A Missing Proof from Section 3

Proof of Theorem 4. Recall that $x = (x_1, x_2, \dots, x_\ell)$ is defined as *group representation* where x_j is the number of ranks assigned to group j for all $j \in [\ell]$, and $y = (y_1, y_2, \dots, y_k)$ is defined as *group assignment* where y_i is the group assigned to rank i for all $i \in [k]$.

For Axiom 1 to be satisfied, the distribution should consist only of rankings where the items from the same group are ranked in the order of their merit. Clearly $\mathcal{D}$ satisfies Axiom 1.

To satisfy Axiom 2 all the group fair representations need to be sampled uniformly at random, and all the non-group fair rankings need to be sampled with probability zero. Hence, $\mathcal{D}$ also satisfies Axiom 2.

We now use strong induction on the prefix length i to show that any distribution over group assignments that satisfies Axiom 3 has to sample each group assignment y, conditioned on a group representation x, with equal probability. We note that whenever we say common prefix, we refer to the longest common prefix.

Induction hypothesis. Any two rankings with a common prefix of length i, for some $0 \leqslant i \leqslant k-2$, have to be sampled with equal probability.

Base case ($i = k-2$). Let y and y' represent a pair of group assignments with fixed group representation x and common prefix till ranks $k-2$. Then there exist exactly two groups $j, j' \in [\ell]$ such that

$$y_{k-1} = y'_k = j \quad \text{and} \quad y_k = y'_{k-1} = j'.$$

Therefore, to satisfy Axiom 3, these two group assignments y and y' need to be sampled with equal probability. Therefore we can conclude that for a fixed x, any two group assignments with the same prefix of length $k-2$ have to be sampled with equal probability. We note here that there do not exist two or more group assignments with group representation x and common prefix of length exactly $k-1$.

Induction step. Assume that for some $i < k-2$, any two group assignments with group representation x and common prefix of length $i' \in \{i+1, i+2, \dots, k-2\}$ are equally likely. Then we want to show that any two group assignments with group representation x and common prefix of length i are also equally likely. Let $y^{(s)}$ and $y^{(t)}$ be two different group assignments with group representation x and common prefix of length i. Let $w = (w_1, w_2, \dots, w_i)$ represent this common prefix of length i, that is,

$$w_1 := y_1^{(s)} = y_1^{(t)}, w_2 := y_2^{(s)} = y_2^{(t)}, \cdots, w_i := y_i^{(s)} = y_i^{(t)}.$$

Observe that if x'_j represents the number of ranks assigned to group j in ranks $(i+1, i+2, \dots, k)$ in $y^{(s)}$, then the number of ranks assigned to group j in ranks $(i+1, i+2, \dots, k)$ in $y^{(t)}$ is also x'_j for all $j \in [\ell]$, since $y^{(s)}$ and $y^{(t)}$ have common prefix of length i, and both have group representation x.

Since w is of length exactly i we also have that $y_{i+1}^{(s)} \neq y_{i+1}^{(t)}$. But the observation above give us that the group assigned to rank $i+1$ in $y^{(t)}$ appears in one of the ranks between $i+2$ and k in $y^{(s)}$. Let $\mathcal{P}$ be the set of all permutations of the elements in the multi-set

$$\left\{y_{i+2}^{(s)}, y_{i+3}^{(s)}, \dots, y_k^{(s)}\right\} \Big\backslash \left\{y_{i+1}^{(t)}\right\},$$

that is, we remove one occurrence of the group assigned to rank $i+1$ in the group assignment $y^{(t)}$ from the multi-set $\left\{y_{i+2}^{(s)}, y_{i+3}^{(s)}, \dots, y_k^{(s)}\right\}$. We then have that each element of $\mathcal{P}$ is a tuple of length $k-i-2$. We now construct two sets of group assignments $M^{(s)}$ and $M^{(t)}$ as follows,

$$M^{(s)} := \left\{ \left\{ \underbrace{w_1, w_2, \dots, w_i}_{\text{first } i}, \underbrace{y_{i+1}^{(s)}}_{i+1}, \underbrace{y_{i+1}^{(t)}}_{i+2}, \underbrace{\hat{w}_1, \hat{w}_2, \dots, \hat{w}_{k-i-2}}_{\text{last } k-i-2} \right\}, \forall \hat{w} \in \mathcal{P} \right\},$$

$$M^{(t)} := \left\{ \left\{ \underbrace{w_1, w_2, \dots, w_i}_{\text{first } i}, \underbrace{y_{i+1}^{(t)}}_{i+1}, \underbrace{y_{i+1}^{(s)}}_{i+2}, \underbrace{\hat{w}_1, \hat{w}_2, \dots, \hat{w}_{k-i-2}}_{\text{last } k-i-2} \right\}, \forall \hat{w} \in \mathcal{P} \right\}.$$

For a fixed $\hat{w} \in \mathcal{P}$ there is exactly one group assignment in $M^{(s)}$ and one group assignment in $M^{(t)}$ such that their $i+1$st and $i+2$nd coordinates are interchanged, and their first i and last $k-i-2$ coordinates are same. Therefore, $\left|M^{(s)}\right| = \left|M^{(t)}\right|$.

We also have from the induction hypothesis that all the group assignments in $M^{(s)}$ are equally likely since they have a common prefix of length $i+2$. Similarly all the group assignments in $M^{(t)}$ are equally likely. For any group assignment in $M^{(s)}$ let $\delta^{(s)}$ be the probability of sampling it. Similarly, for any group assignment in $M^{(t)}$ let $\delta^{(t)}$ be the probability of sampling it. Then,

$$\begin{aligned} &\Pr\left[Y_{i+1} = y_{i+1}^{(s)}, Y_{i+2} = y_{i+1}^{(t)} \mid Y_0, (Y_1, \dots, Y_i) = w, \boldsymbol{X} = x\right] \\ &= \Pr\left[\text{sampling a group assignment from } M^{(s)}\right] = \left|M^{(s)}\right| \delta^{(s)}, \end{aligned} \tag{2}$$

$$\begin{aligned} &\Pr\left[Y_{i+1} = y_{i+1}^{(t)}, Y_{i+2} = y_{i+1}^{(s)} \mid Y_0, (Y_1, \dots, Y_i) = w, \boldsymbol{X} = x\right] \\ &= \Pr\left[\text{sampling a group assignment from } M^{(t)}\right] = \left|M^{(t)}\right| \delta^{(t)}. \end{aligned} \tag{3}$$

Fix two group assignments $y^{(s')} \in M^{(s)}$ and $y^{(t')} \in M^{(t)}$. By the induction hypothesis $y^{(s)}$ and $y^{(s')}$ are equally likely since they have a common prefix of length $i+1$. Similarly $y^{(t)}$ and $y^{(t')}$ are also equally likely. Therefore, for $y^{(s)}$ and $y^{(t)}$ to be equally likely we need $y^{(s')}$ and $y^{(t')}$ to be equally likely.

