Optimal Decision Trees for Interpretable Clustering with Constraints

POUYA SHATI, ELDAN COHEN, SHEILA MCILRAITH

UNIVERSITY OF TORONTO

VECTOR INSTITUTE FOR ARTIFICIAL INTELLIGENCE

CP'23 DOCTORAL PROGRAM

ORIGINALLY APPEARED AT IJCAI-23

Overview

Constrained Clustering

- Semi-supervised

Decision Trees

- Interpretable Classifiers

Decision Tree Clustering

- No constraint support
- No optimality guarantee

Overview

Constrained Clustering

Semi-supervised

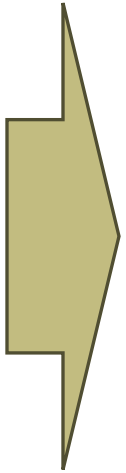
Decision Trees

Interpretable Classifiers

Decision Tree Clustering

No constraint support

No optimality guarantee

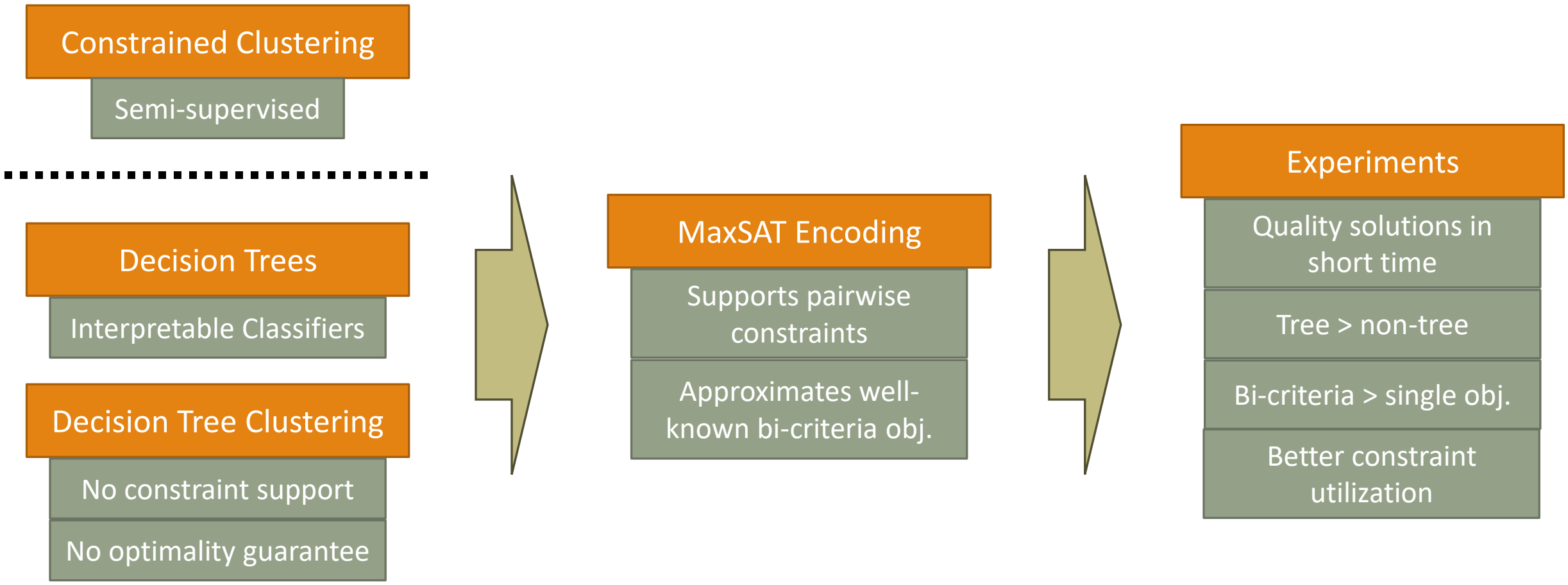


MaxSAT Encoding

Supports pairwise constraints

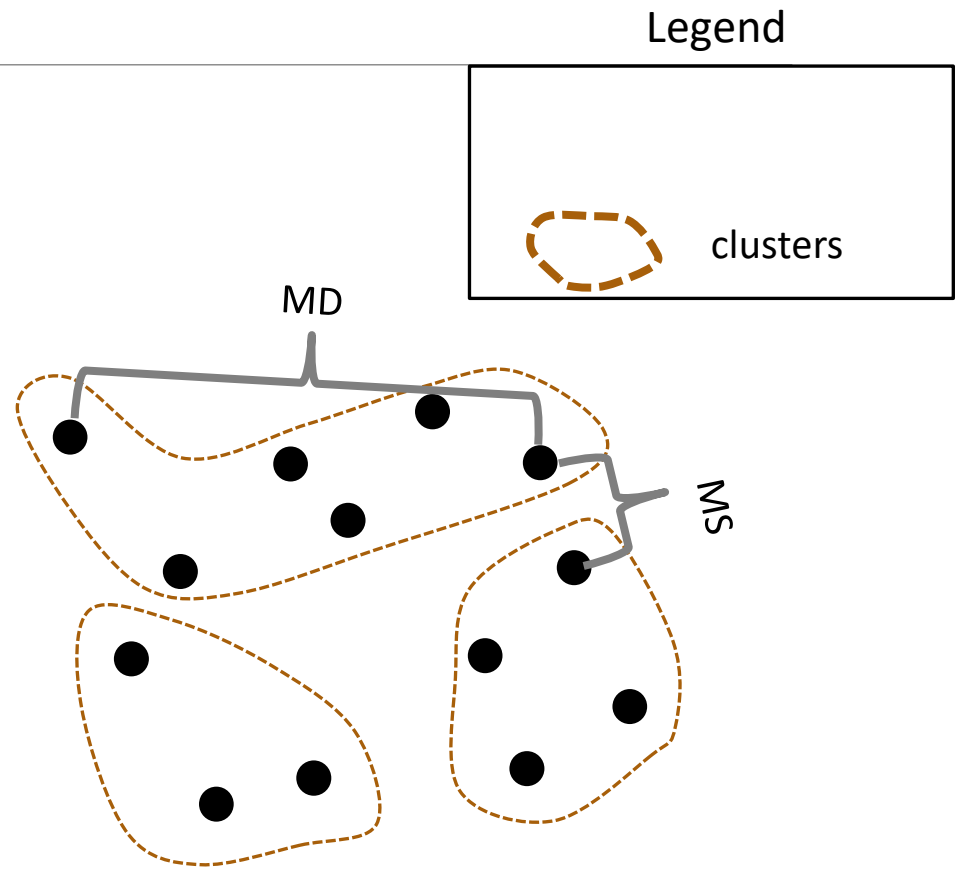
Approximates well-known bi-criteria obj.

Overview


Constrained Clustering
Semi-supervised
Decision Trees
Interpretable Classifiers
Decision Tree Clustering
No constraint support
No optimality guarantee
MaxSAT Encoding
Supports pairwise constraints
Approximates well-known bi-criteria obj.
Experiments
Quality solutions in short time
Tree > non-tree
Bi-criteria > single obj.
Better constraint utilization

