The Alan Turing Institute

Fairness on Graph

Dr. Taufiq Asyhari

https://www.arxiv-vanity.com/papers/2004.12265/

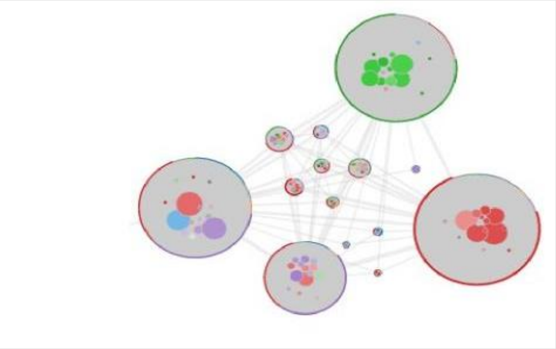
Courtesy

- J Kang and H Tong, Fair Graph Mining, CIKM Tutorial, Nov. 2021.
- I Spinelli, et al., Biased Edge Droput for Enhancing Fairness in Graph Representation, Sapienza Universita di Roma, 2021.
- R Ladysz, Graph Mining: A general overview of some mining techniques, George Mason University.

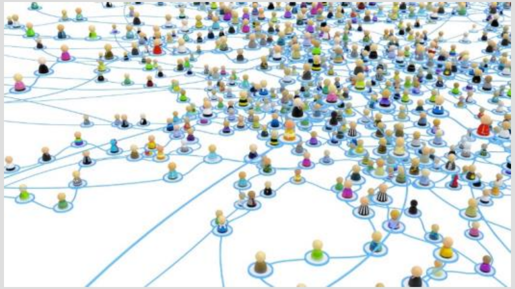
Outline

- Overview
- Graph Preliminaries
- Fairness on Graph Ranking
- Fairness on Graph Embedding
- Remarks and Summary

Prevalence of Graphs / Networks



Collaboration Networks



Social Networks

Biological Networks

Mining from Graph

Social Influence / Policy Treatment

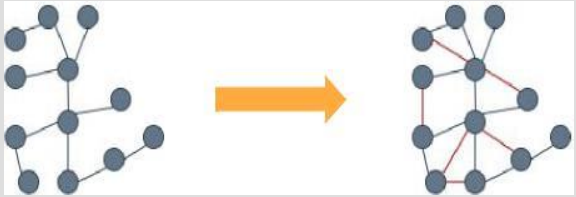
Biology, e.g., Brain Network

[1] Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G.. Network Analysis in the Social Sciences. Science 2009.

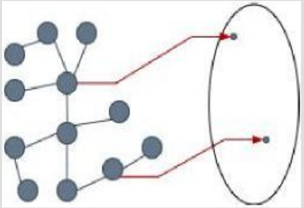
[2] Wang, S., He, L., Cao, B., Lu, C. T., Yu, P. S., & Ragin, A. B.. Structural Deep Brain Network Mining. KDD 2017.

