In Appendix, we present proofs in Section A, illustrative diagrams for Theorem 4.1 in Section B, experimental configurations in Section C, deferred results in Section D, and the demonstration of fair vertex representation in Section E.

A PROOF

Proposition 4.1. For a link prediction function $g(\cdot,\cdot)$ modeled as inner product $g(v,u)=v^\top\Sigma u$, where $\Sigma\in\mathbb{S}^M_{++}$ is a positive-definite matrix, $\exists Q>0,\forall v\sim\mathcal{V}$, $\|v\|_2\le Q$, for $\mathbb{E}_{v\sim U}[v]\in\mathbb{R}^M$, for dyadic fairness based on demographic parity, if $\|\mathbb{E}_{v\sim U}[v\mid v\in S_0]-\mathbb{E}_{v\sim U}[v\mid v\in S_1]\|_2\le\delta$,

$$\Delta_{\text{DP}}:=|\mathbb{E}_{(v,u)\sim U\times U}[g(v,u)\mid S(v)=S(u)]-\mathbb{E}_{(v,u)\sim U\times U}[g(v,u)\mid S(v)\ne S(u)]|\le Q\|\Sigma\|_2\cdot\delta. \tag{2}$$

Proof. To simplify the notations, we use $p:=\mathbb{E}_{v\sim U}[v\mid v\in S_0]\in\mathbb{R}^M$ and $q:=\mathbb{E}_{v\sim U}[v\mid v\in S_1]\in\mathbb{R}^M$ to denote the expectations in representations for S_0 and S_1 respectively.

$$\begin{aligned}|\mathbb{E}_{\text{intra}}-\mathbb{E}_{\text{inter}}|&=\left|\mathbb{E}[v^\top\Sigma u\mid v\in S_0,u\in S_1]-\mathbb{E}[v^\top\Sigma u\mid v\in S_0,u\in S_0\vee v\in S_1,u\in S_1]\right|\\&=\left|p^\top\Sigma q-\left(\frac{|S_0|^2}{|S_0|^2+|S_1|^2}p^\top\Sigma p+\frac{|S_1|^2}{|S_0|^2+|S_1|^2}q^\top\Sigma q\right)\right|\\&=\left|(q-p)^\top\left(\frac{|S_0|^2}{|S_0|^2+|S_1|^2}\Sigma p-\frac{|S_1|^2}{|S_0|^2+|S_1|^2}\Sigma q\right)\right|\end{aligned}$$

To simplify the notation, we will use $\alpha:=|S_0|^2/(|S_0|^2+|S_1|^2)$ and $\beta:=|S_1|^2/(|S_0|^2+|S_1|^2)$

$$\begin{aligned}&\le\|q-p\|_2\cdot\|\alpha\Sigma p-\beta\Sigma q\|_2\\&\le\delta\cdot\|\Sigma\|_2\cdot(\|\alpha p\|_2+\|\beta q\|_2)\\&=Q\|\Sigma\|_2\cdot\delta,\end{aligned}$$

which completes the proof. The first inequality above is due to Cauchy-Schwarz, and the second one is by the definition of spectral norm. The last equality holds by the linearity of expectation: if $\forall v\in\mathcal{V},\|v\|_2\le Q$, then $\|\mathbb{E}[v]\|_2\le\mathbb{E}[\|v\|_2]\le Q$. ■

Theorem 4.1. For an arbitrary graph with nonnegative link weights, after conducting one mean-aggregation over the graph, the consequent representation discrepancy between two sensitive groups $\Delta^{\text{Aggr}}_{\text{DP}}:=\|\mathbb{E}_{v\sim U}[\text{Agg}(v)\mid v\in S_0]-\mathbb{E}_{v\sim U}[\text{Agg}(v)\mid v\in S_1]\|_2$ is bounded by

$$\max\{\alpha_{\min}\|\mu_0-\mu_1\|_\infty-2\sigma,0\}\le\Delta^{\text{Aggr}}_{\text{DP}}\le\alpha_{\max}\|\mu_0-\mu_1\|_2+2\sqrt{M}\sigma, \tag{3}$$

where $\alpha_{\min}=\min\{\alpha_1,\alpha_2\}$, $\alpha_{\max}=\max\{\alpha_1,\alpha_2\}$, $\alpha_1=|1-\frac{m_w}{D_{\text{max}}}(\frac{1}{|S_0|}+\frac{1}{|S_1|})|$, $\alpha_2=|1-\frac{|\widetilde{S_0}|}{|S_0|}-\frac{|\widetilde{S_1}|}{|S_1|}|$.

