1. (20 points; from MIW) Consider a simple toy model of recommendation for simulating large sparse binary datasets with \( n \) rows (users) and \( m \) columns (items). The model operates as follows:

- Each user acts independently of all other users.
- For a given user \( i \), he or she considers each of the \( m \) items in turn, and independently makes a decision with probability \( p \) as to whether to purchase the item or not. If they decide to purchase the item, then they enter a ‘1’ for that item, otherwise a zero.
- We set \( p = \frac{k}{m} \) where \( k \) is a parameter of the model and denotes the expected number of items purchase by each user. Typically \( k \) is much smaller than \( m \), e.g., \( k = 4, m = 1000 \).

Calculate the probability, as a function of \( k \), \( m \), and \( n \), that a new user acting under the same model will have no items at all in common with any of the \( n \) users in an \( n \times m \) matrix generated under the model above. Plot this probability for different values of \( k \), \( m \), and \( n \) and make observations.

2. (20 points) Repeat the above exercise, except that items are not purchased with uniform probability. Instead a power-law distribution holds on the number of purchases when the items are ordered by the purchase ranks. In other words, if we plot the number of purchases \( y \) of a given item as a function of the frequency rank \( x \) of items (i.e., the most frequently ranked item gets a rank of 1, the next frequently ranked item is ranked 2, and so on), we get a relationship of the form: \( y \propto x^{-\alpha} \). This function models a ‘rich get richer’ effect in purchase patterns. Make comparisons with this model and the previous model. Research the literature to determine suitable values of \( \alpha \).

3. (10 points) Read the [Herlocker et al. 1999] paper and consider equation (6) in the ‘producing a prediction’ section. Describe a rating model (similar to the questions above) where this equation would produce ‘perfect’ recommendations. Explain what ‘perfect’ means here.

4. (50 points) Read the [Hofmann and Puzicha, 1999] paper; in class, we developed the equations given in Section 2.2. Continue this saga and develop EM-style equations for the models given in Fig. 1, part (d) and part (f) (we described these in class). Then, present two small synthetic datasets, one where (d) is a better model than (f), and another where (f) is better. For this question, you are expected to implement the EM updates in some language of your choice and unleash it on your chosen datasets and present concrete, quantitative results.