COL866: Special Topics in Algorithms	Fall 2023
Assignment 4	
Total points: 100	Deadline: Nov 17 (Friday)

1. [**20 points**] Recall that computing a ranking that minimizes the Kemeny score is NP-hard. Also recall that we saw a deterministic 2-approximation algorithm using the Footrule distance. In this exercise, we will design a *randomized* algorithm with the same approximation factor.

   For any natural number $m \in \mathbb{N}$, let $[m] \coloneqq \{1, 2, \ldots, m\}$ and $\Pi([m])$ be the space of all permutations (or rankings) on the set $[m]$. Say we are given a set $S = \{\pi_1, \pi_2, \ldots, \pi_n\}$ consisting of n rankings over a set of candidates $[m]$ and any ranking $\tau \in \Pi([m])$ not necessarily in S. The Kemeny score of τ with respect to S is defined as $\mathtt{d}^{\mathtt{Kt}}(\tau, S) \coloneqq \sum_{k \in [n]} \mathtt{d}^{\mathtt{Kt}}(\tau, \pi_k)$, where $\mathtt{d}^{\mathtt{Kt}}(\cdot)$ denotes the Kendall's tau distance between two rankings.

   Consider a randomized algorithm `ALG` whose output is ranking drawn uniformly at random from S. Show that `ALG` is a randomized 2-approximation algorithm. That is,

   $$\mathbb{E}_{\pi \sim \text{Unif}(S)}[\mathtt{d}^{\mathtt{Kt}}(\pi, S)] \leqslant 2 \cdot \mathtt{d}^{\mathtt{Kt}}(\sigma^*, S),$$

   where σ^* is a Kemeny optimal ranking for S, i.e.,

   $$\sigma^* \in \arg\min_{\sigma \in \Pi([m])} \mathtt{d}^{\mathtt{Kt}}(\sigma, S).$$

   Note that σ^* may not belong to the set S.

   For any pair of candidates $i, j \in [m]$, let $w_{i,j} \coloneqq \frac{1}{n} \cdot |\{\pi \in S : i \text{ is ranked above } j \text{ in } \pi\}|$ denote the *fraction* of rankings in S where i is preferred over j, and $w_{j,i}$ analogously denote the fraction of rankings in S where j is preferred over i. Note that $w_{i,j}, w_{j,i} \in [0, 1]$ and $w_{i,j} + w_{j,i} = 1$.

   Let σ^* be a Kemeny optimal ranking for S, and suppose that under σ^*, candidate i is preferred over candidate j.

   Consider a randomized algorithm `ALG` that picks a ranking in S uniformly at random. Then, the probability that the ranking chosen by `ALG` agrees with σ^* for the candidate pair (i, j) is $w_{i,j}$. Likewise, the probability it disagrees is $w_{j,i}$.

   If the ranking chosen by `ALG` agrees (respectively, disagrees) with σ^* over the pair (i, j), then the 'cost' incurred by `ALG` in terms of the Kendall's tau objective is $w_{j.i} \cdot n$ (respectively, $w_{i,j} \cdot n$); this is the number of rankings in S that disagree with the choice of `ALG` for the pair (i, j).

By linearity of expectation, the expected Kemeny score (equivalently, the expected cost) of ALG is given by:

$$\begin{aligned}
\mathbb{E}[\mathtt{d}^{\mathtt{Kt}}(\mathtt{ALG}, S)] &= \sum_{k \in [n]} \frac{1}{n} \cdot \mathbb{E}[\mathtt{d}^{\mathtt{Kt}}(\mathtt{ALG}, \pi_k)] \\
&= \frac{1}{n} \cdot \sum_{k \in [n]} \sum_{(i,j) \in [n] \times [n]} \mathbb{E}[\mathbf{1}(i \succ j \text{ in } \pi_k \text{ and } j \succ i \text{ in } \mathtt{ALG})] \\
&\text{where } \mathbf{1}(\cdot) \text{ is the indicator variable} \\
&= \frac{1}{n} \cdot \sum_{(i,j) \in [n] \times [n]} \sum_{k \in [n]} \Pr(i \succ j \text{ in } \pi_k \text{ and } j \succ i \text{ in } \mathtt{ALG}) \\
&= \frac{1}{n} \sum_{(i,j) \in [n] \times [n]} w_{i,j} \cdot w_{j,i} \cdot n \\
&\leqslant \sum_{(i,j) \in [n] \times [n]} \min\{w_{i,j}, w_{j,i}\} \\
&\text{which holds because } w_{i,j}, w_{j,i} \in [0,1] \\
&= \sum_{(i,j) \in [n] \times [n] :\, i \succ j \text{ in } \sigma^*} \min\{w_{i,j}, w_{j,i}\} + \sum_{(i,j) \in [n] \times [n] :\, j \succ i \text{ in } \sigma^*} \min\{w_{i,j}, w_{j,i}\} \\
&\leqslant \mathtt{d}^{\mathtt{Kt}}(\sigma^*, S) + \mathtt{d}^{\mathtt{Kt}}(\sigma^*, S) \\
&= 2 \cdot \mathtt{d}^{\mathtt{Kt}}(\sigma^*, S).
\end{aligned}$$

Thus, ALG is a randomized 2-approximation algorithm for Kemeny score.