Proof. The feature representation of v after conducting one mean-aggregation is

$$\text{Agg}(v)=\frac{1}{\deg_w(v)}\sum_{u\in\Gamma(u)}a_{vu}u=\frac{1}{\deg_w(v)}(\sum_{u\in\Gamma(u)\cap S_0}a_{vu}u+\sum_{u\in\Gamma(u)\cap S_1}a_{vu}u).$$

Here we separate the summation of neighbor features into two parts in terms of the sensitive attribute.

We use the bracket notation to abbreviate the range of a vector. That is, if a vector u satisfies $\mu-\sigma\le u\le\mu+\sigma$, we abbreviate this as $u\in[\mu\pm\sigma]$.

Consider the unilateral case $v\in S_0$, we have

$$\begin{aligned}\text{Agg}(v)&\in[\frac{\sum_{u\in\Gamma(v)\cap S_0}}{\deg_w(v)}a_{vu}\mu_0+\frac{\sum_{u\in\Gamma(v)\cap S_1}}{\deg_w(v)}a_{vu}\mu_1\pm\sigma\cdot\mathbb{1}]\\&\in[(\mu_0+\frac{\sum_{u\in\Gamma(v)\cap S_1}a_{vu}}{\deg_w(v)}(\mu_1-\mu_0))\pm\sigma\cdot\mathbb{1}]\end{aligned}$$

where $\mathbb{1}$ is the all-one vector with proper size.

The first derivation is due to the fact that each $u \in S_0$ lies in the range of $[\mu_0 \pm \sigma \cdot \mathbb{1}]$ and each $u \in S_1$ lies in the range of $[\mu_1 \pm \sigma \cdot \mathbb{1}]$. The second one is by the definition of weighted degree.

Using $\beta_v = \sum_{v \in \Gamma(v) \cap S_{\text{opp}(v)}} a_{vu}/\text{deg}_w(v)$ where $S_{\text{opp}(v)}$ is the opposite sensitive group where v belongs. The expectation of $\text{Agg}(v)$ for S_0 is

$$\begin{aligned}\mathbb{E}_{v \sim U}[\text{Agg}(v) \mid v \in S_0] &\in [(\frac{1}{|S_0|} \sum_{v \in S_0} (\mu_0 + \beta_v(\mu_1 - \mu_0))) \pm \sigma \cdot \mathbb{1}] \\ &\in [(\mu_0 + \frac{1}{|S_0|} \sum_{v \in S_0} \beta_v(\mu_1 - \mu_0)) \pm \sigma \cdot \mathbb{1}].\end{aligned}$$

And for $v \in S_1$ we have

$$\mathbb{E}_{v \sim U}[\text{Agg}(v) \mid v \in S_1] \in [(\mu_1 + \frac{1}{|S_1|} \sum_{v \in S_1} \beta_v(\mu_0 - \mu_1)) \pm \sigma \cdot \mathbb{1}].$$

Based on the above two terms, the gap in expectation of two groups after passing one mean-aggregation layer becomes

$$\mathbb{E}_{v \sim U}[\text{Agg}(v) \mid v \in S_0] - \mathbb{E}_{v \sim U}[\text{Agg}(v) \mid v \in S_1] \in [(1 - (\frac{1}{|S_0|} \sum_{v \in S_0} \beta_v + \frac{1}{|S_1|} \sum_{v \in S_1} \beta_v)) \cdot (\mu_0 - \mu_1) + 2\sigma \cdot \mathbb{1}].$$

Next we study the range of $\alpha' := 1 - (|S_0|^{-1} \sum_{v \in S_0} \beta_v + |S_1|^{-1} \sum_{v \in S_1} \beta_v)$. First we consider the term $|S_0|^{-1} \sum_{v \in S_0} \beta_v$. Since $\text{deg}_w(v) \leq D_{\text{max}}, \forall v \in \mathcal{V}$, we have

$$\sum_{v \in S_0} \beta_v = \sum_{v \in S_0} \frac{\sum_{u \in \Gamma(v) \cap S_1} a_{vu}}{\text{deg}_w(v)} \geq \frac{1}{D_{\text{max}}} \sum_{v \in S_0} \sum_{u \in \Gamma(v) \cap S_1} a_{vu} = \frac{m_w}{D_{\text{max}}}.$$