Common Tasks

Link Prediction



Node Representation Learning

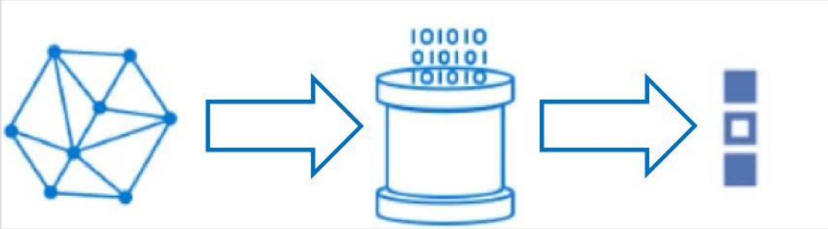


Graph spaceEmbedding

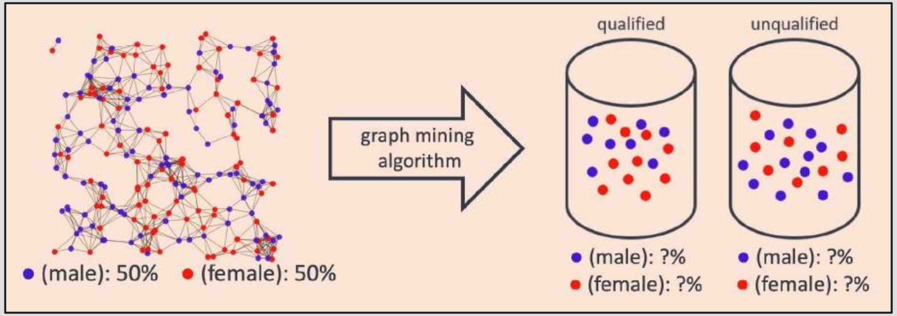
Pipeline of Graph Mining

- Pipeline

Input Model Results



- Example:



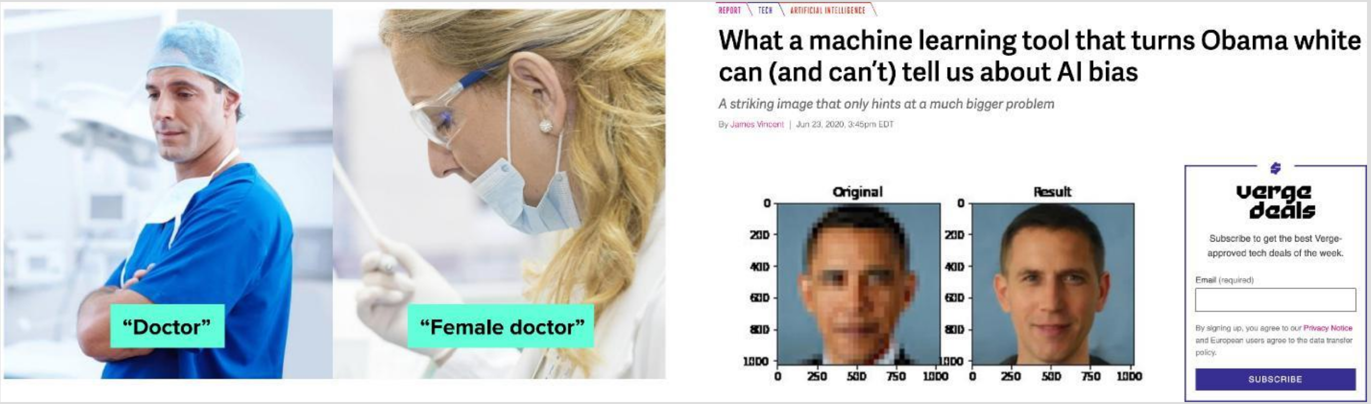
Fairness in Graph Mining

- How to ensure the mining process is fair?

  - Why are two seemingly similar applicants belong to different admission groups?
  - Why does the algorithm ‘decide’ a certain individual likely receive benefits?
  - Why is a particular content more likely to go viral than the others?
  - How does the prior connectivity affect the algorithm results?
  - How do bot inputs change trending topics?

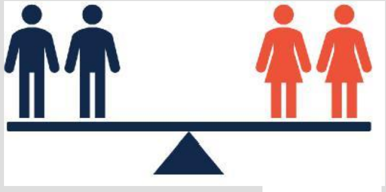
Algorithmic Fairness

– **Motivation**: Mitigate unintentional bias caused by machine learning (ML) algorithms

