Multi-fair capacitated students-topics grouping problem

Tai Le Quy[1*], Gunnar Friege[2], Eirini Ntoutsi[3]

[1] L3S Research Center – Leibniz University Hannover, Germany

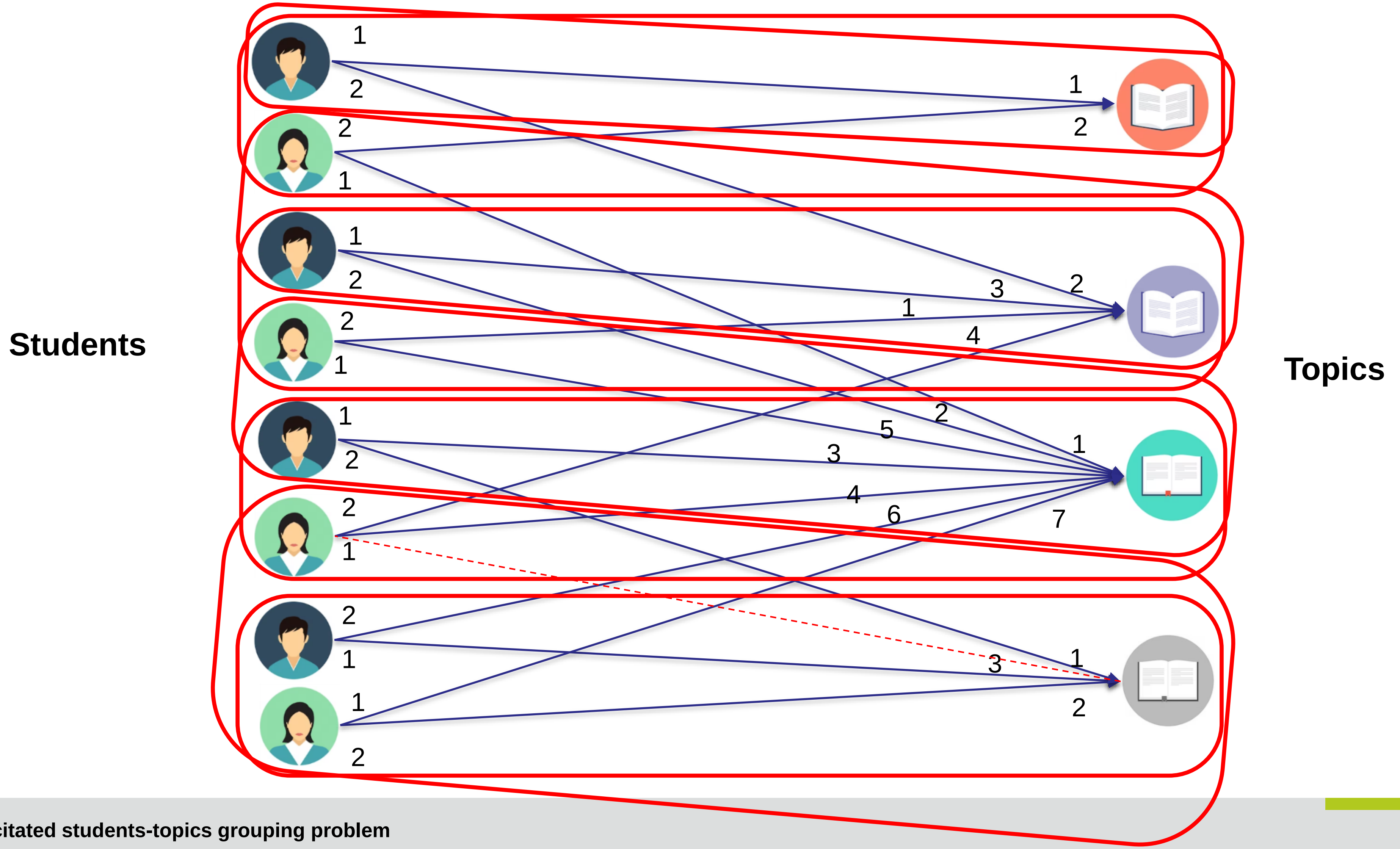
[2] Faculty of Mathematics and Physics – Leibniz University Hannover, Germany

