FairDen: Fair Density-based Clustering

Lena Krieger*[1,2], Anna Beer*[3], Pernille Matthews[4], Anneka Myrup Thiesson[4], Ira Assent[1,4]

International Conference on Learning Representations 2025

[1] IAS-8, Forschungszentrum Jülich
[2] Institute for Computer Science, LMU Munich
[3] Faculty of Computer Science, University of Vienna
[4] Department of Computer Science, Aarhus University
*Equal Contribution.

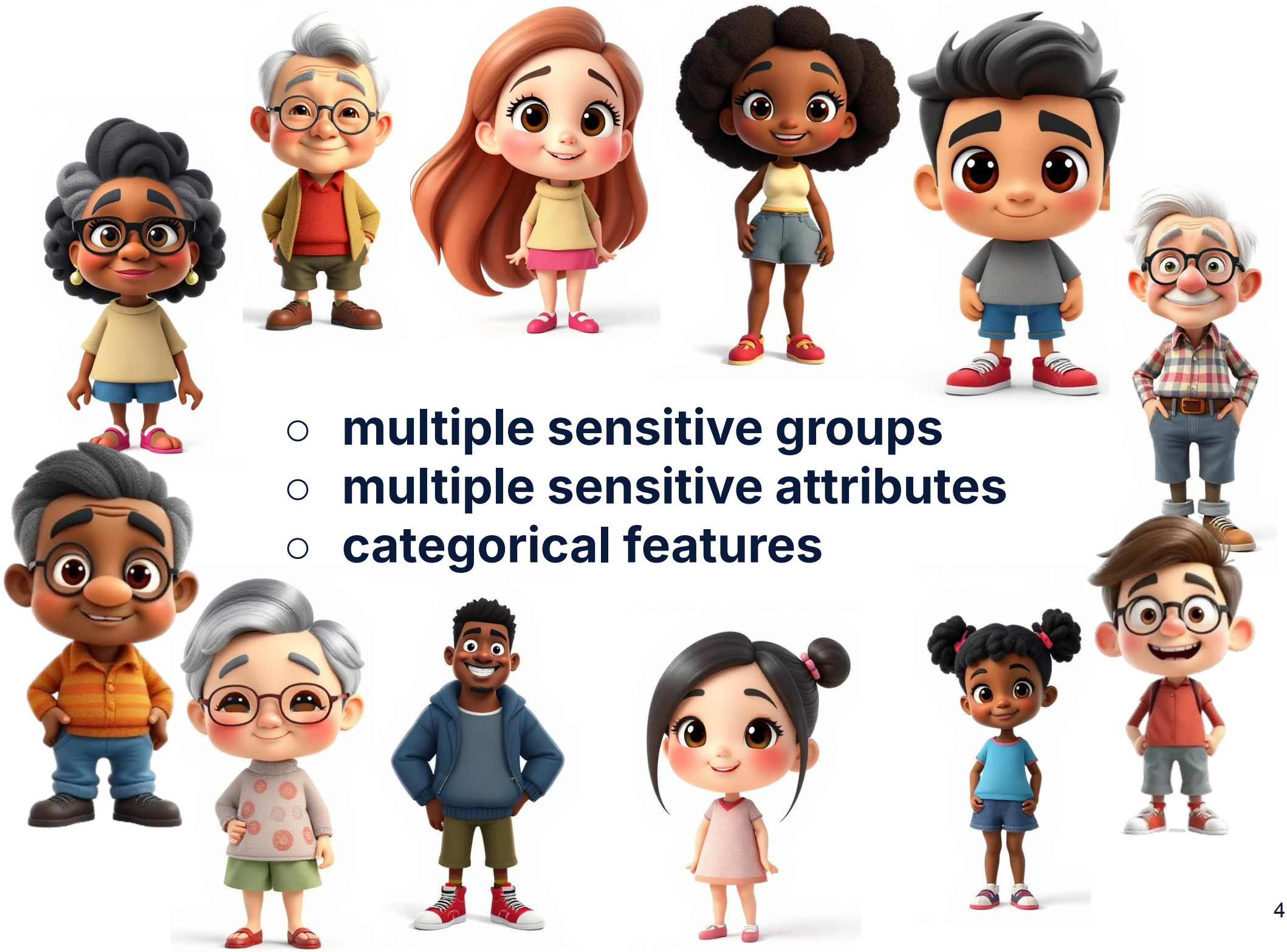
ICLR AARHUS UNIVERSITY universität wien LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN JÜLICH Forschungszentrum