2. [**20 points**] Consider the problem of determining whether, given a preference profile and a nonnegative integer k, there exists a ranking with Kemeny score at most k. Design an $\mathcal{O}(2^k \cdot \text{poly}(n, m))$ algorithm for this problem, where n is the number of the voters and m is the number of candidates.

(Proof sketch.) The desired running time can be achieved by a *branching* algorithm.

For any candidate pair (a, b), let $R_{a,b}$ denote the set of rankings in the given preference profile where a is ranked above b, and similarly let $R_{b,a}$ denote the set of rankings in the profile where b is ranked above a.

Let D ("the disagreement set") denote the set of unordered candidate pairs $\{a, b\}$ for which both $R_{a,b}$ and $R_{b,a}$ are nonempty; we will call any such pair of candidates a *conflicting* pair. Thus, any ranking will incur a hit of at least 1 in its Kemeny score for each conflicting pair. Note that the set D can be constructed in $\mathcal{O}(\text{poly}(n, m))$ time.

If there are more than k conflicting pairs (i.e., if $|D| > k$), our algorithm can safely return NO, as any ranking will have a Kemeny score of at least $|D|$. Thus, we will assume from here onwards that $|D| \leqslant k$.

For each conflicting pair $\{a, b\}$, our algorithm will consider the two possibilities of ordering it, namely, whether to rank a above b or b above a. This branching strategy can be visualized as a complete binary tree with $|D|$ levels, wherein each level corresponds to a unique conflicting pair, and the left and right branches at each node correspond to the two ordering choices. Each leaf of this tree corresponds to a fixed choice for each conflicting pair. Note that the set of assignments leading up to a leaf node may not induce a transitive ordering.

By the unanimity property of Kemeny rule, we can assume that the candidate pairs not included in D are ordered in the same way as in the given preference profile.

There are $2^{|D|}$ (hence, at most 2^k) leaf nodes in total. For each leaf node, the algorithm checks whether the resulting assignment of the conflicting pairs (together with the natural assignment of the non-conflicting pairs) induces a valid ranking. If so, the algorithm checks whether the Kemeny score of the induced ranking is at most k. If, for any leaf node, there exists such a ranking, the algorithm returns YES, otherwise it returns NO.

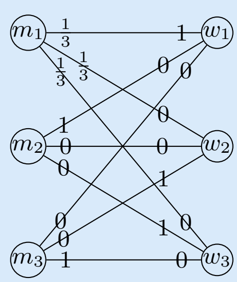
3. [**20 points**] Consider the following two-sided matching instance:

$$
\begin{array}{llll}
m_1\colon & w_3 & w_2 & w_1 \\
m_2\colon & w_1 & w_2 & w_3 \\
m_3\colon & w_3 & w_1 & w_2
\end{array}
\qquad\qquad
\begin{array}{llll}
w_1\colon & m_1 & m_3 & m_2 \\
w_2\colon & m_3 & m_2 & m_1 \\
w_3\colon & m_2 & m_1 & m_3
\end{array}
$$

What is the distortion of the men-proposing and women-proposing deferred-acceptance algorithms on the above instance? Note that the optimal (i.e., utilitarian welfare maximizing) matching does not have to be stable.

The men-proposing outcome is $\mu_1 := (m_1, w_3), (m_2, w_2), (m_3, w_1)$ and the women-proposing outcome is $\mu_2 := (m_1, w_1), (m_2, w_3), (m_3, w_2)$.

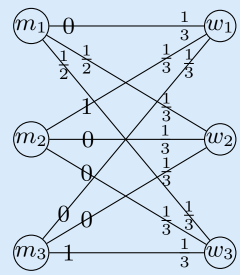
The welfare loss of the men-proposing algorithm is maximized for the following cardinal utility profile:



and the corresponding welfare loss (or distortion) is:

$$\frac{\mathcal{W}(\{(m_1, w_1), (m_2, w_3), (m_3, w_2)\})}{\mathcal{W}(\mu_1)} = \frac{\frac{1}{3} + 3}{\frac{1}{3}} = 10.$$

The welfare loss of the women-proposing algorithm is maximized for the following cardinal utility profile:



and the corresponding welfare loss (or distortion) is:

$$\frac{\mathcal{W}(\{(m_1, w_2), (m_2, w_1), (m_3, w_3)\})}{\mathcal{W}(\mu_2)} = \frac{1 + 1 + \frac{1}{2} + 1}{1} = 3.5.$$

4. [**20 points**] Consider the approval-based multiwinner voting problem. Here, we are given a set C of m candidates, a set V of n voters, and a positive integer k. Each voter $i \in V$ approves a subset $A_i \subseteq C$ of the candidates. The goal is to find a committee $W \subseteq C$ of k candidates (i.e., $|W| = k$) that satisfies some notion of representation. We will focus on the notion called *justified representation* (JR) which says that no large, cohesive group of voters should be unrepresented. Formally, a k-sized committee $W \subseteq C$ satisfies JR if there is no subset of voters $S \subseteq V$ with $|S| \geqslant n/k$ and $\cap_{i \in S} A_i \neq \emptyset$ such that W contains no candidate from $\cup_{i \in S} A_i$.