[3] Research Institute CODE – Bundeswehr University Munich, Germany

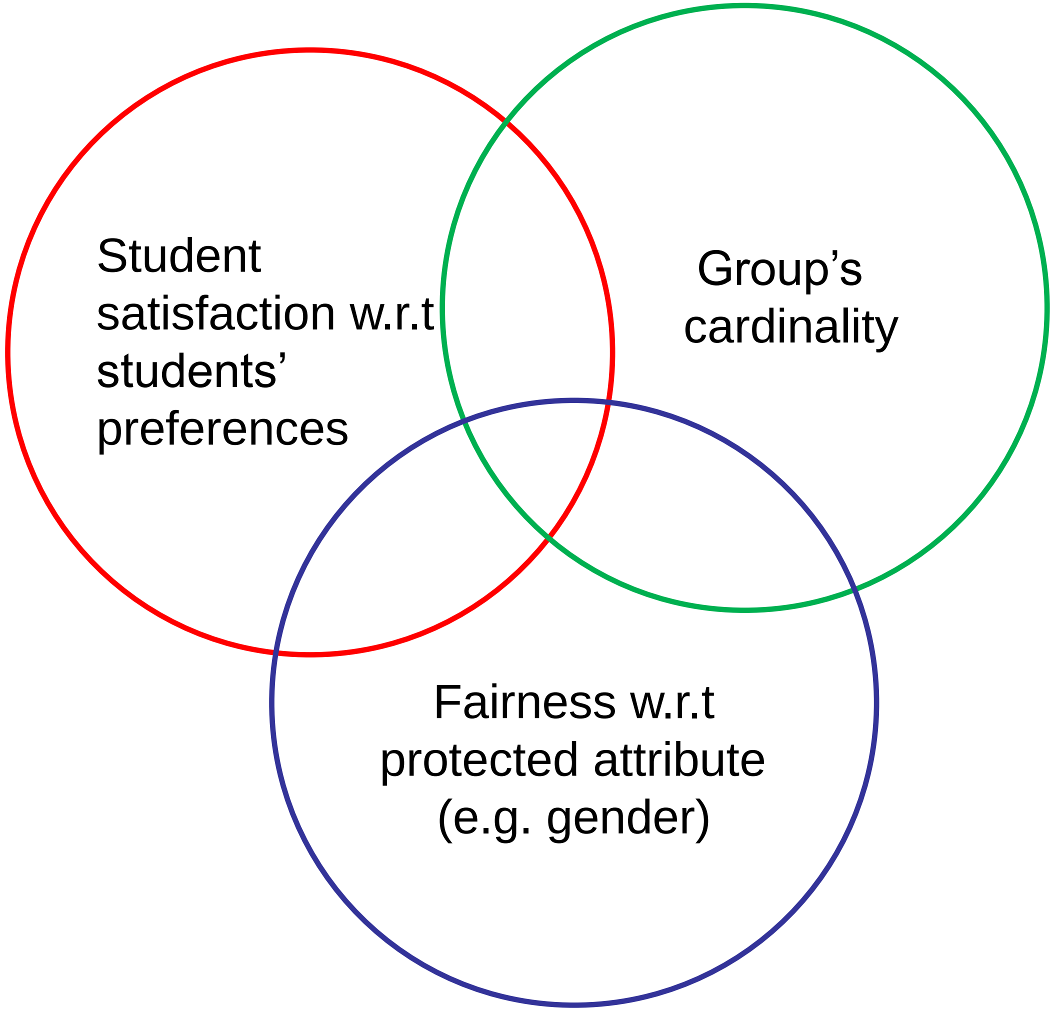
Osaka, 27/05/2023

* Corresponding author: tai@l3s.de

Motivation (1/2)



