InFoRM: Individual Fairness on Graph Mining

Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong
Team : University of Illinois at Urbana-Champaign
Conference : KDD ‘20

CAU
Junseo, Yu

DMAIS Lab Meeting
09.20.2024

Contents

1. Introduction

- background
- literature review
- research objectives

1. Introduction

Background

- Need for fairness graph mining
- Non-trivial features about non-i.i.d
- Research in graph mining about group fairness has been relatively active recently
- However, Is the group fairness approach the best way?

Risk of group fairness

- Ignoring Individual Differences within Groups
- Representativeness Issues
- Potential for Reverse Discrimination

1. Introduction

Individual Fairness

- Any two individuals who are similar should receive similar algorithmic outcome
- In some situation, finer granularity and reasonable metric. (Not all situation)

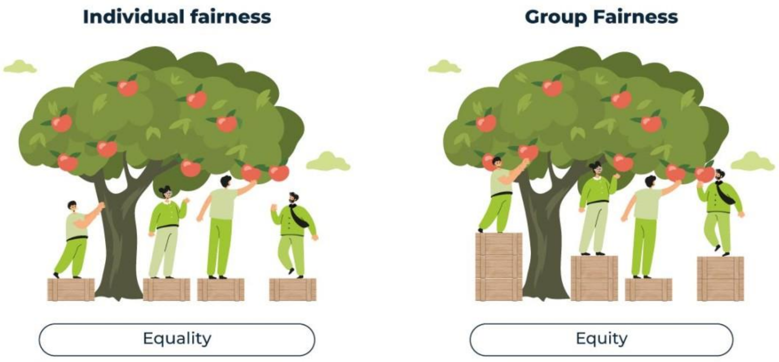
How to measure?

$$D_1(f(x), f(y)) \leq L D_2(x, y) \quad \forall\, (x, y),$$

(D1, D2)-Lipschitz property

$$\|\mathbf{Y}[i,:] - \mathbf{Y}[j,:]\|_F^2 \leq \frac{\epsilon}{\mathbf{S}[i,j]} \; \forall i, j = 1, \ldots, n$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{Y}[i,:] - \mathbf{Y}[j,:]\|_F^2 \mathbf{S}[i,j] = 2\mathrm{Tr}(\mathbf{Y}'\mathbf{L}_\mathbf{S}\mathbf{Y}) \leq m\epsilon = \delta$$



1. Introduction

Literature Review

- Some group fairness graph mining literatures
- Some individual fairness on IID data literatures
- However, No individual fairness graph mining literatures

1. Introduction

Research Objective

- Individual Fairness : Similar individuals -> Similar Outcome

1. Need to tell if the mining results are fair in arbitrary similarity measures.

   (The notion of similarity concepts depends on the context)

2. Need to develop bias mitigating algorithms

3. Need to calculate cost about efficiency and effectiveness.

2. Methodology

- InFoRM Measure
- InFoRM Algorithms
- InFoRM Cost

2. Methodology

InFoRM Measure

Individual Fairness measure

$$\|\mathbf{Y}[i,:] - \mathbf{Y}[j,:]\|_F^2 \leq \frac{\epsilon}{\mathbf{S}[i,j]} \quad \forall i, j = 1, \ldots, n$$

- Quite intuitive : The greater the similarity, the smaller the difference
- However, need to calculate every pair of nodes

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{Y}[i,:] - \mathbf{Y}[j,:]\|_F^2 \mathbf{S}[i,j] = 2\mathrm{Tr}(\mathbf{Y}'\mathbf{L}_\mathbf{S}\mathbf{Y}) \leq m\epsilon = \delta$$