Comparing $y^{(s')}$ and $y^{(t')}$ instead of $y^{(s)}$ and $y^{(t)}$. We know from above that $y^{(s')}$ and $y^{(t')}$ are sampled with probability $\delta^{(s)}$ and $\delta^{(t)}$ respectively. Therefore for any distribution satisfying Axiom 3 we have,

$$\begin{aligned}
&\Pr\left[Y_{i+1} = y^{(s)}_{i+1}, Y_{i+2} = y^{(t)}_{i+1} \mid Y_0, Y_{1:i} = w, \boldsymbol{X} = x\right] \\
&= \Pr\left[Y_{i+1} = y^{(t)}_{i+1}, Y_{i+2} = y^{(s)}_{i+1} \mid Y_0, Y_{1:i} = w, \boldsymbol{X} = x\right] \\
&\implies \left|M^{(s)}\right| \delta^{(s)} = \left|M^{(t)}\right| \delta^{(t)}, \text{ from Equations (2) and (3)} \\
&\implies \delta^{(s)} = \delta^{(t)}, \qquad \because \left|M^{(s)}\right| = \left|M^{(t)}\right|.
\end{aligned}$$

Note that the converse is also easy to show, which means that Axiom 3 is satisfied if and only if $y^{(s')}$ and $y^{(t')}$ are equally likely. Therefore, Axiom 3 is satisfied if and only if $y^{(s)}$ and $y^{(t)}$ are equally likely.

For a fixed group representation x, for any two group assignments with corresponding group representation x, there exists an $i \in \{0, 1, \ldots, k-2\}$ such that they have a common prefix of length i. Therefore, any two group assignments, for a fixed group representation x, have to be equally likely. Therefore $\mathcal{D}$ is the unique distribution that satisfies all three axioms. □

Proof of Theorem 5. Given an $\delta > 0$ and a distribution $\widehat{\mathcal{D}}$ that is at total-variation distance of δ from $\mathcal{D}$ defined in Theorem 4, when sampling group representation. Therefore,

$$\sup_{A \subseteq \mathcal{X}} \left| \Pr_{\mathcal{D}}(A) - \Pr_{\widehat{\mathcal{D}}}(A) \right| = \delta. \tag{4}$$

Now, fix a group $j \in [\ell]$ and a rank $i \in [k]$. Let $\mathcal{X}$ be the set of all group fair representations for given constraints, $L_j, U_j, \forall j \in [\ell]$. Then,

$$\begin{aligned}
\Pr_{\widehat{\mathcal{D}}}[Y_i = j] &= \sum_{x \in \mathcal{X}} \Pr_{\widehat{\mathcal{D}}}[X = x] \Pr_{\widehat{\mathcal{D}}}[Y_i = j | X] \\
&\quad \text{(by the law of total probability)} \\
&= \sum_{x \in \mathcal{X}} \Pr_{\widehat{\mathcal{D}}}[X = x] \frac{x_j}{k} \\
&\geqslant \frac{L_j}{k} \sum_{x \in \mathcal{X}} \Pr_{\widehat{\mathcal{D}}}[X = x] \\
&= \frac{L_j}{k}.
\end{aligned}$$

Similarly we get $\Pr_{\widehat{\mathcal{D}}}[Y_i = j] \leqslant \frac{U_j}{k}$. □

Proof of Corollary 6. Given an $\delta > 0$ and a distribution $\widehat{\mathcal{D}}$ that is at total-variation distance of δ from $\mathcal{D}$ defined in Theorem 4, when sampling group representation. Fix a group $j \in [\ell]$ and rank $i, i' \in [k]$ such that $i \leqslant i'$. Let $\mathcal{X}$ be the set of all group fair representations for given constraints, $L_j, U_j, \forall j \in [\ell]$.

$$\begin{aligned}
\mathbb{E}_{\widehat{\mathcal{D}}}\left[Z^j_{i,i'}\right] &= \mathbb{E}_{\widehat{\mathcal{D}}}\left[\sum_{\hat{i}=i}^{i'} \mathbb{I}\left[Y_{\hat{i}} = j\right]\right] \\
&= \sum_{\hat{i}=i}^{i'} \mathbb{E}_{\widehat{\mathcal{D}}}\left[\mathbb{I}\left[Y_{\hat{i}} = j\right]\right] \qquad \text{by linearity of expectation} \\
&= \sum_{\hat{i}=i}^{i'} \Pr_{\widehat{\mathcal{D}}}\left[Y_{\hat{i}} = j\right] \\
&\geqslant \frac{i' - i + 1}{k} \cdot L_j. \qquad \text{from Theorem 5}
\end{aligned}$$

Similarly $\mathbb{E}_{\widehat{\mathcal{D}}}\left[Z^j_{i,i'}\right] \leqslant \frac{i'-i+1}{k} \cdot U_j$. □

B Missing Proofs from Section 4

Proof of Theorem 7. We first show by mathematical induction on i that for any $k' \in \{0, 1, \ldots, k\}$,

$$\begin{aligned}
D[k', i] = |\{&(x_1, x_2, \ldots, x_i) \mid L_j \leqslant x_j \leqslant U_j, \forall j \in [i] \\
&\text{and } x_1 + x_2 + \ldots + x_i = k'\}|
\end{aligned} \tag{5}$$

In the base case, $D[k', 1] = 1$ if $L_1 \leqslant k' \leqslant U_1$ because choosing $x_1 = k'$ gives us exactly one feasible integer solution. Let us assume that the hypothesis is true for every k' and for every i' such that $i' \leqslant i < \ell$. Then for $i + 1$, and for any k', the feasible values of x_{i+1} are in $[L_{i+1}, U_{i+1}]$. For each of these values of x_{i+1}, all the feasible solutions with the first i groups that sum to $k' - x_{i+1}$ are feasible solutions for that value of x_{i+1}. By the induction hypothesis, this is exactly what $D[k' - x_{i+1}, i]$ stores. Therefore, for any $k' \in \{0, 1, \ldots, k\}$, $D[k', i{+}1] = \sum_{L_i \leqslant x_i \leqslant U_i} D[k'{-}x_i, i{-}1]$ is the number of feasible integer solutions with the first $i + 1$ groups that sum to k', which is exactly what Step 4 is counting. Therefore, $D[k, \ell]$ counts the number of integer solutions in the polytope K.