Motivation (2/2)



Problem definition (1/2)

- X = {x_1,…, x_n}: *n* students, T = {t_1,…, t_m}: a set of *m* topics

a) Matrix $wishes_{n \times h}$ ($h = 3$ preferences, $n = 5$ students)

	$wish_1$	$wish_2$	$wish_3$
x_1	t_1	t_3	t_2
x_2	t_3	t_1	t_4
x_3	t_2	t_3	t_1
x_4	t_4	t_2	t_3
x_5	t_1	t_4	t_3

Students choose *h* topics as their wishes

b) Matrix $V_{n \times m}$ ($m = 4$ topics, $n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	3	1	1.5	0
x_2	1.5	0	3	1
x_3	1	3	1.5	0
x_4	0	1.5	1	3
x_5	3	0	1	1.5

V: level of interest in the topic

c) Matrix $W_{n \times m}$ ($m = 4$ topics, $n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	4	1	3	0
x_2	3	0	5	3
x_3	2	2	4	0
x_4	0	3	2	1
x_5	1	0	1	2

W: time-weight matrix based on registration time

d) Matrix $welfare_{n \times m}$ ($m = 4$ topics, $n = 5$ students)

	t_1	t_2	t_3	t_4
x_1	2.0	0.67	1.1	0
x_2	1.25	0	2.0	1.08
x_3	0.83	1.67	1.3	1.0
x_4	0	1.5	0.73	1.25
x_5	1.25	0	0.53	1.0

$$\boxed{welfare_{ij} = \alpha v_{ij} + \beta w_{ij}}$$

- Protected attribute, e.g., gender, $\psi(x_i) = \{p, \bar{p}\}$, i.e. {female, male}

Problem definition (2/2)

- The goal is to divide all students into k groups $\mathcal{G} = \{G_1, \ldots, G_k\}$, $k \leq m$, which maximizes the objective function:

$$L(X, \mathcal{G}) = \prod_{r=1}^{k} \left(1 + \sum_{i=1}^{n} welfare_{ij_r} * y_{ij_r}\right)$$

$L(X, \mathcal{G})$ is the Nash social welfare function*

  - The group assignment is fair, i.e., maximizing the objective function (students' satisfaction)
  - $balance(G_r)$ is maximized: fairness constraint w.r.t protected attribute
  - $C^l \leq |G_r| \leq C^u$: capacity constraint

where: $J = \{j_1, \ldots, j_k\} = \{j \mid x_i \in G_r, welfare_{ij} > 0\}$, r = 1..k

$$y_{ij_r} = \begin{cases} 1 \; if \; x_i \; is \; assigned \; to \; topic \; t_{j_r} \\ 0 \qquad\qquad\qquad if \; not \end{cases}$$

$$balance(G_r) = \min\left(\frac{\{x \in G_r | \psi(x) = p\}}{\{x \in G_r | \psi(x) = \bar{p}\}}, \frac{\{x \in G_r | \psi(x) = \bar{p}\}}{\{x \in G_r | \psi(x) = p\}}\right)$$

Multi-fair capacitated (MFC) grouping problem

* Fluschnik et al., Fair knapsack. In *AAAI, 2019*

Proposed methods

- Greedy heuristic approach **Student's preferences**
  - Assign students to the most preferred topic among their preferences
- Knapsack-based approach **Group's cardinality**
  - Search the most suitable students for each topic by a maximal knapsack problem
- MFC knapsack approach **MFC constraints**
  - Search the most suitable students for each topic by a new MFC knapsack satisfying constraints of the MFC problem

Greedy heuristic approach

- 2-step approach
  - Assign students to groups
    - Assign students to their most preferred topic
    - If many students choose the same topic, we assign the student with the highest *welfare* value to the topic
  - Group adjustment
    - To satisfy constraints (fairness w.r.t. protected attribute, cardinality).
    - If there are ungrouped students, we will try to assign them to existing groups