Problem Definition

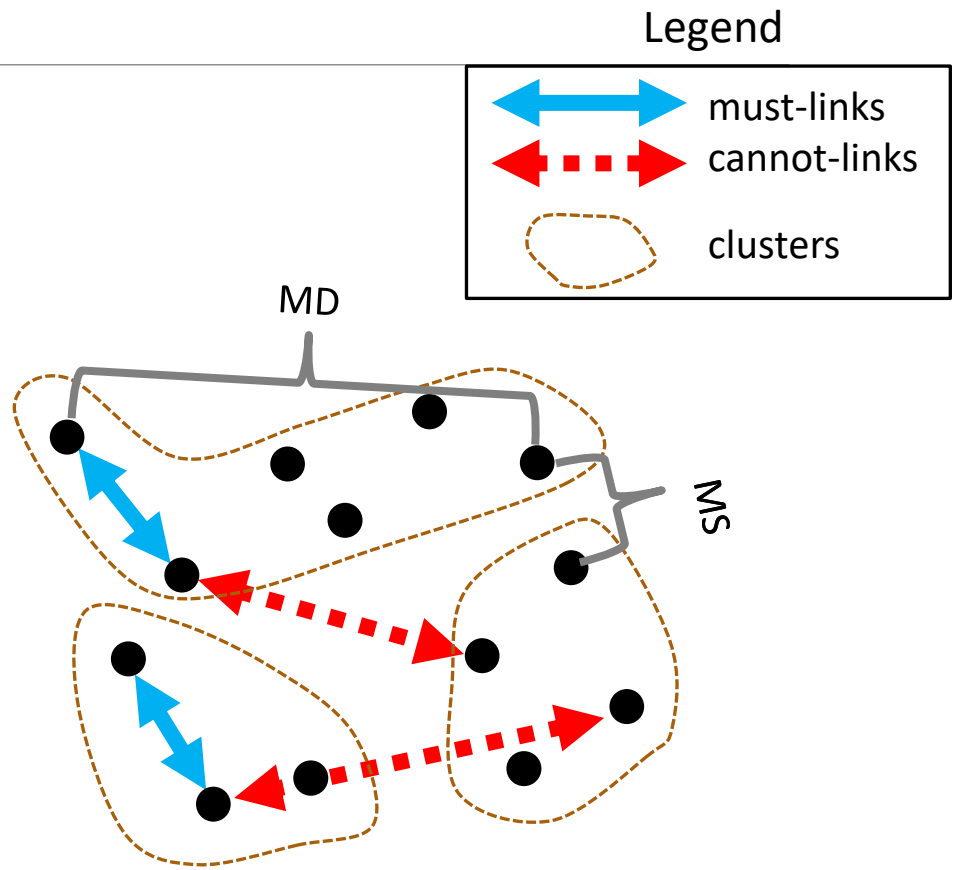
- **Bi-criteria** objective:
  - **Maximize** minimum split **(MS)** between clusters
  - **Minimize** maximum diameter **(MD)** within clusters



Problem Definition

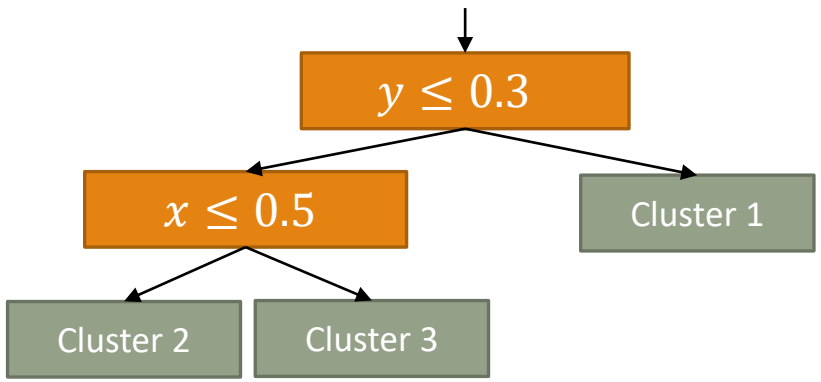
- **Bi-criteria** objective:
  - **Maximize** minimum split **(MS)** between clusters
  - **Minimize** maximum diameter **(MD)** within clusters

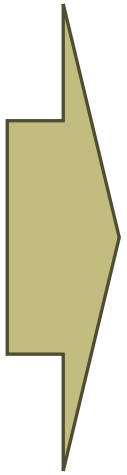
- **Pairwise** Constraints:
  - **Must-links:** pairs that should be in the same cluster
  - **Cannot-links:** pairs that should be in different clusters