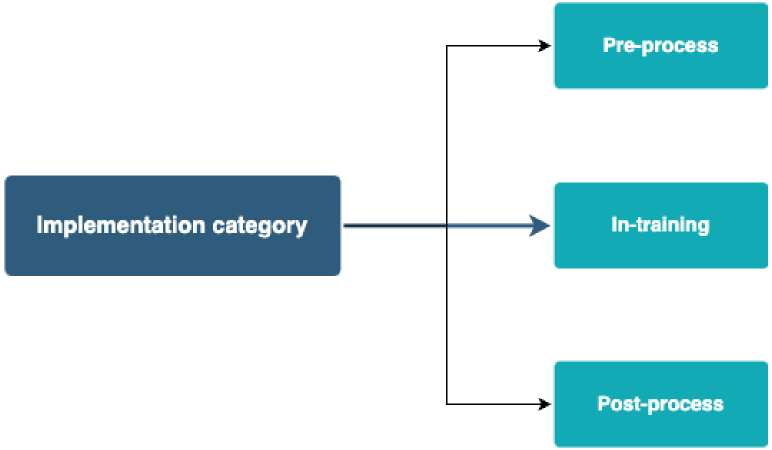
- Transform the equation
- Need to just Calculate Tr() term.

2. Methodology

InFoRM Algorithms

Overview

- Formulate the framework as an optimization problem about individual fairness across all three steps. (The Key idea is quite simple, However..)
  - Debiasing the Input Graph
  - Debiasing the Mining Model
  - Debiasing the Mining Results



2. Methodology

InFoRM Algorithms

Debiasing the Input Graph

$$\min_{\tilde{\mathbf{A}}} \quad \|\tilde{\mathbf{A}} - \mathbf{A}\|_F^2 + \alpha \mathrm{Tr}(\mathbf{Y}'\mathbf{L}_S\mathbf{Y}) \quad \text{s.t.} \quad \mathbf{Y} = \operatorname*{argmin}_{\mathbf{Y}} l(\tilde{\mathbf{A}}, \mathbf{Y}, \theta)$$

- Bi-level optimization problem -> Hard to solve
- Need to change this into lower-level optimization problem by with KKT conditions

$$\min_{\tilde{\mathbf{A}}} \quad \|\tilde{\mathbf{A}} - \mathbf{A}\|_F^2 + \alpha \mathrm{Tr}(\mathbf{Y}'\mathbf{L}_S\mathbf{Y}) \quad \text{s.t.} \quad \partial_{\mathbf{Y}} l(\tilde{\mathbf{A}}, \mathbf{Y}, \theta) = 0$$

2. Methodology

InFoRM Algorithms

Debiasing the Input Graph

Algorithm 1: Debiasing the Input Graph

Input : Adjacency matrix $\mathbf{A}$, similarity matrix $\mathbf{S}$, a mining algorithm $l(\mathbf{A}, \mathbf{Y}, \theta)$, regularization parameter α, learning rate η;

Output : modified topology $\tilde{\mathbf{A}}$ and debiasd mining results $\mathbf{Y}^*$.

initialize $\tilde{\mathbf{A}} = \mathbf{A}$;

while *not converge* **do**

  find $\tilde{\mathbf{Y}} = \text{argmin}_{\mathbf{Y}}\, l(\tilde{\mathbf{A}}, \mathbf{Y}, \theta)$;

  calculate partial derivative $\frac{\partial J}{\partial \tilde{\mathbf{A}}}$ by Eq. (5);

  calculate derivative $\frac{dJ}{d\tilde{\mathbf{A}}}$ based on partial derivative $\frac{\partial J}{\partial \tilde{\mathbf{A}}}$;

  update $\tilde{\mathbf{A}} = \tilde{\mathbf{A}} - \eta \frac{dJ}{d\tilde{\mathbf{A}}}$;

return $\tilde{\mathbf{A}}$ and $\mathbf{Y}^* = \text{argmin}_{\mathbf{Y}}\, l(\tilde{\mathbf{A}}, \mathbf{Y}, \theta)$;

$$J = \|\tilde{\mathbf{A}} - \mathbf{A}\|_F^2 + \alpha \mathrm{Tr}(\tilde{\mathbf{Y}}' \mathbf{L}_{\mathbf{S}} \tilde{\mathbf{Y}}).$$

$$\min_{\tilde{\mathbf{A}}} \quad \|\tilde{\mathbf{A}} - \mathbf{A}\|_F^2 + \alpha \mathrm{Tr}(\mathbf{Y}' \mathbf{L}_{\mathbf{S}} \mathbf{Y}) \quad \text{s.t.} \quad \partial_{\mathbf{Y}} l(\tilde{\mathbf{A}}, \mathbf{Y}, \theta) = 0$$

- Iterate gradient descent!