Knapsack-based approach (1/2)

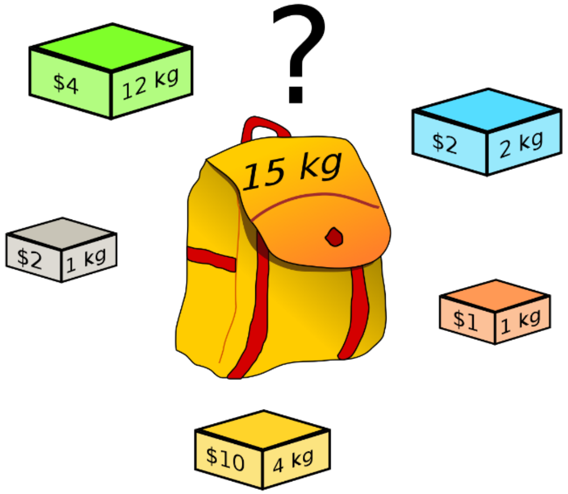
- Select suitable students for a group by a *maximal knapsack* problem
  - For each topic $t_{j_r} \in T$, r is the index of k selected topic $J = \{j_1, j_2, \ldots, j_k\}$, select a subset of students (G_r):

$$\text{maximize} \sum_{i=1} welfare_{ij_r} * y_{ij_r}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^l \end{cases}$$

where $y_{ij_r} = 1$ if x_i is assigned to topic t_{j_r}, otherwise $y_{ij_r} = 0$

  - value ~ welfare, weight ~ capacity



Source: https://en.wikipedia.org/wiki/Knapsack_problem

Knapsack-based approach (2/2)

- 2-step approach
  - Assign students to groups
    - Select suitable candidates among unassigned students by the result of a vanilla maximal knapsack problem
    - Use dynamic programming to solve the knapsack problem
  - Group adjustment
    - Apply the same procedure as in the greedy heuristic approach

MFC knapsack approach (1/3)

- MFC knapsack algorithm
  - Search the group of suitable student w.r.t. MFC constraints: select a subset G_r:

$$\text{maximize} \sum_{i=1}^{n} welfare_{ij_r} * y_{ij_r}$$

$$\text{subject to} \quad \begin{cases} \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^u \text{ or} \\ \sum_{i=1}^{n} capacity_i * y_{ij_r} \leq C^l \\ balance(G_r) \text{ is maximized} \end{cases}$$

MFC knapsack approach (2/3)

- 2-step approach
  - Assign students to groups
    - Select suitable candidates among unassigned students by the result of a group fairness MFC knapsack problem
    - Use dynamic programming to solve the MFC knapsack problem (inspired by knapsack problem with group fairness constraints of Patel et al. (2021)*
  - Group adjustment
    - Apply the same procedure as in the greedy heuristic approach

* Patel, D., Khan, A., & Louis, A. (2021). Group fairness for knapsack problems. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* (Vol. 2, pp. 989-997).

MFC knapsack approach (3/3)

Algorithm 4: MFC knapsack algorithm

Input: $\mathcal{S} = \{x_1, x_2, \dots, x_z\}$: a set of unassigned students; C^l, C^u: capacities; $welfare_{n\times m}$: a welfare matrix; θ: balance score

Output: An optimal total welfare value

$avg = \dfrac{\sum_{i=1}^{n} welfare_{ij_r}}{(C^l + C^u)/2}$;

Let $\mathcal{A}(p, s, w), \forall p \in \{0, 1\}$, be the total welfare of the first s students in the set $\mathcal{S}$ with capacity w on group p ;

Initialize $\mathcal{A}(p, 0, w) \leftarrow 0$; $\mathcal{A}(p, s, 0) \leftarrow 0$;

$\boxed{\mathcal{A}(p, s, w) \leftarrow max\{\mathcal{A}(p, s-1, w), \mathcal{A}(p, s-1, w-1) + \sum_{i=1}^{s} welfare_{ij_r}\}}$;