Now let X be an integer random vector $(X_1, X_2, \ldots, X_\ell) \in [k]^\ell$ representing the group representation.

$$\begin{aligned}
&\Pr\left[\text{DP outputs } X = (x_1, \ldots, x_\ell)\right] \\
&\quad = \Pr\left[\text{DP outputs } x_1 \wedge \text{DP outputs } x_2 \wedge \cdots \wedge \text{DP outputs } x_\ell\right] \\
&\quad = \prod_{i=1}^{\ell} \Pr\left[\text{DP outputs } x_i \mid \text{DP output } x_{i+1}, \cdots, \text{DP output } x_\ell\right] \\
&\quad = \prod_{i=1}^{\ell} \frac{D[k - x_i - x_{i+1} - \cdots - x_\ell,\ i-1]}{D[k - x_{i+1} - \cdots - x_\ell,\ i]}.
\end{aligned} \tag{6}$$

We first show that the DP never samples infeasible solutions. To see this, observe that any tuple $(x_1, x_2, \ldots, x_\ell)$ can be infeasible in two cases. One when there exists a group $j \in [\ell]$ such that the condition $L_j \leqslant x_j \leqslant U_j$ is not satisfied. Other case is when the summation constraint

$x_1+x_2+\cdots+x_\ell = k$ is not satisfied. The former does not occur in the DP because for every $j \in [\ell]$ it only samples the values of $x_j \in [L_j, U_j]$. For the latter, the product term in Equation (6) will have the count of the entry $D[k - \sum_{j\in[\ell]} x_j, 0]$, which is 0 due to our initialization. Hence, such an x is sampled with probability 0.

When x is feasible, for any $k' \in \{0, 1, \ldots, k\}$ and for each sampling step i, the DP samples $x_i \in [L_i, U_i]$ from a valid probability distribution because $\sum_{L_i \leqslant x_i \leqslant U_i} D[k' - x_i, i-1]/D[k', i] = 1$. Moreover, we have $x_1 + x_2 + \ldots + x_\ell = k$. Therefore the telescopic product in Equation (6) always gives $D[0,0]/D[k,\ell]$. Due to our initialization, $D[0,0] = 1$. Hence, the probability of sampling any feasible x is $1/D[k,\ell]$. Therefore, this DP gives uniform random samples.

Since the DP table is of size $k\ell$ and computing each entry takes time $\mathcal{O}(k)$, the *counting* step takes time $\mathcal{O}\left(k^2\ell\right)$. Sampling from categorical distribution of size at most k in Line 9 takes time $\mathcal{O}(k)$ and this step is run ℓ times. Hence, *sampling* takes $\mathcal{O}(k\ell)$ amount of time. $\square$

Proof of Theorem 8

For completeness, we restate the Theorem 1.2 in [Cousins and Vempala, 2018] that states the running time and success probability of their uniform sampler, in Appendix B.

Theorem 9 (Theorem 1.2 in [Cousins and Vempala, 2018])**.** *There is an algorithm that, for any $\delta > 0$, $p > 0$, and any convex body $C \in \mathbb{R}^d$ that contains the unit ball and has $\mathbb{E}_C(\|X\|^2) = R^2$, with probability $1-p$, generates random points from a density ν that is within total variation distance δ from the uniform distribution on C. In the membership oracle model, the complexity of each random point, including the first, is $\mathcal{O}^*\left(\max\left\{R^2d^2, d^3\right\}\right)$*[8].

Lemma 10. $B(0,\Delta) \subseteq P'$.

Proof. From the definition of Δ we have the following inequalities.

$$\begin{aligned}
&\Delta \leqslant \left\lfloor \frac{k - \left(\sum_{j\in[\ell]} L_j\right)}{\ell} \right\rfloor \\
&\implies \Delta \leqslant \frac{k - \left(\sum_{j\in[\ell]} L_j\right)}{\ell} \\
&\implies \ell\cdot\Delta + \sum_{j\in[\ell]} L_j \leqslant k \\
&\implies \sum_{j\in[\ell]} (L_j + \Delta) \leqslant k, &(7)\\
&\Delta \leqslant \left\lfloor \frac{\left(\sum_{j\in[\ell]} U_j\right) - k}{\ell} \right\rfloor \\
&\implies \Delta \leqslant \frac{\left(\sum_{j\in[\ell]} U_j\right) - k}{\ell} \\
&\implies -\ell\cdot\Delta + \sum_{j\in[\ell]} U_j \geqslant k \\
&\implies \sum_{j\in[\ell]} (U_j - \Delta) \geqslant k, &(8)
\end{aligned}$$

and

$$\begin{aligned}
&\Delta \leqslant \left\lfloor \frac{U_j - L_j}{2} \right\rfloor \\
&\implies \Delta \leqslant \frac{U_j - L_j}{2} \\
&\implies 2\Delta \leqslant U_j - L_j \\
&\implies L_j + \Delta \leqslant U_j - \Delta. &(9)
\end{aligned}$$

To show that Steps 4 to 7 find the correct center we use the following loop invariant.

Loop invariant. At the start of every iteration of the **for** loop x^* is an integral point such that $L_j + \Delta \leqslant x^*_j \leqslant U_j - \Delta, \forall j \in [\ell]$.

Initialization: In Step 4 each x^*_j is initialized to $L_j + \Delta$. From Equation (9) we know that $L_j + \Delta \leqslant U_j - \Delta$. Moreoever, L_j, U_j, and Δ are all integers. Therefore, x^* is integral and satisfies $L_j + \Delta \leqslant x^*_j \leqslant U_j - \Delta, \forall j \in [\ell]$.

Maintenance: If the condition in Step 7 fails, the value of x^* is not updated. Therefore the invariant is maintained. If the condition succeeds we have that,

$$\sum_{j'\in[\ell]} x^*_{j'} < k \qquad (10)$$

The value x^*_j is set to $\min\left\{k - \sum_{j'\in[\ell]:j'\neq j} x^*_{j'}\ ,\quad U_j - \Delta\right\}$ in Step 7. The following two cases arise based on the minimum of the two quantities.

