The 34th ACM International Conference on Information and Knowledge Management (CIKM 2025)

FairAD: Computationally Efficient Fair Graph Clustering via Algebraic Distance

Minh Phu Vuong[1], Young-Ju Lee[1], Iván Ojeda-Ruiz[2], and Chul-Ho Lee[1]

[1]Texas State University

[2]Lamar University

Introduction

- Machine learning (ML) has become an integral part of modern life, influencing various aspects of technology, finance, healthcare, and law enforcement.

Financial Risk Analysis

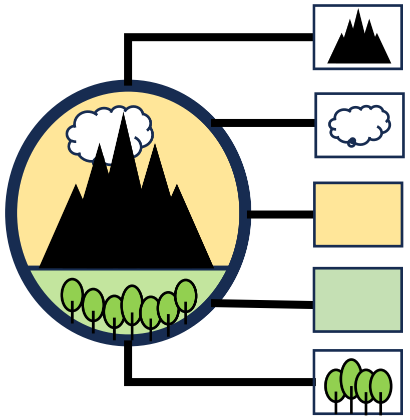
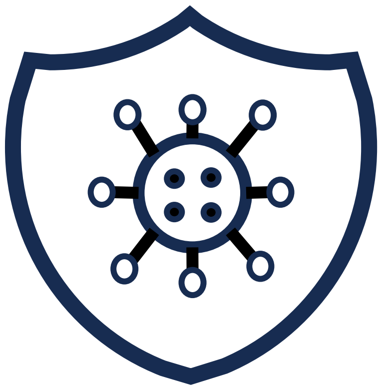
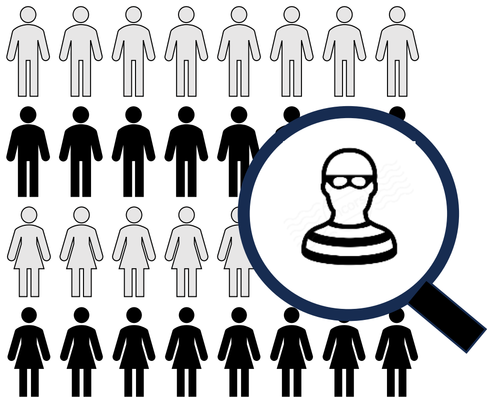


Image Segmentation



Epidemic Control



Law Enforcement

Motivation

Amazon, Apple, Google, IBM, and Microsoft worse at transcribing black people's voices than white people's with AI voice recognition, study finds

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Gender Bias In AI: Addressing Technological Disparities

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

Artificial Intelligence has a gender bias problem – just ask Siri

Racially-biased medical algorithm prioritizes white patients over black patients

The algorithm was based on the faulty assumption that health care spending is a good proxy for wellbeing. But there seems to be a quick fix.

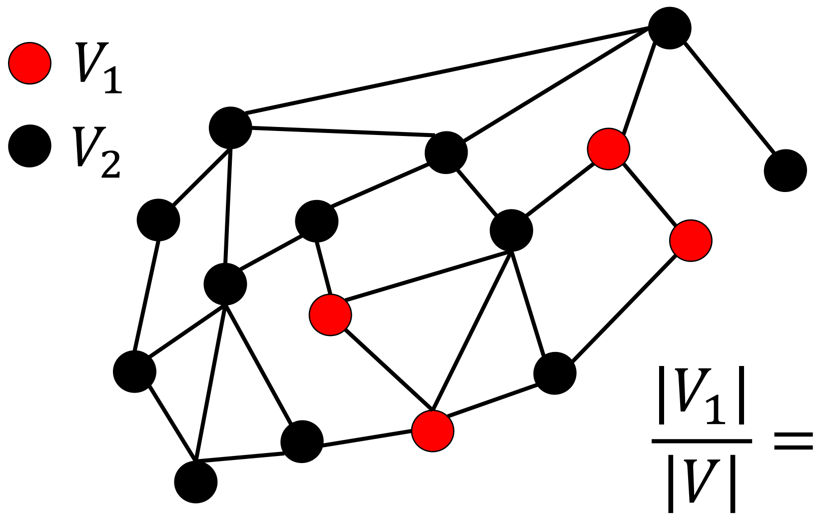
AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

Problem Formulation

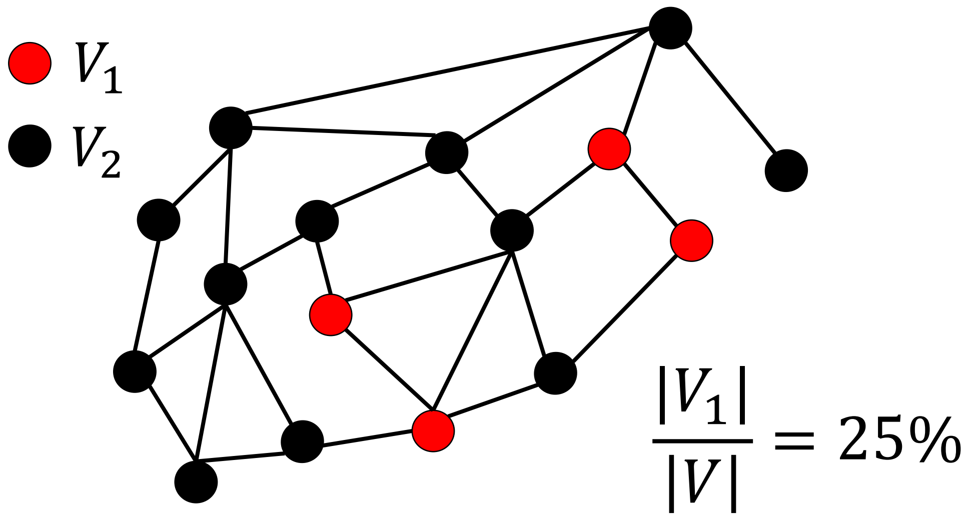
- Fair graph clustering: Partition a graph such that the distribution of protected groups within each cluster is the same as their distribution in the entire graph (while minimizing the cut between clusters).