Group-Level Fairness: Non-binary groups

Gender

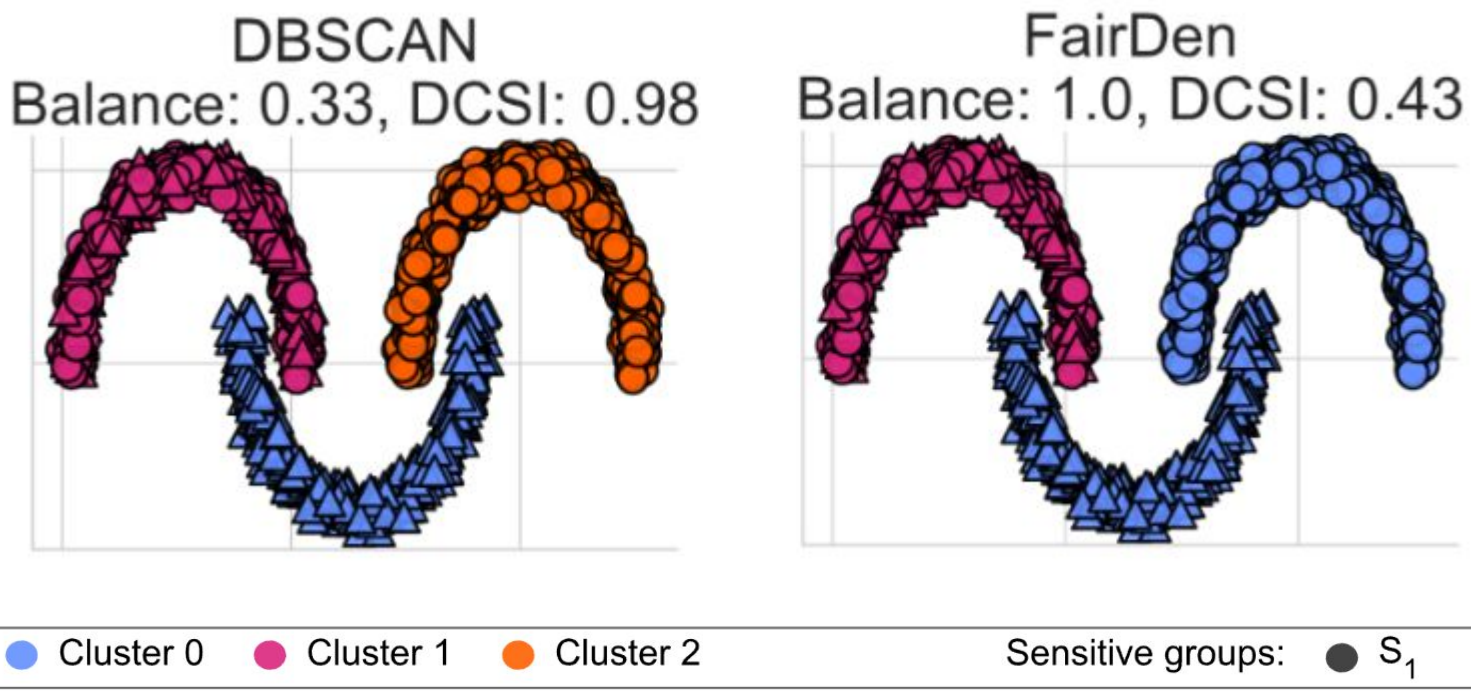
Age

Group-Level Fairness: Multiple sensitive attributes



- multiple sensitive groups
- multiple sensitive attributes
- categorical features

Group-level fairness for density-based clusters



Use the density-connectivity distance [1] to capture structure.

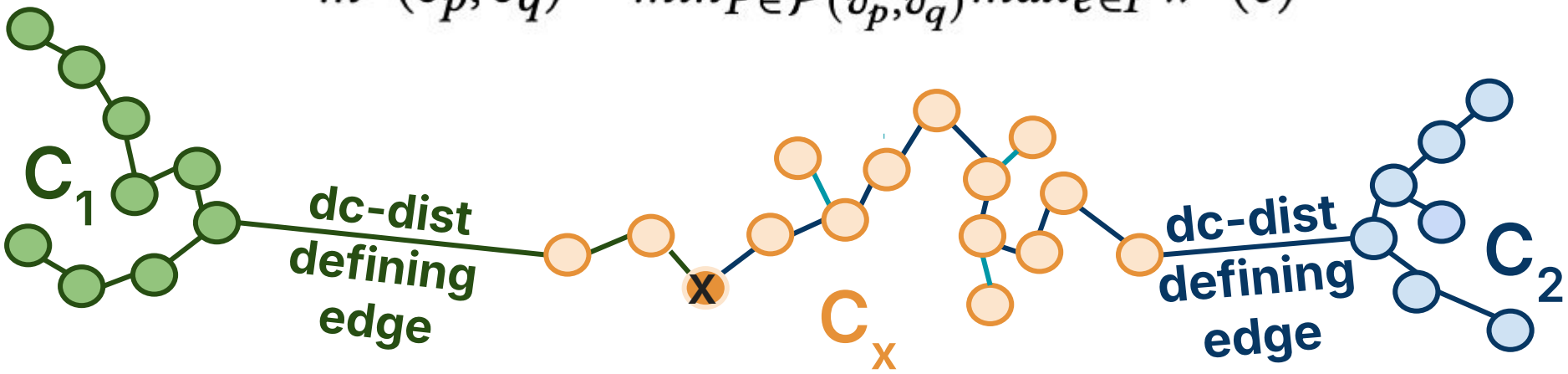
$$d_{dc}^{\mu}(p,q) = m^{d_r^{\mu}}(v_p, v_q) = min_{P \in \mathcal{P}(v_p,v_q)} max_{e \in P} w^{d_r^{\mu}}(e)$$

