MATHEMATICAL MODEL IN MARKETING

EXAM (DECEMBER 3, 2017)

This is a take-home exam. Your answers must be your own work: you cannot search the web for answers, or consult other students.

Please return your *typeset* answers (preferably using LaTex) to Rajeev Kohli or Khaled Boughanmi by Monday, December 11, 2017.

Question 1 (100 points)

Facebook offers its users a customized newsfeed that includes status updates, photos, videos, links, app activity and likes from people, pages and groups that a user follows on the social network. Facebook uses multiple algorithms to evaluate the relevance of the content for a user. Each algorithm uses data from different sources, such as past behavior (e.g., likes and dislikes), surveys in which users compare pairs of side-by-side posts, and information from a "feed quality panel." The recommendations of the individual algorithms are combined using a final ranking algorithm that determines which stories appear and their positions in a user's newsfeed.[1]

Suppose each of m algorithms ranks n posts/stories in decreasing relevance order for a Facebook user. The *linear ordering problem* (LOP) is to combine the rankings of the m algorithms into a single ranking that best represents them all. One criterion that is often used to combine the rankings maximizes the number of *non-reversals* over all pairs of stories, summed across all algorithms. A non-reversal occurs if both the combined ordering and the ordering provided by an algorithm rank story j higher than story k; a reversal occurs if one of these ranks j higher than k, and the other ranks k higher than j. For example, suppose there are three news stories, 1, 2, and 3, and two ranking algorithms. Algorithm 1 ranks the stories in the order $1-2-3$, and algorithm 2 ranks them in the order $2-3-1$ (a lower rank means the story is considered to be more relevant for a user). Consider a candidate combined ranking, $1-3-2$. There are two non-reversals between the candidate ranking and the ranking given by algorithm 1: both rank 1 before 2, and 1 before 3 (but one ranks 2 before 3, and the other 3 before 2). And there is no non-reversal between the candidate ranking and the ranking given by algorithm 2. Thus, the total number of non-reversals for this candidate ranking is 2+0=2.

[1]Twitter uses a similar method to compose "while you were away" lists of tweets; and search engines combine the relevance rankings of the results to a query obtained using multiple search algorithms.

1. Formulate the LOP as a 0-1 integer program. Explicitly specify the decision variables, the objective function and the constraints. Explain your formulation. How many constraints does it have as function of m and n? (25 points)

2. LOP is known to be NP Hard. What does this imply about the problem? (5 points)

3. Associate a (person specific) utility $u_k = v_k + \epsilon_k$ with story k. Assume that each ϵ_k has an independent, extreme value distribution, with density function $f_k(\epsilon_k) = \exp(-\epsilon_k e^{-\epsilon_k})$ and cumulative distribution function $F_k(\epsilon_k) = \exp(-e^{-\epsilon_k})$. Give an expression for the probability of obtaining a reversal. Use this probability to specify an expression for the expected number of non-reversals across newsfeed algorithms. (25 points)

4. Show that the solution to the problem of maximizing the expected number of reversals is equivalent to finding the optimal solution to the discrete linear ordering problem. (20 points)

5. Redo question 3, except that instead of maximizing the expected number of non-reversals, use the objective of maximizing a likelihood function. What is the interpretation of this formulation? Is it a convex optimization problem? Explain. (25 points).

Question 2 (100 points)

Consider a group of n people, each of whom owns a different house. Let h_i denote person i's house, $i = 1, \ldots, n$. Assume that each person has a strict preference ordering over all n houses. For example, person i may have the preference ordering $h_2 \succ_i h_1 \succ_i h_3$, which means that person i likes person 2's house the most, then his/her own house, and then person 3's house.

Our interest is in the possible re-assignment of houses to people. Define a *coalition* to be a subgroup of the n people who can reassign houses among themselves in a way that makes each of them better off. An assignment of houses to people is *coalition proof* if there is no such coalition. An assignment of houses to people is *stable* if it is coalition-proof.

1. Devise a polynomial-time algorithm that produces a stable assignment of the n houses to the n people. Prove that the algorithm terminates and obtains a stable assignment. (60 points)

2. Is it a dominant strategy for each person to truthfully reveal his or preferences? If yes, give a proof. If no, give an example. (40 points)

QUESTION 3 (100 POINTS)

Twitter is a social network with one-sided relations: one person can follow another without being followed back. If person a is a follower of person b, then b is a leader, but not necessarily a follower, of a. Followers receive tweets from their leaders, which they can retweet to their own followers. Table 1 provides selected summary statistics on the average numbers of leaders and followers; the average number of followers of leaders; and the average number of leaders of followers on Twitter.

TABLE 1. Average number of followers, leaders, followers of leaders and leaders of followers for Twitter

Average number of:	
Followers	18.67
Leaders	18.67
Followers of leaders	$1,252.31$
Leaders of followers	$281,107.65$

The first two lines in Table 1 give the average number of followers per person, and the average number of leaders per person.

The third line in Table 1 shows the average number of followers of leaders ($= 1,252.31$). It is computed as follows. Let $L(u)$ denote the set of leaders of a Twitter user u. Let $F(l)$ denote the set of followers of $l \in L(u)$. Then

$$A = \frac{\sum_u \sum_{l \in L(u)} |F(l)|}{\sum_u |L(u)|}$$

is the average number of followers per leader. The third line of Table 1 reports the value of A.

The fourth line in Table 1 shows the average number of leaders of followers ($= 281,107.65$). It is computed as follows. Let $F(u)$ denote the set of followers of a Twitter user u. Let $L(f)$ denote the set of leaders of follower $f \in F(u)$. Then

$$B = \frac{\sum_u \sum_{f \in F(u)} |L(f)|}{\sum_u |F(u)|}$$

is the average number of leaders per follower. The fourth line of Table 1 reports the value of B.

Answer the following questions.

1. In Table 1, the average number of followers is equal to the average number of leaders. Show that this must be always true. (10 points)

2. The average number of followers of leaders (1,252.31) is much higher than the average number of followers (18.67). That is, on average a person's tweet

goes out to less than 19 people, but on average a tweet sent by a person's leaders goes out to over 1,000 people. Explain how this could be related to the fact that some celebrities, like Katy Perry, Justin Beiber and Barack Obama, have close to 100 million followers each on Twitter. (Twitter had $n = 330$ millions active users in November 2017. Twitter counter gives current information on the people who have the most followers and the most leaders; see https://twittercounter.com/pages/100.) (30 points)

3. The average number of leaders of followers (281,107.65) is much higher than the average number of leaders (18.67). That is, if you were an average Twitter user, you would receive tweets from about nineteen people, but each of your followers would receive tweets from (on average) more than a quarter million people. Explain how this result could be related to the fact that some people, like MarQuis Trill and Megamix Champion, follow millions of people on Twitter. (30 points)

4. The average number of leaders of followers (281,107.65) is much greater than the average number of followers of leaders (1,252.31). What might explain this large discrepancy in the two averages? What is the interpretation/implication of this discrepancy? (30 points)