Fair Walk & Cross Walk

Fairwalk: Towards Fair Graph Embedding
Tahleen Rahman, Bartlomiej Surma,
Michael Backes and Yang Zhang
CISPA Helmholtz Center for Information Security
Conference : IJCAI' 19.

CrossWalk: Fairness-Enhanced Node Representation Learning
Ahmad Khajehnejad, Moein Khajehnejad, Mahmoudreza Babaei,
Krishna P. Gummadi, Adrian Weller, Baharan Mirzasoleiman
Conference : WWW' 23

CAU
Junseo, Yu

DMAIS Lab Meeting
09.20.2024

Contents

01 Fairness Notion
- Why do we need?
- How can we measure?
- Examples

02 Related Work
- Existing methods
- Implementation category on graph
- random walk based methods
  (deep walk, random walk)

03 Fair Walk
- implementation
- experiment
- contribution & limitation

04 Cross Walk
- implementation
- experiment
- contribution & limitation

Fair Walk & Cross Walk

1. Fairness Notion

- Why do we need?
- How can we measure?
- Examples

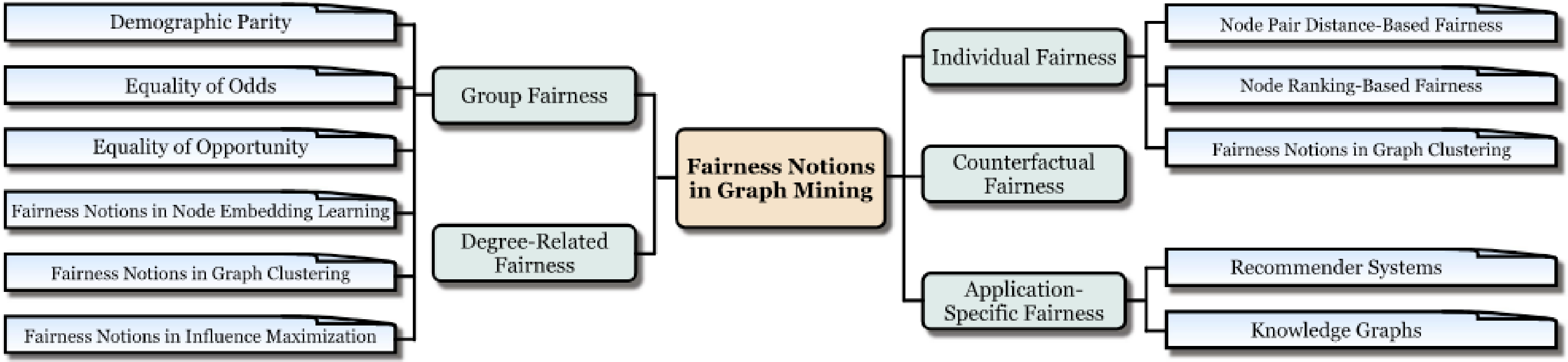
1. Fairness Notion

Why do we need?

- Fairness : the quality of treating people equally or in a way that is right or reasonable
- ML could unreasonably discriminate people
  - Learned from data
  - may have captured undesired sensitive information
  - Can not distinguish between correlation and causality
- Even, this can reinforce the social bias(discrimination) into undesired way
- It is directly related with human centered task
  - Job recommender systems
  - Disaster response
  - Criminal justice
  - Loan approval

1. Fairness Notion

Existing Methods



Taxonomy of algorithmic fairness notions in graph mining algorithms.

Yushun Dong, Fairness in Graph Mining: A Survey, TKD 2023

1. Fairness Notion

Examples

Statistical Parity Statistical Parity(a.k.a. Demographic Parity)

- all groups, regardless of sensitive attributes like gender or race, should have an equal probability of receiving a positive outcome.

$$P(\hat{Y} = 1 | S = 0) = P(\hat{Y} = 1 | S = 1).$$

For easy understanding

$$P(G_{ij}^{\mathcal{S}}) = \left|\{(u,v) : v \in \rho(u) \wedge (u,v) \in G_{ij}^{\mathcal{S}}\}\right| / \left|G_{ij}^{\mathcal{S}}\right|$$

$$\text{bias}^{\text{SI}}(\mathcal{G}^{\mathcal{S}}) = \text{Var}(\{P(G_{ij}^{\mathcal{S}})\} : G_{ij}^{\mathcal{S}} \in \mathcal{G}^{\mathcal{S}})$$

In the fair walk paper $G_{ij}^{\mathcal{S}} = \{(u,v) : \zeta(u) = i \wedge \zeta(v) = j \wedge u, v \in U\}.$

1. Fairness Notion

Examples

Equality of Representation

all groups should be evenly distributed across the predicted outcomes.

$$\text{bias}^{\text{ERg}}(\mathcal{G}^{\mathcal{S}}) = \text{Var}(\{N(G_{ij}^{\mathcal{S}})\} : G_{ij}^{\mathcal{S}} \in \mathcal{G}^{\mathcal{S}})$$

$$\text{bias}^{\text{ERu}}(z) = \frac{1}{|\mathcal{Z}^{\mathcal{S}}|} - \frac{\sum_{u \in U} z\text{-share}(u)}{|U|} \qquad z\text{-share}(u) = \frac{|\rho_z(u)|}{|\rho(u)|}$$

