Fairness-Aware Graph Filter Design

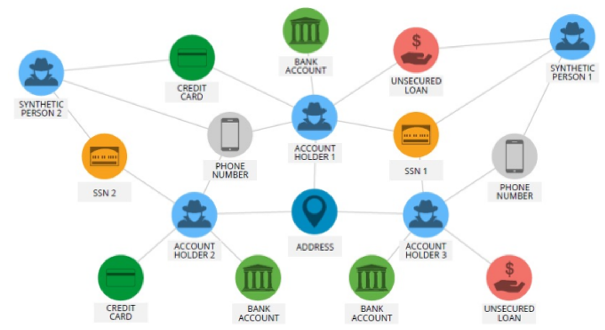
O. Deniz Kose[1], Yanning Shen[1], and Gonzalo Mateos[2]

[1]University of California, Irvine, Dept. of Electrical Engineering and Computer Science

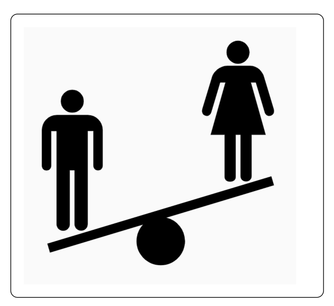
[2]University of Rochester, Dept. of Electrical and Computer Engineering

Motivation

- Connectivity era: Growing amount of data describing interconnected systems



- Graphs are utilized to model such complex data
  - Graph nodes: users in social networks, accounts holding money
  - Graph edges: friendship between users, money transactions
  - Nodal features: education level of users, locations of accounts
- Processing & learning from graph data can provide significant advancements
  - Increasing attention towards graph signal processing & ML over graphs
  - Cross-pollination of GSP and ML over graphs provides new insights [1]
- ML algorithms propagate algorithmic bias
  - Impact of ethnicity in crime prediction
  - Impact of gender in ad recommendation



- Use of network connectivity in learning amplifies existing bias [2]
- Motivation: Consideration of bias is necessary for graph-based learning
- Limitation of current works: Task/algorithm-specific, no theoretical analysis
- Intuition: Can we leverage GSP-based tools to design a general-purpose bias mitigation strategy?

Preliminaries & Problem Statement

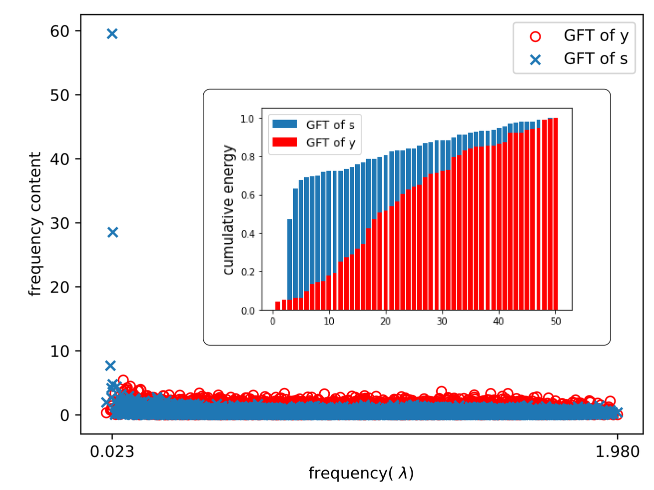
- Focus on undirected graphs, $\mathcal{G} := (\mathcal{V}, \mathcal{E})$
- Connectivity information described via graph adjacency $\mathbf{A} \in \{0,1\}^{N \times N}$ and normalized Laplacian matrices $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$
- Sensitive attributes $\mathbf{s} \in \{-1,1\}^N$, nodal features $\mathbf{X} \in \mathbb{R}^{N \times F}$ and labels $\mathbf{y} \in \{-1,1\}^N$ for node classification
- Graph Fourier Transform of signal $\mathbf{z} \in \mathbb{R}^N$ is $\tilde{\mathbf{z}} = \mathbf{V}^\top \mathbf{z}$, where $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$
- Filtering graph signal $\mathbf{z} \in \mathbb{R}^N$ via a filter with frequency response $\tilde{\mathbf{h}} := [\tilde{h}_1, \dots, \tilde{h}_N]^\top$ yields the output signal $\mathbf{z}_{out} = \mathbf{V}\,\mathrm{diag}(\tilde{h}_1, \dots, \tilde{h}_N)\tilde{\mathbf{z}}$.