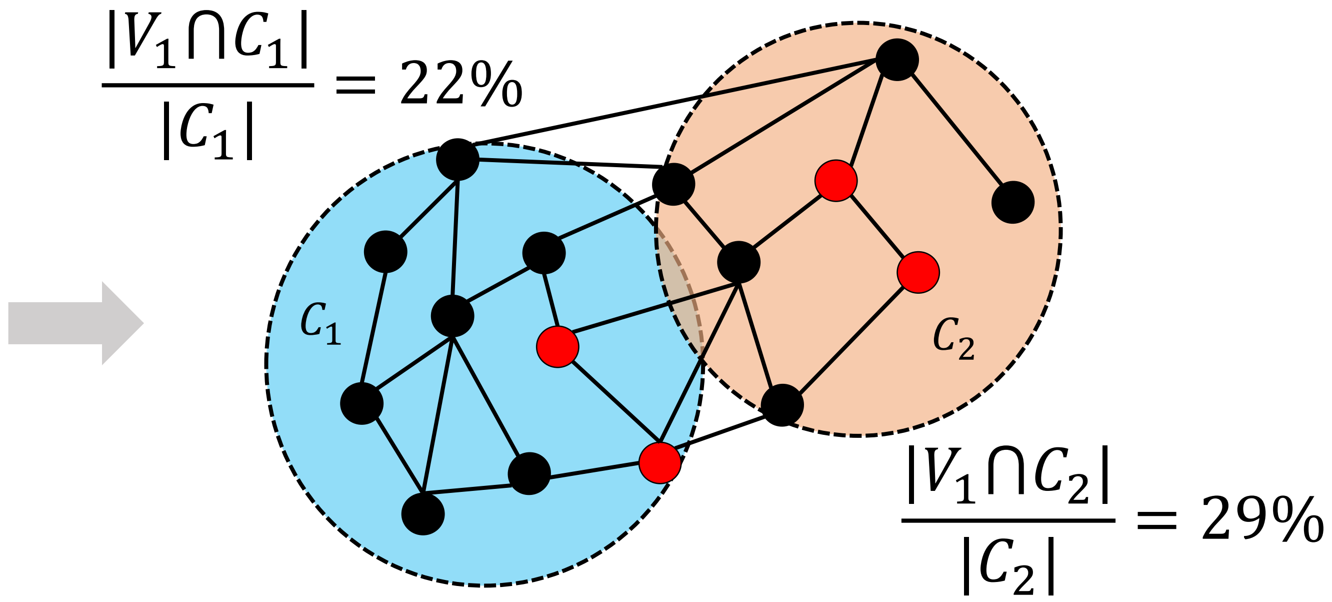


$$\frac{|V_1|}{|V|} = 25\%$$

Problem Formulation

- Fair graph clustering: Partition a graph such that the distribution of protected groups within each cluster is the same as their distribution in the entire graph (while minimizing the cut between clusters).

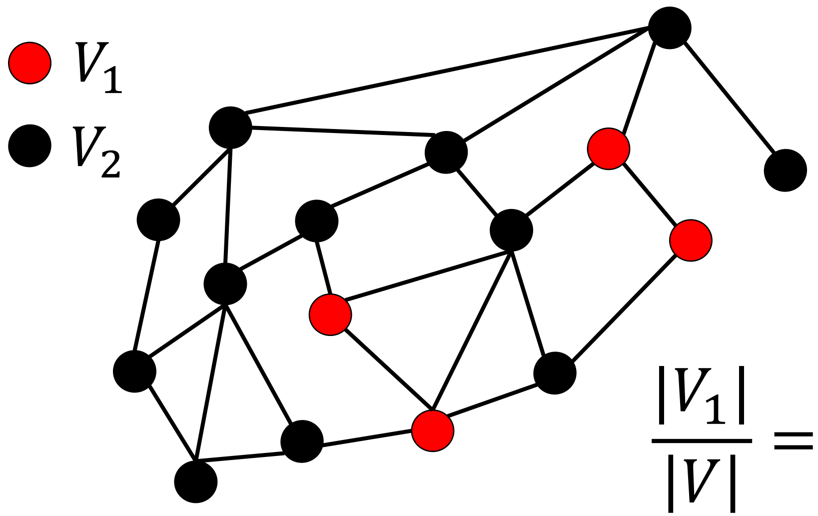




TEXAS STATE UNIVERSITY

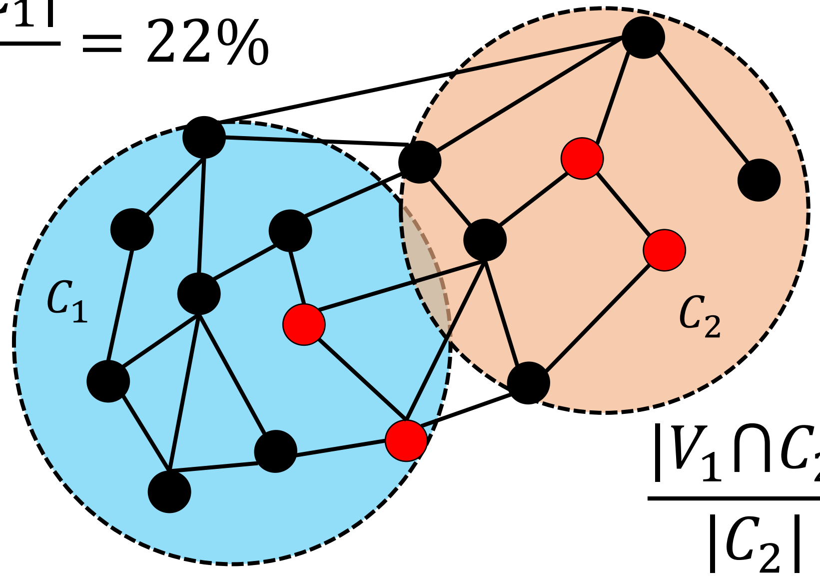
Problem Formulation

- Fair graph clustering



$$\frac{|V_1|}{|V|} = 25\%$$

$$\frac{|V_1 \cap C_1|}{|C_1|} = 22\%$$



$$\frac{|V_1 \cap C_2|}{|C_2|} = 29\%$$

$$\boxed{{}^{*}\text{Goal: } \frac{|V_s \cap C_l|}{|C_l|} = \frac{|V_s|}{|V|}}$$

* For every group and every cluster.