Beware, thins fairness notion is not suitable for all situations

Disparity

Differences in specific metrics (e.g., positive prediction rate, error rate) across different groups.

$$\text{disparity}(A) = Var(\{Q_i\} : i \in [C])$$

Q : fraction of positive decisions
i : specific group e.g., male and female

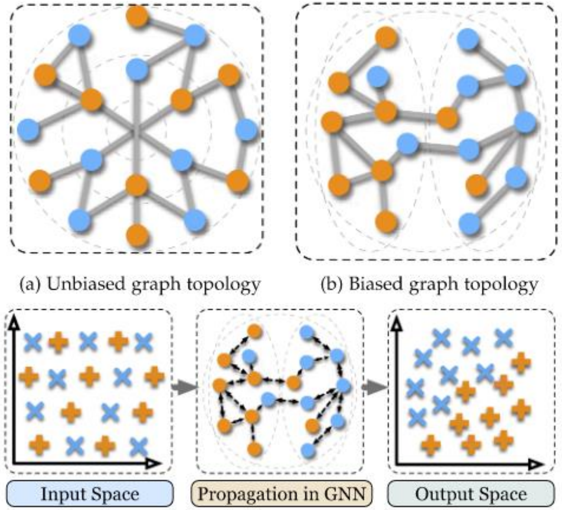
2. Related Work

- Existing methods
- Implementation category on graph
- random walk based methods
  (deep walk, random walk)

2. Related Work

Existing Methods

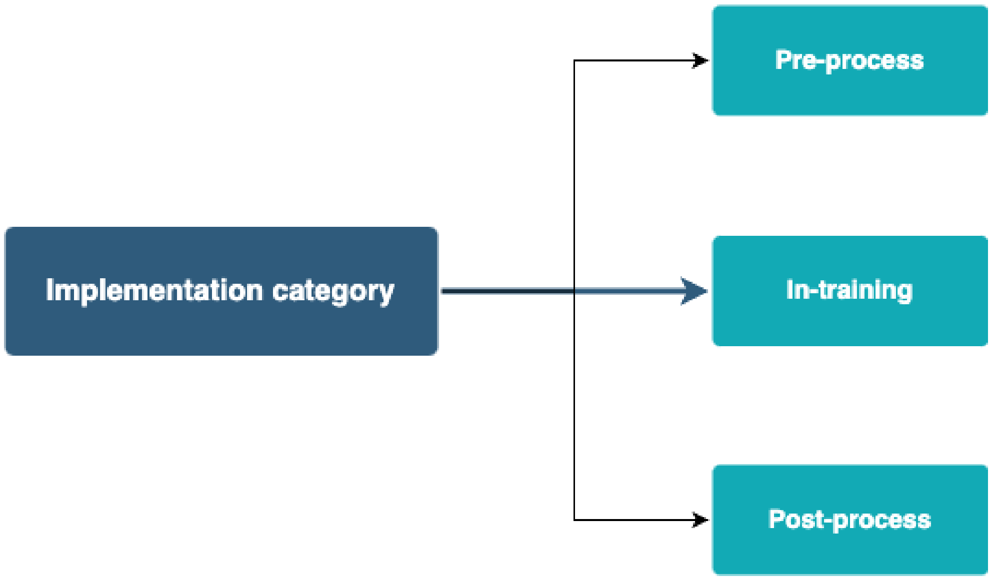
- Fairness in machine learning notions have been significantly developed since 2014.
- However, unlike i.i.d. data mining algorithms Most of graph mining algorithms lack fairness consideration due to non-i.i.d. characteristc (non-iid features)
- Why fulfilling fairness in graph mining is non-trivial?
  - Graph structure
  - Relational information (topology)
  - Need to prevent the graph mining algorithms from inheriting the bias exhibited in the input relational info