Problem Statement

Given $\mathcal{G}$ and $\mathbf{s}$, design of graph filters with frequency response $\tilde{\mathbf{h}} \in \mathbb{R}^N$, so that algorithmic bias sourced from graph topology can be attenuated with the application of such filters.

Preliminary Spectrum Analysis

- Topology bias due to homophily: Denser connectivity within sensitive groups
  - Higher energy concentration is expected for both $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{y}}$ over lower frequencies
  - Few frequencies where the magnitudes of $\tilde{\mathbf{s}}$ are markedly higher than those of $\tilde{\mathbf{y}}$

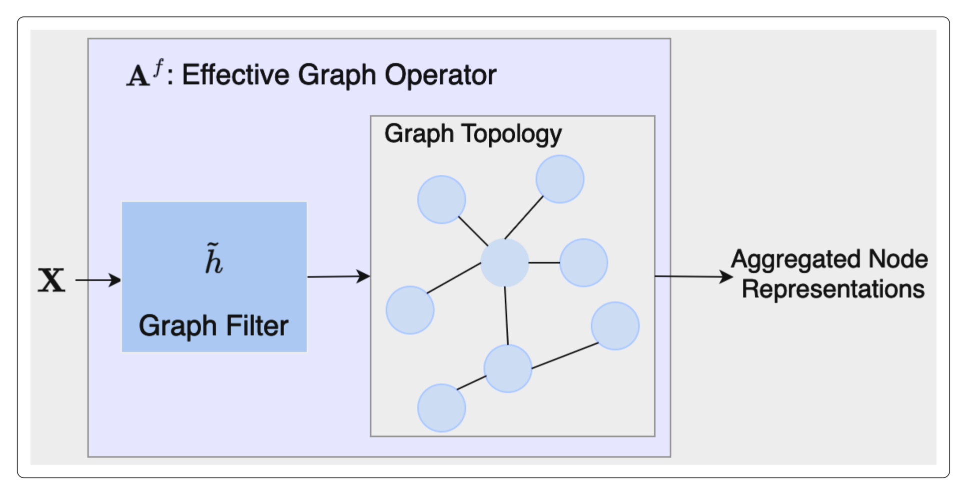


- Upshot: We can design a graph filter to attenuate sensitive information while preserving the data structure necessary for downstream ML tasks

Bias Metric

- Aggregation over graph

$$\mathbf{R}^{\mathrm{f}} = \hat{\mathbf{A}}\mathbf{V}\mathrm{diag}(\tilde{\mathbf{h}})\mathbf{V}^\top\mathbf{X} = \mathbf{A}^f\mathbf{X},$$
$$\mathbf{A}^f := \mathbf{V}(\mathbf{I}_N - \mathbf{\Lambda})\mathrm{diag}(\tilde{\mathbf{h}})\mathbf{V}^\top$$



- Novel unsupervised bias measure: $\rho := \|\mathbf{s}^\top\mathbf{A}^f\|_2$, can be manipulated via filter design

Linear Programming-based Optimal Filter Design

- Proposition: Bias metric ρ, can be upper bounded by:

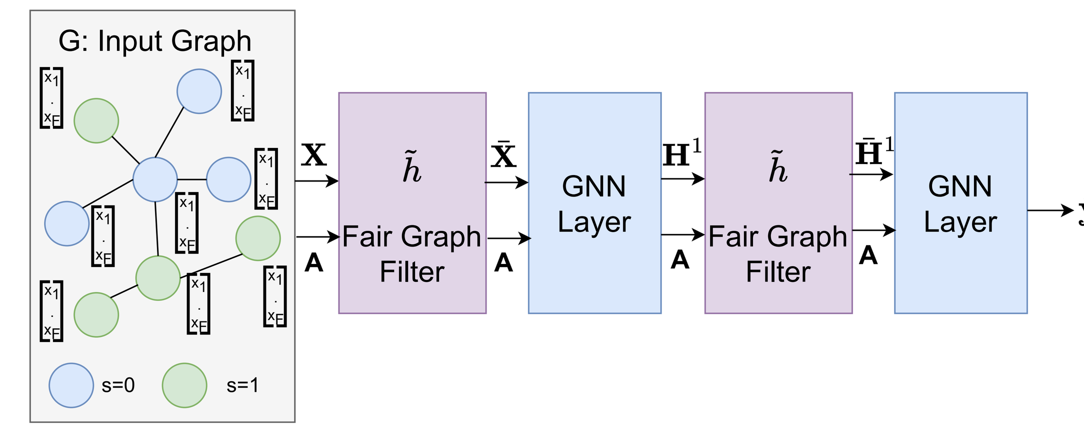
$$\rho \leq \sqrt{N}\sum_{i=1}^{N}|\tilde{s}_i||(1-\lambda_i)||\tilde{h}_i|. \tag{1}$$