- **Case 1:** $k - \sum_{j'\in[\ell]:j'\neq j} x^*_{j'} \leqslant U_j - \Delta$**.** In this case x^*_j is set to $k - \sum_{j'\in[\ell]:j'\neq j} x^*_{j'} \leqslant U_j -$

[8]The $\mathcal{O}^*$ notation suppresses error terms and logarithmic factors.

Δ, which is an an integer value since both x_j^* and $k - \sum_{j' \in [\ell]: j' \neq j} x_{j'}^*$ are integers before the iteration. From (10) we have that

$$k - \sum_{j' \in [\ell]: j' \neq j} x_{j'}^* = x_j^* + k - \sum_{j' \in [\ell]} x_{j'}^* > x_j^*. \quad (11)$$

Since $x_j^* \geqslant L_j + \Delta$ before the iteration, (11) gives us that x_j^* is greater than $L_j + \Delta$ even after the update.

– **Case 2:** $k - \sum_{j' \in [\ell]: j' \neq j} x_{j'}^* > U_j - \Delta$**.** Since $U_j - \Delta \geqslant L_j + \Delta$ from Equation (9) and since $U_j - \Delta$ is an integer, the value of x_j^* after the update is an integer such that $U_j - \Delta \geqslant x_j^* \geqslant L_j + \Delta$.

Therefore in both the cases the invariant is maintained.

Termination: At termination $j = \ell$. The invariant gives us that x^* is an integral point such that $L_j + \Delta \leqslant x_j^* \leqslant U_j - \Delta, \forall j \in [\ell]$.

From Equation (7) we have that before the start of the **for** loop $\sum_{j \in [\ell]} x_j^* = \sum_{j \in [\ell]} L_j + \Delta \leqslant k$. After the termination of the **for** loop we have that $x_j^* = U_j - \Delta$, forall $j \in [\ell]$, when the **if** condition in Step 7 fails for all $j \in [\ell]$, or the **if** condition in Step 7 succeeds for some j, in which case $\sum_{j \in [\ell]} x_j^* = k$, and the value of x^* does not change after this iteration. Therefore, after the **for** loop we get $\sum_{j \in [\ell]} x_j^* = \min \left\{ \sum_{j \in [\ell]} U_j - \Delta, k \right\}$. But Equation (8) gives us that $\sum_{j \in [\ell]} U_j - \Delta \geqslant k$. Therefore, the **for** loop finds an integral point x^* such that $L_j + \Delta \leqslant x_j^* \leqslant U_j - \Delta, \forall j \in [\ell]$, and $\sum_{j \in [\ell]} x_j^* = k$.

Therefore there is an l_1 ball of radius Δ in P centered at the integral point $x^* \in H$ (that is, $\sum_{j \in [\ell]} x_j^* = k$). Consequently there exists an l_1 ball of radius Δ centered at the origin in the polytope P'. Since an l_1 ball of radius Δ centered at origin encloses an l_2 ball of radius Δ centered at origin we get that an l_2 ball of radius Δ centered at the origin, $B(0, \Delta)$, is in the polytope P'.

$\square$

Let $C(x, \beta) \subseteq \mathbb{R}^\ell$ represent a cube of side length β centered at x. For any integral point $x \in K'$ let F_x represent the set of points in $\left(1 + \frac{\sqrt{\ell}}{\Delta}\right) K'$ that are rounded to x.

Lemma 11. *For any integral point $x \in K'$, $F_x \subseteq H' \cap C(x, 2)$.*

Proof. Let z be the point sampled in Step 11. Since $z \in \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) K'$ we have that $\sum_{j \in [\ell]} z_j = 0$. Therefore,

$$\sum_{j \in [\ell]} \lfloor z_j \rfloor \leqslant 0 \qquad \text{and} \qquad \sum_{j \in [\ell]} \lceil z_j \rceil \geqslant 0.$$

Then,

$$m = \left| \sum_{j \in [\ell]} \lfloor z_j \rfloor \right| = \left| \sum_{j \in [\ell]} \lfloor z_j \rfloor - \sum_{j \in [\ell]} z_j \right|$$
$$= \left| \sum_{j \in [\ell]} (\lfloor z_j \rfloor - z_j) \right| \leqslant \sum_{j \in [\ell]} |\lfloor z_j \rfloor - z_j| \leqslant \ell,$$

where the second equality is because $\sum_{j \in [\ell]} z_j = 0$. Hence, starting from $x_j = \lfloor z_j \rfloor, \forall j \in [\ell]$, the algorithm has to round at most ℓ coordinates to $x_j = \lceil z_j \rceil$. Since $j \in [\ell]$ this is always possible. Therefore, the rounding in Step 12 always finds an integral point x that satisfies the following,

$$\sum_{j \in [\ell]} x_j = 0 \qquad \text{and} \qquad (\forall j \in [\ell], x_j = \lfloor z_j \rfloor \text{ or } x_j = \lceil z_j \rceil). \quad (12)$$

Therefore, the set of points $z \in \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) K'$ that are rounded to the integral point $x \in K'$ satisfying (12) is a strict subset of

$$\left\{ z : (\forall j \in [\ell], x_j = \lfloor z_j \rfloor \vee x_j = \lceil z_j \rceil) \wedge \sum_{j \in [\ell]} z_j = 0 \right\},$$

which is contained in $H' \cap C(x, 2)$ since $|z_j - \lfloor z_j \rfloor| \leqslant 1$ and $|\lceil z_j \rceil - z_j| \leqslant 1, \forall j \in [\ell]$. $\square$

Lemma 12. *For any $x \in K'$, $H' \cap C(x, 2) \subseteq \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) K'$.*

Proof. Fix a point $x \in P'$. Then for any $x' \in C(x, 2)$, $\|x' - x\|_2 \leqslant \sqrt{\ell}$. Lemma 10 gives us that the translated polytope P' contains a ball of radius Δ centered at the origin. Then the polytope $\frac{\sqrt{\ell}}{\Delta} P'$ contains a ball of radius $\sqrt{\ell}$ centered at the origin, which implies that the polytope $\frac{\sqrt{\ell}}{\Delta} P'$ contains every vector of length at most $\sqrt{\ell}$. Therefore, $x' - x \in \frac{\sqrt{\ell}}{\Delta} P'$. Now since $x \in P'$ we get that $x' \in \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) P'$. Therefore, $C(x, 2) \subseteq \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) P'$. Consequently, $H' \cap C(x, 2) \subseteq H' \cap \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) P' = \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) (H' \cap P')$ since $\alpha H' = H'$ for any scalar $\alpha \neq 0$. Hence, $H' \cap C(x, 2) \subseteq \left(1 + \frac{\sqrt{\ell}}{\Delta}\right) K'$ $\square$