Yushun Dong, Fairness in Graph Mining: A Survey, TKD 2023

2. Related Work

Implementation category on graph



- Pre-process

  Manipulate biased graph data

  e.g., edge rewiring
- In-training

  Control in model learning process

  e.g., optimization with constraint(s)
- Post-process

  Regulate prediction output

  e.g., orthogonal projection

2. Related Work

Random walk based methods

Random walk

- Goal

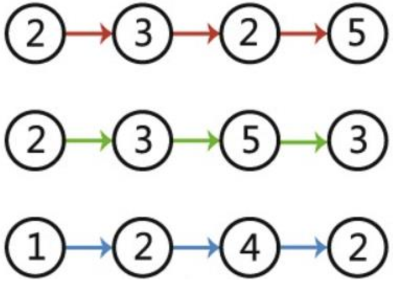
  Capture graph structure information

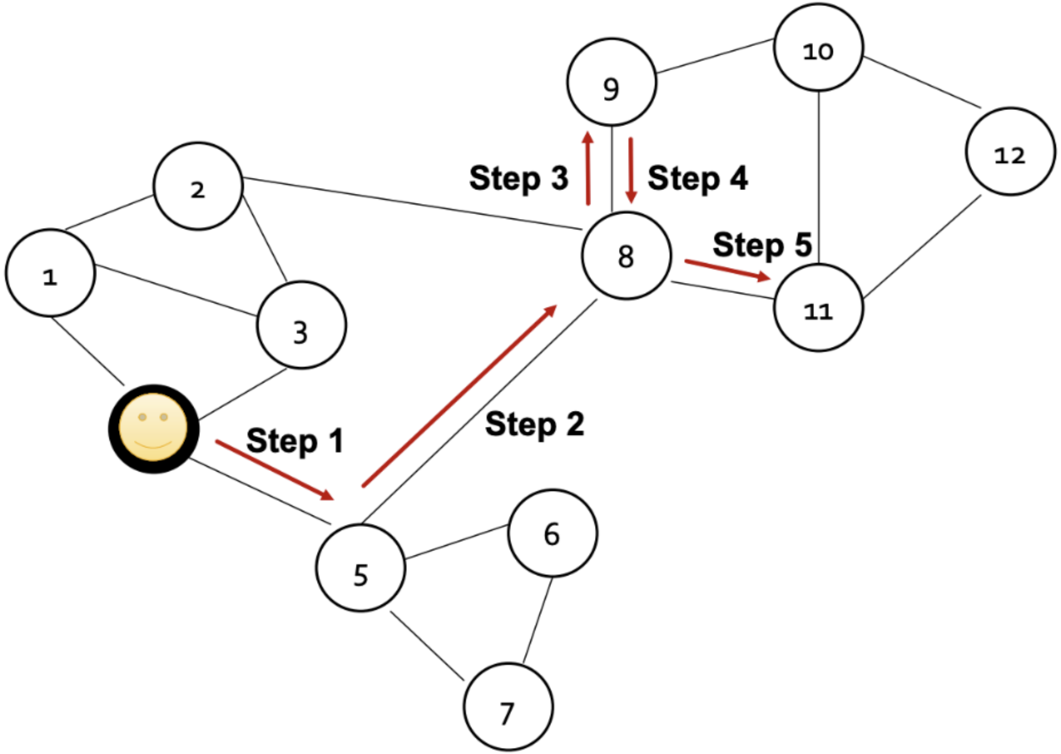
- Method

  Visit neighbor node “randomly”

  and record the visting order.

  Iterate!





Jure Leskovec, Stanford CS224W: Machine Learning with Graphs

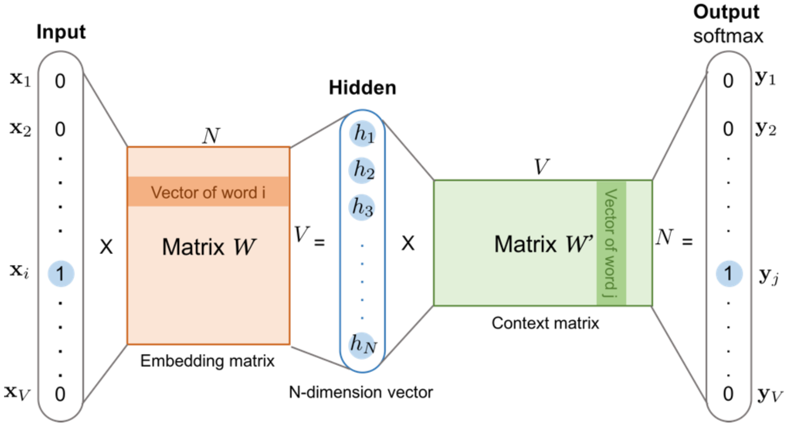
2. Related Work

Random walk based methods

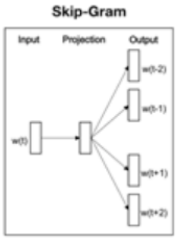
Deep walk

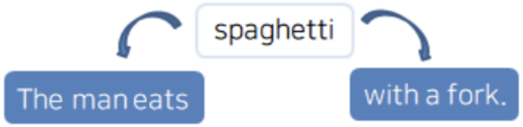
- Goal : Generate node embeddings that incorporate structural information by using output of random walk
- Methods : use the skip-gram model of Word2Vec (NLP field methods)

  This model enables to generate the node embeddings that will be input of the downstream tasks such as link prediction
- Key point : The visiting nodes and their order during the random walk process directly influence the resulting node embeddings.