- Define $\mathbf{m} \in \mathbb{R}^N$, where $m_i := |\tilde{s}_i||(1-\lambda_i)|, \forall i = 1, \dots, N$
- Optimal fairness-aware filter design:

$$\tilde{\mathbf{h}}^{fair} := \underset{\tilde{\mathbf{h}}}{\operatorname{argmin}}\ \mathbf{m}^\top\tilde{\mathbf{h}}$$
$$\text{s. to } \sum_{i=1}^{N}\tilde{h}_i \geq N\tau,\ 0 \leq \tilde{h}_i \leq 1, \forall i \in \{1, \dots, N\}. \tag{2}$$

- Closed-form solution: $(\tilde{h}^{\mathrm{fair}})_{\alpha_i} = \left[1 - \left[N(1-\tau) - \sum_{j=1}^{i-1}\left(1 - (\tilde{h}^{\mathrm{fair}})_{\alpha_j}\right)\right]_+\right]_+$, where $\boldsymbol{\alpha} = \mathrm{argsort}(-\mathbf{m})$
- A flexible design that can be **pre-computed once** for different learning algorithms, and can be used at different stages of learning (i.e., pre-processing, post-processing)

Experimental Settings & Results



- Datasets: Real social networks, region is sensitive attribute & job is label
- Task: Node classification, classification accuracy is reported
- Fairness metrics (lower values are desired):
  - $\Delta_{SP} = |P(\hat{y} = 1 \mid s = 0) - P(\hat{y} = 1 \mid s = 1)|$
  - $\Delta_{EO} = |P(\hat{y} = 1 \mid y = 1, s = 0) - P(\hat{y} = 1 \mid y = 1, s = 1)|$

	Pokec-z			Pokec-n		
	Accuracy (%)	Δ_{SP} (%)	Δ_{EO} (%)	Accuracy (%)	Δ_{SP} (%)	Δ_{EO} (%)
GNN	66.52 ± 0.27	6.79 ± 2.45	7.26 ± 3.29	64.96 ± 0.19	6.79 ± 2.45	7.26 ± 3.29
Adversarial	64.26 ± 1.79	4.85 ± 2.16	5.99 ± 2.71	64.22 ± 0.71	4.34 ± 3.87	3.84 ± 2.71
EDITS	62.67 ± 2.64	3.17 ± 2.49	4.54 ± 2.99	62.67 ± 0.51	4.40 ± 2.41	5.38 ± 1.92
FairDrop	$\mathbf{66.79} \pm 0.65$	9.11 ± 1.89	8.35 ± 3.81	64.33 ± 0.44	4.46 ± 1.67	5.02 ± 1.84
$\tilde{\mathbf{h}}^{\mathrm{fair}}$+ GNN	66.34 ± 0.27	$\mathbf{1.23} \pm 1.43$	$\mathbf{2.15} \pm 1.96$	$\mathbf{65.05} \pm 0.21$	$\mathbf{2.13} \pm 0.93$	$\mathbf{2.39} \pm 1.78$

- Similar utility performance compared to fairness-agnostic GNN model
- Enhanced stability for both utility and fairness measures
- Typically better fairness, utility compared to SOTA fairness-aware baselines
- An explanation for effective bias mitigation:

Figure 1: Distribution of the intra-edges (green) and inter-edges (red) in the effective network topology without (left)/ with (right) the application of $\tilde{\mathbf{h}}^{\mathrm{fair}}$.

Conclusions

- A novel, unsupervised bias measure dependent on filter parameters
- Theory-based surrogate loss allowing efficient, LP-based design
- Closed-form solution, leading to optimal and efficient graph filter design
- Versatile use and pre-trained computation
- All results are reproducible: `http://bit.ly/Kose_FairFilterDesign`
- Future work: Computationally efficient (eigendecomposition-free) designs

[1] F. Gama et al., "Stability properties of graph neural networks", *IEEE Transactions on Signal Processing*, 2020.

[2] E. Dai and S. Wang, "Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information." *Proc. International Conference on Web Search and Data Mining*, 2021.