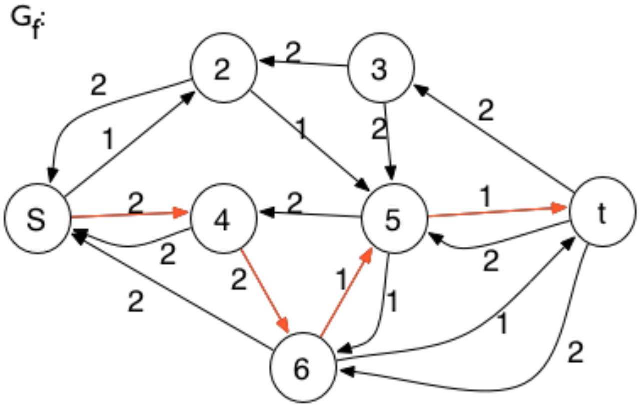
2
3
S
4
5
t
6
2
2
1
1
2
2
4
2
3
2
1
2
1
1
2

G:



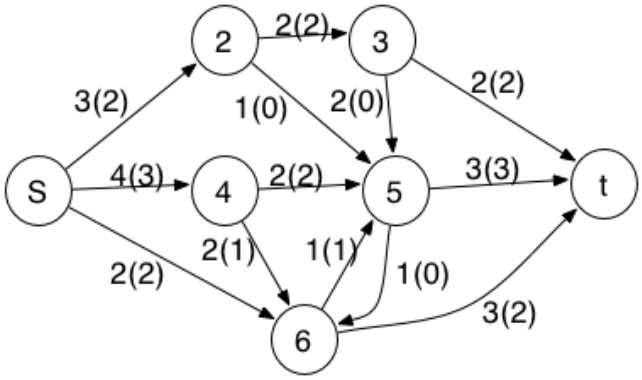
2
3
S
4
5
t
6
2(2)
3(2)
1(0)
2(0)
2(2)
4(2)
2(2)
3(2)
2(0)
1(0)
2(2)
1(0)
3(2)

G_f:

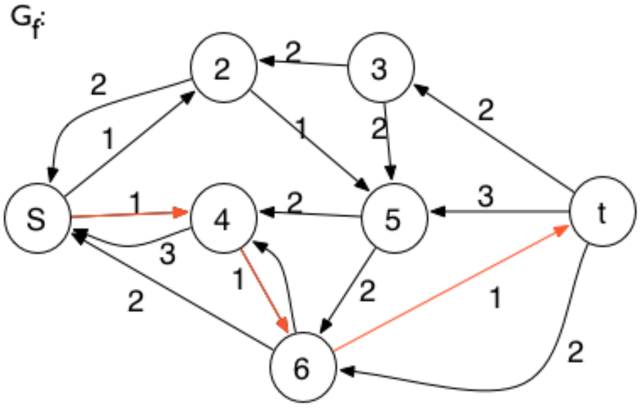


2
3
S
4
5
t
6
2
2
1
1
2
2
2
2
1
2
2
2
1
2
1
2
1
2

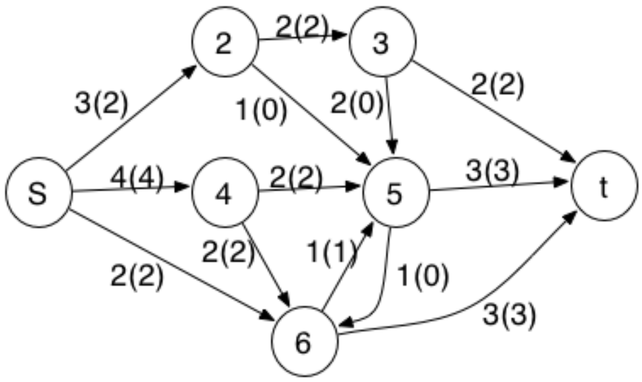
G:


