# CSCI4022 Midterm 2 Review F21

These problems form a mini-exam. The exam problems will be similar, and may also include short answers or T/F type questions.

1. Recommendation Systems

   a) You want to design a recommendation system for an online bookstore that has been launched recently. The bookstore has over 1 million book titles, but its rating database has only 10,000 ratings. Which of the following would be a better recommendation system?
   i) User-user collaborative filtering
   ii) Item-item collaborative filtering
   iii) Content-based recommendation.
   In One sentence justify your answer.

   b) Suppose the bookstore is using the recommendation system you suggested above. A customer has only rated two books: "Linear Algebra" and "Differential Equations" and both ratings are 5 out of 5 stars. Which of the following books is less likely to be recommended?
   i) "Operating Systems"
   ii) "A Tale of Two Cities"
   iii) "Convex Optimization"
   iv) It depends on other users' ratings.

   c) After some years, the bookstore has enough ratings that it starts to use a more advanced recommendation system like the one won the Netflix prize. Suppose the mean rating of books is 3.4 stars. Alice, a faithful customer, has rated 350 books and her average rating is 0.4 stars higher than average users' ratings. Animals Farm, is a book title in the bookstore with 250,000 ratings whose average rating is 0.7 higher than global average. What would be a baseline estimate of Alice's rating for Animal Farms?

   d) Consider the following utility matrix:

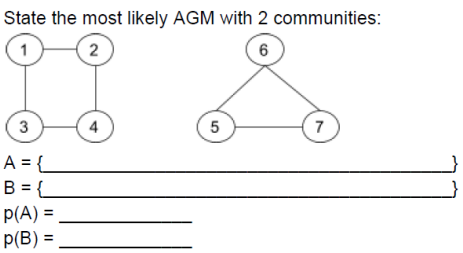
| | $User\,1$ | $User\,2$ | $User\,3$ | $User\,4$ | $User\,5$ |
|---|---|---|---|---|---|
| $Item\,1$ | 4 | 4 | 4 | 1 | 1 |
| $Item\,2$ | 3 | 1 | | 4 | |
| $Item\,3$ | 4 | 2 | | 2 | 3 |
| $Item\,4$ | | 2 | 3 | | 1 |
| $Item\,5$ | | | 1 | 4 | 3 |
| $Item\,6$ | 1 | 1 | | | 2 |

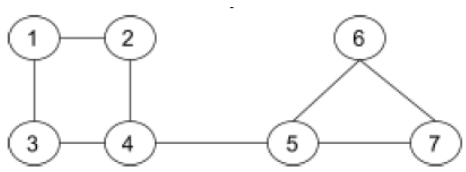
   (i) Assume that we use the Pearson correlation coefficient as the similarity measure and that we predict a rating by averaging the two nearest (most similar) neighbors. Which two users do we use to predict the rating of Item 4 by User 1?

   (ii) What is this predicted rating?

   (iii) What is the correlation coefficient for each?

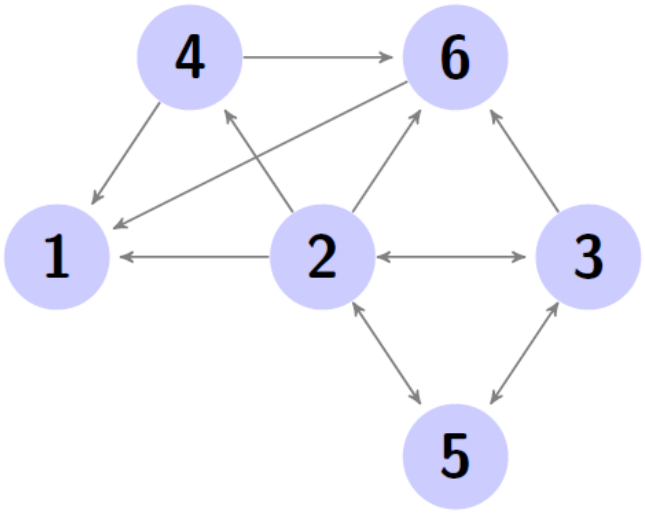
2. On Social Graphs.

   (a)
   
   

   (b)
   
   
   What would be the *cut* value for the optimal cut of this graph?

(c) Suppose we wanted to answer 2b formally by creating adjacency $A$, degree $D$, and Laplacian $L$ matrices. What is the $L$ matrix? What computations we would take to solve 2b? Make an approximate depiction of the plot we would make to determine where to cut.

3. On Directed Graphs. Consider the following graph:




(a) Set up the PageRank balanced-state equations, assuming $\beta = 0.8$ or a jump probability of 0.2. Denote the rank of node $a$ with $r(a)$. Use the variant that forces teleportation at dead ends.

(b) Order the PageRanks for nodes $1, 2$ and 4, from lowest to highest. You do not need to compute their actual values, your intuition will suffice.

(c) Set up the hubs and authorities algorithm on the graph $G$.

(d) What is the outcome of running a PageRank power iteration on with no teleporting and no normalization? What will it converge to?

4.

$$M = \begin{array}{c} \\ Diane \\ Ethan \\ Frank \\ Grace \\ Hank \\ Ingrid \\ Joe \end{array} \begin{array}{c} \begin{array}{ccccc} 4022 & ML & LinAlg & HCI & WebDesign \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

$M$ represents the ratings of courses taken by some CSCI courses. Each row is a student's ratings ,each column a course. The SVD for M is found to be:

$$USV^T = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

Looks like there are two clear concepts of classes: mathy and data classes against graphic and design classes.

Suppose a new students named Tony has the following reviews: 4 for ML, 5 for LinAlg, 2 for Web Design.

(a) What is the representation of Tony in concept space?

(b) What does this representation predict about how much Tony would like 4022 and HCI?

(c) Another student, Zach, has reviews of 5 for 4022, 2 for ML, 4 for HCI, and 5 for WebDesign. What is the representation for Zach in concept space?

(d) What is the cosine similarity for Tony and Zach using their concept space vectors?

n. A few things that are not here but worth considering for the exam:

(a) A-priori: review first exam sol'n...

(b) BigCLAM: how do we take one step? Especially for a single $F_{u,A}$ element?

(c) PCA and UV decompositions, plus our element-wise optimization scheme for UV.

(d) How do random forests work?