For non-negative weights,

$$D_{\text{max}} \geq \text{deg}_w(v) = \sum_{u \in \Gamma(v) \cap S_0} a_{vu} + \sum_{u \in \Gamma(v) \cap S_1} a_{vu} \geq \sum_{u \in \Gamma(v) \cap S_1} a_{vu}.$$

This means for $v \in S_0$,

$$\beta_v = \frac{\sum_{u \in \Gamma(u) \cap S_1} a_{vu}}{\text{deg}_w(u)} \leq 1,$$

thus,

$$\sum_{v \in S_0} \beta_v = \sum_{v \in \widetilde{S_0}} \beta_v \leq |\widetilde{S_0}|.$$

The first equality holds because $\beta_v = 0$ when $v \in S_0/\widetilde{S_0}$, meaning v doesn't contain any inter-edges.

Since the analysis for S_1 is similar, we derive the lower and upper bounds for $|S_i|^{-1} \sum_{v \in S_i} \beta_v,\ i = 0, 1$

$$\frac{1}{|S_i|} \cdot \frac{m_w}{D_{\text{max}}} \leq \frac{1}{|S_i|} \sum_{v \in S_i} \beta_v \leq (\frac{|\widetilde{S_i}|}{|S_i|}), \quad i = 0, 1.$$

Based on the above results, we give the bound for α' as follows:

$$\alpha' \in [\, 1 - (\frac{|\widetilde{S_0}|}{|S_0|} + \frac{|\widetilde{S_1}|}{|S_1|}),\ 1 - \frac{m_w}{D_{\text{max}}}(\frac{1}{|S_0|} + \frac{1}{|S_1|})\,],$$

Let $\alpha_{\min}$ and $\alpha_{\max}$ be lower bound and upper bower of $|\alpha'|$, we have

$$\alpha_{\max} = \max\{1 - (\frac{|\widetilde{S_0}|}{|S_0|} + \frac{|\widetilde{S_1}|}{|S_1|}),\ 1 - \frac{m_w}{D_{\text{max}}}(\frac{1}{|S_0|} + \frac{1}{|S_1|})\}$$

$$\alpha_{\min} = \min\{1 - (\frac{|\widetilde{S_0}|}{|S_0|} + \frac{|\widetilde{S_1}|}{|S_1|}),\ 1 - \frac{m_w}{D_{\text{max}}}(\frac{1}{|S_0|} + \frac{1}{|S_1|})\}$$

Thus we give the upper bound of $\Delta_{\text{DP}}^{\text{Aggr}}$:

$$\Delta_{\text{DP}}^{\text{Aggr}} \leq \alpha_{\max}\|\mu_0 - \mu_1\|_2 + 2\sqrt{M}\sigma \tag{8}$$

where the second part in RHS is due to $2\sigma \cdot \|\mathbb{1}\|_2 = 2\sqrt{M}\sigma$.

Next we consider i-th entrance of μ_0 and μ_1, denoted as μ_0^i and μ_1^i respectively. The i-the entrance of $\mathbb{E}_{v\sim U}[\text{Agg}(v) \mid v \in S_0] - \mathbb{E}_{v\sim U}[\text{Agg}(v) \mid v \in S_1]$ take nonzero values if and only if

$$|(1 - (\frac{1}{|S_0|}\sum_{v\in S_0}\beta_v + \frac{1}{|S_1|}\sum_{v\in S_1}\beta_v)) \cdot (\mu_0^i - \mu_1^i)| \geq 2\sigma$$

Thus we obtain the lower bound of $\Delta_{\text{DP}}^{\text{Aggr}}$:

$$\Delta_{\text{DP}}^{\text{Aggr}} \geq \max\{\alpha_{\min}\|\mu_0 - \mu_1\|_\infty - 2\sigma, 0\} \tag{9}$$

which completes the proof. ■

B COMPLEMENTARY DIAGRAMS TO THEOREM 4.1

We provide diagrams to help better understand the upper bound in Theorem 4.1.

Figure 3 provides a common case that the gap in expectation between two sensitive groups shrinks after mean-aggregation. Here the maximal deviation term σ can be neglected since it is much smaller than the expectation gap.