Problem Formulation

- Fairness as linear constraints (for two groups and two clusters)

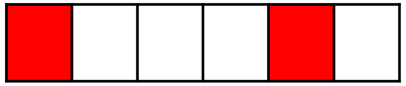
$$\boldsymbol{f}^\top \boldsymbol{v} = 0$$

Problem Formulation

- Fairness as linear constraints (for two groups and two clusters)

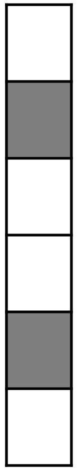
$$\boldsymbol{f}^\top \boldsymbol{v} = 0$$

Fairness matrix



$$f_i = \begin{cases} 1 - \frac{|V_1|}{|V|}, \text{if } i \in V_1 \\ -\frac{|V_1|}{|V|}, \text{otherwise} \end{cases}$$

Cluster indicator $\boldsymbol{v}$ for C_1

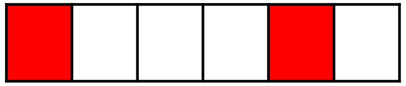


Problem Formulation

- Fairness as linear constraints (for two groups and two clusters)

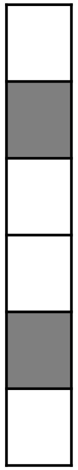
$$\boxed{\boldsymbol{f}^{\top}}\boxed{\boldsymbol{v}} = 0$$

Fairness matrix



$$f_i = \begin{cases} 1 - \dfrac{|V_1|}{|V|}, \text{if } i \in V_1 \\ -\dfrac{|V_1|}{|V|}, \text{otherwise} \end{cases}$$

Cluster indicator $\boldsymbol{v}$ for C_1



$$\boxed{\frac{|V_1 \cap C_1|}{|C_1|} = \frac{|V_1|}{|V|}}$$

The same for C_2

Problem Formulation

- Fair graph clustering problem becomes

$$\min_{\mathbf{V}\in\mathbb{R}^{n\times k}} \mathrm{Tr}(\mathbf{V}^\top\overline{\mathbf{L}}\mathbf{V})$$

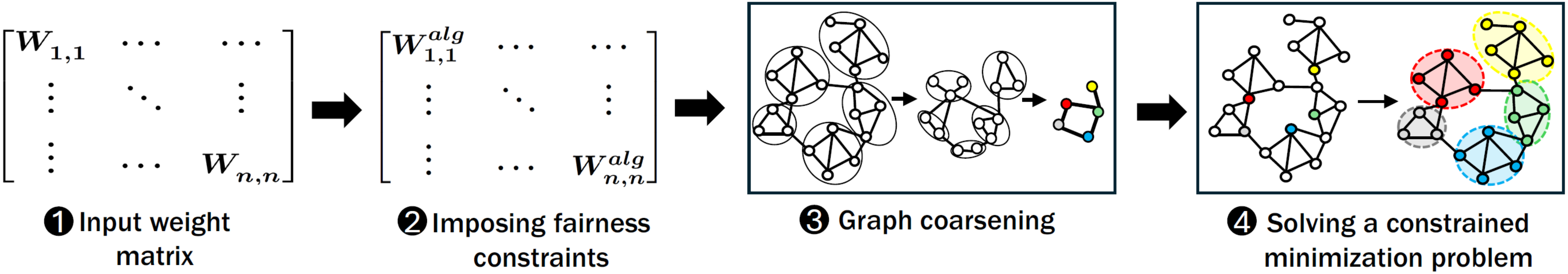
$$\text{subject to } \mathbf{V}^\top\mathbf{V} = \mathbf{I} \text{ and } \mathbf{F}^\top\mathbf{V} = \mathbf{0}.$$

- FairSC[1] and sFairSC[2] add fairness constraints as linear constraints into the spectral clustering problem.
- However, they require solving constrained eigenvalue problems through computationally expensive operations.

[1]Kleindessner, Matthäus, et al. "Guarantees for spectral clustering with fairness constraints." International Conference on Machine Learning. PMLR, 2019.
[2]Wang, Ji, et al. "Scalable spectral clustering with group fairness constraints." International Conference on Artificial Intelligence and Statistics. PMLR, 2023.

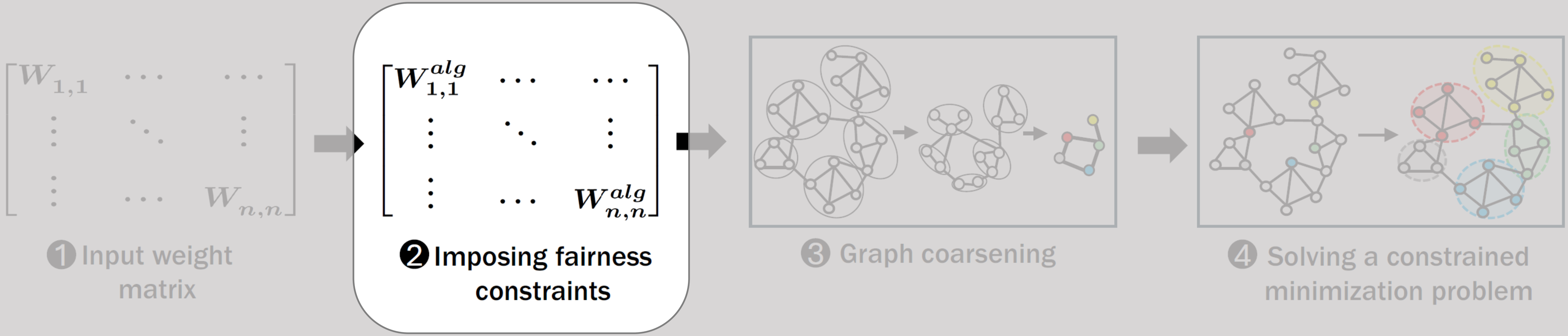
TEXAS STATE UNIVERSITY

Overview of FairAD



* Please refer to our paper for more details.