Let $\mathcal{B}(p, w)$ be the total welfare of group p with capacity w;

$p_0^l \leftarrow \left[\dfrac{C^l}{\frac{1+\theta}{\theta}} \right]$; $p_0^u \leftarrow \left[\dfrac{C^u}{\frac{1+\theta}{\theta}} \right]$; $S_0 \leftarrow \{x \in \mathcal{S} | \varphi(x) = 0\}$; $S_1 \leftarrow \{x \in \mathcal{S} | \varphi(x) = 1\}$;

$\mathcal{B}(0, w) \leftarrow max\{\mathcal{A}(0, |S_0|, w) | p_0^l \le w \le p_0^u\}$;

$\mathcal{B}(1, w) \leftarrow max\{\mathcal{B}(0, w') + \mathcal{A}(1, |S_1|, w - w') | C^l - p_0^l \le w - w' \le C^u - p_0^u, p_0^l \le w' \le p_0^u$, and $\frac{w'}{w - w'} \ge \theta\}$;

return $argmax\{\mathcal{B}(1, w) | min\{\mathcal{B}(1, w) - avg\}\}$;

The total welfare of the first s students in the set $\mathcal{S}$ with capacity w on group p ∈ {0,1}

The total welfare with capacity w w.r.t. the protected attribute

Evaluation

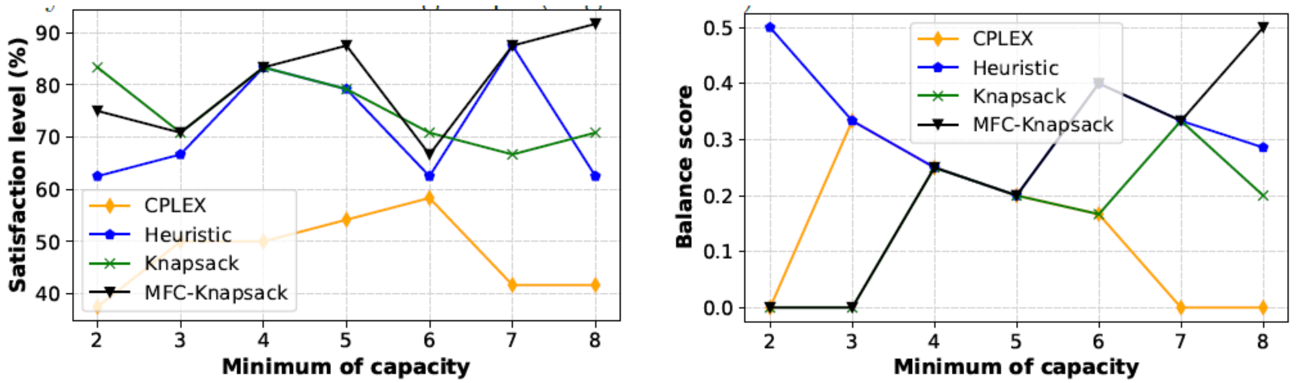
- Dataset
  - Real data science dataset: Students have to register 3 desired topics out of 16 topics
  - Student performance: generate student's preferences (semi-synthetic dataset)

Dataset	#instances	#attributes	Protected attribute	Balance score
Real data science	24	23	Gender (F: 8, M: 16)	0.5
Student-Mathematics	395	33	Gender (F: 208, M: 187)	0.899
Student-Portuguese	649	33	Gender (F: 383; M: 266)	0.695

- Measures
  - Nash social welfare
  - Balance
  - Satisfaction level: $$Satisfaction = \frac{\mid \{i \mid wishes_{ip} = k, i \in groups_k, p \in [h]\} \mid}{n}$$
- Baseline
  - The CPLEX integer programming model