Figure 4 provides a case that the term σ is not negligible against the expectation gap between two sensitive groups. Here $\sigma = 100$ and the gap equals to 0. After aggregation, we see the new expectation gap becomes 20, showing that the discrepancy in representations increases.

Figure 5 provides another case that the contraction coefficient α equals to 1 due to the resistance of α_2. Here all vertices possess inter links, and the graph is a complete bipartite graph. Then the aggregation fully exchanges the sensitive information, and thus the representation discrepancy remains unchanged.

Cases in Figure 4 and 5 are also pointed out by the analysis in Section 4.

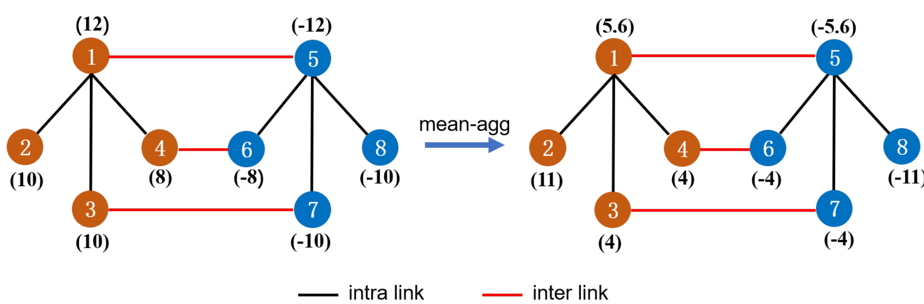

Figure 3: An illustrative graph example with two protected groups S_0 and S_1. All vertices have self-loop. The expectation gap shrinks after mean aggregation. Here, $|\mathbb{E}_{v\sim U}[v|v \in S_0] - \mathbb{E}_{v\sim U}[v|v \in S_1]| = 20$, $\sigma = 2$ and all link weights are equal. After aggregation, $|\mathbb{E}_{v\sim U}[\text{Agg}(v)|v \in S_0] - \mathbb{E}_{v\sim U}[\text{Agg}(v)|v \in S_1]| = |6.15 - (-6.15)| = 12.3 < 20$.

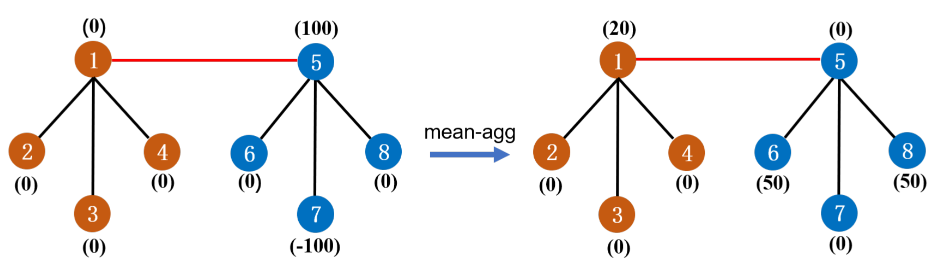

Figure 4: Case 1: The maximal deviation term $O(\sigma)$ is not negligible. Here $\sigma = 100$ and all link weights are equal. All vertices have self-loop. $|\mathbb{E}_{v\sim U}[v|v \in S_0] - \mathbb{E}_{v\sim U}[v|v \in S_1]| = 0$. But after mean-aggregation, $|\mathbb{E}_{v\sim U}[\text{Agg}(v)|v \in S_0] - \mathbb{E}_{v\sim U}[\text{Agg}(v)|v \in S_1]| = |5 - 25| = 20 > 0$.

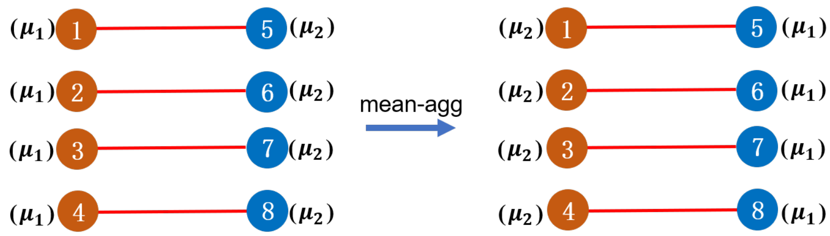

Figure 5: Case 2: The contraction coefficient α equals to 1. This happens when the graph is a complete bipartite graph. Mean-aggregation fully exchanges the sensitive information and the gap of two groups remains unchanged.