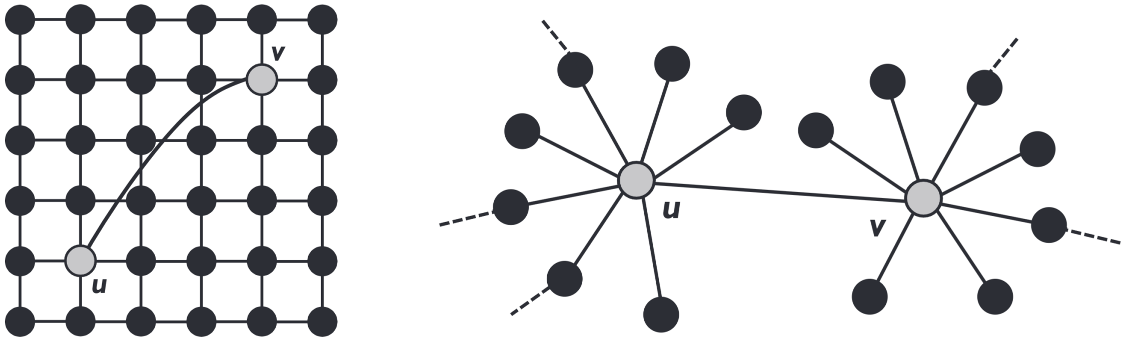
Overview of FairAD



* Please refer to our paper for more details.

Imposing Fairness Constraints

- Algebraic distance is a measure that quantifies the “closeness" between two nodes.



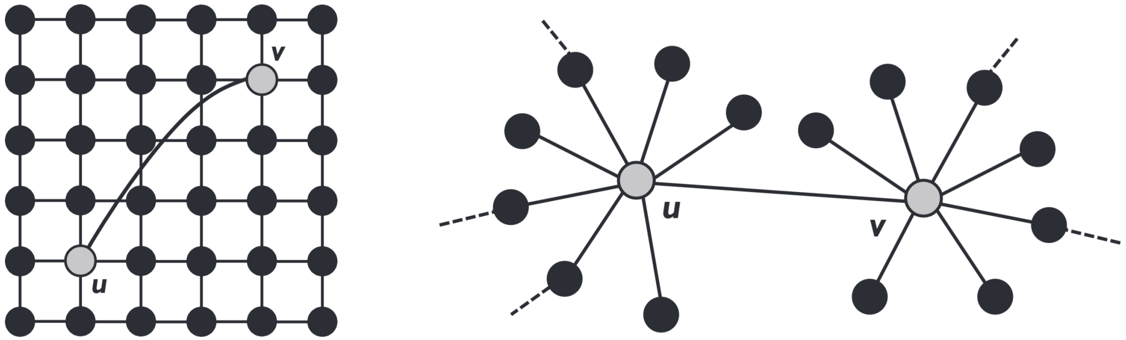
$$s(i,j) = \max_{r=1,2,\ldots,R} \left| x_{r,i} - x_{r,j} \right|$$

$$W_{i,j}^{\mathrm{alg}} = \exp(-s(i,j))$$

Oren E. Livne and Achi Brandt. “Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver.” SIAM Journal on Scientific Computing, 2012.

Imposing Fairness Constraints

- Algebraic distance is a measure that quantifies the "closeness" between two nodes.



$$s(i, j) = \max_{r=1,2,\ldots,R} \left| x_{r,i} - \boxed{x_{r,j}} \right|$$

Test vectors from Jacobi relaxation on $Lx = 0$.

New affinity matrix

$$\boxed{W_{i,j}^{\text{alg}}} = \exp(-s(i, j))$$

Oren E. Livne and Achi Brandt. "Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver." SIAM Journal on Scientific Computing, 2012.

Imposing Fairness Constraints

- Imposing fairness constraint into the algebraic distance matrix

$$\boldsymbol{x}^t = \mathbf{D}^{-1}\mathbf{W}\boldsymbol{x}^{t-1} \longrightarrow \text{Test vector at } t\text{-th Jacobi relaxation iteration}$$

$$\mathbf{D}\boldsymbol{x}^t = \mathbf{W}\boldsymbol{x}^{t-1} \text{ subject to } \boxed{\mathbf{F}^\top \boldsymbol{x}^t = \mathbf{0}} \longrightarrow \text{Fairness constraint}$$

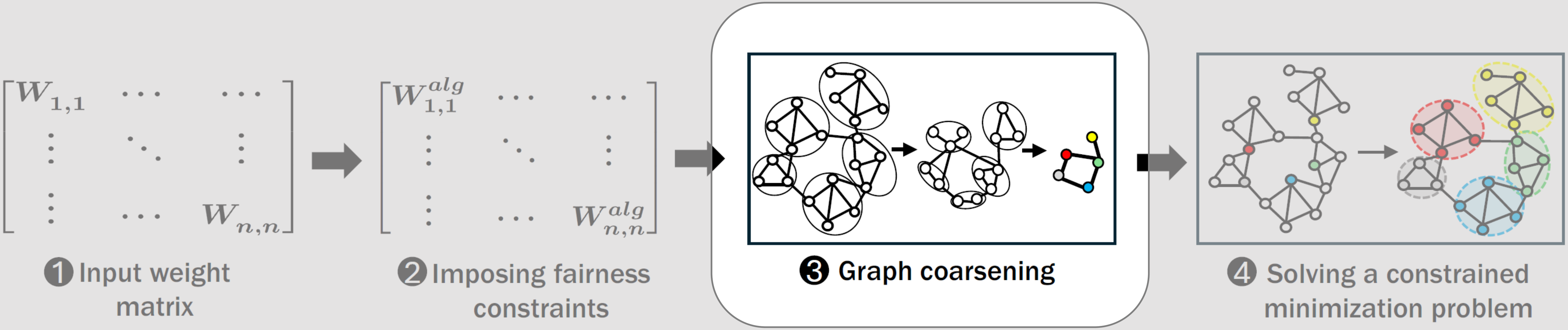
$$\boldsymbol{x}^t = (\mathbf{D} + \mu\mathbf{F}\mathbf{F}^\top)^{-1}\mathbf{W}\boldsymbol{x}^{t-1} \longrightarrow \text{Test vector with fairness constraint}$$

$$W^{\text{alg}}_{i,j} = \exp(-s(i,j)) \longrightarrow \text{New affinity matrix with fairness constraint}$$

* Please refer to our paper for more details.