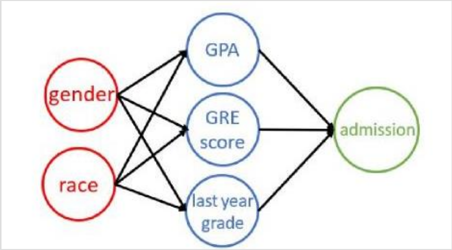


Algorithmic Fairness

- Definition: Lack of favoritism from one side or another



- Group fairness
  - Statistical parity
  - Equal opportunity



- Individual fairness
  - More tailored than group fairness

Fairness in Graph Mining

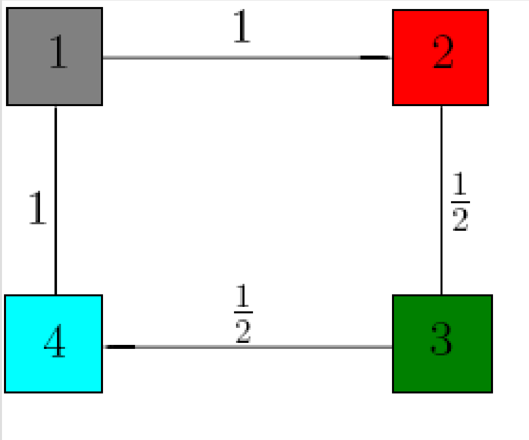
- Graph is a way to represent data or information using a collection of points, called "nodes," and lines connecting them, called "edges."

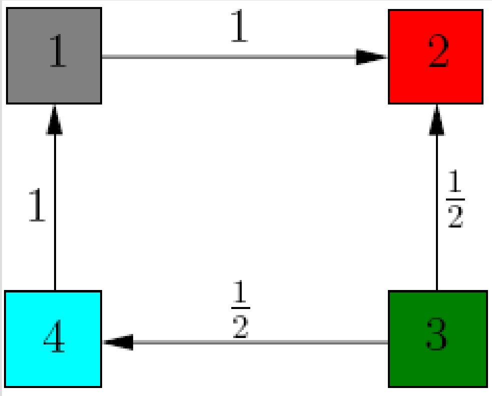
- Nodes can represent anything, such as people in a social network or cities in a transportation system.
  - Vertex is a labeled node in graph.

- Edges can represent relationships or connections between them, such as friendship or a road.
  - Edges are identified by the pair of vertices that they connect

Graph Preliminaries

– Undirected versus Directed Graphs (Modified from [1])

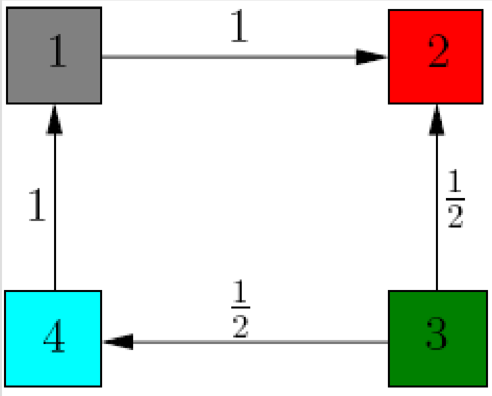




[1] http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture2/lecture2.html

Graph Preliminaries

– Transition Matrix $\boldsymbol{A}$ (Modified from [1])



From Node

$$
\begin{array}{cc}
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
A = & \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}
\end{array}
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}
$$

To Node

[1] http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture2/lecture2.html

Fairness on Graph Ranking: PageRank

- Motivation
  - Finding the most important or influential nodes in graph
  - Used in Google search engine
  - Example: Important webpage -> linked by may others

- PageRank algorithm:
  - Iteratively solve the following equation:

$$\boldsymbol{r} = c\boldsymbol{A}\boldsymbol{r} + (1-c)\boldsymbol{e}$$

PageRank solution:

$$\boldsymbol{r}^* = (1-c)(\boldsymbol{I} - c\boldsymbol{A})^{-1}\boldsymbol{e}$$

$\boldsymbol{r}$: PageRank vector

$\boldsymbol{A}$: transition matrix

c: damping factor

$\boldsymbol{e}$: teleportation vector

[1] Page, L., Brin, S., Motwani, R., & Winograd, T.. The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab 1999.
[2] Haveliwala, T. H.. Topic-sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. TKDE 2003.
[3] Tong, H., Faloutsos, C., & Pan, J. Y.. Fast Random Walk with Restart and Its Applications. ICDM 2006

