Learning Fair Representation for Recommendation: A rapt-based Perspective

WWW’21

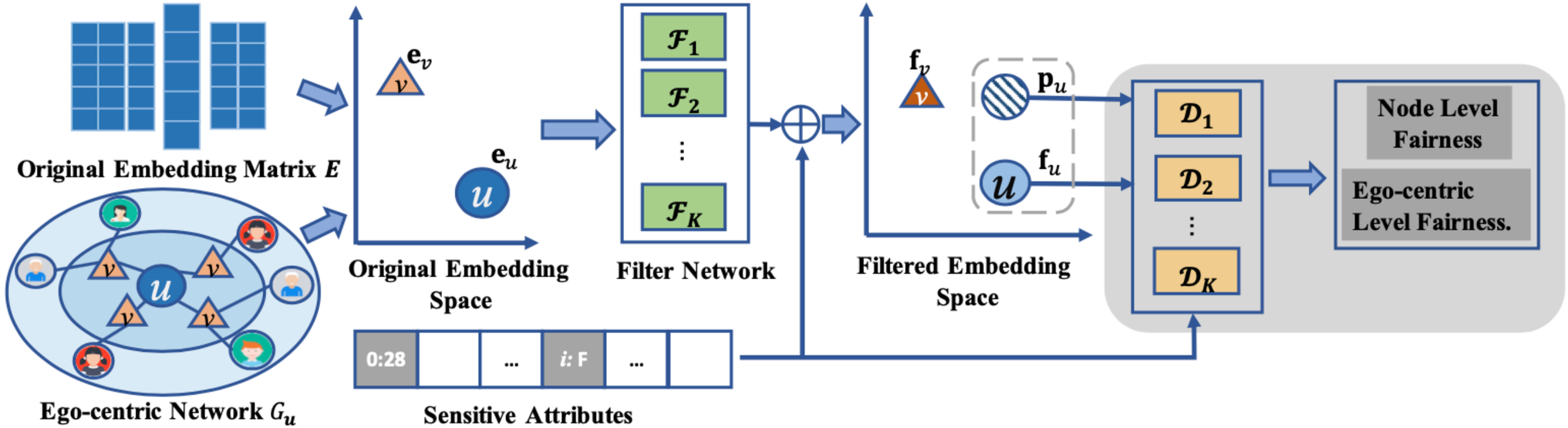
Motivation

- The recommendation results are also vulnerable to biases and unfairness
- Most fairness learning models assume independence of entities, and ignore the correlations among entities.

Contribution

- FairGo can obfuscate sensitive information in the sensitive attribute set and maintains recommendation accuracy.
- The framework is model agnostic and applicable for multiple sensitive attributes.

Model


Original Embedding Matrix E
$\mathbf{e}_v$
$\mathbf{e}_u$
Original Embedding Space
$\mathcal{F}_1$
$\mathcal{F}_2$
$\mathcal{F}_K$
Filter Network
$\mathbf{f}_v$
$\mathbf{p}_u$
$\mathbf{f}_u$
Filtered Embedding Space
$\mathcal{D}_1$
$\mathcal{D}_2$
$\mathcal{D}_K$
Node Level Fairness
Ego-centric Level Fairness.
Ego-centric Network G_u
0:28
...
i: F
...
Sensitive Attributes

Theoretical Analysis

The Overall Value Function

$$\arg\max_{\mathcal{F}} \arg\min_{\mathcal{D}} V(\mathcal{F}, \mathcal{D}) = V_R - \lambda V_G$$
$$= \mathbb{E}_{(u,v,r,x_u)\sim p(\mathbf{E},\mathbf{R},\mathbf{X})}[\ln q_{\mathcal{R}}(r|(\mathbf{f}_u, \mathbf{f}_v) = \mathcal{F}(\mathbf{e}_u, \mathbf{e}_v)) -$$
$$\lambda \ln q_{\mathcal{D}}(x|(\mathbf{f}_u, \mathbf{p}_u) = \mathcal{F}(\mathbf{e}_u, \mathbf{e}_v))],$$

Given enough capacity of the discriminator network

$$\min_{\mathcal{F}} \boxed{H(\mathbf{R}|\mathbf{F})} - \lambda \sum_{k=1}^{K} \boxed{H(\mathbf{X}_k|(\mathbf{F}, \mathbf{P}))}$$

Maximizing the conditional entropy between each sensitive attribute and the filtered embeddings.