C Experimental Configurations

For all experiments, we set $T_1 = 50$ and the total epochs which contain T_1 and T_2 equal to 4. Graph neural networks are applied with two hidden layers with size 32 and 16 respectively. η_θ is set to 0.01. For $\eta_{\tilde{A}}$ for different datasets, we have: *Oklahoma97*: 0.1; *UNC28*: 0.1; *Cora*: 0.2; *Citeseer*: 0.5. Experiments are conducted on Nvidia Titan RTX graphics card.

D Additional Results

We present experimental results for *Citeseer* and *UNC28* in this section. All the results deliver similar conclusions as we state in the main body of this paper. Additionally, we include another dataset *Facebook#1684* in response to the second limitation as indicated in Section 4. In this case, Δ_{DP}, Δ_{true}, Δ_{false} are already small as given by VGAE, and FairAdj is not able to further minimize the gap.

Table 4: Experimental results on *Oklahoma97*.

Method	AUC ↑	AP ↑	Δ_{DP} ↓	Δ_{true} ↓	Δ_{false} ↓	Δ_{FNR} ↓	Δ_{TNR} ↓
VGAE	**90.13** ± 0.32	**91.24** ± 0.37	8.73 ± 0.38	8.56 ± 0.44	0.40 ± 0.32	36.51 ± 1.41	2.26 ± 0.92
node2vec	86.49 ± 0.35	84.09 ± 0.50	7.23 ± 0.64	3.35 ± 0.45	1.08 ± 0.97	32.55 ± 1.32	2.36 ± 0.69
Fairwalk	86.56 ± 0.32	84.23 ± 0.44	7.31 ± 0.62	3.49 ± 0.47	1.13 ± 0.85	32.77 ± 1.20	2.18 ± 0.69
FairAdj$_{\text{T2=5}}$	84.92 ± 0.81	85.07 ± 0.92	3.60 ± 0.35	0.40 ± 0.32	0.33 ± 0.28	**4.00** ± 0.88	**2.02** ± 0.76
FairAdj$_{\text{T2=20}}$	81.01 ± 1.01	80.79 ± 0.93	**2.96** ± 0.30	**0.38** ± 0.31	**0.32** ± 0.25	5.61 ± 1.06	2.03 ± 0.92

Table 5: Experimental results on *Cora*.

Method	AUC ↑	AP ↑	Δ_{DP} ↓	Δ_{true} ↓	Δ_{false} ↓	Δ_{FNR} ↓	Δ_{TNR} ↓
VGAE	**88.48** ± 0.88	**90.81** ± 0.78	26.74 ± 1.51	9.99 ± 2.32	10.26 ± 1.59	28.25 ± 4.46	26.71 ± 3.83
node2vec	87.93 ± 0.75	87.82 ± 1.06	39.99 ± 2.75	6.63 ± 3.58	27.86 ± 4.94	23.66 ± 4.73	32.96 ± 5.24
Fairwalk	88.04 ± 0.84	88.10 ± 1.20	40.49 ± 2.58	7.30 ± 3.28	29.43 ± 4.86	23.74 ± 4.19	33.79 ± 5.08
FairAdj$_{\text{T2=5}}$	86.00 ± 1.12	88.32 ± 0.86	21.05 ± 1.26	6.99 ± 2.24	6.14 ± 1.59	20.72 ± 3.62	19.46 ± 3.62
FairAdj$_{\text{T2=20}}$	83.85 ± 1.07	86.08 ± 0.93	**17.87** ± 1.18	**5.40** ± 2.23	**3.74** ± 1.46	**16.75** ± 4.87	**15.37** ± 3.84

Table 6: Experimental results on *Pubmed*.