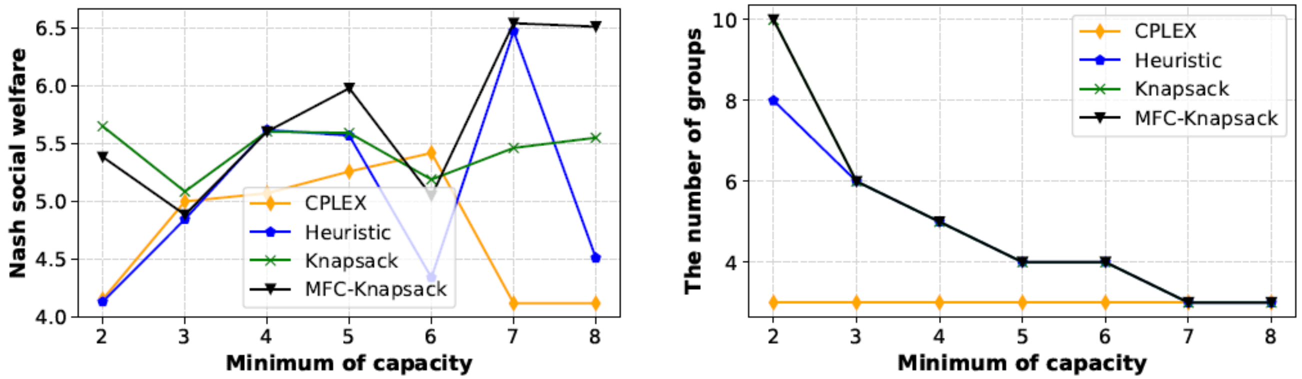
Experimental results (1/3)

- The MFC knapsack method is better:
  - In terms of the Nash social welfare and satisfaction level
  - When a group has at least 4 people
- CPLEX fails to assign students while maintaining only a constant number of groups



(a) Satisfaction level of students' preferences (higher is better)

(b) Balance score w.r.t. Gender (higher is better)

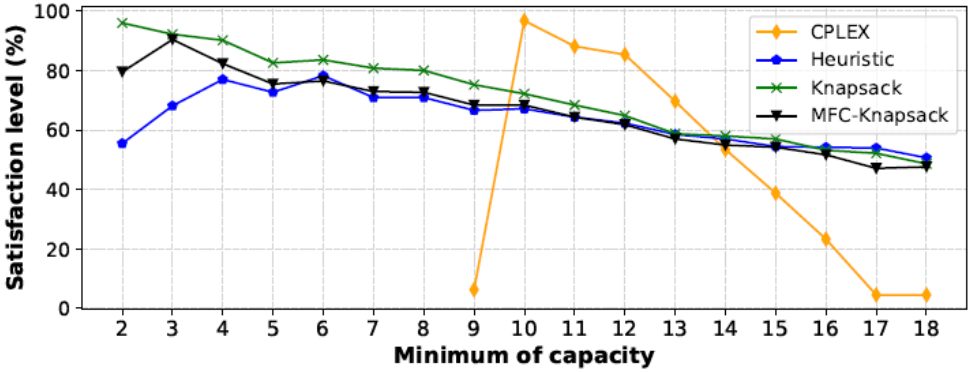


(c) Nash social welfare (higher is better)