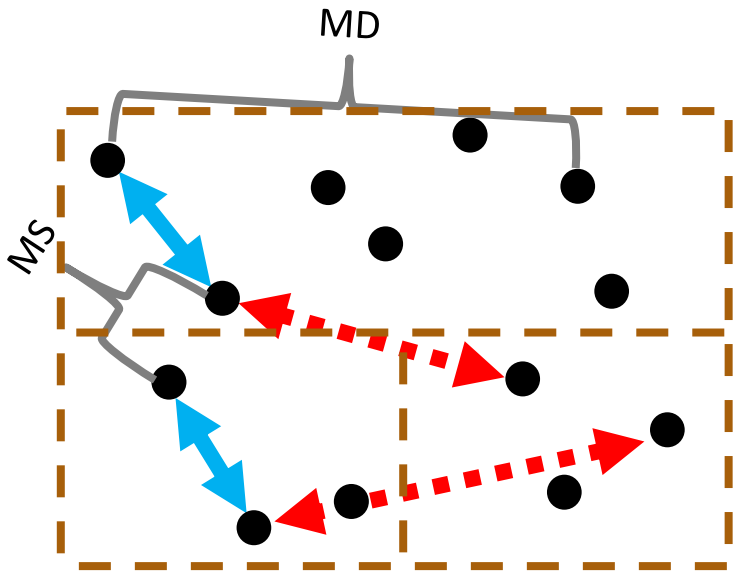


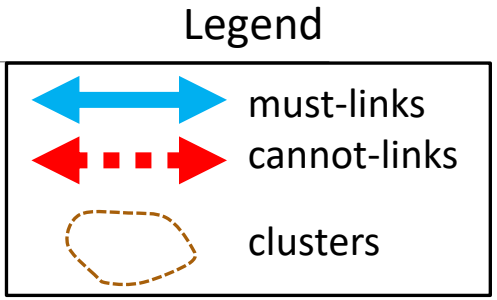
Problem Definition

- **Decision tree** clustering:



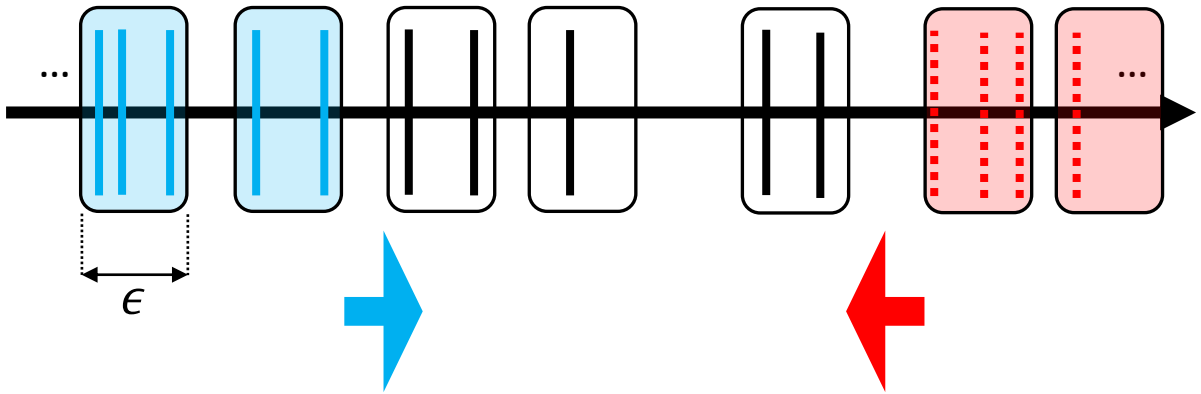






Encoding

- All distances of pairs sorted into **distance classes**



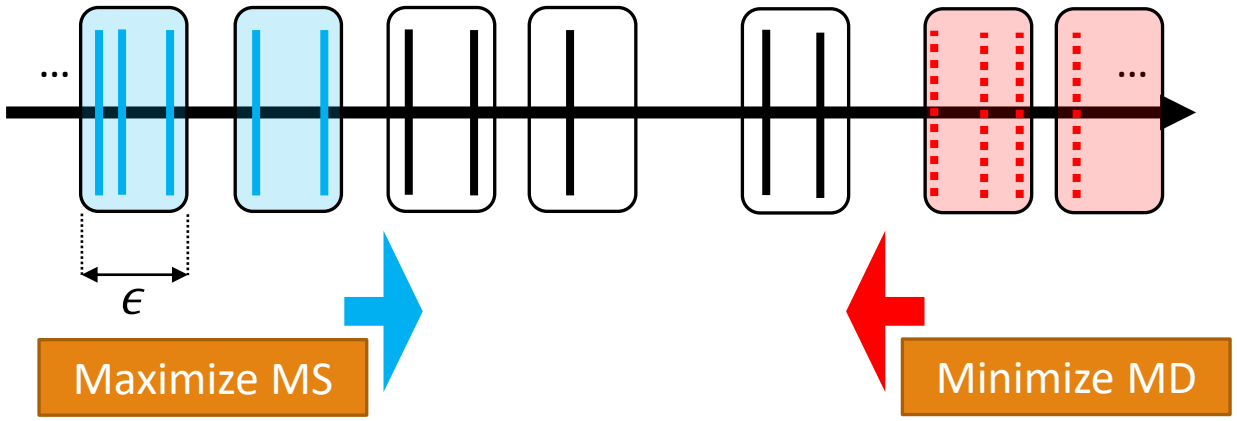
Legend

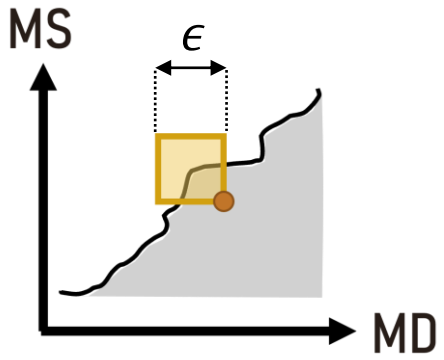
same cluster

diff. clusters

Encoding

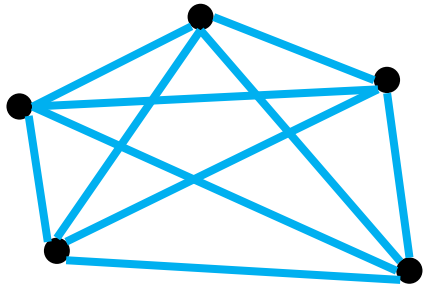
- All distances of pairs sorted into **distance classes**

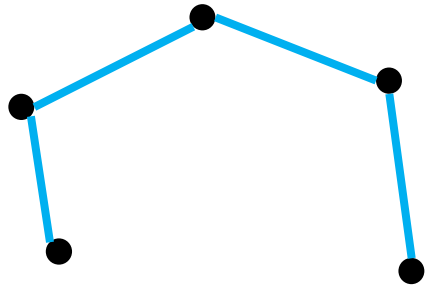




Smart Pairs

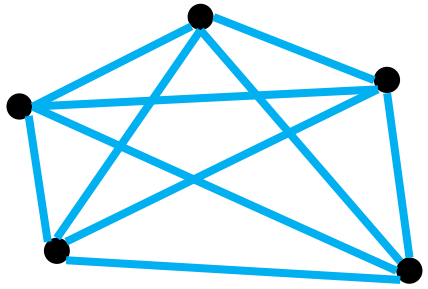
- **Linear** number of clauses enough to enforce **quadratic** number of must-links