- Making sensitive attributes and (F,P) as irrelevant as possible

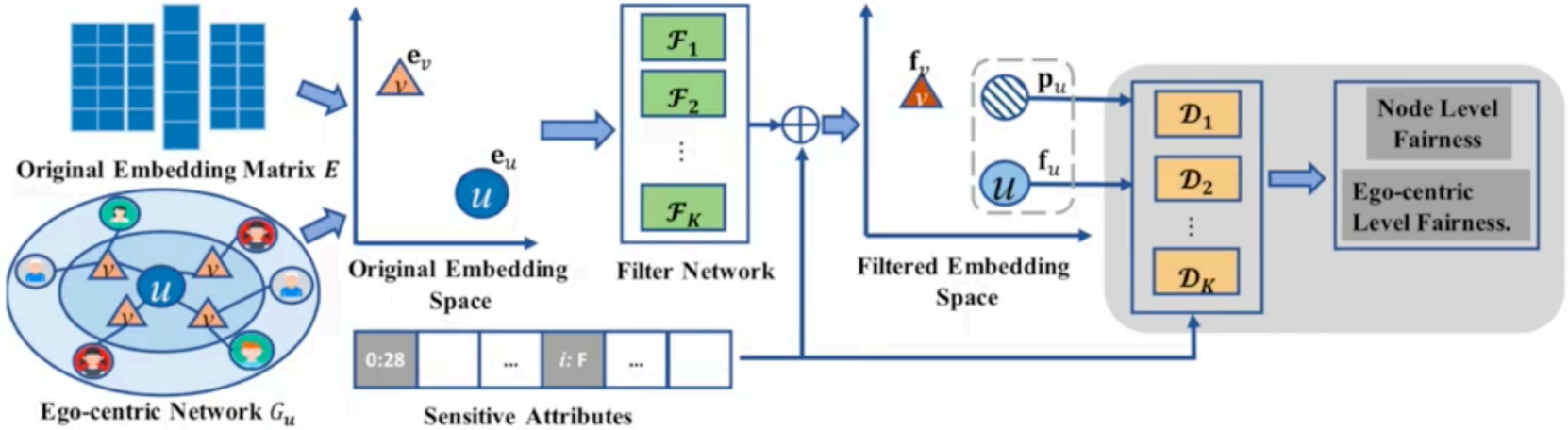
Minimizing the conditional entropy between the ratings and filtered embeddings

- Maintain the necessary information for recommendation

7:59 / 11:29

FairGo

Framework: trade-off between fairness and recommendation



Object Function

$$\arg\max_{\mathcal{F}} \arg\min_{\mathcal{D}} V(\mathcal{F},\mathcal{D}) = V_R - \lambda V_G$$

$$= \mathbb{E}_{(u,v,r,x_u)\sim p(\mathrm{E,R,X})} [\ln q_{\mathcal{R}}(r|(\mathbf{f}_u,\mathbf{f}_v) = \mathcal{F}(\mathbf{e}_u,\mathbf{e}_v)) - \lambda \ln q_{\mathcal{D}}(x|(\mathbf{f}_u,\mathbf{p}_u) = \mathcal{F}(\mathbf{e}_u,\mathbf{e}_v))]$$