TEXAS STATE UNIVERSITY

Overview of FairAD

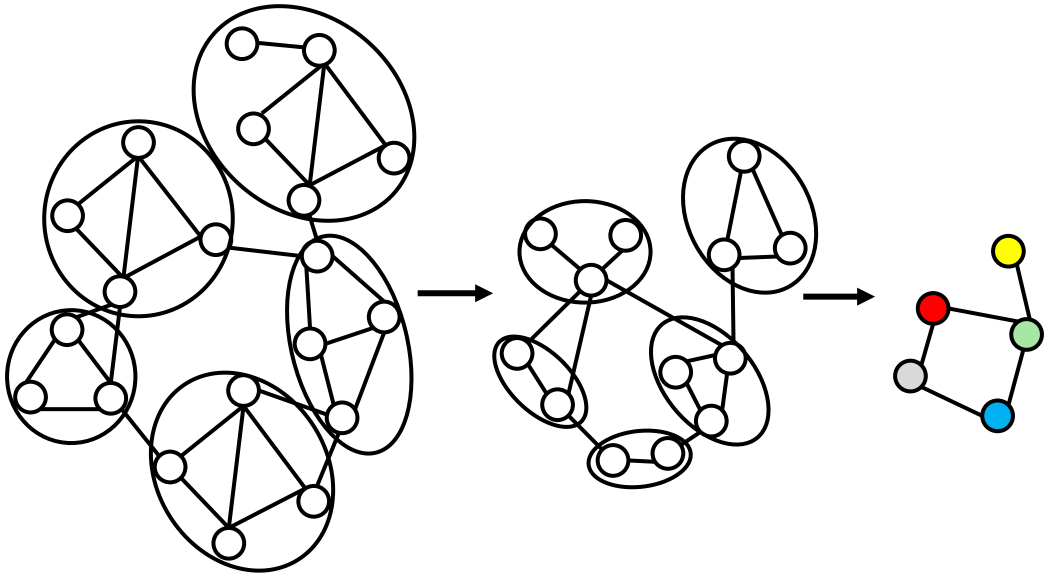


$$\begin{bmatrix} W_{1,1} & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & W_{n,n} \end{bmatrix} \rightarrow \begin{bmatrix} W^{alg}_{1,1} & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & W^{alg}_{n,n} \end{bmatrix}$$

* Please refer to our paper for more details.

Graph Coarsening

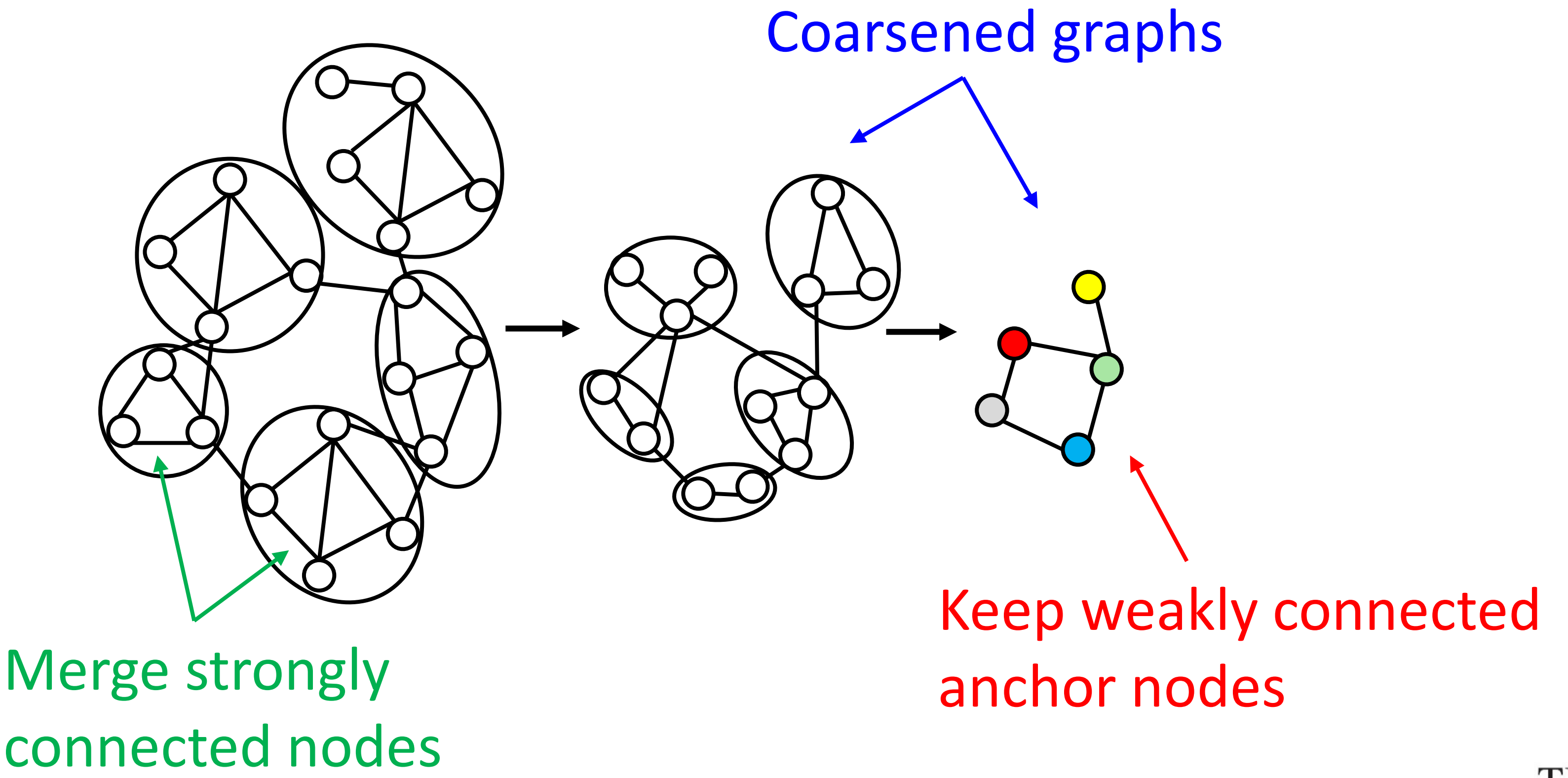
- **Graph coarsening** identifies a small set of representative nodes that serve as anchors to guide the final clustering.