- Condition
  - have the access to modify the graph
  - have knowledge about the mining model
  - As long as $\partial\tilde{Y} / \partial\tilde{A}$ exists

2. Methodology

InFoRM Algorithms

Debiasing the Mining Model

$$\mathbf{Y}^* = \underset{\mathbf{Y}}{\operatorname{argmin}} \quad J = l(\mathbf{A}, \mathbf{Y}, \theta) + \alpha \mathrm{Tr}(\mathbf{Y}'\mathbf{L}_{\mathbf{S}}\mathbf{Y})$$

- Intuition : Directly incorporate the bias term as a regularization term in the loss
- Condition : as long as a (stochastic) gradient descent solution exists

Algorithm 2: Debiasing the Mining Model

Input : Adjacency matrix $\mathbf{A}$, similarity matrix $\mathbf{S}$, a mining model $l(\mathbf{A}, \mathbf{Y}, \theta)$, regularization parameter α, learning rate η;

Output : Debiased mining results $\mathbf{Y}^*$.

1 solve Eq. (10);

2 **return** $\mathbf{Y}^*$;

$$\frac{dJ}{d\mathbf{Y}} = \frac{\partial J}{\partial \mathbf{Y}} = \frac{\partial l(\mathbf{A}, \mathbf{Y}, \theta)}{\partial \mathbf{Y}} + \alpha \frac{\partial \mathrm{Tr}(\mathbf{Y}'\mathbf{L}_{\mathbf{S}}\mathbf{Y})}{\partial \mathbf{Y}}$$
$$= \frac{\partial l(\mathbf{A}, \mathbf{Y}, \theta)}{\partial \mathbf{Y}} + \alpha(\mathbf{L}_{\mathbf{S}} + \mathbf{L}_{\mathbf{S}}')\mathbf{Y} = \frac{\partial l(\mathbf{A}, \mathbf{Y}, \theta)}{\partial \mathbf{Y}} + 2\alpha \mathbf{L}_{\mathbf{S}}\mathbf{Y}$$

2. Methodology

InFoRM Algorithms

Debiasing the Mining Results

$$\mathbf{Y}^* = \underset{\mathbf{Y}}{\operatorname{argmin}} \quad J = \|\mathbf{Y} - \bar{\mathbf{Y}}\|_F^2 + \alpha \mathrm{Tr}(\mathbf{Y}'\mathbf{L_S}\mathbf{Y})$$

- Intuition : mitigate bias via post-processing strategy
- Condition : as long as mining results **Y** are in a matrix form

Algorithm 3: Debiasing the Mining Results

Input : Vanilla graph mining results $\bar{\mathbf{Y}}$, similarity matrix $\mathbf{S}$, regularization parameter α;

Output : Debiased mining results $\mathbf{Y}^*$.

1 calculate $\mathbf{I} + \alpha\mathbf{L_S}$;

2 solve $(\mathbf{I} + \alpha\mathbf{L_S})\mathbf{Y}^* = \bar{\mathbf{Y}}$;

3 **return** $\mathbf{Y}^*$;

$$\frac{\partial J}{\partial \mathbf{Y}} = \frac{\partial \|\mathbf{Y} - \bar{\mathbf{Y}}\|_F^2}{\partial \mathbf{Y}} + \alpha \frac{\partial \mathrm{Tr}(\mathbf{Y}'\mathbf{L_S}\mathbf{Y})}{\partial \mathbf{Y}} = 0$$

$$\Rightarrow 2\mathbf{Y} - 2\bar{\mathbf{Y}} + 2\alpha\mathbf{L_S}\mathbf{Y} = 0 \Rightarrow (\mathbf{I} + \alpha\mathbf{L_S})\mathbf{Y}^* = \bar{\mathbf{Y}}$$