→ Guaranteeing the recommendation performance

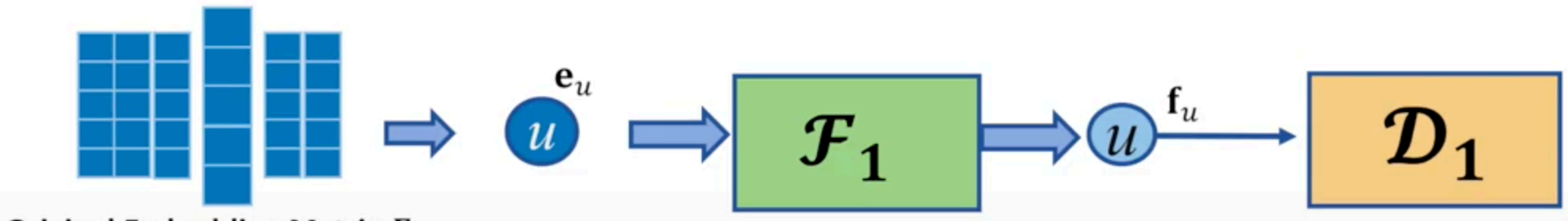
→ Meeting the fairness requirements

FairGo

Framework: fairness part

Node level fairness

$$V_N = \mathbb{E}_{(u,v,r,x_u)} \sum_{k=1}^{K} x_{uk} ln \mathcal{D}^k(\mathbf{f}_u).$$



Removing sensitive attributes

Distinguishing sensitive attributes

6:31 / 11:29

Experiment

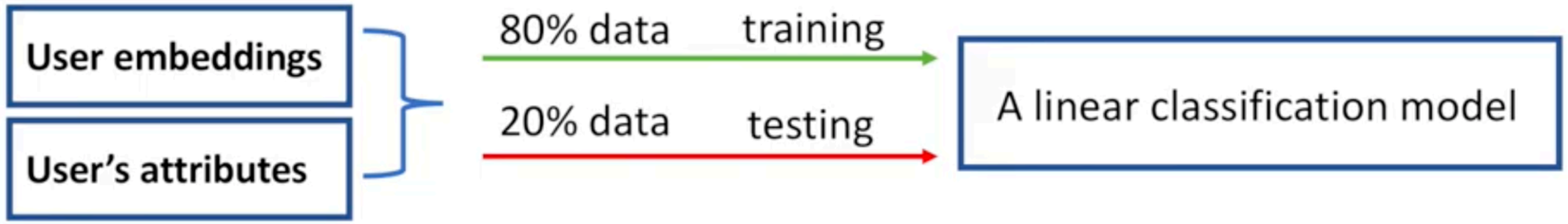
Dataset

	User	Item	Ratings	Attributes
MovieLens-1M	6040	3952	1,000,209	gender (2 classes), age (7 classes), occupation (21 classes)
Lastfm-360K	359,347	292,589	17,559,443	gender (2 classes), age (3 classes)