Skip-gram





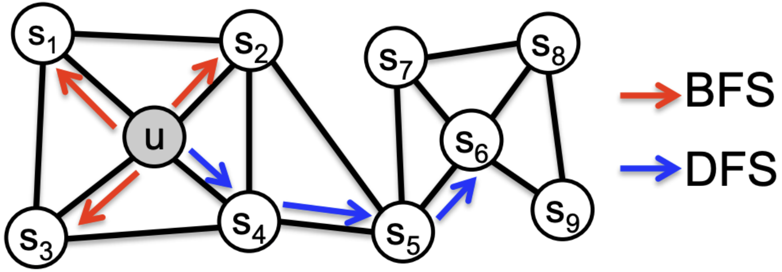
2. Related Work

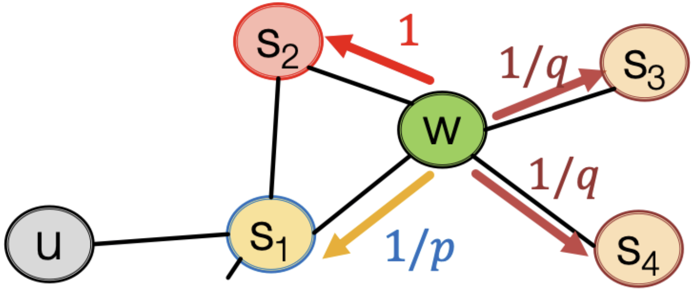
Random walk based methods

node2vec

- Goal : Capture local and global neighborhood information properly
- Methods : Same with Deep walk except for random walk process.

  The Node2vec does biased random walk.
- Output : More concisely capture the structure information by modifying random walk process





Jure Leskovec, Stanford CS224W: Machine Learning with Graphs

3. Fair Walk

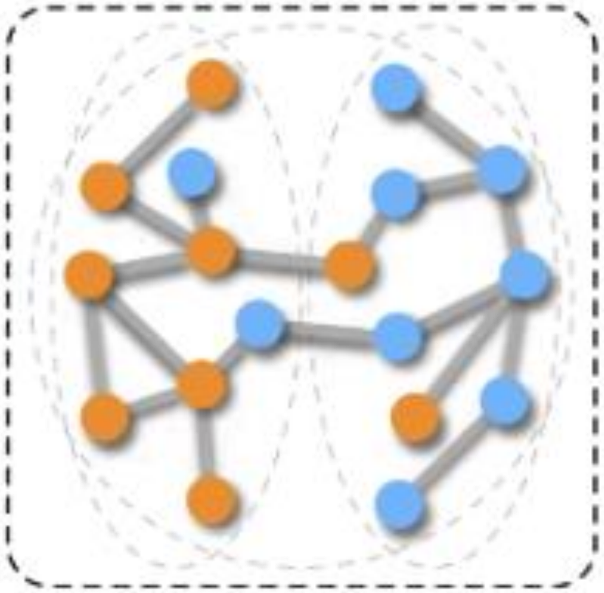
- implementation
- experiment
- contribution & limitation

3. Fair Walk

implementations

Why?

- Even if it reduces the performance, Why do we need to adjust the learning process instead of training on the origin graph as it is.
- Experimentally, this method has shown improvements across multiple fairness criteria.
- Think about random walk process in the biased graph like right one.



(b) Biased graph topology

3. Fair Walk

implementations

Why?

- Thin about some specific tasks. For example in friendship recommender system, the system may sugesst same group. Perhaps, this is a deprivation of opportunity.
- Figure 1 shows that bias are similar with origin network and output of node2vec -> mirrors the gap between minorities and majorities

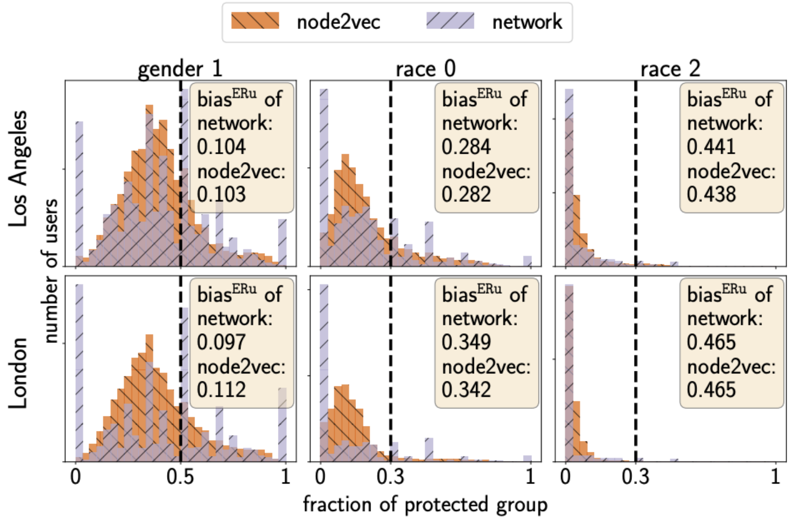


Figure 1: z-share distributions of node2vec and original network. The vertical line shows the *fair fraction* (0.5 and 0.3)

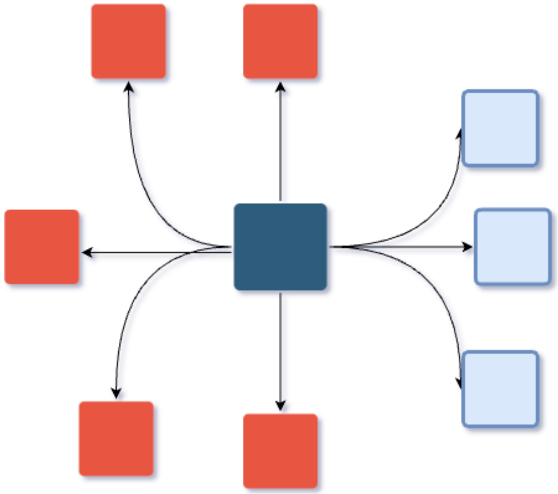
3. Fair Walk

implementations

How?

