Appendix for Fair k-Centers via Maximum Matching

Anonymous Authors[1]

1. Additional experiments

In this section, we provide further experiments on real and stimulated dataset. The experimental setup is the same as the other experiments, we adjust the constraint for required number of centers for each group. For stimulated data, we require one group to have one group to have disproportionately more than the reset. We set one groups to have the number of center to be 8 and set the rest to be 1. For real dataset, we set the number of center to be the same for each group. A-Gender and A-Race have 100 and 50 required centers for each group respectively. S-Sex, S-School, and S-Adress have 50 required centers for each group. Finally, each group in W-location has 20 required centers.

Table 4. Mean and standard deviation of objective value on stimulated data

Algorithm	50 Groups	100 Groups	200 Groups	400 Groups
Alg 2-Seq	6.4 (0.34)	**6.17 (0.31)**	**6.35 (0.46)**	6.41 (0.35)
Alg 2-Heu B	**6.36 (0.32)**	6.22 (0.38)	6.36 (0.47)	**6.31 (0.37)**
Kleindessner	6.61 (0.55)	6.66 (0.76)	7.09 (0.65)	7.19 (0.48)
Heuristic A	18.63 (2.18)	17.08 (1.34)	16.2 (1.52)	14.08 (1.41)
Heuristic B	6.72 (0.28)	7.36 (1.15)	7.79 (0.68)	7.82 (0.65)
Heuristic C	6.47 (0.4)	6.66 (0.45)	7.12 (0.52)	7.39 (0.55)

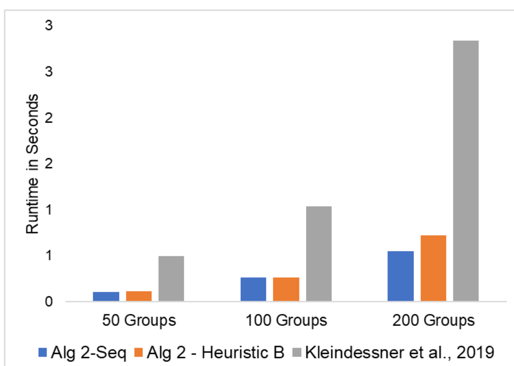
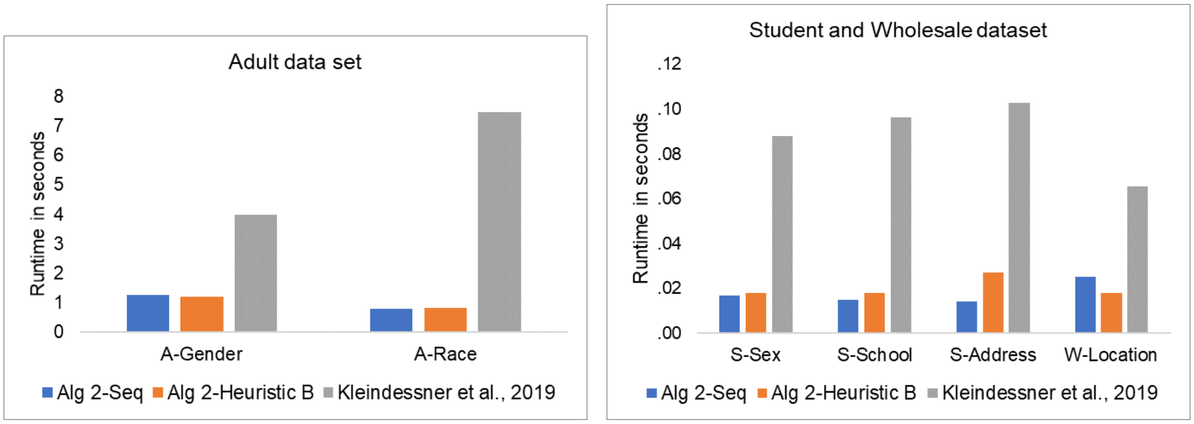


Figure 4. Mean runtime in seconds on stimulated data

Table 5. Mean and standard deviation of objective value on real data

Algorithms	A-Gender	A-Race	S-Sex	S-School	S-Address	W-location
Alg 2-Seq	**0.27 (0.01)**	0.36 (0.024)	**0.98 (0.01)**	**1 (0.02)**	1.03 (0.02)	0.17 (0.01)
Alg 2-Heu B	**0.27 (0.01)**	0.34 (0.03)	**0.98 (0.01)**	**1 (0.02)**	1.03 (0.02)	0.17 (0.01)
Kleindessner	0.31 (0.02)	0.3 (0.02)	1 (0.04)	1.04 (0.05)	1.06 (0.05)	**0.15 (0.01)**
Heuristic A	0.3 (0.01)	0.38 (0.02)	1.04 (0.01)	1.04 (0.01)	1.06 (0.02)	0.21 (0.03)
Heuristic B	0.28 (0.005)	0.37 (0.02)	**0.98 (0.01)**	**1 (0.02)**	1.04 (0.02)	0.19 (0.01)
Heuristic C	0.28 (0.005)	**0.26 (0.002)**	1.03 (0.01)	1.04 (0.01)	**0.99 (0.01)**	**0.15 (0.004)**