Method	AUC ↑	AP ↑	Δ_{DP} ↓	Δ_{true} ↓	Δ_{false} ↓	Δ_{FNR} ↓	Δ_{TNR} ↓
VGAE	**91.20** ± 0.85	**91.26** ± 0.80	20.88 ± 1.48	4.19 ± 0.93	8.04 ± 1.83	12.01 ± 2.92	19.18 ± 4.16
node2vec	74.27 ± 1.23	79.24 ± 1.29	19.14 ± 0.93	3.38 ± 2.57	8.90 ± 2.56	6.65 ± 2.21	10.91 ± 1.88
fairwalk	73.43 ± 1.11	78.96 ± 1.24	18.42 ± 1.65	3.11 ± 1.84	7.79 ± 3.49	**6.61** ± 2.28	10.93 ± 2.54
FairAdj$_{\text{T2 = 5}}$	88.64 ± 1.09	88.21 ± 1.22	16.06 ± 0.98	1.96 ± 0.82	4.40 ± 1.28	8.93 ± 2.90	12.75 ± 1.56
FairAdj$_{\text{T2 = 20}}$	87.53 ± 1.03	87.10 ± 1.17	**14.73** ± 0.98	**1.39** ± 0.92	**3.17** ± 1.10	9.09 ± 2.10	**10.46** ± 1.73

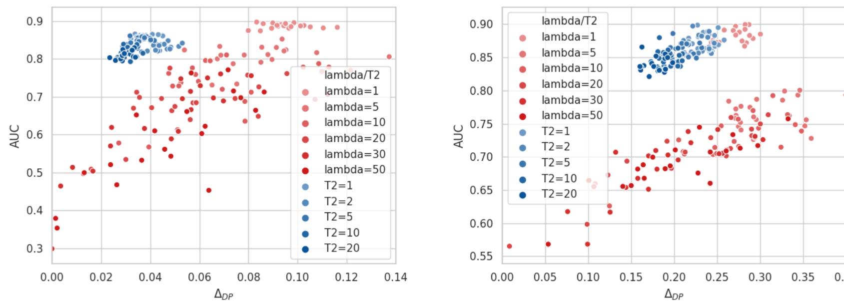

Figure 6: Compare to adversarial training on vertex representations. Left: *Oklahoma97*; Right: *Cora*.

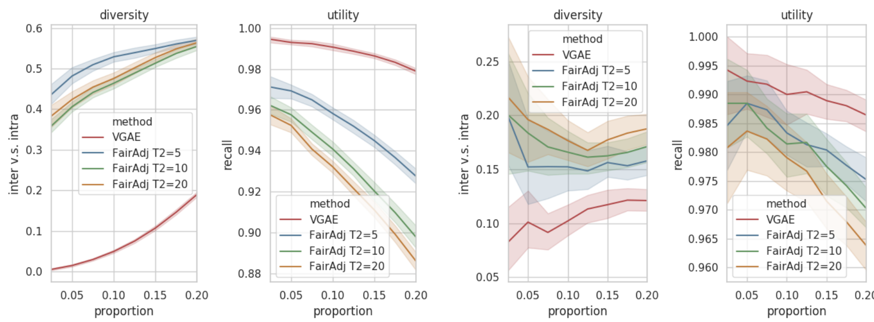

Figure 7: Diversity and utility in recommendations. Left: *Oklahoma97*; Right: *Cora*.

Table 7: Experimental results on *Facebook#1684*.

Method	AUC	AP	Δ_{DP}	Δ_{true}	Δ_{false}	Δ_{FNR}	Δ_{TNR}
VGAE	94.66 $\pm$ .55	93.91 $\pm$ .68	2.03 $\pm$ .81	0.59 $\pm$ .49	**0.90** $\pm$ .57	4.48 $\pm$ 1.57	4.94 $\pm$ 1.32
node2vec	90.57 $\pm$ .74	85.61 $\pm$ 1.09	**1.70** $\pm$ 1.43	**0.52** $\pm$ .49	2.47 $\pm$ 1.52	6.51 $\pm$ 2.04	5.06 $\pm$ 1.36
fairwalk	90.56 $\pm$ .63	85.58 $\pm$ .87	1.97 $\pm$ 1.51	0.62 $\pm$ .47	2.14 $\pm$ 1.77	6.92 $\pm$ 2.19	5.03 $\pm$ 1.46
FairAdj$_{T2=1}$	**94.68** $\pm$ .48	**93.94** $\pm$ .62	2.02 $\pm$ .82	0.60 $\pm$ .50	0.93 $\pm$ .60	**4.42** $\pm$ 1.57	**4.82** $\pm$ 1.54
FairAdj$_{T2=20}$	94.63 $\pm$ .49	93.84 $\pm$ .64	1.77 $\pm$ .81	0.53 $\pm$ .41	0.92 $\pm$ .49	5.00 $\pm$ 1.52	4.86 $\pm$ 1.41

E Fair Vertex Representation

As an intermediate result, we inspect the fairness in vertex representation in Figure 8. To quantify that, we conduct K-means clustering on vertex representation and evaluate the ratio of samples from different sensitive groups within each clusters, and the ratio is called balance. We range the number of clusters from 4 to 8 and report the average balance across all clusters. In general, the higher the balance, the fairer in vertex representations. Overall the series of **FairAdj** achieves a higher balance, which shows the invariant representations on vertices across different sensitive groups.