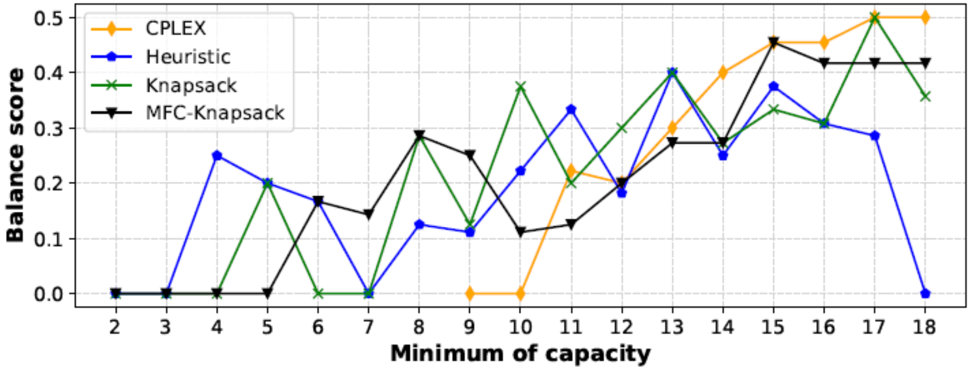
(d) Number of groups

Performance of methods on the real data science dataset

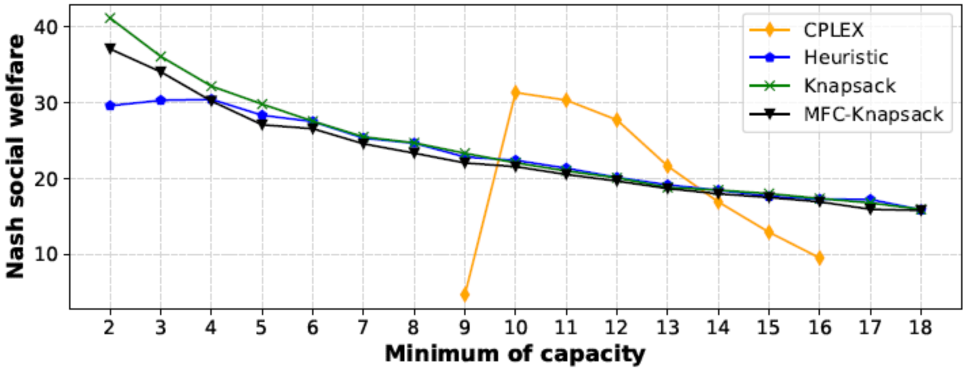
Experimental results (2/3)



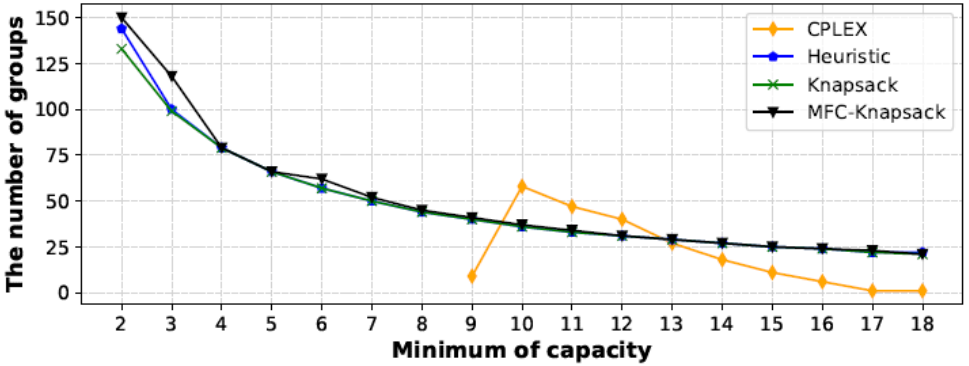
(a) Satisfaction level of students' preferences (higher is better)



(b) Balance score w.r.t. Gender (higher is better)



(c) Nash social welfare (higher is better)