Evaluation

The recommendation performance: RMSE

The fairness performance: AUC/F1

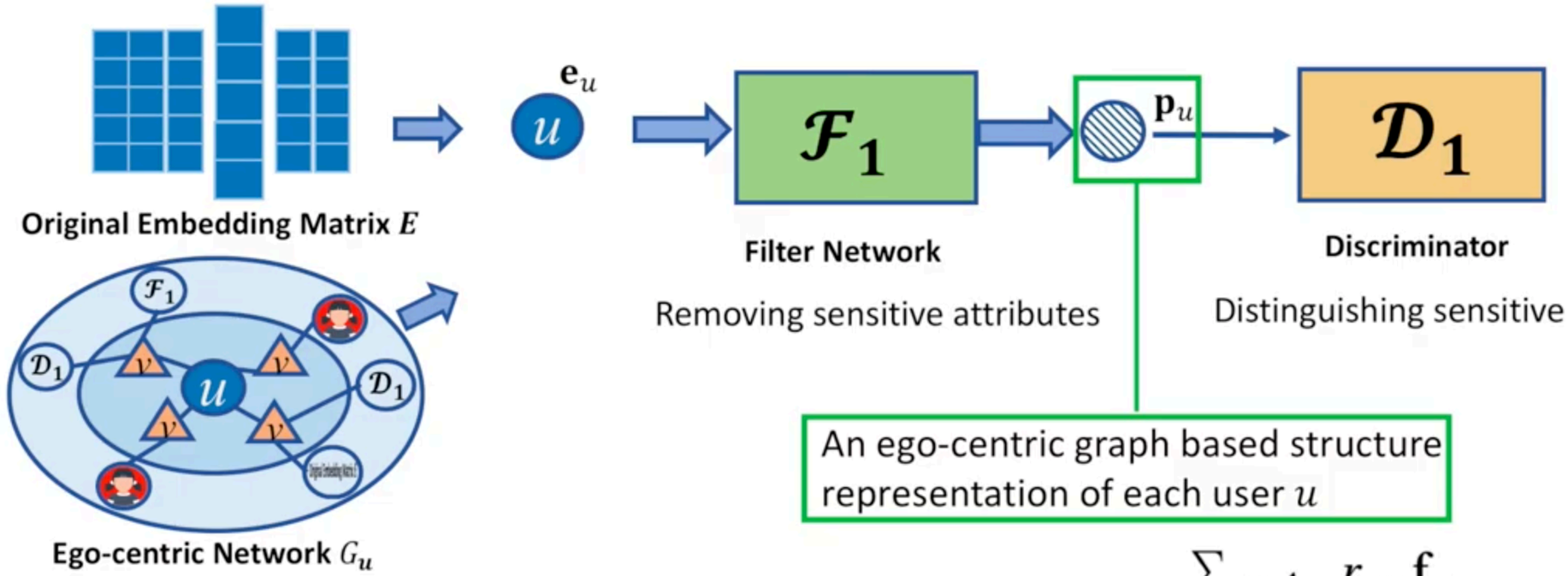


FairGo

Framework: fairness part

Ego-centric level fairness

$$V_S = \mathbb{E}_{(u,v,r,x_u)} \sum_{k=1}^{K} x_{uk} ln \mathcal{D}^k(\mathbf{p}_u)$$



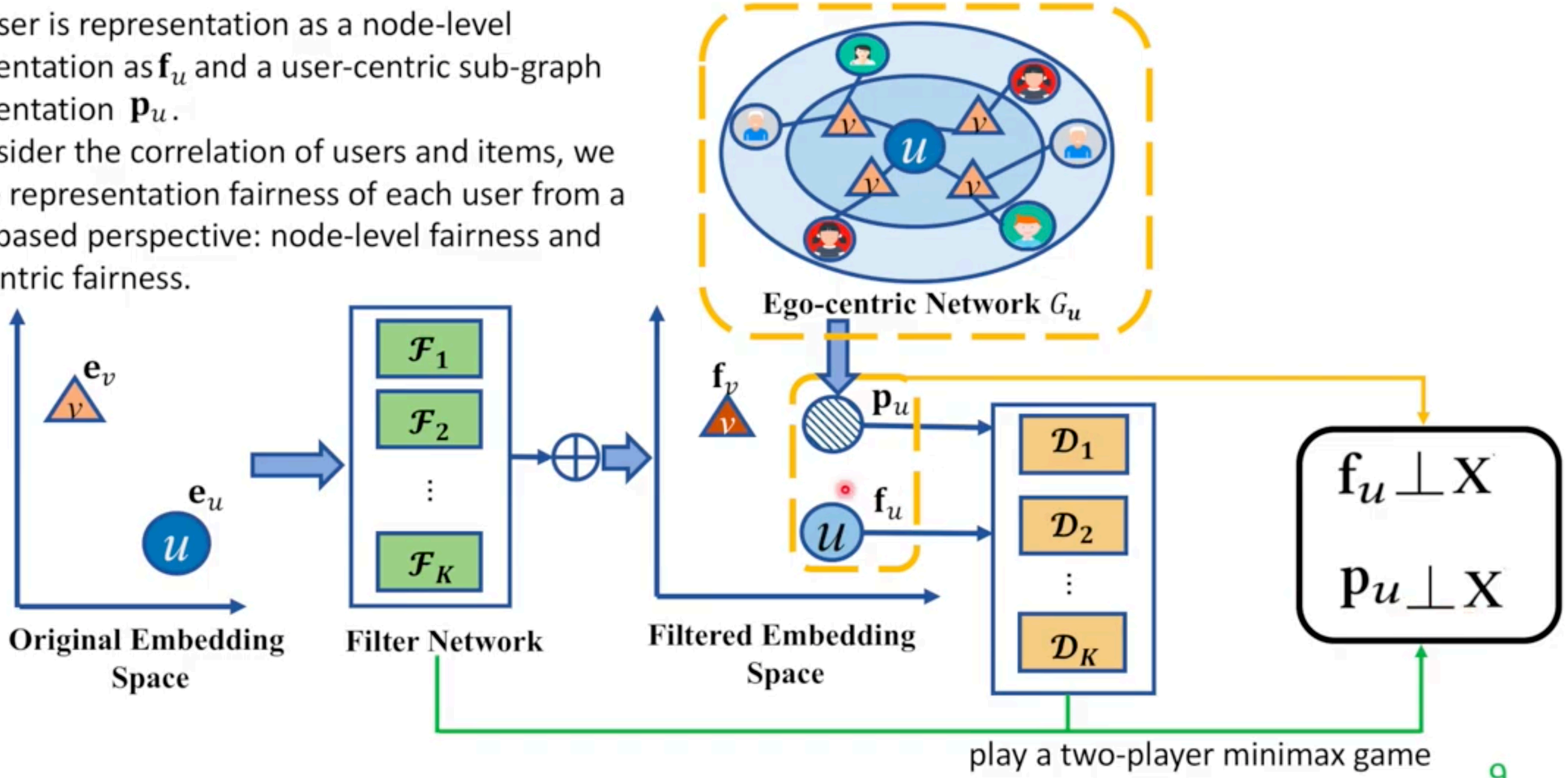
An ego-centric graph based structure representation of each user u