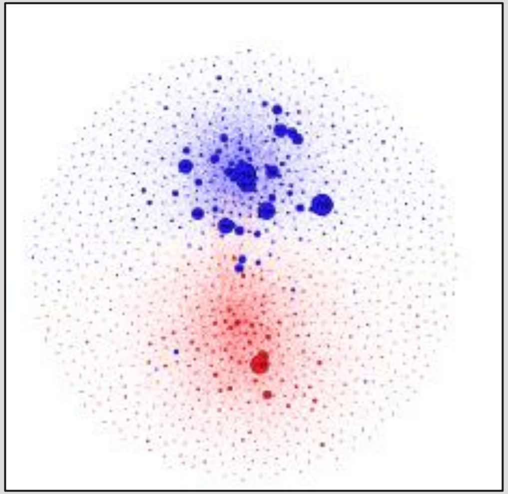
Fairness Measure for PageRank

- ϕ-fair PageRank:
  - Given: Graph G
  - Definition:
    - PageRank vector is ϕ-fair if ϕ fraction of total PageRank mass is allocated to the protected group
- Variants:
    - Statistical Parity: ϕ= fraction of protected group
    - Affirmative Action: ϕ= a desired ratio

[1] Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., & Mamoulis, N.. Fairness-Aware PageRank. WWW 2021.

Fairness Measure for PageRank

- Given:
  - A graph with transition matrix A
  - Partitions of nodes
    - Red nodes (R): protected group
    - Blue nodes (B): unprotected group



  - Find: A fair PageRank vector $\tilde{\boldsymbol{r}}$ that is
    - ϕ-fair
    - Close to the original PageRank vector $\boldsymbol{r}^*$

[1] Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., & Mamoulis, N.. Fairness-Aware PageRank. WWW 2021.

Fairness-aware PageRank: Mechanics

- From PageRank Solution:

$$\boldsymbol{r}^* = (1 - c)(\boldsymbol{I} - c\boldsymbol{A})^{-1}\boldsymbol{e}$$

- Teleportation vector $\boldsymbol{e}$: Control the starting node where a random walker restarts the work over the graph
  - Can we let the walker restart at a protected node or its neighborhood?
- Transition matrix $\boldsymbol{A}$: Control the next step where the walker goes to
  - Can we let the walker go to the protected nodes more frequently?
- Dumping factor c: Avoid sinks in the random walk (i.e., nodes without outgoing links)

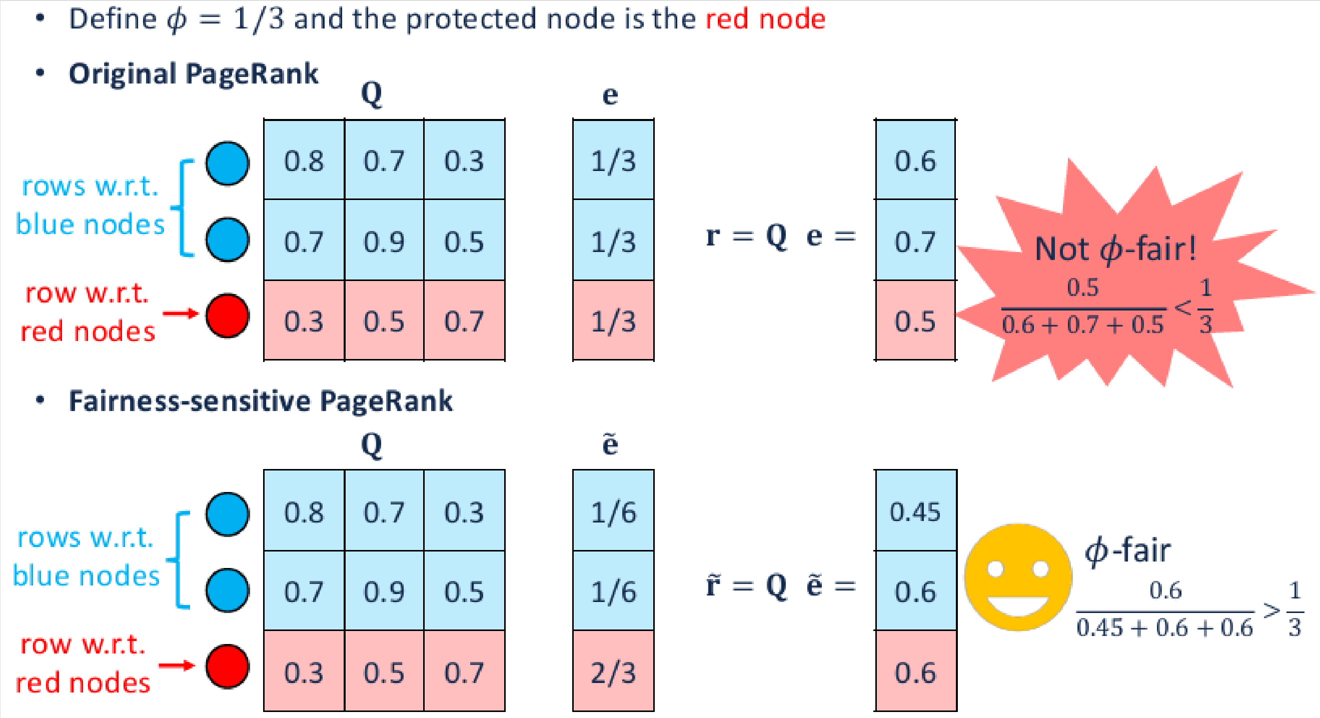
[1] Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., & Mamoulis, N.. Fairness-Aware PageRank. WWW 2021.