- Change random walk procedure from random to biased way
- Detail
  1. Partition neighbors into groups
  2. Give each group the same probability of being chosen (regardless of their size)

VERY SIMPLE!



Probability of next walk

- Red : 5/8 -> 1/2
- Blue : 3/8 -> 1/2

3. Fair Walk

Experiments

Ratio

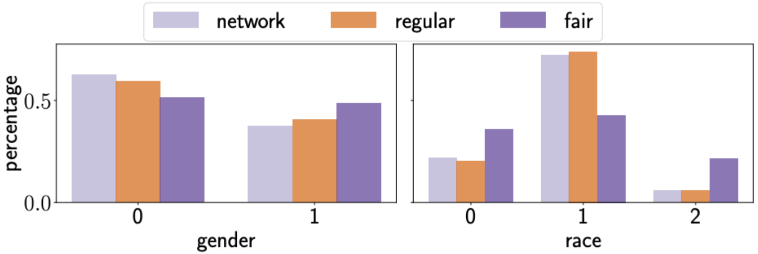


Figure 2: Ratio of each gender and race in the original network and regular and fair random walk traces in Los Angeles dataset

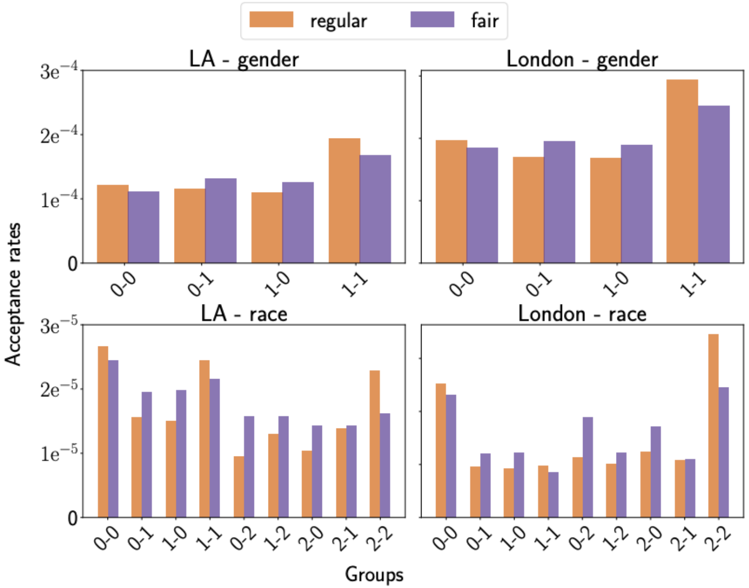
Setup

- Data : Instagram API
- Sensitive attributes : race and gender
- How classify? : Face++ (Face recognition)

  -> Pick with highest confidence at least 20 percentage points higher than the second
- Task : Friendship recommendation

3. Fair Walk

Experiments

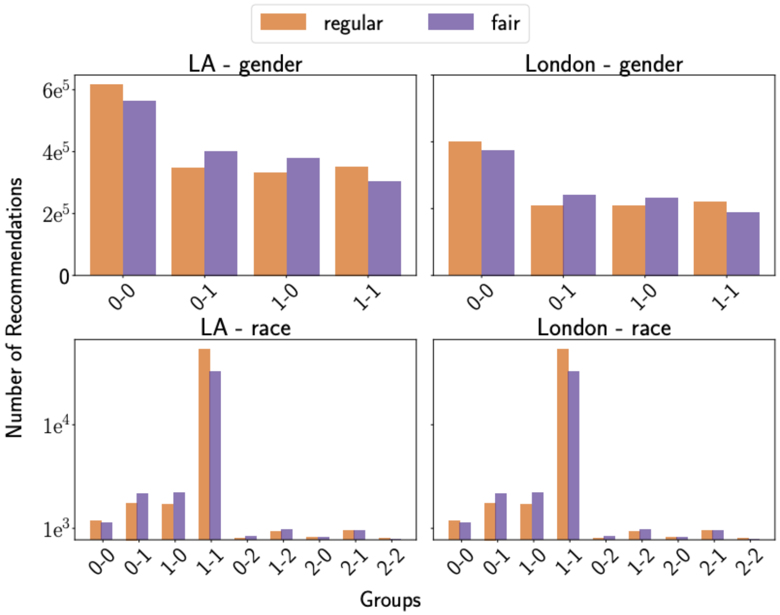


		LA		London	
		gender	race	gender	race
ERg	regular	$1.3e^{10}$	$2.5e^{7}$	$6.5e^{9}$	$2.4e^{7}$
	fair	$0.8e^{10}$	$1.9e^{7}$	$4.8e^{9}$	$1.9e^{7}$
SI	regular	$4.7e^{-9}$	$1.4e^{-12}$	$1.1e^{-8}$	$7.1e^{-11}$
	fair	$1.7e^{-9}$	$0.4e^{-12}$	$0.2e^{-8}$	$2.8e^{-11}$