$$\mathbf{p}_u = \mathcal{P}(G_u, \mathbf{F}) = \frac{\sum_{v \in A_u} r_{uv} \mathbf{f}_v}{\sum_{v \in A_u} r_{uv}}$$

FairGo

Framework: meeting fairness

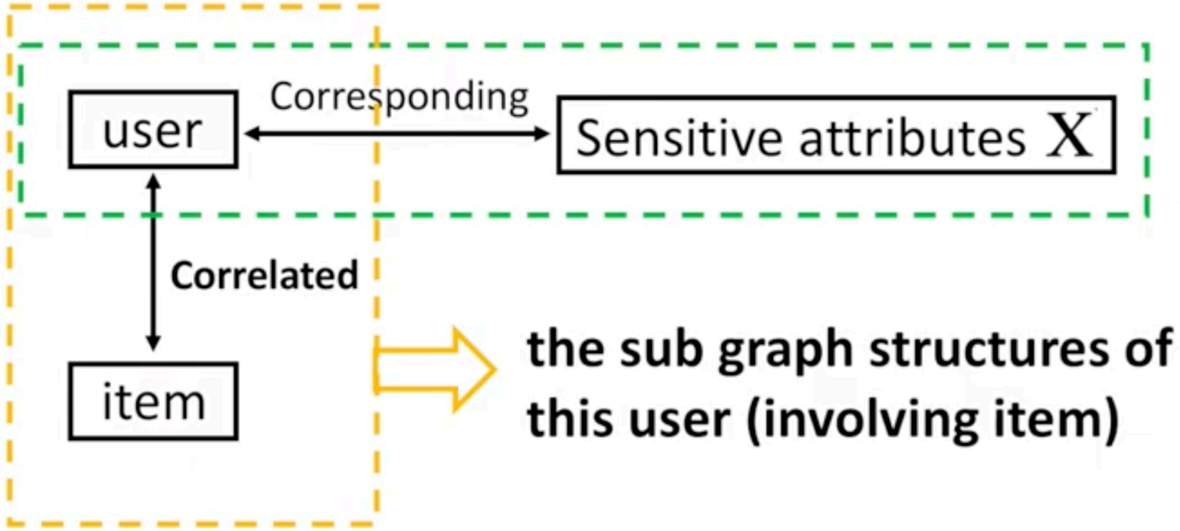
- Each user is representation as a node-level representation as $\mathbf{f}_u$ and a user-centric sub-graph representation $\mathbf{p}_u$.
- To consider the correlation of users and items, we ensure representation fairness of each user from a graph based perspective: node-level fairness and ego-centric fairness.



Solution

Fair Graph based RecOmmendation (FairGo)

For each user, fairness is achieved under the **adversarial learning** of a user-centric graph, in order to obfuscate each sensitive feature between both **the filtered user embedding and the sub graph structures of this user.**



$$\mathbf{f}_u \perp \mathbf{X}$$

➢ Filtering sensitive attributes from user embedding

$$\mathbf{p}_u \perp \mathbf{X}$$

➢ Filtering sensitive attributes from the sub graph structures of this user.