Lemma 13. *For any point $z \in \frac{1}{\left(1 + \frac{\sqrt{\ell}}{\Delta}\right)} K'$ the integral point it is rounded to belongs to the polytope K'.*

Proof. From Lemma 11 we know that for any integral point $x \in K'$, $F_x \subseteq H' \cap C(x, 2)$. Due to convexity of K', $\frac{1}{\left(1 + \frac{\sqrt{\ell}}{\Delta}\right)} K'$ is contained entirely inside K'. Therefore, Lemma 11 is true for all the points in $\frac{1}{\left(1 + \frac{\sqrt{\ell}}{\Delta}\right)} K'$. By arguing similarly as in the proof of Lemma 12, we can show that for any any $x \in \frac{1}{\left(1 + \frac{\sqrt{\ell}}{\Delta}\right)} K'$, $H' \cap C(x, 2) \subseteq K'$. This proves the lemma. $\square$

Let μ be a uniform probability measure on the convex rational polytope $\left(1+\frac{\sqrt{\ell}}{\Delta}\right)K'$.

Lemma 14. *Fix any two distinct integral points $x, x' \in K'$, $\mu\left(F_x\right)=\mu\left(F_{x'}\right)$.*

Proof. Given two distinct integral points $x, x' \in K'$, let $c = x' - x$. Clearly c is an integral point and $\sum_{j\in[\ell]} c_j = \sum_{j\in[\ell]} x'_j - \sum_{j\in[\ell]} x_j = 0 - 0 = 0$. Let $z \in F_x$ and $z' = z + c$. Then

$$\left|\sum_{j\in[\ell]} \lfloor z'_j \rfloor\right| = \left|\sum_{j\in[\ell]} \lfloor z_j \rfloor + c_j\right| = \left|\sum_{j\in[\ell]} \lfloor z_j \rfloor + \sum_{j\in[\ell]} c_j\right| = \left|\sum_{j\in[\ell]} \lfloor z_j \rfloor\right| = m.$$

Therefore, for both z and z' the first m coordinates are rounded up, and the remaining are rounded down, in Step 12. Since $\lfloor z'_j \rfloor = \lfloor z_j \rfloor + c_j$ and $\lceil z'_j \rceil = \lceil z_j \rceil + c_j$, the point z' is rounded to is nothing but x'. Therefore, for every point $z \in F_x$ there is a unique point $z' \in F_{x'}$ such that they are rounded to x and x' respectively. This gives us a bijection between the sets F_x and $F_{x'}$. Therefore, $\mu\left(F_x\right)=\mu\left(F_{x'}\right)$. ☐

Proof of Theorem 8. Let $K'' = \left(1+\frac{\sqrt{\ell}}{\Delta}\right)K'$.

Let ν be the distribution from which the SAMPLING-ORACLE samples a point from K''. That is for a given $\delta > 0$,

$$\sup_{A\subseteq K''} |\nu(A) - \mu(A)| \leqslant \delta. \tag{13}$$

Close to uniform sample. From Lemmas 11 and 12 we know that for any point $x \in K'$, $F_x \subseteq K''$. Therefore, from (13) we have that

$$|\nu(F_x) - \mu(F_x)| \leqslant \delta.$$

Let ν' be the density from which Algorithm 2 samples an integral point from K. We know that $\forall x,\ x \in K' \iff x + x^* \in K$. Since x^* is an integral point, for any integral point x we also have that $x \in K' \iff x + x^* \in K$. This gives us a bijection between the integral points in K' and K.

Therefore for any integral point $x \in K'$, $\nu'(x) = \nu(F_x)$. Let $\mu'(x) := \mu(F_x)$ for any integral point x in K'.

For any two integral points $x, x' \in K'$, Lemma 14 gives us that $\mu\left(F_x\right)=\mu\left(F_{x'}\right)$. Moreover, since F_x and $F_{x'}$ are both fully contained in K'', we get that $\mu'(F_x) = \mu(F_x) = \mu(F_{x'}) = \mu'(F_{x'})$. Therefore, μ' is a uniform measure over all the integral points in K'.

Moreover, for any subset of integral points in K, say I, we have that

$$\nu'(I) = \nu\left(\cup_{x\in I} F_x\right).$$

From Equation (13) we have that

$$\left|\nu\left(\cup_{x\in I} F_x\right) - \mu\left(\cup_{x\in I} F_x\right)\right| \leqslant \delta.$$

Consequently,

$$|\nu'(I) - \mu'(I)| \leqslant \delta.$$

Therefore ν' over the integral points in K' is at a total variation distance of at most δ from the uniform probability measure μ' over the integral points in K'.

Probability of acceptance. Algorithm 2 samples points from $\left(1+\frac{\sqrt{\ell}}{\Delta}\right)K'$ in each iteration. Due to Lemma 13 we know for sure that whenever the algorithm samples a point from $\frac{1}{\left(1+\frac{\sqrt{\ell}}{\Delta}\right)}K'$, it will be rounded to an integral point in K'. Therefore, the probability of sampling an integral point in K' is

$$\begin{aligned}
&\geqslant \nu\left(\frac{1}{\left(1+\frac{\sqrt{\ell}}{\Delta}\right)}K'\right)\\
&\geqslant \mu\left(\frac{1}{\left(1+\frac{\sqrt{\ell}}{\Delta}\right)}K'\right) - \delta\\
&\qquad \text{from } Equation\ (13)\\
&= \frac{\mathsf{Vol}_{\ell-1}\left(\frac{1}{\left(1+\frac{\sqrt{\ell}}{\Delta}\right)}K'\right)}{\mathsf{Vol}_{\ell-1}\left(K''\right)} - \delta\\
&\qquad \because \mu \text{ is a uniform distribution over } K''\\
&= \frac{\mathsf{Vol}_{\ell-1}\left(\frac{1}{\left(1+\frac{\sqrt{\ell}}{\Delta}\right)}K'\right)}{\mathsf{Vol}_{\ell-1}\left(\left(1+\frac{\sqrt{\ell}}{\Delta}\right)K'\right)} - \delta\\
&\qquad \text{by the definition of } K''\\
&= \left(1+\frac{\sqrt{\ell}}{\Delta}\right)^{-2(\ell-1)} - \delta\\
&\qquad \mathsf{Vol}_{\ell-1} \text{ is volume in } \ell-1 \text{ dimensions}\\
&\geqslant e^{-\frac{\sqrt{\ell}}{\Delta}2(\ell-1)} - \delta\\
&\qquad \text{using } (1+x) \leqslant e^x, \forall x
\end{aligned}$$