Table 3: bias^{SI} and bias^{ERg} for both cities (lower, the better)

3. Fair Walk

Experiments

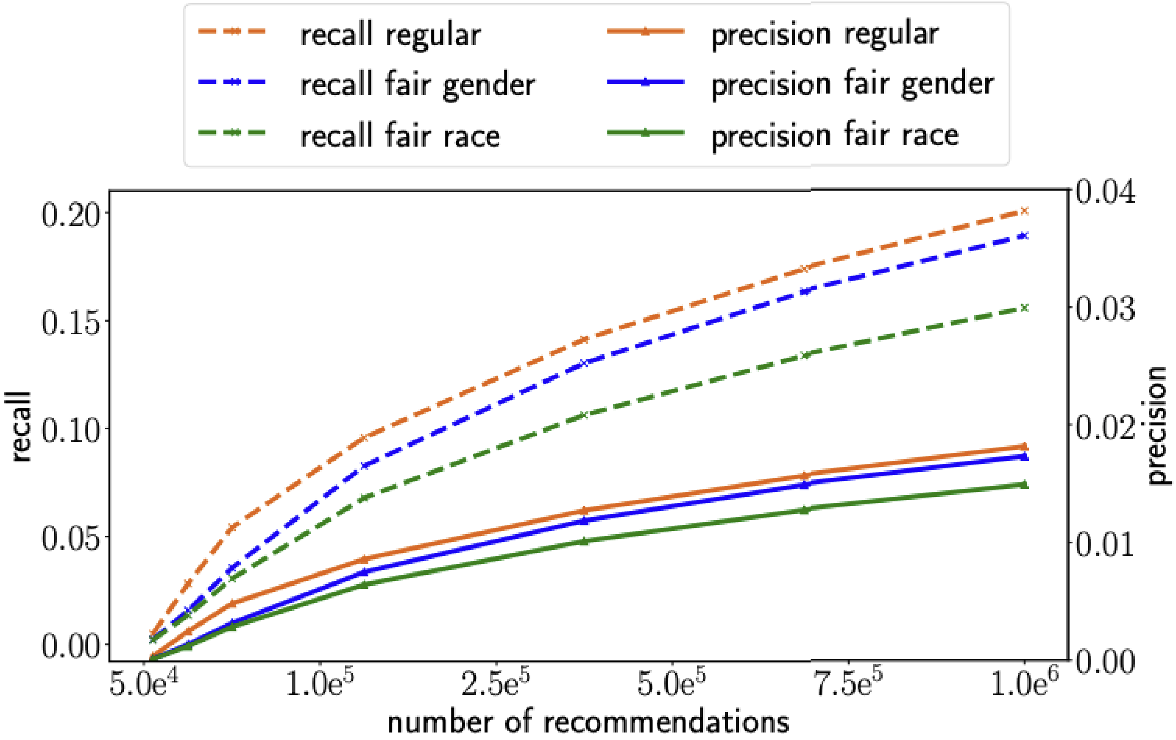


		gender 0	gender 1	race 0	race 1	race 2
LA	network	0.104	0.104	0.117	0.392	0.275
	node2vec	0.103	0.103	0.115	0.387	0.272
	fairwalk	0.068	0.068	0.054	0.288	0.234
London	network	0.097	0.097	0.183	0.481	0.298
	node2vec	0.112	0.112	0.176	0.474	0.298
	fairwalk	0.095	0.095	0.135	0.417	0.282

Table 4: Bias by *Equality of Representation* at user level for both genders and all three races (lower, the better).

3. Fair Walk

Experiments



3. Fair Walk

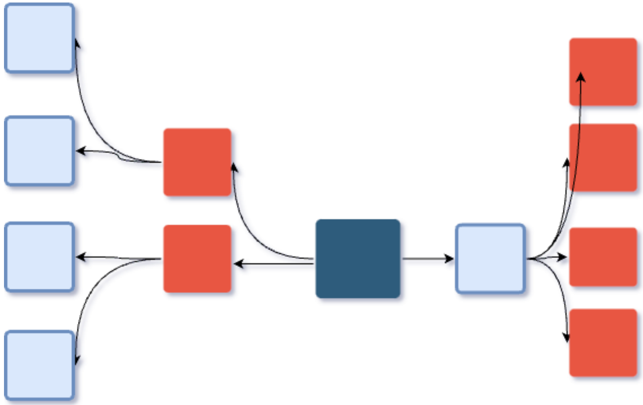
contribution & limitation

Contribution

- First study of fairness in the grping embedding methods
- Can generalize if the methods is based on random walk

Limitations

- Fail to caputre the neighborhoods located in more than one-hop away information