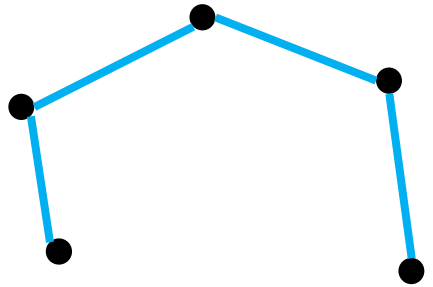




Smart Pairs

- **Linear** number of clauses enough to enforce **quadratic** number of must-links





	Constraints	Conditional
Force to be in the same cluster	Must-link	MS obj.
Force to be in different clusters	Cannot-link	MD obj.

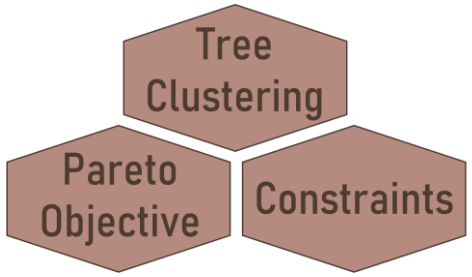
Smart Pairs

Detect **redundant** edges

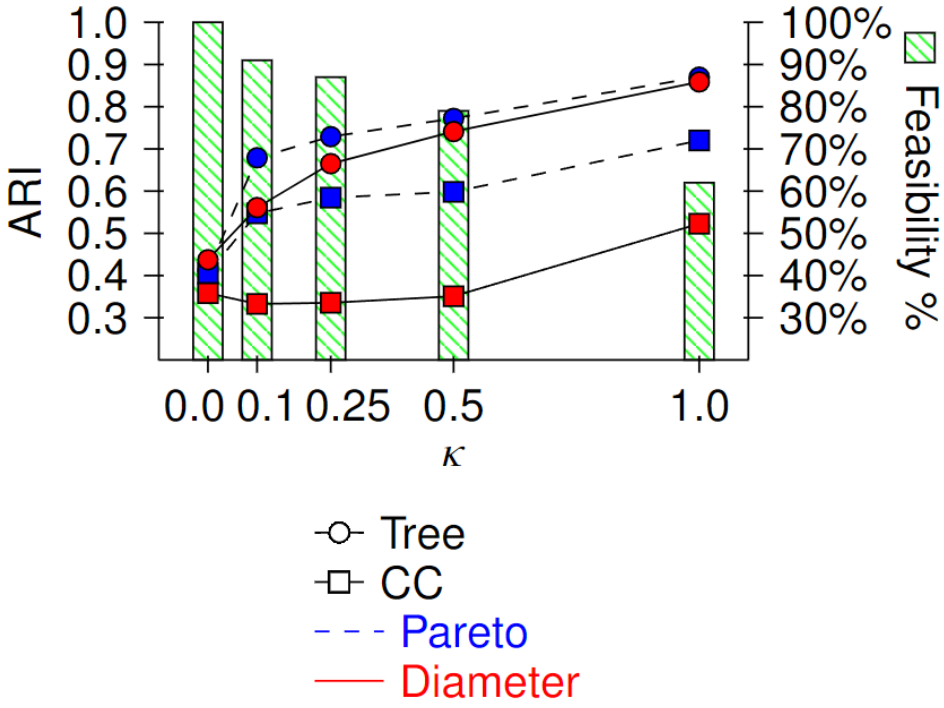
Detect **infeasible** edges

Better Score + Better Interpretability

- High quality solutions in a short time



- Improve the solution individually
- Complement each other



- Trade-off between quality and feasibility

Optimal Decision Trees for Interpretable Clustering with Constraints

Thank you for your time!

POUYA SHATI, ELDAN COHEN, SHEILA MCILRAITH

UNIVERSITY OF TORONTO

VECTOR INSTITUTE FOR ARTIFICIAL INTELLIGENCE

CP'23 DOCTORAL PROGRAM

ORIGINALLY APPEARED AT IJCAI-23