2. Methodology

InFoRM Cost

Scope

- Question : How the debiased graph mining results Y* would deviate from the vanilla ones
- Scope
  - Debiasing the input graph and mining models are dependent on the context
  - Debiasing the mining results does not. -> Focus on this

3. Results

- setup
- experiment
- contribution & limitation

3. Results

Setup

Downstream Tasks

- Page Rank
- Clustering (spectral clustering)
- Graph Embedding (LINE)

Setup

- Use similarities as Jaccard Index and Cosine Similarity
- The key point of difference is how much it differs from the original.

3. Results

Experiments

PageRank

Table 5: Effectiveness results for PageRank. Lower is better in gray columns. Higher is better in the others.

Debiasing the Input Graph

Datasets	Jaccard Index						Cosine Similarity					
	Diff	KL	Prec@50	NDCG@50	Reduce	Time	Diff	KL	Prec@50	NDCG@50	Reduce	Time
AstroPh	0.059	4.61×10^{-4}	0.840	0.887	16.3%	3632	0.117	1.99×10^{-3}	0.680	0.738	31.9%	3844
CondMat	0.008	1.06×10^{-5}	0.980	0.986	2.16%	1817	0.031	1.57×10^{-4}	0.940	0.957	9.37%	1922
Facebook	0.031	1.83×10^{-4}	0.920	0.943	7.01%	3442	0.072	9.38×10^{-4}	0.760	0.827	16.6%	3623
Twitch	0.109	5.37×10^{-4}	1.000	1.000	24.7%	564.9	0.299	5.41×10^{-3}	0.860	0.899	62.9%	649.3
PPI	0.185	1.90×10^{-3}	0.920	0.944	43.4%	584.4	0.328	8.07×10^{-3}	0.780	0.838	68.7%	636.8

Debiasing the Mining Model

Datasets	Jaccard Index						Cosine Similarity					
	Diff	KL	Prec@50	NDCG@50	Reduce	Time	Diff	KL	Prec@50	NDCG@50	Reduce	Time
AstroPh	0.133	3.28×10^{-3}	0.820	0.871	51.0%	23.08	0.143	4.16×10^{-3}	0.880	0.912	50.4%	26.92
CondMat	0.117	2.43×10^{-3}	0.880	0.915	51.6%	12.02	0.149	4.01×10^{-3}	0.860	0.901	54.6%	12.83
Facebook	0.149	3.33×10^{-3}	0.840	0.884	47.7%	32.41	0.179	4.65×10^{-3}	0.840	0.883	53.3%	33.31
Twitch	0.182	4.97×10^{-3}	0.940	0.958	62.0%	16.18	0.315	1.05×10^{-2}	0.940	0.957	73.9%	12.73
PPI	0.211	4.78×10^{-3}	0.920	0.942	50.8%	10.76	0.280	9.56×10^{-3}	0.900	0.928	67.5%	10.50

Debiasing the Mining Results

Datasets	Jaccard Index						Cosine Similarity					
	Diff	KL	Prec@50	NDCG@50	Reduce	Time	Diff	KL	Prec@50	NDCG@50	Reduce	Time
AstroPh	0.055	1.40×10^{-3}	0.960	0.971	37.4%	0.038	0.094	4.46×10^{-3}	0.960	0.972	49.2%	0.054
CondMat	0.040	8.26×10^{-4}	0.940	0.959	34.4%	0.021	0.082	3.01×10^{-3}	0.780	0.839	48.9%	0.025
Facebook	0.047	1.12×10^{-3}	0.900	0.930	32.6%	0.048	0.086	3.87×10^{-3}	0.960	0.972	44.6%	0.062
Twitch	0.035	9.75×10^{-4}	0.980	0.986	33.9%	0.033	0.101	5.84×10^{-3}	0.940	0.958	44.6%	0.024
PPI	0.045	1.22×10^{-3}	0.940	0.958	27.0%	0.020	0.112	6.97×10^{-3}	0.940	0.958	45.0%	0.019