2
2(2)
3
3(2)
1(0)
2(0)
2(2)
S
4(3)
4
2(2)
5
3(3)
t
2(1)
1(1)
1(0)
2(2)
6
3(2)

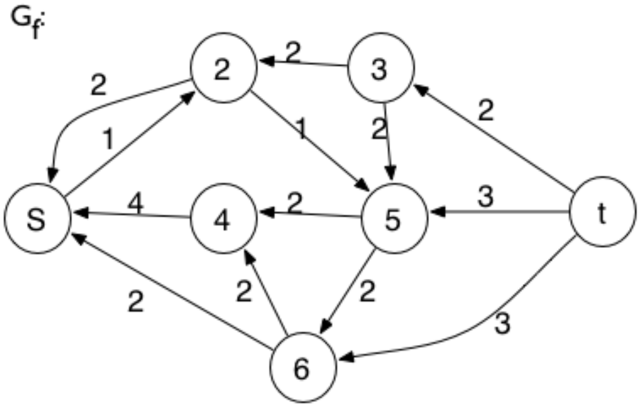
G_f:


2
2
3
2
1
1
2
2
S
1
4
2
5
3
t
3
1
2
2
1
6
2

G:


2
2(2)
3
3(2)
1(0)
2(0)
2(2)
S
4(4)
4
2(2)
5
3(3)
t
2(2)
1(1)
1(0)
2(2)
6
3(3)

G_f:


2
2
3
2
1
1
2
2
S
4
4
2
5
3
t
2
2
2
6
3

(3) Show the minimum cut.

The minimum cut is shown in below image. Therefore, the minimum cut has capacity 8 and the maximum flow has flow 8.

Minimum Cut:

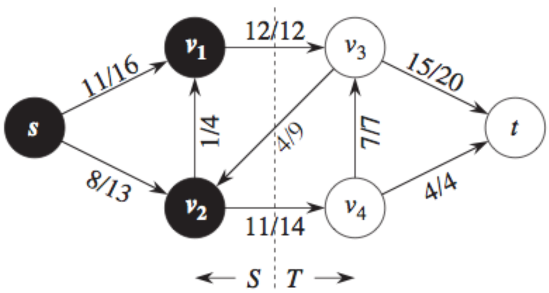
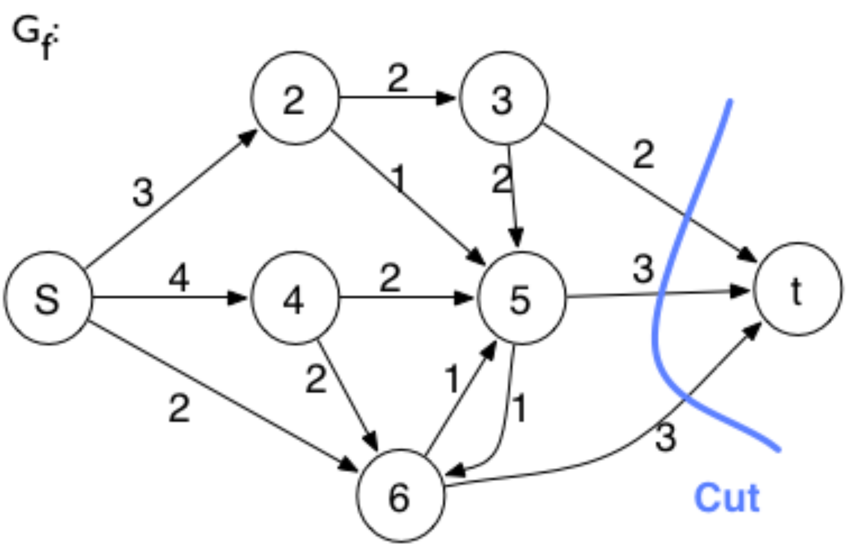


Figure 26.5 A cut (S, T) in the flow network of Figure 26.1(b), where $S = \{s, v_1, v_2\}$ and $T = \{v_3, v_4, t\}$. The vertices in S are black, and the vertices in T are white. The net flow across (S, T) is $f(S, T) = 19$, and the capacity is $c(S, T) = 26$.