* Please refer to our paper for more details.

Graph Coarsening

- **Graph coarsening** identifies a small set of representative nodes that serve as anchors to guide the final clustering.



* Please refer to our paper for more details.

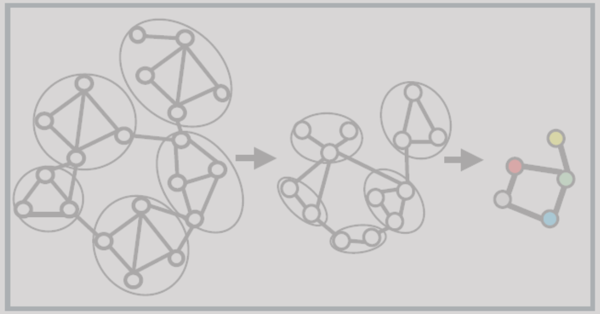
Overview of FairAD

$$\begin{bmatrix} W_{1,1} & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & W_{n,n} \end{bmatrix}$$

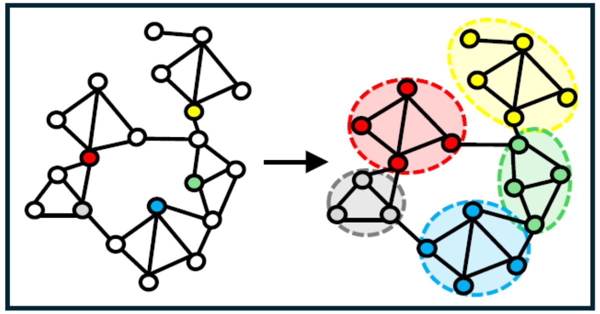
❶ Input weight matrix

$$\begin{bmatrix} W_{1,1}^{alg} & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & W_{n,n}^{alg} \end{bmatrix}$$

❷ Imposing fairness constraints



❸ Graph coarsening

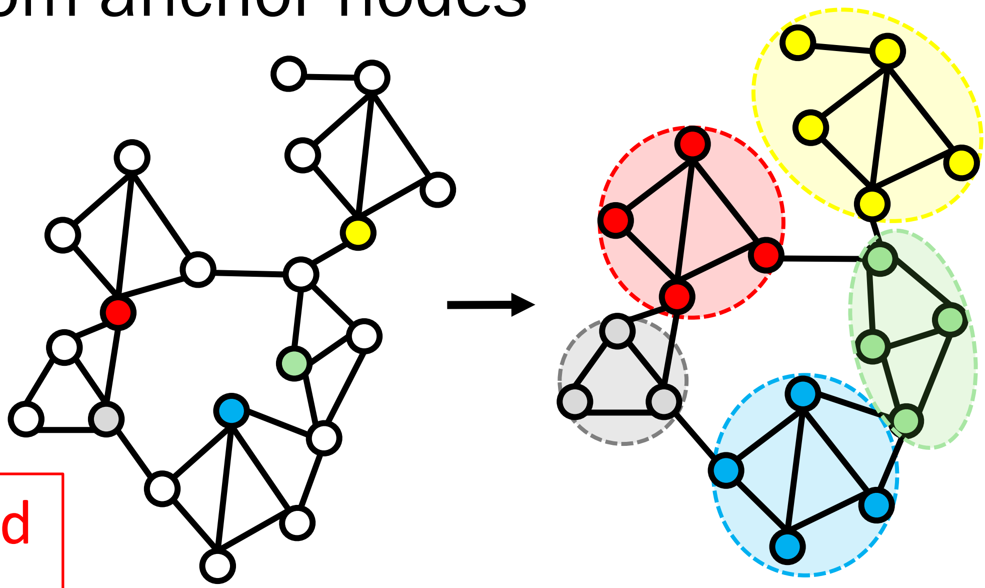


❹ Solving a constrained minimization problem

* Please refer to our paper for more details.

Solving a Constrained Minimization Problem

- Finding solution from anchor nodes



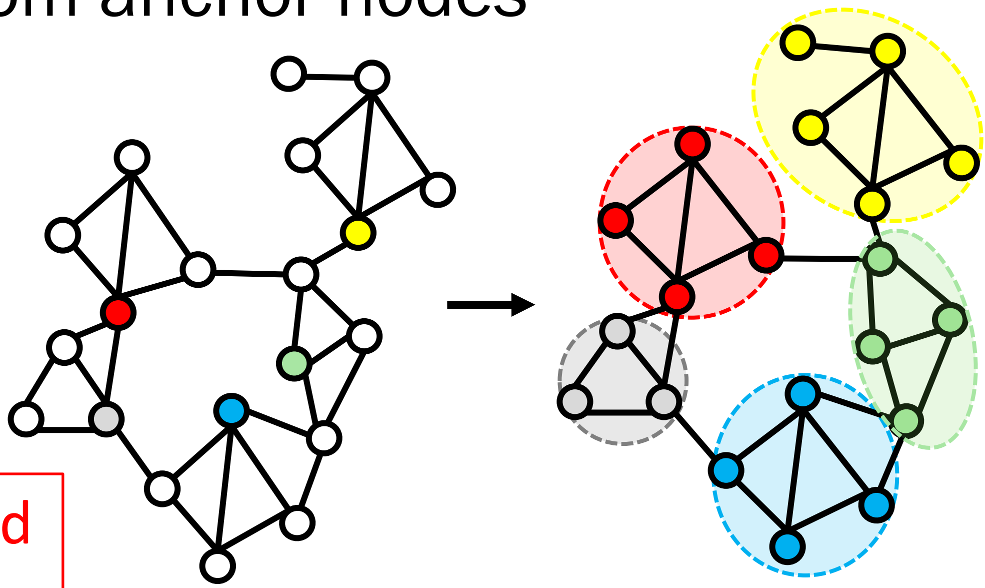
Formulate as a constrained minimization problem

$$\min_{\mathbf{B}v_i=\mathbf{c}_i} \frac{1}{2} v_i^\top \overline{\mathbf{L}}_{\text{alg}} v_i$$

* Please refer to our paper for more details.