Mutual Reachability distance

$$d_r^{\mu}(p,q) = max(d_E(p,q), d_{core}^{\mu}(p), d_{core}^{\mu}(q))$$

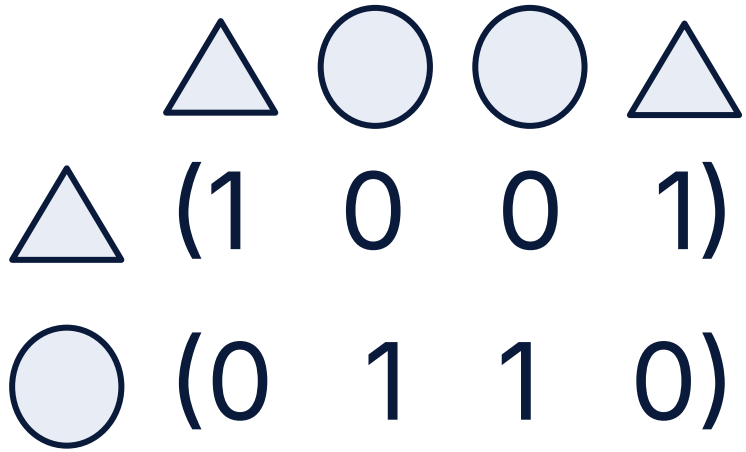
Minimax Path

$$m^{\delta}(v_p, v_q) = min_{P \in \mathcal{P}(v_p,v_q)} max_{e \in P} w^{\delta}(e)$$



Construct fairness matrix [2] to ensure balance.

- Binary encoding for each *sensitive group:*



- Columns of fairness matrix:

$$\triangle\,(1 \quad 0 \quad 0 \quad 1) - \frac{|\triangle|}{n} \cdot \mathbf{1}_n$$

Combining Density-Connectivity and Fairness leads to a problem we can solve with spectral clustering.

Density-connectivity objective

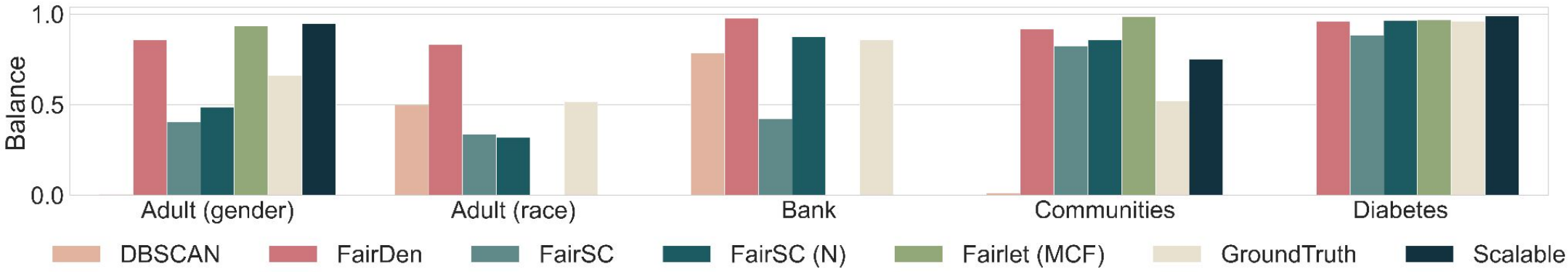
$$\min_{\mathcal{H} \in \mathbb{R}^{n \times k}} Tr(\mathcal{H}^\top \mathcal{L} \mathcal{H})$$