3. Results

InFoRM

Experiments

LINE

Debiasing the Mining Model														
Datasets	Jaccard Index							Cosine Similarity						
	Diff	Orig. ROC	Fair ROC	Orig. F1	Fair F1	Reduce	Time	Diff	Orig. ROC	Fair ROC	Orig. F1	Fair F1	Reduce	Time
AstroPh	0.462	0.973	0.970	0.924	0.914	51.6%	934.7	0.913	0.973	0.966	0.924	0.906	49.5%	923.0
CondMat	0.302	0.963	0.962	0.922	0.920	44.1%	1130	0.439	0.963	0.961	0.922	0.918	41.6%	1133
Facebook	0.323	0.946	0.954	0.888	0.902	49.6%	1099	0.442	0.946	0.957	0.888	0.906	56.0%	1100
Twitch	0.099	0.687	0.690	0.625	0.625	0.64%	333.8	0.152	0.687	0.694	0.625	0.628	0.83%	340.3
PPI	0.238	0.682	0.715	0.618	0.642	5.85%	180.3	0.418	0.682	0.740	0.618	0.669	7.71%	181.6

Debiasing the Mining Results														
Datasets	Jaccard Index							Cosine Similarity						
	Diff	Orig. ROC	Fair ROC	Orig. F1	Fair F1	Reduce	Time	Diff	Orig. ROC	Fair ROC	Orig. F1	Fair F1	Reduce	Time
AstroPh	0.365	0.973	0.962	0.924	0.898	83.3%	3.284	0.539	0.973	0.963	0.924	0.902	91.1%	6.461
CondMat	0.215	0.963	0.961	0.922	0.918	71.8%	1.464	0.322	0.963	0.960	0.922	0.915	78.4%	2.213
Facebook	0.304	0.946	0.950	0.888	0.890	88.5%	4.122	0.416	0.946	0.953	0.888	0.891	92.4%	7.394
Twitch	0.457	0.687	0.681	0.625	0.629	95.2%	2.320	0.603	0.687	0.658	0.625	0.616	97.6%	4.343
PPI	0.508	0.682	0.713	0.618	0.642	90.1%	1.031	0.722	0.682	0.634	0.618	0.589	97.0%	2.245

3. Results

Experiments

Spectral Clustering

Debiasing the Mining Model								
Datasets	Jaccard Index				Cosine Similarity			
	Diff	NMI	Reduce	RT	Diff	NMI	Reduce	RT
AstroPh	0.885	0.948	10.2%	333.9	1.085	0.868	23.6%	323.4
CondMat	1.108	0.856	26.4%	383.7	1.186	0.742	35.9%	360.7
Facebook	0.972	0.816	31.9%	549.3	0.897	0.810	37.9%	545.0
Twitch	1.147	0.838	88.3%	26.50	1.145	0.875	87.4%	26.62
PPI	0.994	0.658	67.0%	6.047	0.897	0.667	75.2%	6.244

Debiasing the Mining Results								
Datasets	Jaccard Index				Cosine Similarity			
	Diff	NMI	Reduce	Time	Diff	NMI	Reduce	Time
AstroPh	0.071	1.000	24.3%	10.22	0.123	0.984	39.5%	16.46
CondMat	0.071	1.000	34.5%	2.076	0.108	0.985	46.2%	3.196
Facebook	0.056	0.994	24.8%	8.425	0.102	0.994	35.9%	12.81
Twitch	0.150	1.000	90.9%	4.820	0.204	1.000	91.7%	6.513
PPI	0.242	0.811	77.5%	2.896	0.343	0.731	87.4%	4.288

3. Results

Contributions & Limitations

Contribution

- First try of individual fairness in graph mining
- Propose a generic algorithms, InFoRM
- Provide a cost analysis

Limitations

- Is the process explainable?

4. Discussion

- implications
- further directions

4. Discussions

Implications

- Converting fairness problems into optimization problems

Future directions

- Can apply other fairness notion into optimization problems
- Combine each methods in pre-processing, in-training, and post-processing