   Prove that if each candidate is approved by at least one voter (i.e., for every candidate $c \in C$, there exists some voter $i \in V$ such that $c \in A_i$), then there are at least $m - k + 1$ committees of size k each that satisfy JR.

Starting with an empty committee, consider a greedy procedure that, at each step, includes a candidate in the existing committee that is approved by the largest set of "uncovered" voters, i.e., voters for whom none of the approved candidates has been included in the committee so far.

Notice that after k steps, the committee constructed by this procedure, say $c_1, c_2, \ldots, c_k$, satisfies JR.

Now consider the execution of the greedy procedure for the first $k - 1$ steps (i.e., the committee is $W = \{c_1, \ldots, c_{k-1}\}$). If all voters are covered by W, then it already satisfies justified representation. There are $m - (k - 1)$ choices for the k^{th} member of the committee, and thus there are at least $m - k + 1$ committees that satisfy JR.

Otherwise, suppose W does not cover all the voters. Thus, there must be a group of at least $\frac{n}{k}$ uncovered voters, say U. By the greedy selection rule, each candidate c_i in W must cover at least $\frac{n}{k}$ voters that are not covered by $c_1, \ldots, c_{i-1}$. This implies that the *voter block* for each candidate c_i (i.e., the set of voters who approve c_i) must be *disjoint*. Let us call these voter blocks $B_1, B_2, \ldots, B_{k-1}$.

Thus, none of the voters in the remaining uncovered set U approve any candidate in W. We know that the greedy procedure successfully computes a JR committee in k steps. Thus, there must be a candidate c_k that covers all voters in U. Therefore, we can consider U also as a "block" $B_k = U$.

By assumption, each of the remaining $m-k$ candidates (other than $c_1, \ldots, c_k$) must be approved by at least one voter in the disjoint sets $B_1, B_2, \ldots, B_k$. If that voter belongs to B_i, we replace the corresponding candidate c_i with this external candidate, and obtain $m-k$ different committees satisfying JR. Furthermore, $W \cup c_k$ also satisfies JR. This gives the desired bound.

Acknowledgments: This problem was based on the SAGT 2022 paper titled "Justifying Groups in Multiwinner Approval Voting" by Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. See Theorem 2 in the full version at `https://arxiv.org/pdf/2108.12949.pdf`.

5. [**20 points**] In this exercise, we will think about the multiwinner voting problem where the voters' preferences are given not as *approvals* but as *rankings*. Let $C = \{c_1, c_2, \ldots, c_m\}$ denote the set of m candidates, and $V = \{v_1, v_2, \ldots, v_n\}$ denote the set of n voters. Each voter v_i has a strict and complete ranking R_i over the candidates in C.

   Consider the following voting rule f for selecting a committee of k candidates: Given any k-sized committee of candidates $W \subseteq C$, the score that voter v_i assigns to the committee W is equal to $m - j$ if voter v_i's *favorite* candidate in W is ranked at j^{th} position in its ranking R_i. The score of the committee W is the *minimum* of the scores it receives from all voters. The voting rule returns a committee with the highest score.

   For example, suppose there are three voters v_1, v_2, v_3 and five candidates $c_1, \ldots, c_5$. The rankings of the voters are $v_1 : c_1 \succ c_2 \succ c_3 \succ c_4 \succ c_5$, $v_2 : c_4 \succ c_2 \succ c_1 \succ c_5 \succ c_3$, and $v_3 : c_5 \succ c_4 \succ c_3 \succ c_2 \succ c_1$. Then, the committee $\{c_2, c_5\}$ gets a score of 3 from voter v_1 because its favorite candidate in the committee, namely c_2, is ranked second. Likewise, the committee gets scores of 3 and 4 from voters v_2 and v_3, respectively, resulting in an overall score of $\min\{3, 3, 4\} = 3$.

   Show that when the rankings $R_1, \ldots, R_n$ are single-peaked (with respect to a known axis σ), the output of the voting rule f can be computed in polynomial time.

We will provide an algorithm that, for any given threshold $\theta \in \mathbb{Q}$, determines whether there exists a committee with score at least θ. By iterating over all values of θ between 1 and $m-1$, such an algorithm can be used to evaluate the given voting rule f.

Towards this goal, we will use the contiguous segments property, which, as discussed in class, is equivalent to single-peakedness. Recall that the contiguous segments property says that for any k, the set of top k favorite candidates of any voter constitute a connected set with respect to the given axis σ.

Note that there exists a committee with score at least θ if and only if there exists one where each voter derives a utility of at least θ. This, in turn, happens if the committee contains a candidate from the top $(m-\theta)$ candidates in each voter's ranking. By the contiguous segments property, the set of top $(m-\theta)$ candidates for each voter constitute a connected segment with respect to the axis σ. Thus, the task of determining the existence of a committee with score at least θ turns out to be equivalent to finding a "hitting set" of size k with respect to the voters' intervals. The latter problem admits a well-known greedy algorithm.