$$\text{subject to } \mathcal{H}^\top \mathcal{D} \mathcal{H} = \mathcal{I}_k$$

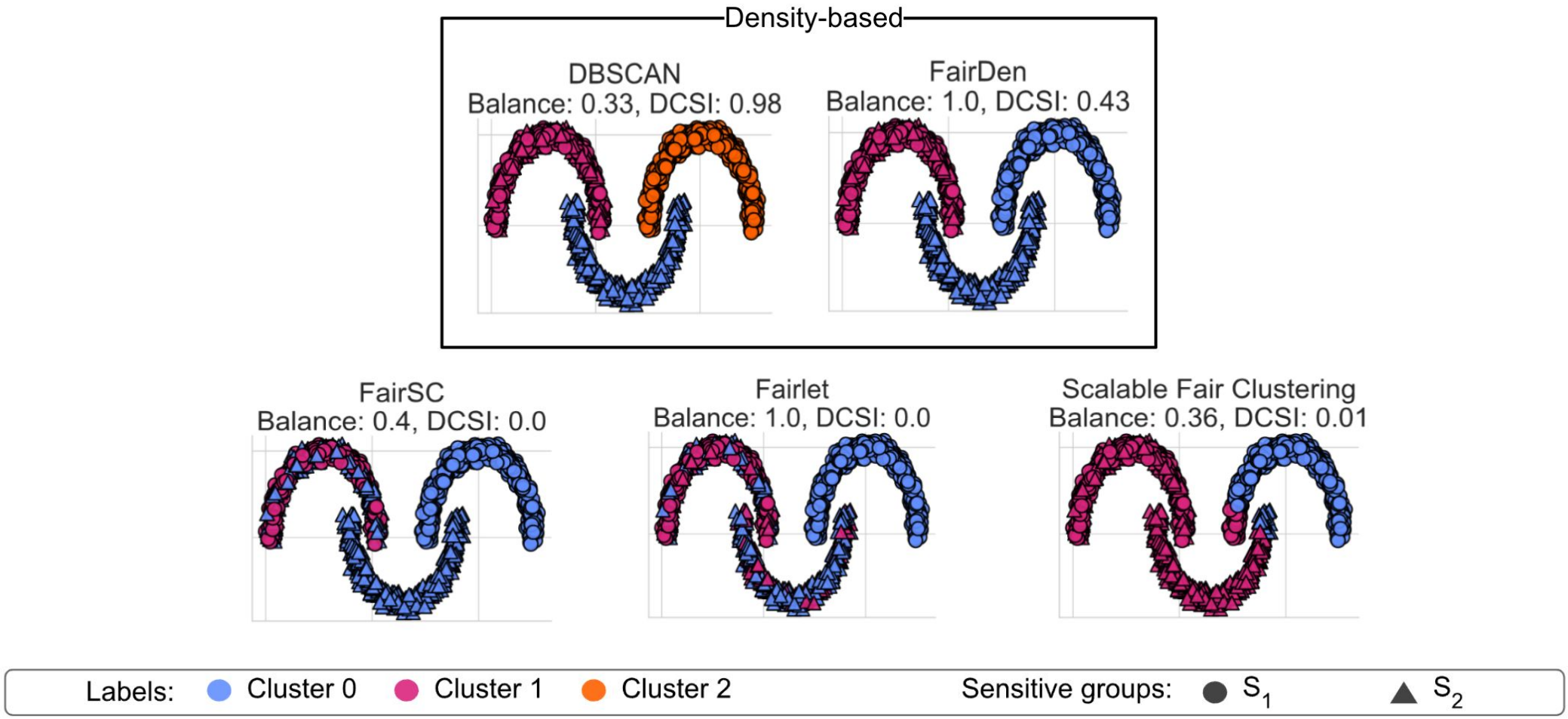
Fairness Constraint

$$\text{and } \mathcal{F}^\top \mathcal{H} = 0$$

FairDen achieves balanced clusterings w.r.t. sensitive attributes.



FairDen detects density-based clusters better than other fair methods.



FairDen

Fair Density-based Clustering

Lena Krieger
l.krieger@fz-juelich.de
Anna Beer
anna.beer@univie.ac.at
Pernille Matthews
matthews@cs.au.dk
Anneka Myrup Thiesson
amt@post.au.dk
Ira Assent
ira@cs.au.dk

Code

Paper

FairDen is the first **density-based group-level fair** clustering algorithm.

It can even cluster data with:

- **multiple sensitive groups**
- **multiple sensitive attributes**
- **categorical features**

References

[1] Anna Beer, Andrew Draganov, Ellen Hohma, Philipp Jahn, Christian MM Frey, and Ira Assent. "Connecting the Dots--Density-Connectivity Distance unifies DBSCAN, k-Center and Spectral Clustering." In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 80-92. 2023.

[2] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. "Guarantees for spectral clustering with fairness constraints." In *International conference on machine learning*, pp. 3458-3467. PMLR, 2019.

Pictures of people generated with https://deepai.org/machine-learning-model/3d-cartoon-generator

ICLR AARHUS UNIVERSITY universität wien LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN JÜLICH Forschungszentrum