The ***capacity*** of the cut (S, T) is

$$c(S, T) = \sum_{u \in S} \sum_{v \in T} c(u, v) . \quad (26.10)$$

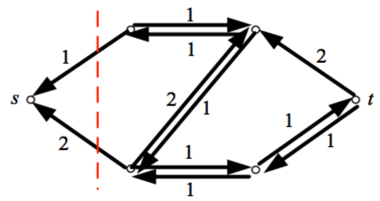
A minimum cut of a network is a cut whose capacity is minimum over all cuts of the network.

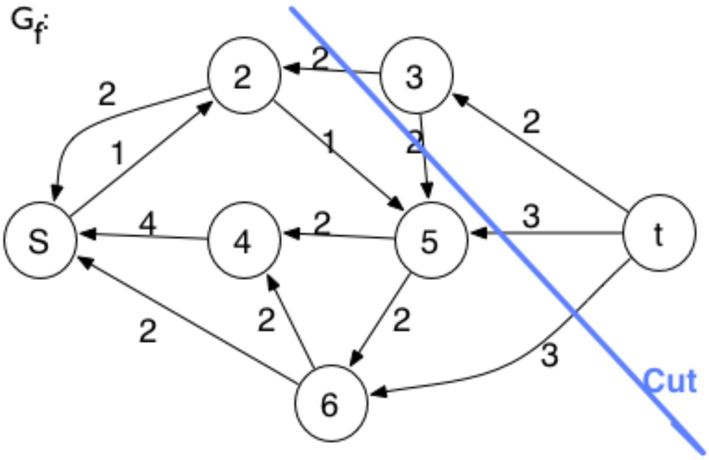


Another way to find minimum cut:

- After the max-flow is found, the minimum cut is determined by

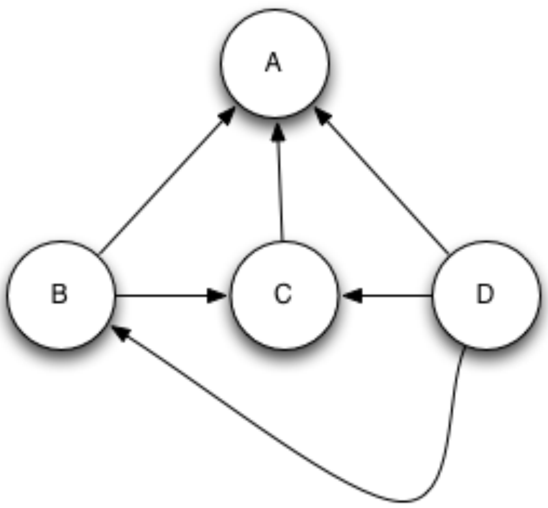
$$\begin{aligned} S &= \{\text{All vertices reachable from } s\} \\ T &= G \backslash S \end{aligned}$$



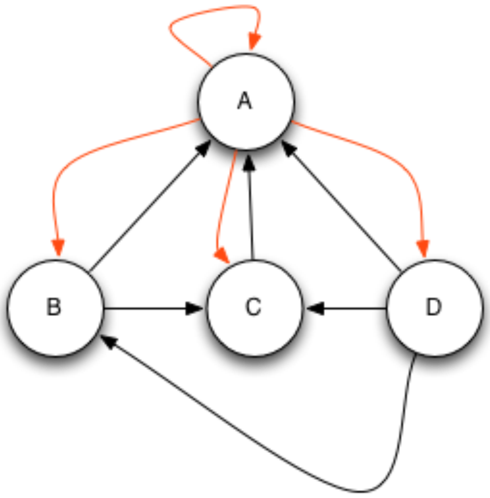


2. PageRank and HITS.

The link structure of four web pages is shown in the following figure.



(1) Suppose d = 0.85, please calculate PageRank score of each state in the first and second iterations. The initiate score of each state is 0.25.