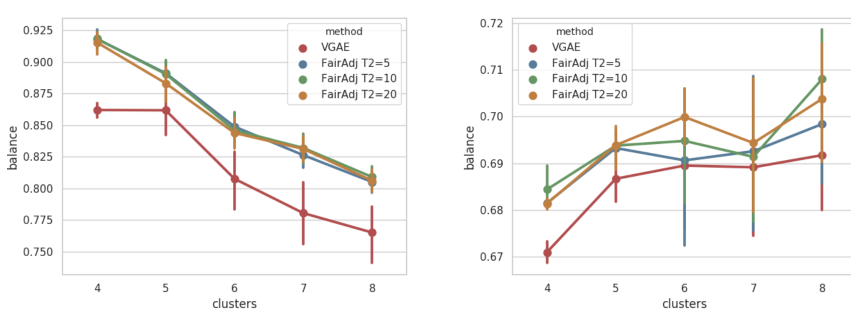

Figure 8: Evaluations on balance of clusters. Left: *Oklahoma97*; Right: *UNC28*.

F EXTEND COROLLARY 4.1 TO TWO-LAYER GNNS

For $\Delta_{DP}^{(2)}$ on vertices after passing two layer $\text{GNN}_\theta^{(2)}(X,\widetilde{A}) := (\widetilde{A}(\text{GNN}_\theta^{(1)}(X,\widetilde{A}))W_\theta^{(2)}) = \rho(\widetilde{A}\rho(\widetilde{A}XW_\theta^{(1)})W_\theta^{(2)})$. Here we denote $\mu_i' := \mathbb{E}_{v\sim U}[\text{GNN}_\theta^{(1)}(v,\widetilde{A}) \mid v \in S_i]$, $i = 0,1$, $Q' := \sup\{\|\text{GNN}_\theta^{(1)}(v,\widetilde{A})\|_2 \mid v \in \mathcal{V}\}$. $\mu_i'' := \mathbb{E}_{v\sim U}[\text{GNN}_\theta^{(2)}(v,\widetilde{A}) \mid v \in S_i]$, $i = 0,1$, $Q'' := \sup\{\|\text{GNN}_\theta^{(2)}(v,\widetilde{A})\|_2 \mid v \in \mathcal{V}\}$. Let σ_1 be the maximal deviation of $\{v|v \in \mathcal{V}\}$, and σ_2 be the maximal deviation of $\{\text{GNN}_\theta^{(1)}(v,\widetilde{A})|v \in \mathcal{V}\}$. We have $Q' \leq L\|W_\theta^{(1)}\|_2 Q$, $Q'' \leq L\|W_\theta^{(2)}\|_2 Q' \leq L^2\|W_\theta^{(1)}\|_2\|W_\theta^{(2)}\|_2 Q$.

Then

$$\Delta_{DP}^{(2)} \leq Q''||\Sigma||_2 \cdot \|\mu_0'' - \mu_1''\|_2 \leq QL^3||W_\theta^{(2)}||_2^2\|W_\theta^{(1)}\|_2(\alpha\|\mu_0' - \mu_1'\|_2 + 2\sqrt{M}\sigma_2)$$

And

$$\|\mu_0' - \mu_1'\|_2 \leq L||W_\theta^{(1)}||_2(\alpha\|\mu_0 - \mu_1\|_2 + 2\sqrt{M}\sigma_1)$$

Finally we have:

$$\Delta_{DP}^{(2)} \leq QL^4||W_\theta^{(2)}||_2^2\|W_\theta^{(1)}\|_2^2\alpha^2\|\mu_0 - \mu_1\|_2 + 2\sqrt{M}QL^3||W_\theta^{(2)}||_2^2\|W_\theta^{(1)}\|_2(L||W_\theta^{(1)}||_2 \cdot \sigma_1 + \sigma_2)$$