Balancing Fairness and Accuracy in Graph Learning via Fairness-Constrained Rewiring

Jason Wang, Lukas Fesser, Melanie Weber

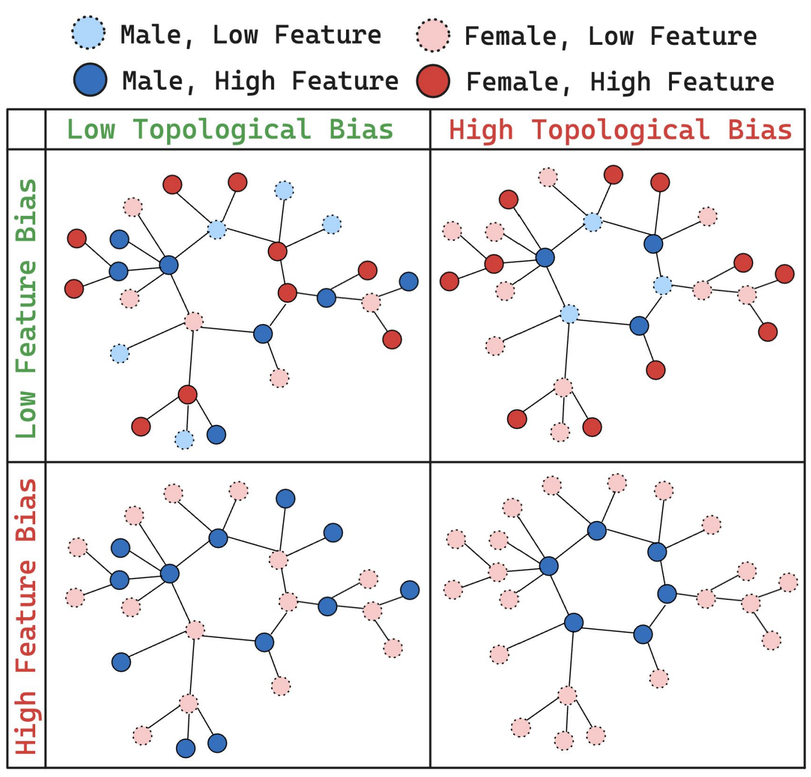
Introduction

Algorithmic decision making is **prone to bias**

- Perpetuation of stereotypes in training data
- Hiring, loans, sentencing, recommendations

Fairness has traditionally been applied to tabular data; **graphs pose a new challenge**

- Non-i.i.d.; knowledge about neighbors gives information about you
- Difficult to describe all topological features



A common method for altering the topology of graphs is **graph rewiring**—the adding and removing of edges. This is typically done to combat oversmoothing and/or oversquashing. **But may this be at the cost of worse fairness?**

Fairness-Constrained Rewiring

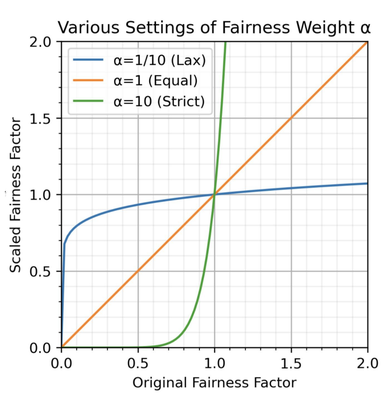
Observing this tradeoff, we propose a **soft fairness constraint** that can be applied to any classical rewiring approach to varying strengths. Rewiring approaches score edges, so we can simply multiply our weighted fairness factor in.

We define the regularization by the percent change in counterfactual topological bias:

$$\mathcal{F}^{+}(u,v) = \frac{\mathcal{L}(V,E\cup\{(u,v)\})}{\mathcal{L}(V,E)} \quad \mathcal{F}^{-}(u,v) = \frac{\mathcal{L}(V,E\setminus\{(u,v)\})}{\mathcal{L}(V,E)}$$

Thus, edges are scored by:

- FairFoSR: $\frac{2x_u x_v}{\sqrt{(1+d_u)(1+d_v)}}\mathcal{F}^{+}(u,v)^{\alpha}$
- FairBORF: $\kappa(u,v)\,(\mathcal{F}^{-}(u,v))^{-\alpha}$ $\pi(p,q)d(p,q)\,(\mathcal{F}^{+}(p,q))^{-\alpha}$
- FairDropEdge: $p\cdot\mathcal{F}^{-}(u,v)^{-\alpha}$



Conclusion

- The fairness-accuracy tradeoff exists, and our fairness-constrained rewiring allows for fine-grained control of this tradeoff.
- Future work may extend to graph-level tasks.

Contributions

Research Question: How does graph rewiring impact fairness?

Contributions

1. We define a **principled fairness metric** (topological bias) that captures relational bias separately from feature bias.
2. We demonstrate that existing rewiring approaches either boost accuracy at the cost of fairness or vice versa, leading to a **fairness-accuracy trade-off**.
3. We propose **fairness-constrained rewiring**, which allows for improving accuracy in downstream task without reinforcing topological biases.

Defining Fairness

Demographic Statistical Parity

Seeks equal positive outcome rate for each demographic group A.

$$\left|\Pr\left[\hat{Y}=1|A=0\right]-\Pr\left[\hat{Y}=1|A=1\right]\right|$$

Topological Statistical Parity (Ours)

Seeks equal positive outcome rate for each topological group T.

$$\left|\Pr\left[\hat{Y}=1|T=0\right]-\Pr\left[\hat{Y}=1|T=1\right]\right|$$

The advantage of this definition is its flexibility in choice of topological attribute; e.g., node popularity, neighborhood composition, or community affiliation. We use node popularity for our experiments.

Fairness-Accuracy Tradeoffs for Graph Rewiring

We test three rewiring approaches and analyze their accuracy and topological statistical parity, We also analyze FairEdit, an existing fairness rewiring baseline.

Approach	Over-smooth	Over-squash	Type	Complexity
FoSR (Karhadkar et al., 2022)	✗	✓	spectral	$O(n^2)$
BORF (Nguyen et al., 2023)	✓	✓	curvature-based	$O(m\cdot d_{\max}^3)$
DropEdge (Rong et al., 2019)	✓	✗	probabilistic	$O(m\cdot n^2)$

Data	Baseline	FoSR	BORF	DropEdge	FairEdit
German	0.684	0.708	0.711	0.693	0.628
	0.0004	0.054	0.117	0.158	0.019
Credit	0.725	0.732	-	0.713	0.727
	0.046	0.054	-	0.070	0.050
Recidivism	0.882	0.888	-	0.887	0.884
	0.321	0.333	-	0.313	0.339
AMiner-S	0.912	0.918	0.927	0.922	-
	0.177	0.212	0.225	0.196	-

Rewirings increase accuracy (top), but also increase unfairness (bottom). FairEdit has a rigid definition of fairness and no way to control the tradeoff.

Experiments

Setting

- Six Node Classification Datasets
- Alpha: [0, 0.1, 1, 10]

Results

- We see a relatively consistent interpolation between no rewiring, rewiring with high fairness factors, and rewiring with low fairness factors.