$PR_1(A) = \frac{1-d}{N} + d * (\frac{PR_0(B)}{L(B)} + \frac{PR_0(C)}{L(C)} + \frac{PR_0(D)}{L(D)} + \frac{PR_0(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.25}{2} + \frac{0.25}{1} + \frac{0.25}{3} + \frac{0.25}{4}) = 0.480$

$PR_1(B) = \frac{1-d}{N} + d * (\frac{PR_0(D)}{L(D)} + \frac{PR_0(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.25}{3} + \frac{0.25}{4}) = 0.161$

$PR_1(C) = \frac{1-d}{N} + d * (\frac{PR_0(B)}{L(B)} + \frac{PR_0(D)}{L(D)} + \frac{PR_0(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.25}{2} + \frac{0.25}{3} + \frac{0.25}{4}) = 0.268$

$PR_1(D) = \frac{1-d}{N} + d * (\frac{PR_0(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.25}{4}) = 0.091$

$PR_2(A) = \frac{1-d}{N} + d * (\frac{PR_1(B)}{L(B)} + \frac{PR_1(C)}{L(C)} + \frac{PR_1(D)}{L(D)} + \frac{PR_1(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.161}{2} + \frac{0.268}{1} + \frac{0.091}{3} + \frac{0.480}{4}) = 0.462$

$PR_2(B) = \frac{1-d}{N} + d * (\frac{PR_1(D)}{L(D)} + \frac{PR_1(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.091}{3} + \frac{0.480}{4}) = 0.165$

$PR_2(C) = \frac{1-d}{N} + d * (\frac{PR_1(B)}{L(B)} + \frac{PR_1(D)}{L(D)} + \frac{PR_1(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.161}{2} + \frac{0.091}{3} + \frac{0.480}{4}) = 0.234$

$PR_2(D) = \frac{1-d}{N} + d * (\frac{PR_1(A)}{L(A)}) = \frac{0.15}{4} + 0.85 * (\frac{0.480}{4}) = 0.140$

(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.

```
G := set of pages
for each page p in G do
  p.auth = 1 // p.auth is the authority score of the page p
  p.hub = 1 // p.hub is the hub score of the page p
function HubsAndAuthorities(G)
  for step from 1 to k do // run the algorithm for k steps
    norm = 0
    for each page p in G do  // update all authority values first
      p.auth = 0
      for each page q in p.incomingNeighbors do // p.incomingNeighbors is the set of pages that link to p
         p.auth += q.hub
      norm += square(p.auth) // calculate the sum of the squared auth values to normalise
    norm = sqrt(norm)
    for each page p in G do  // update the auth scores
      p.auth = p.auth / norm  // normalise the auth values
    norm = 0
    for each page p in G do  // then update all hub values
      p.hub = 0
      for each page r in p.outgoingNeighbors do // p.outgoingNeighbors is the set of pages that p links to
        p.hub += r.auth
      norm += square(p.hub) // calculate the sum of the squared hub values to normalise
    norm = sqrt(norm)
    for each page p in G do  // then update all hub values
      p.hub = p.hub / norm   // normalise the hub values
```

Let x represent authority score and y represent hub score.