Solving a Constrained Minimization Problem

- Finding solution from anchor nodes



Formulate as a constrained minimization problem

$$\min_{\mathbf{B}\boldsymbol{v}_i=\mathbf{c}_i} \frac{1}{2}\boldsymbol{v}_i^\top \overline{\mathbf{L}}_{\text{alg}}\boldsymbol{v}_i$$

$$\boxed{\boldsymbol{v}_i \approx \mu(\overline{\mathbf{L}}_{\text{alg}} + \mu\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{c}_i}$$

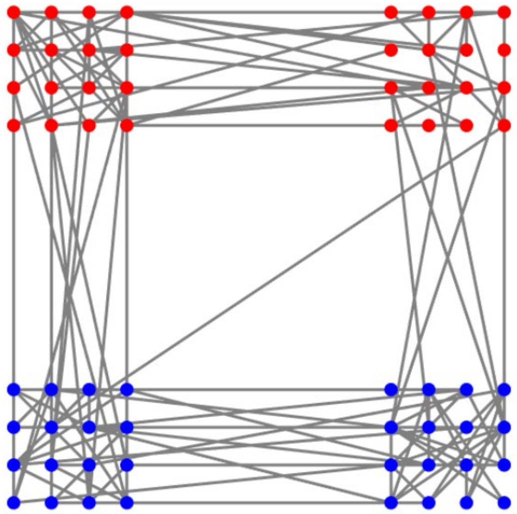
Approximate closed-form expression

TEXAS STATE UNIVERSITY

* Please refer to our paper for more details.

Experiment Setup

- Datasets: We consider both synthetic and public real-world datasets for performance evaluation.



Modified Stochastic Block Model (m-SBM)

Dataset	|V|	|E|	Sensitive Attribute	h
NBA	403	10,621	Country	2
German	1,000	21,742	Gender	2
LastFM	7,624	27,806	Country	4
Recidivism	18,876	311,870	Race	2
Deezer	28,281	92,752	Gender	2
Credit	29,460	136,196	Education	3

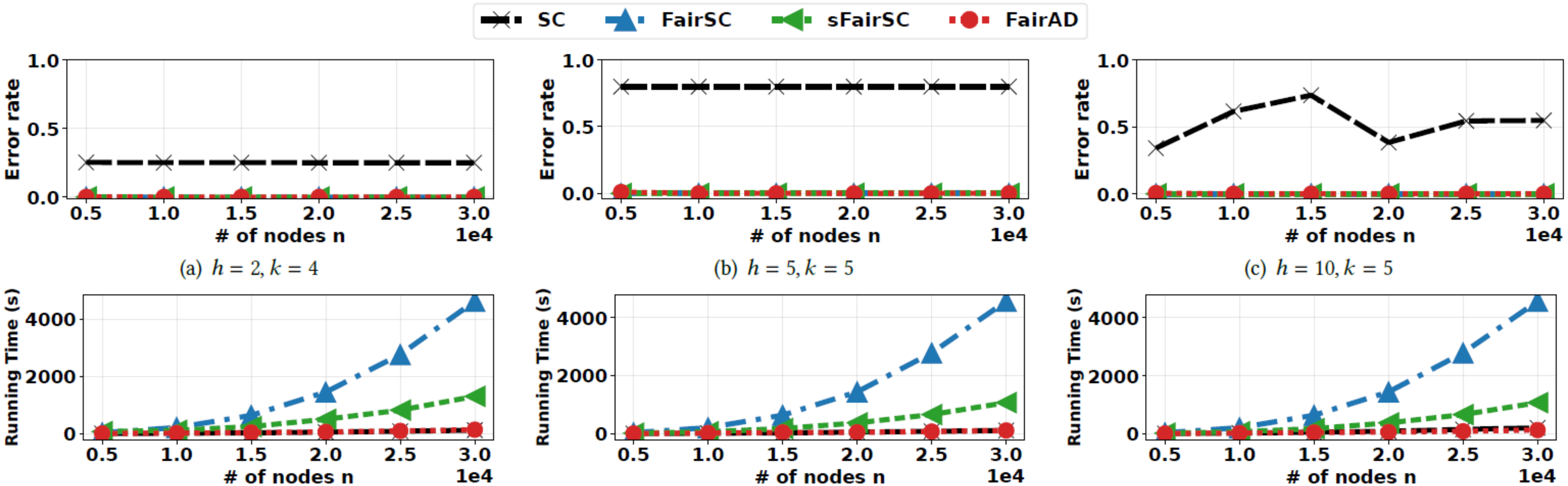
Experiment Setup

- Baselines: Spectral clustering (SC), FairSC[1], and sFairSC[2].
- Performance metrics: Error rate, average balance, and running time.
    - Error rate: measure the discrepancy between computed and ground truth clustering labels.
    - Average balance: measure how evenly different groups are represented across clusters, with a higher score indicating fairer clustering.
    - Running time: measure the total running time of an algorithm.

[1]Kleindessner, Matthäus, et al. "Guarantees for spectral clustering with fairness constraints." International conference on machine learning. PMLR, 2019.
[2]Wang, Ji, et al. "Scalable spectral clustering with group fairness constraints." *International conference on artificial intelligence and statistics*. PMLR, 2023.

Simulation Results

- Error rate and running time for mSBM with varying h and k.