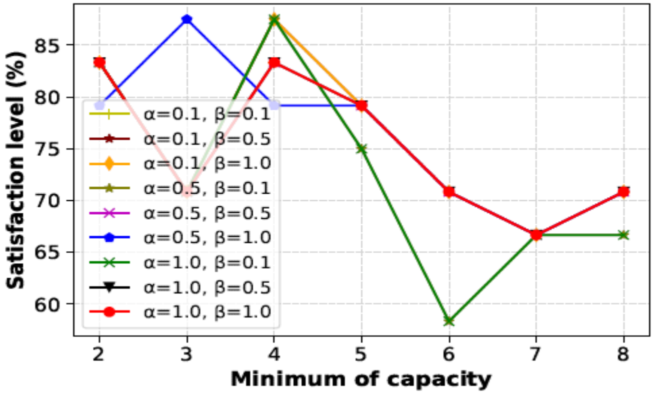


(d) Number of groups

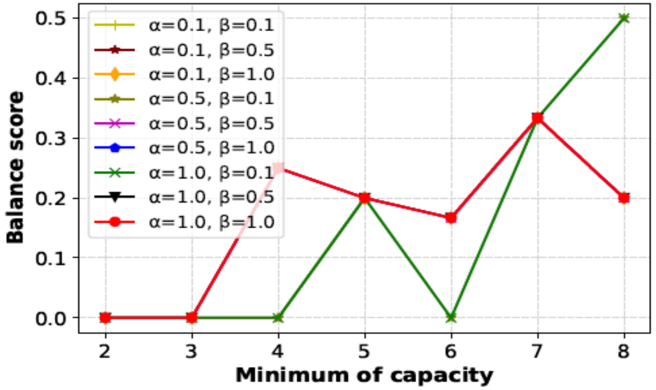
Performance of methods on Student performance – Mathematics dataset

Experimental results (3/3)

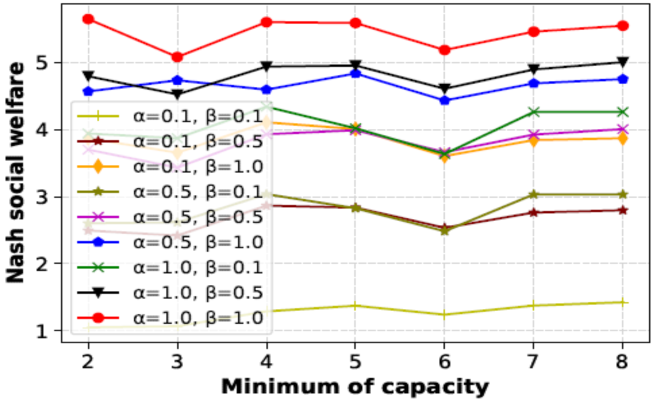
- In all datasets, the knapsack-based model shows the best performance with α = 1.0 and β = 1.0



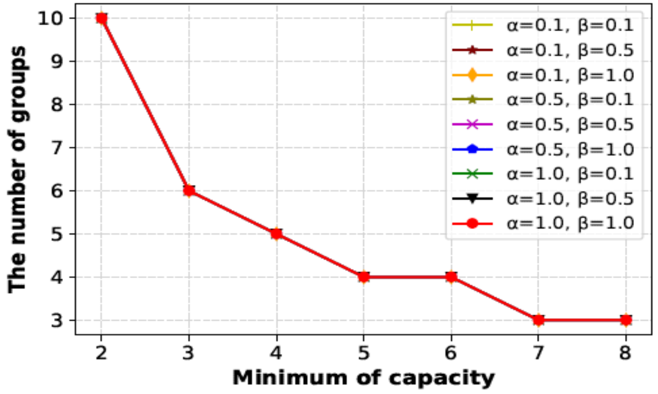
(a) Satisfaction level of students' preferences (higher is better)



(b) Balance score w.r.t. Gender (higher is better)



(c) Nash social welfare (higher is better)



(d) Number of groups

Real data science: Impact of α, β parameters on the knapsack-based model

Conclusion

- We introduced the MFC grouping problem:
  - Ensures fairness in multiple aspects: i) student satisfaction and ii) protected attribute
  - Maintains groups' cardinality within the given bounds.
- We proposed three methods:
  - The greedy heuristic approach
  - The knapsack-based approach
  - The MFC knapsack approach
- The experiments show that our methods are effective regarding student satisfaction and fairness w.r.t. the protected attribute while maintaining cardinality within the given bounds.

PAKDD 2023

Thank you for your attention!

LernMINT

The work is supported by LernMINT - The Lower Saxony doctoral program (Ministry for Science and Culture, Lower Saxony, funding period 2019-2024)
https://lernmint.org/

tailequy
tai@l3s.de
tailequy.github.io