CMSC5733 Social Computing

Exercise 2

Deadline: 23:59:59, October 24(Thursday), 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.

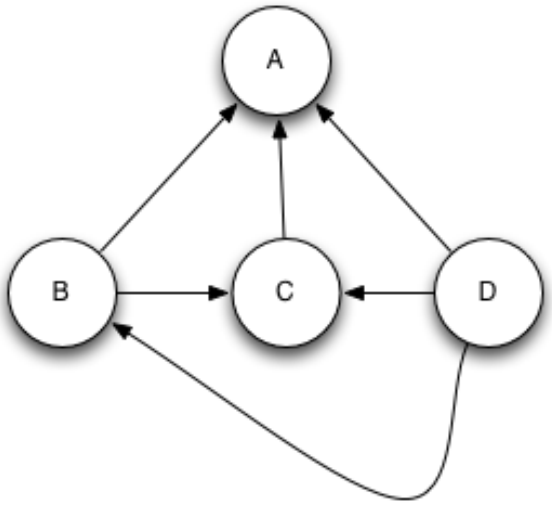
(1) Draw the network.

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

(3) Show the minimum cut.

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. (Note: please refer to the PageRank definition in tutorial notes)

(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.

Note: please also refer to PageRank in http://en.wikipedia.org/wiki/PageRank or http://beowulf.csail.mit.edu/18.337-2012/MapReduce-book-final.pdf. And please with special attention to dangling nodes. About hits algorithm, please refer to http://en.wikipedia.org/wiki/HITS_algorithm.

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, Cosine 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.

(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.

(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.

(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.

Note: please also refer to the definition in http://en.wikipedia.org/wiki/Collaborative_filtering, and notice the trick of average calculation. When calculating prediction, please use the following equation in wikipedia.

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