Therefore the expected running time of the algorithm before it outputs an acceptable point is inversely proportional to the probability of acceptance, that is, $1/\left(e^{-2\frac{\ell\sqrt{\ell}}{\Delta}} - \delta\right)$. When $\delta < e^{-2}$ and is a non-negative constant, and if $\Delta = \Omega\left(\ell^{1.5}\right)$, this probability is at least a constant. Hence, repeating the whole process a polynomial number of times in expectation guarantees we sample an integral point from K'. ☐

B.1 Exact Uniform Sampling for small ℓ

In this section we give another exact sampling algorithm to sample a uniform random group-fair representation. In Equation (1) the convex polytope K is described using an $\mathcal{H}$ description defined as follows,

Definition 15 (**$\mathcal{H}$-description of a polytope**)**.** A representation of the polytope as the set of solutions of finitely many linear inequalities.

We can also have a representation of the polytope as described by its vertices, defined as follows,

Definition 16 (**$\mathcal{V}$-description of a polytope**)**.** The representation of the polytope by the set of its vertices.

[Barvinok, 2017] gave an algorithm to count exactly, the number of integral points in K, as re-stated below.

Theorem 17 (Theorem 7.3.3 in [Barvinok, 2017])**.** *Let us fix the dimension ℓ. Then there exists a polynomial time algorithm that, for any given rational $\mathcal{V}$-polytope $P \subset \mathbb{R}^\ell$, computes the number $|P \cap \mathbb{Z}^\ell|$. The complexity of the algorithm in terms of the dimension ℓ is $\ell^{\mathcal{O}(\ell)}$.*

We also have the algorithm by [Pak, 2000] gives us an exact uniform random sampler for the integral points in K.

Theorem 18 (Theorem 1 in [Pak, 2000])**.** *Let $P \subset \mathbb{R}^\ell$ be a rational polytope, and let $B = P \cap \mathbb{Z}^\ell$. Assume an oracle can compute $|B|$ for any P as above. Then there exists a polynomial-time algorithm for sampling uniformly from B, which calls this oracle $\mathcal{O}\left(\ell^2 L^2\right)$ times where L is the bit complexity of the input.*

Using the counting algorithm given by Theorem 17 as the counting oracle in Theorem 18 gives us our second algorithm that samples a uniform random group representation exactly.

Theorem 19. *For given fairness parameters $L_j, U_j \in \mathbb{Z}_{\geqslant 0}$ and an integer $k > 0$, there is an algorithm that samples an exact uniform random integral point in K and runs in time $\ell^{\mathcal{O}(\ell)}\mathcal{O}\left(\log^2 k\right)$.*

Proof of Theorem 19. The proof essentially follows from the proof of Theorem 1 in [Pak, 2000]. They assume access to an oracle that counts the number of integral points in any convex polytope that their algorithm constructs. We show that Barvinok's algorithm can be used as an oracle for all the polytopes that are constructed in the algorithm to sample a uniform random integral point from our convex rational polytope K.

The algorithm in Theorem 18 intersects the polytope by an axis-aligned hyperplane and recurses on one of the smaller polytopes (to be specified below). In the deepest level of recursion where the polytope in that level contains only one integral point, the algorithm terminates the halving process and outputs that point. The proof of their theorem shows that this gives us a uniform random integral point from the polytope we started with.

Let us consider the dimension 1 w.l.o.g. The algorithm finds a value c such that $L_1 < c < U_1$, $|H_+ \cap B|/|B| \leqslant 1/2$, $|H_- \cap B|/|B| \leqslant 1/2$, where H_+ and H_- are two halves of the space separated by the hyperplane H defined by $x_1 = c$. That is, H_+ is the halfspace $x_1 \geqslant c$ and H_- is the halfspace $x_1 \leqslant c$. Therefore, there are three possible polytopes for the algorithm to recurse on, $H_+ \cap B$, $H_- \cap B$, and $H \cap B$. Here $|H \cap B|/|B|$ can be $\geqslant 1/2$. Let

$$f_+ = \frac{|H_+ \cap B|}{|B|},\ f_- = \frac{|H_- \cap B|}{|B|},\ \text{and } f = \frac{|H \cap B|}{|B|}.$$

Then the algorithm recurses on the polytope $H_+ \cap B$ with probability f_+, on $H_- \cap B$ with probability f_-, and on $H \cap B$ with probability f.

Observe that K is also defined by the axis aligned hyperplanes, $x_1 = L_1$ and $x_1 = U_1$, amongst others. Therefore, $x_1 \geqslant L_1$ will become a redundant constraint if the algorithm recurses on $H_+ \cap B$. Else $x_1 \leqslant U_1$ will become a redundant constraint if the algorithm recurses on $H_- \cap B$. In both these cases, the number of integral points reduces by more than $1/2$. If the algorithm recurses on $H \cap B$, it fixes the value of x_1 to c, and the dimension of the problem reduces by 1. Since the number of integral points is $\exp(dL)$, the number of halving steps performed by the algorithm is at most $\mathcal{O}(dL)$.

Observe that in all levels of recursion, the polytopes constructed are of d dimensions ($1 \leqslant d \leqslant \ell$) and are of the following form,

$$\Big\{ (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d \ \Big|\ \sum_{j \in [d]} x_j = k' \text{ and } c'_j \leqslant x_j \leqslant c''_j, \forall j \in [\ell] \Big\},$$

where k', c', c'' are some constants. This gives us the $\mathcal{H}$-description of each of the polytopes the algorithm constructs in each level of recursion. The vertices of such a polytope are formed by the intersection of d hyperplanes. Therefore, there could be at most $\binom{2d+2}{d} = 2^{\mathcal{O}(d)}$ number of vertices for such a polytope in d dimensions, which gives us that the $\mathcal{V}$-description can be computed from the $\mathcal{H}$-description in time $2^{\mathcal{O}(d)}$. Therefore, for all these intermediate polytopes, we can use the counting algorithm given by Theorem 17 whose run time depends on $d^{(\mathcal{O}(d)}$. Using $d \leqslant \ell$, we get that the counting algorithm given by Theorem 17 takes time $\ell^{(\mathcal{O}(\ell)}$ for all the polytopes constructed by the algorithm. Further, the algorithm makes at most $\mathcal{O}(dL)$ calls to this counting algorithm in each recursion step.