Fair Walk & Cross Walk

4. Cross Walk

- implementation
- experiment
- contribution & limitation

4. Cross Walk

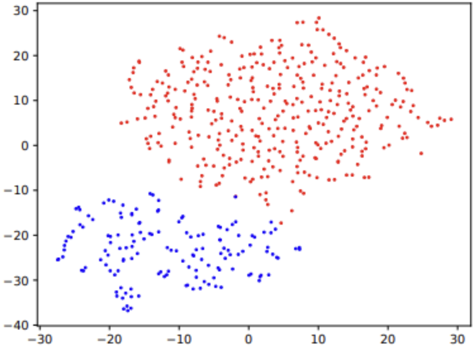
implementations

Goal

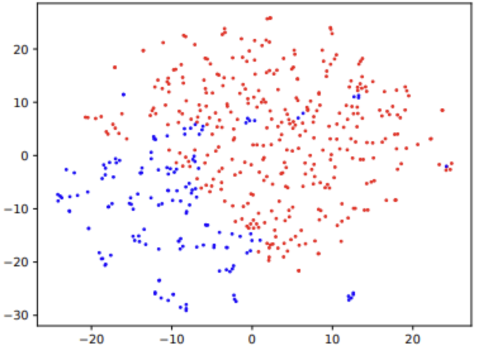
- manipulate random walk process in more detail

How?

- Upweighting edges
  - Closer the groups' peripheries
  - Connecting different groups



(a) Synthetic Layered Dataset DeepWalk



(b) Synthetic Layered Dataset CrossWalk ($\alpha = 0.5$, $p = 2$)

4. Cross Walk

implementations

Bias towards Group Boundaries

$$m(v) = \frac{\sum_{j \in [r]} \sum_{u \in \mathcal{W}_v^j} \mathbb{I}[l_v \neq l_u]}{r \times d}.$$

m(v)	: measure of proximity
W_u	: The path of random walk
W_v	: The group to which v belongs
r	: Number of random walk
d	: length of random walk

- Every node have m(v)
- Intuitively, this indicated the fraction of nodes from other groups in v's close proximity.

  -> How close to other groups (The higher, the closer)
- Can capture more distant neighborhood information due to considering random walk path

4. Cross Walk

implementations

Bias towards Other Groups

$$w'_{vu} = \begin{cases} w_{vu}(1-\alpha) \times \frac{m(u)^p}{\sum_{z \in N_v} w_{vz} m(z)^p} & \text{if } l_v = l_u \\ w_{vu}\alpha \times \frac{m(u)^p}{|R_v| \sum_{z \in N_v^c} w_{vz} m(z)^p} & \text{if } l_v \neq l_u = c. \end{cases}$$

- If same group -> upper case
- If different group -> lower case
- α : manipulate strengths to the other group (The higher, the powerful)
- p : manipulate strengths to the group boundaries (The higher, the powerful)

4. Cross Walk

implementations

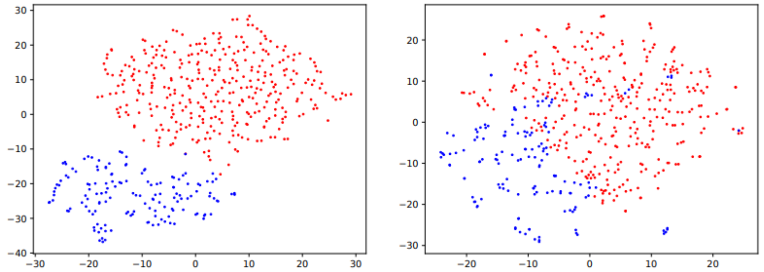
Which things are better than fair walk

- Can consider one-hop away nodes

  -> Cross walk consider few steps away nodes
- This pulls the final embeddings of the nodes
- Can consider carefully the degree of inner and outer reweights values.

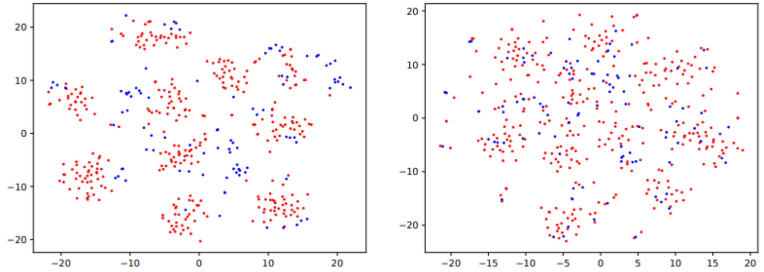
  (correspond to α and p respectively.)

$$w'_{vu} = \begin{cases} w_{vu}(1-\alpha) \times \frac{m(u)^p}{\sum_{z \in N_v} w_{vz} m(z)^p} & \text{if } l_v = l_u \\ w_{vu}\alpha \times \frac{m(u)^p}{|R_v| \sum_{z \in N_v^c} w_{vz} m(z)^p} & \text{if } l_v \neq l_u = c. \end{cases}$$



(a) Synthetic Layered Dataset DeepWalk

(b) Synthetic Layered Dataset CrossWalk ($\alpha = 0.5, p = 2$)