(a) (b)

Figure 5. Mean runtime in seconds on real data

2. Full algorithm

Algorithm 4 gives the detailed algorithm for algorithm 2 in the paper. The algorithm starts by running the Gonzalez's algorithm to compute k centers sequentially. The algorithm proceeds by running binary search to find the largest integer h such that the first h items in the sequence returned by Gonzalez's algorithm satisfy the fair shift constraint. Then, another binary search procedure is to find the smallest radius that allow a fair shift. The centers C corresponding to this radius guarantees the 3-approximation according to Lemma 5.3. Finally, the algorithm arbitrarily adds any additional centers to C for the fairness constraint.

Algorithm 4 3-approximation algorithm for k-centers with fairness

Input: a set of points $S = \{s_1, ..., s_n\}$ each with a demographic group value $f_i \in [m]$, a distance metric d, the values k_f where $\sum_f k_f = k$, and a metric d

Output: A set C such that $C \subseteq S$ and $|\{s_i | s_i \in C \bigwedge f_i = f\}| = k_f$ for all demographic group values f

Run Algorithm 1 with the set of points S, the number of centers k, the metric d, and set $(a_i)_{i=1}^k$ to be the returned centers as a sequence. Record the values d_i.
$h_{lo} \leftarrow 0, h_{hi} \leftarrow k$
$G_0 \leftarrow \{V = \{s,t\} \sqcup \{v_f \mid$ each dem. group value $f\}, E = \{(v_f, t),$ capacity $= k_f, r_e = 0 \mid$ each dem. group value $f\}\}$
$G' \leftarrow G_0$
while $h_{lo} \neq h_{hi}$ **do**
    $G \leftarrow G', \ell \leftarrow \lceil (h_{lo} + h_{hi})/2 \rceil$
    for $j \leftarrow h_{lo} + 1$ **to** ℓ **do**
        $G \leftarrow \{V(G) \sqcup \{v_j\}, E(G) \sqcup \{(s, v_j),$ capacity $= 1, r_e = 0\}\}$
        Calculate the closest point to a_j for each demographic group f by a single sweep of S
        for *each group f with* $d(a_j, S_f) \leq d_\ell/2$ **do**
            $G \leftarrow \{V(G), E(G) \sqcup \{(v_j, v_f),$ capacity $= 1, r_{(v_j,v_f)} = d(a_j, S_f)\}$
        end
    end
    Obtain flow F by running Dinic's algorithm on G from s to t
    if $|F| = \ell$ **then**
        $G' \leftarrow G, h_{lo} \leftarrow \ell$
    end
    else
        $h_{hi} \leftarrow \ell - 1$
    end
end
$G' \leftarrow \{V = V(G_0) \sqcup \{v_j \mid j = 1$ to $h_{lo}\}, E(G) \sqcup \{(s, v_j),$ capacity $= 1, r_e = 0 \mid j = 1$ to $h_{lo}\}\}$
$R \leftarrow \emptyset$
for $j = 1$ **to** h **do**
    Calculate the distance (denoted $r_{(v_j,v_f)}$) of the closest point (denoted by $p_{j,f}$) to a_j for each demographic group f by a single sweep on S
    for *each group f with* $r_{(v_j,v_f)} \leq d_{h_{lo}}/2$ **do**
        $R \leftarrow R \sqcup \{(r_{(v_j,v_f)}, p_{j,f})\}$
    end
end
$F' \leftarrow \emptyset$
while $|\{r \mid (r,p) \in R\}| > 1$ **do**
    $G \leftarrow G', (r', p') \leftarrow$ median $r_{(v,f)}$ in R
    for $(r_{(v_j,v_f)}, p_{j,f}) \in R \mid r_{(v_j,v_f)} \leq r'$ **do**
        $G \leftarrow \{V(G), E(G) \sqcup \{(v_j, v_f),$ capacity $= 1, r_e = r_{(v_j,f)},$ label $= p_{j,f}\}\}$
    end
    Obtain flow F by running Dinic's algorithm on G from s to t
    if $|F| = h$ **then**
        $F' \leftarrow F, R \leftarrow R \setminus \{(r,p) \in R \mid r \geq r'\}$
    end
    else
        $G' \leftarrow G, R \leftarrow R \setminus \{(r,p) \in R \mid r \leq r'\}$
    end
end
Obtain C from the labels of edges used in the flow F'
Arbitrarily add centers to C to satisfy the fairness constraint to equality
Return C