Since the input to the algorithm consists of a number $k > 0$ and fairness parameters $0 \leqslant L_j \leqslant U_j \leqslant k, \forall j \in [\ell]$, where all these parameters are integers, the bit complexity of the input is $\log k$.

Therefore, the total running time of our algorithm is $\mathcal{O}\left(d^2 L^2\right)\ell^{\mathcal{O}(\ell)} = \mathcal{O}\left(\ell^2 \log^2 k\right)\ell^{\mathcal{O}(\ell)} = \mathcal{O}\left(\log^2 k\right)\ell^{\mathcal{O}(\ell)}$. ☐

Additional References

[Barvinok, 2017] Barvinok. Handbook of discrete and computational geometry (3rd ed.). 2017

[Pak, 2000] Igor Pak. On sampling integer points in polyhedra. Foundations of Computational Mathematics, 2000

C Additional Experimentals Results

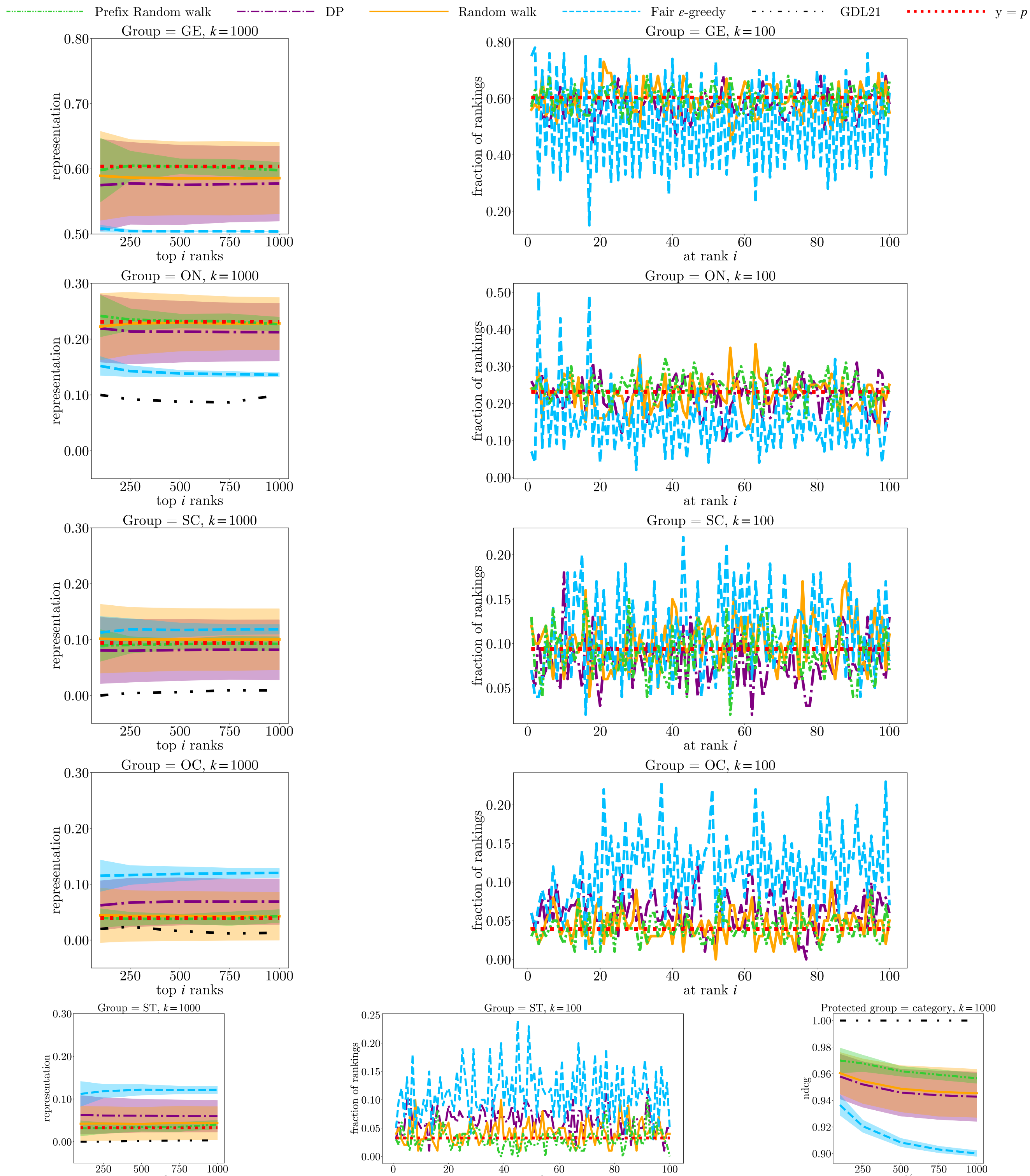

Figure 4: Results on the JEE 2009 dataset with *birth category* as the protected group (with 5 groups). For Fair ϵ-greedy we use $\epsilon = 0.3$.