(a) $h = 2, k = 4$ (b) $h = 5, k = 5$ (c) $h = 10, k = 5$

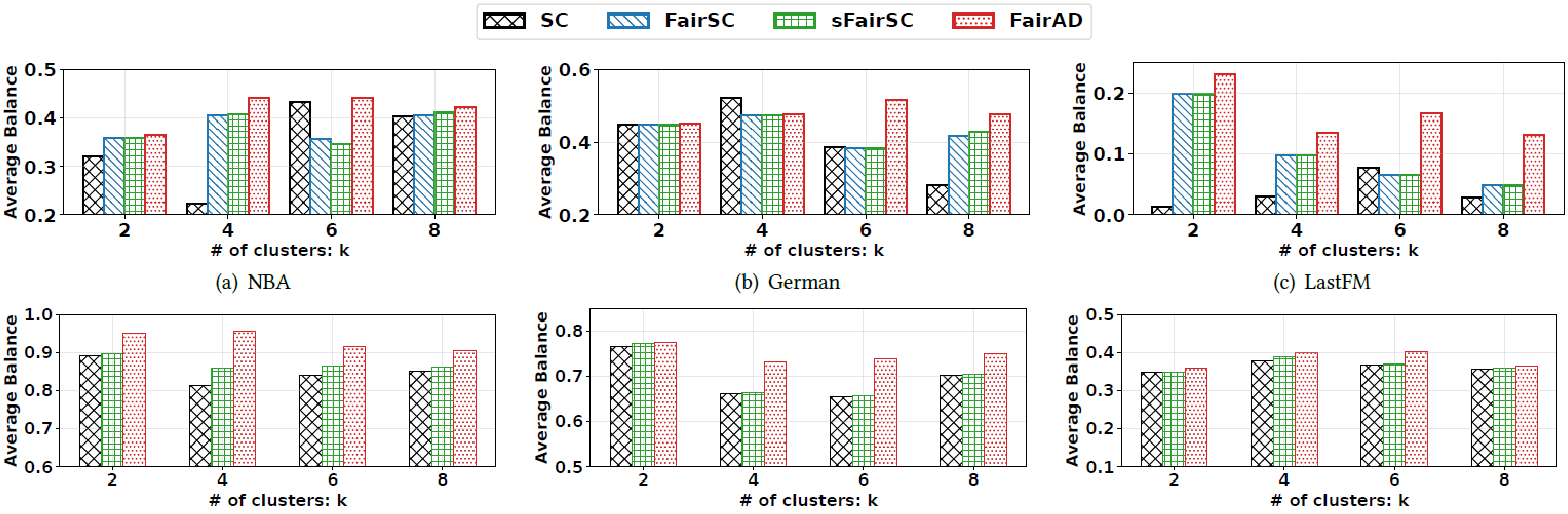
(d) $h = 2, k = 4$ (e) $h = 5, k = 5$ (f) $h = 10, k = 5$

- Observation 1: FairSC, sFairSC, and FairAD successfully recover the ground-truth labels, while SC fails with high error rate.
- Observation 2: FairAD is significantly faster, achieving up to a 42x speedup over FairSC and a 12x speedup over sFairSC.

TEXAS STATE UNIVERSITY

Simulation Results

- Average balance on real-world datasets.



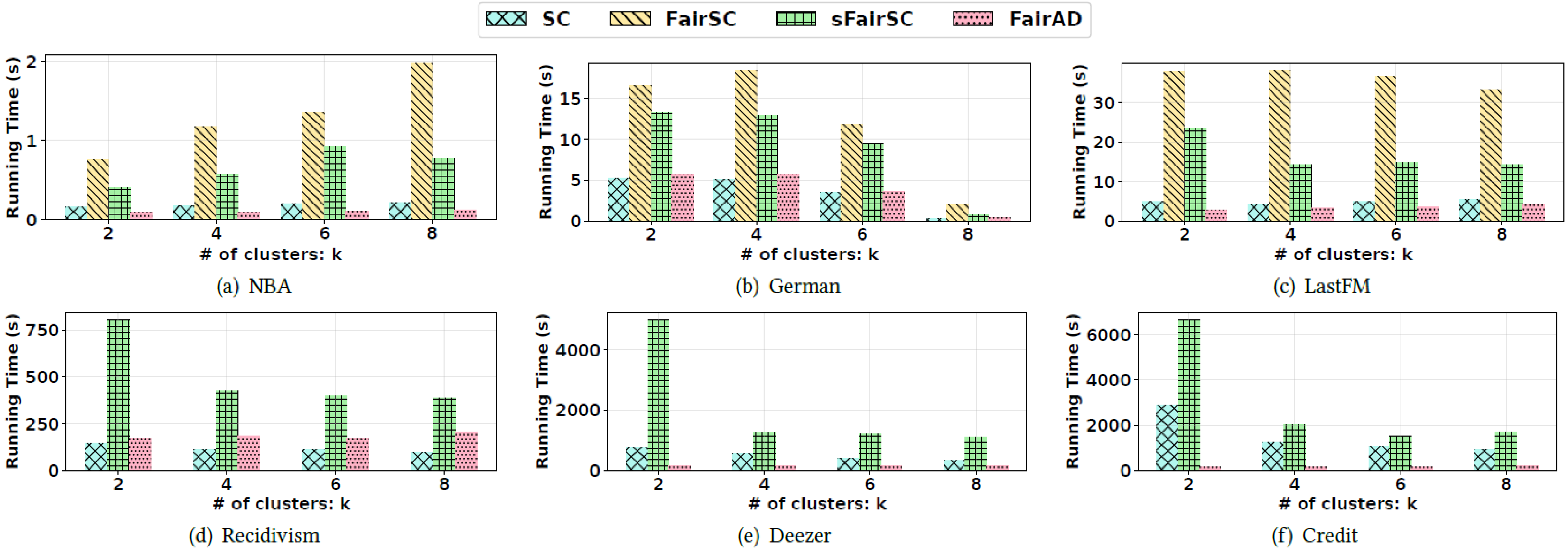
(a) NBA (b) German (c) LastFM

(d) Recidivism (e) Deezer (f) Credit

➢ Observation: FairAD consistently delivers the most balanced clusters, outperforming baselines by 10-15% on large graphs and up to 100% on smaller ones.

Simulation Results

- Running time on real-world datasets.



➢ Observation: FairAD is significantly more efficient than its counterparts, delivering up to 3x speedup on small graphs and a speed-up of up to 40x on large graphs.

TEXAS STATE UNIVERSITY

Conclusion

- We have developed FairAD, a computationally efficient fair graph clustering method.
- We have proposed a framework that imposes fairness constraints directly in the affinity matrix via algebraic distance.
- We have conducted extensive experiments to demonstrate the correctness and effectiveness of FairAD.
- We expect that FairAD can be an effective approach for fair graph clustering on large graphs.

TEXAS STATE UNIVERSITY

Thank you!!

Questions & Answers