(c) Rice-Facebook Dataset DeepWalk

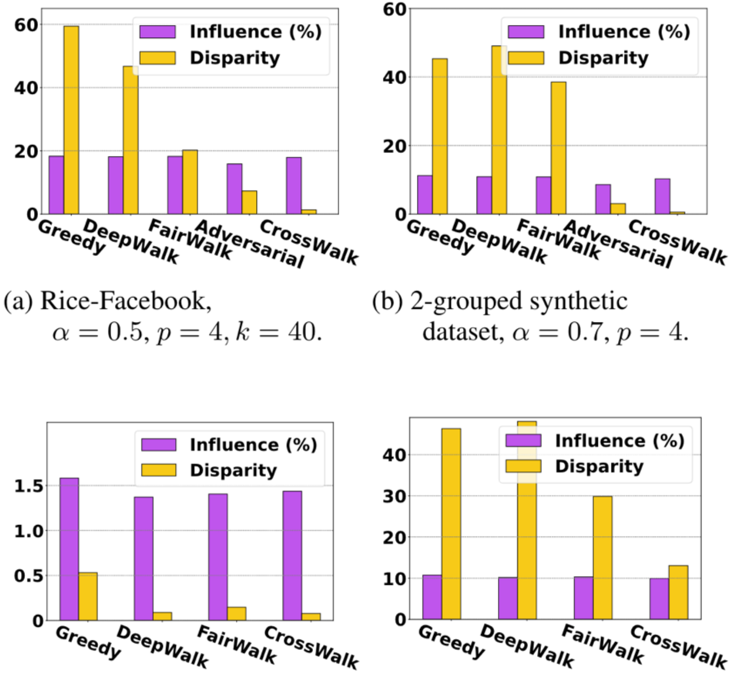
(d) Rice-Facebook Dataset CrossWalk ($\alpha = 0.5, p = 4$)

4. Cross Walk

Experiments

Setup

- Use Deep Walk embeddings
- Also compare with Adversarial Embedding (Khajehnejad et al. 2020). (only applicable to network of 2 groups.)



(a) Rice-Facebook, $\alpha = 0.5, p = 4, k = 40$.

(b) 2-grouped synthetic dataset, $\alpha = 0.7, p = 4$.

(c) Twitter dataset, $\alpha = 0.5, p = 2$.

(d) 3-grouped synthetic dataset, $\alpha = 0.7, p = 4$.

Figure 4: Influence Maximization

4. Cross Walk

Experiments

Setup

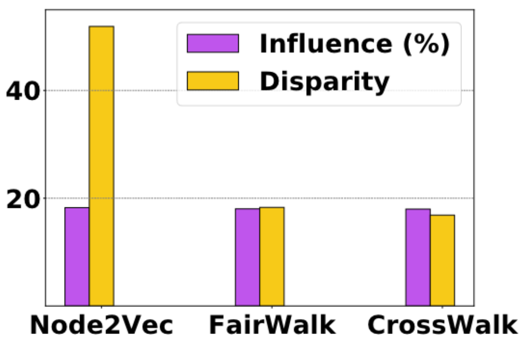


Figure 5: Influence Maximization - CrossWalk and FairWalk on Node2vec

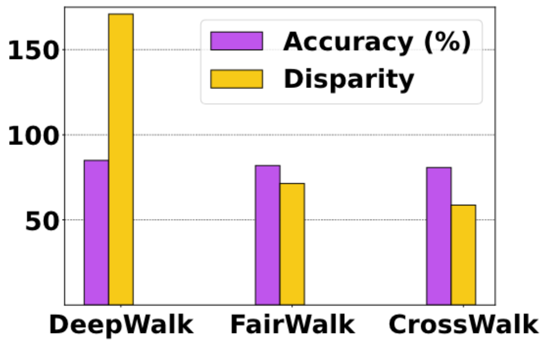


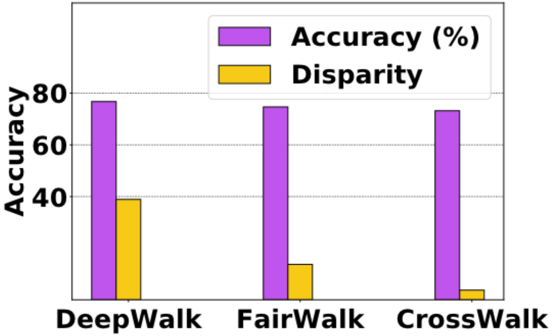
Figure 6: Node Classification - Rice-Facebook dataset. $\alpha = 0.5, p = 1.$

- Use Node2vec (p=0.5, q=0.5)

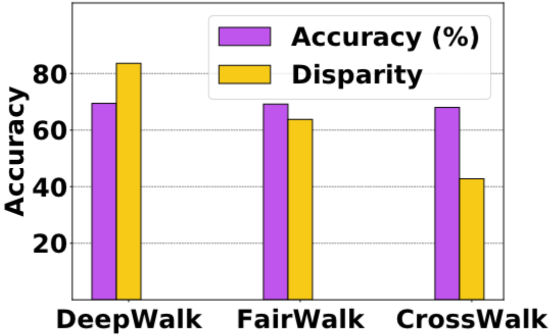
4. Cross Walk

Experiments

Setup



(a) Rice-Facebook Dataset, $\alpha = 0.5, p = 2.$



(b) Twitter Dataset, $\alpha = 0.5, p = 2$

Figure 7: Link Prediction

4. Cross Walk

contribution & limitation

Contribution

- More precise calculation about fair random walk process

Limitations

- Is the result output truly surprising?