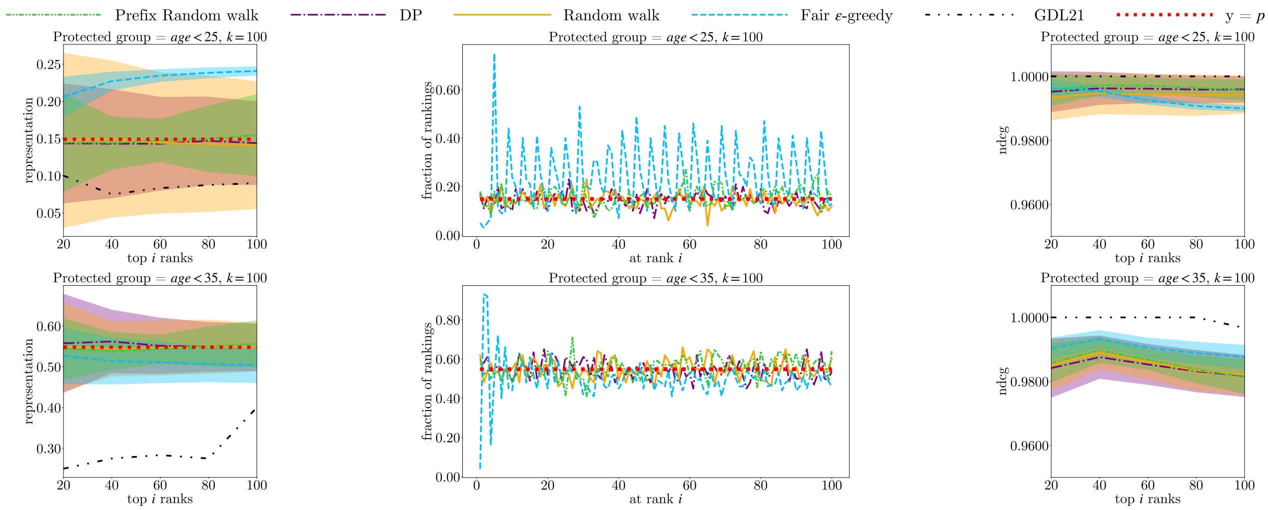

Figure 5: Results on the German Credit Risk dataset with *age* < 25 as the protected group in the first row and *age* < 35 as the protected group in the first row. For Fair ϵ-greedy we use $\epsilon = 0.15$.

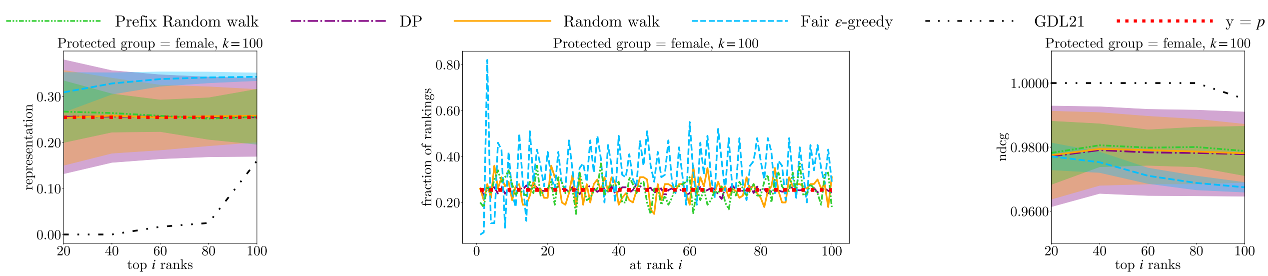

Figure 6: Results on the JEE 2009 dataset with *gender* as the protected group. For Fair ϵ-greedy we use $\epsilon = 0.15$.

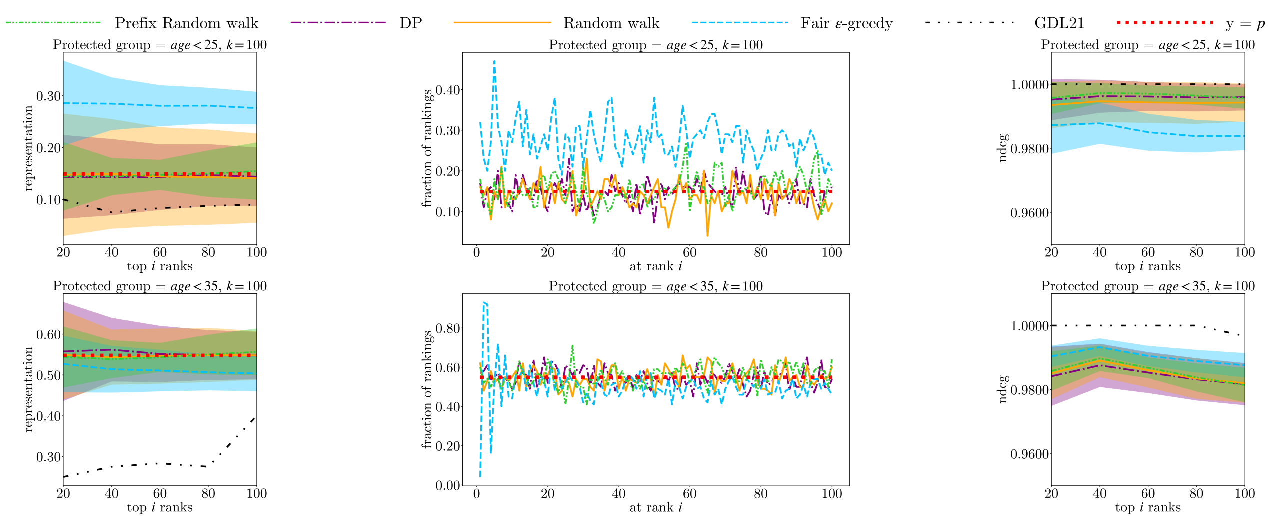

Figure 7: Results on the German Credit Risk dataset with *age* < 25 as the protected group in the first row and *age* < 35 as the protected group in the first row. For Fair ϵ-greedy we use $\epsilon = 0.5$.

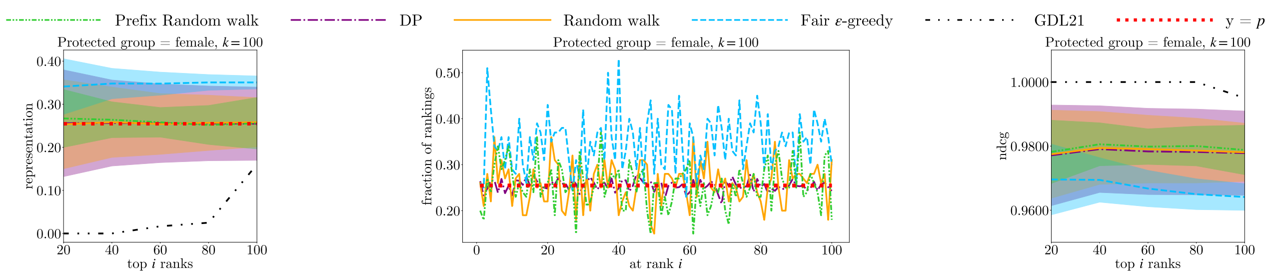

Figure 8: Results on the JEE 2009 dataset with *gender* as the protected group. For Fair ϵ-greedy we use $\epsilon = 0.5$.