$x_1(A) = y_0(B) + y_0(C) + y_0(D) = 0.2 + 0.2 +0.2 = 0.6$

$x_1(B) = y_0(D) = 0.2$

$x_1(C) = y_0(B) + y_0(D) = 0.2 + 0.2 = 0.4$

$x_1(D) = 0$

$y_1(A) = 0$

$y_1(B) = x_1(A) + x_1(C) = 0.6 + 0.4 = 1$

$y_1(C) = x_1(A) = 0.6$

$y_1(D) = x_1(A) + x_1(B) + x_1(C) = 0.6 + 0.2 + 0.4 = 1.2$

$x_2(A) = y_1(B) + y_1(C) + y_1(D) = 1 + 0.6 + 1.2 = 2.8$

$x_2(B) = y_1(D) = 1.2$

$x_2(C) = y_1(B) + y_1(D) = 1 + 1.2 = 2.2$

$x_2(D) = 0$

$y_2(A) = 0$

$y_2(B) = x_2(A) + x_2(C) = 2.8 + 2.2 = 5$

$y_2(C) = x_2(A) = 2.8$

$y_2(D) = x_2(A) + x_2(B) + x_2(C) = 0.6 + 0.2 + 0.4 = 6.2$

3. Memory-based Collaborative Filtering.

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

	i_1	i_2	i_3	i_4	i_5	i_6	Average
u_1	0	1	5	3	1	0	$\frac{3+1}{2}=2$
u_2	3	5	4	3	0	2	
u_3	3	0	0	4	2	2	$\frac{4+2}{2}=3$
u_4	3	0	4	5	5	3	
u_5	1	3	5	0	2	2	

$$PCC(u_3,u_1) = \frac{\left(3-\frac{4}{2}\right)*\left(4-\frac{6}{2}\right)+\left(1-\frac{4}{2}\right)*\left(2-\frac{6}{2}\right)}{\sqrt{\left(3-\frac{4}{2}\right)^2*\left(1-\frac{4}{2}\right)^2}*\sqrt{\left(4-\frac{6}{2}\right)^2*\left(2-\frac{6}{2}\right)^2}} = 1.0$$

$$PCC(u_3,u_2) = 0.866$$

$$PCC(u_3,u_3) = 1.0$$

$$PCC(u_3,u_4) = 0.302$$

$$PCC(u_3,u_5) = -1.0$$

, from which we know u_1 and u_2 are the top 2 most similar users of u_3.

	i_1	i_2	i_3	i_4	i_5	i_6	Average
u_1	0	1	5	3	1	0	$\frac{1+5+3+1}{4}=2.5$
u_2	3	5	4	3	0	2	$\frac{3+5+4+3+2}{5}=3.4$
u_3	3	0	0	4	2	2	$\frac{3+4+2+2}{4}=2.75$
u_4	3	0	4	5	5	3	
u_5	1	3	5	0	2	2	

$$r_{u,i} = \bar{r_u} + k \sum_{u' \in U} simil(u,u')(r_{u',i} - \bar{r_{u'}})$$

$$r_{32} = \overline{r_3} + \frac{PCC(u_3,u_2)*(r_{22}-\overline{r_2})+PCC(u_3,u_1)*(r_{12}-\overline{r_1})}{PCC(u_3,u_2)+PCC(u_3,u_1)} = 2.75+\frac{0.866*(5-3.4)+1*(1-2.5)}{0.866+1}=2.689$$

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

$$simil(x,y) = cos(\vec{x},\vec{y}) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}||_2 \times ||\vec{y}||_2} = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_{xy}} r_{x,i}^2}\sqrt{\sum_{i \in I_{xy}} r_{y,i}^2}}$$

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

$$Cosine(u_3,u_1) = \frac{3*4+1*2}{\sqrt{3^2+1^2}*\sqrt{4^2+2^2}} = 0.990$$

$$Cosine(u_3,u_2) = 0.990$$

$$Cosine\ (u_3,u_3) = 1.0$$

$$Cosine\ (u_3,u_4) = 0.950$$

$$Cosine\ (u_3,u_5) = 0.889$$

, from which we know u_1 and u_2 are the top 2 most similar users of u_3.

$$r_{32} = \bar{r_3} + \frac{PCC(u_3,u_2)*(r_{22}-\bar{r_2})+PCC(u_3,u_1)*(r_{12}-\bar{r_1})}{PCC(u_3,u_2)+PCC(u_3,u_1)} = 2.75 + \frac{0.990*(5-3.4)+0.990*(1-2.5)}{0.990+0.990} = 2.8$$

(3) find top 2 most similar items of i_5 and estimate u_2's rating on i_5 using PCC similarity and item-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,

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

$$PCC(i_5,i_1) = \frac{\left(1-\frac{1+3}{2}\right)*\left(1-\frac{1+2}{2}\right)+\left(3-\frac{1+3}{2}\right)*\left(2-\frac{1+2}{2}\right)}{\sqrt{\left(1-\frac{4}{2}\right)^2*\left(3-\frac{4}{2}\right)^2}*\sqrt{\left(1-\frac{3}{2}\right)^2*\left(2-\frac{3}{2}\right)^2}} = 0.5$$

$$PCC(i_5,i_2) = 1.0$$

$$PCC(i_5,i_3) = -0.971$$

$$PCC(i_5,i_4) = 0.961$$

$$PCC(i_5,i_5) = 1.0$$

$$PCC(i_5,i_6) = 1.0$$

, from which we know i_6 and i_2 are the top 2 most similar items of i_5.

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

$$r_{ai} = \frac{\sum_j s_{ij} r_{aj}}{\sum_j s_{ij}}$$

$$r_{25} = \frac{PCC(i_5,i_6)*r_{26}+PCC(i_5,i_2)*r_{22}}{PCC(i_5,i_6)+PCC(i_5,i_2)} = \frac{1*2+1*5}{1+1} = 3.5$$

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

$$Cosine(i_5,i_1) = 0.919$$

$$Cosine\ (i_5,i_2) = 0.990$$

$$Cosine\ (i_5,i_3) = 0.787$$

$$Cosine\ (i_5,i_4) = 0.930$$

$$Cosine\ (i_5,i_5) = 1.0$$

$$Cosine\ (i_5,i_6) = 0.971$$

, from which we know i_6 and i_2 are the top 2 most similar items of i_5.

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

$$r_{25} = \frac{PCC(i_5,i_6)*r_{26}+PCC(i_5,i_2)*r_{22}}{PCC(i_5,i_6)+PCC(i_5,i_2)} = \frac{0.971*2+0.990*5}{0.971+0.990} = 3.515$$

Betweenness and Closeness solution in HW#1:
(Please note we don't use this definition in mid-term. We will use another definition talked in tutorial)

(1) Find the closeness of vertex D and F in the graph.

For D: ABDF ACDF ABDFG ACDFG BDC BDE BDF BDFG CDF CDFG
So the closeness of vertex D is 10.

For F: ACEFG ABDFG ACDFG BDFG CEFG CDFG DFG EFG

So the closeness of vertex F is 8.

(2) Find the betweenness of vertex D and F in the graph.

The betweenness of vertex D is 63.

The betweenness of vertex F is 41.

For vertex D:

From A	ACDB ACEDB ACEFDB ABDC ABDEC ABDFEC ACDE ACDFE ABDCE ABDE ABDFE ABDF ABDCEF ABDEF ACDF ACDEF ACEDF ABDFG ABDCEFG ABDEFG ACDEFG ACDFG ACEDFG
From B	BDC BDEC BDFEC BACDE BACDFE BDCE BDE BDFE BACDF BACEDF BACDEF BDF BDCEF BDEF BACDFG BACDEFG BACEDFG BDCCEFG BDEFG BDFG
From C	CABDE CABDFE CDE CDFE CABDF CABDEF CDF CDEF CEDF CABDFG CABDEFG CDEFG CDFG CEDFG
From D	ECABDF ECDF EDF ECABDFG ECDFG EDFG

For vertex F:

From A	ACEFDB ABDFEC ACEFD ACDFE ABDFE ACEFG ABDFG ABDCEFG ABDEFG ACDEFG ACDFG ACEDFG
From B	BDFEC BACEFD BACDFE BDFE BDFG BDCEFG BDEFG BACDFG BACDEFG BACEDFG BACEDG
From C	CEFD CABDFE CDFE CABDFG CABDEFG CDEFG CDFG CEDFG CEFG
From D	DFE DFG DBACEFG DCEFG DEFG
From E	EFG EDFG ECDFG ECABDFG