Fairness-sensitive PageRank: Teleportation Vector Based

- Intuition
  - Find a teleportation vector $\boldsymbol{e}$ to enable ϕ-fair PageRank vector
  - Keep transition matrix $\boldsymbol{A}$
  - Define: $\boldsymbol{Q} = (1-c)(\boldsymbol{I} - c\boldsymbol{A})^{-1}$
  - $\boldsymbol{e}$ is found such that the distance between $\boldsymbol{Qe}$ and $\boldsymbol{r}^*$ is minimized whilst meeting ϕ-fair requirement
  - This can be done through convex optimization solvers

[1] Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., & Mamoulis, N.. Fairness-Aware PageRank. WWW 2021.

Example of Teleportation Vector Based



Fairness-aware PageRank: Transition Matrix Based

- From PageRank Solution:

$$\boldsymbol{r}^* = (1 - c)(\boldsymbol{I} - c\boldsymbol{A})^{-1}\boldsymbol{e}$$

- Teleportation vector $\boldsymbol{e}$: Control the starting node where a random walker restarts the work over the graph

Can we let the walker restart at a protected node or its neighborhood?

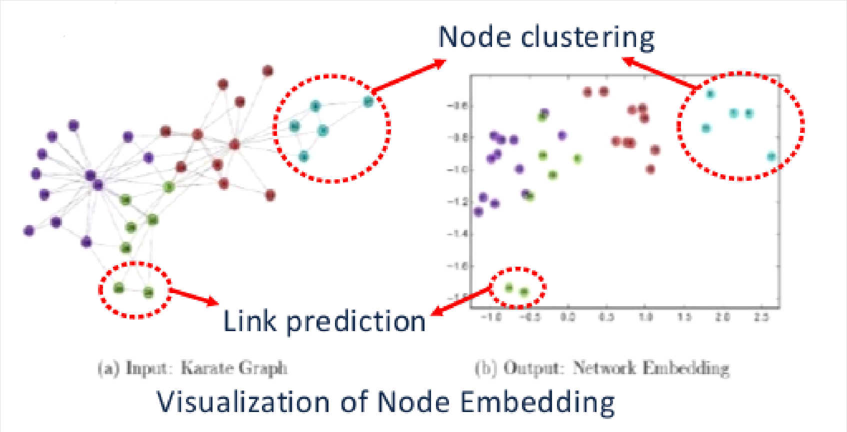
- Transition matrix $\boldsymbol{A}$: Control the next step where the walker goes to
  - Can we let the walker go to the protected nodes more frequently?
  - Locally fair PageRank
- Dumping factor c: Avoid sinks in the random walk (i.e., nodes without outgoing links)

[1] Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., & Mamoulis, N.. Fairness-Aware PageRank. WWW 2021.

Fairness on Graph Embedding

– Motivation: Learn low-dimensional node representations that preserve structural/attributive information

– Common tasks:
  – Node classification
  – Link prediction
  – Node visualization

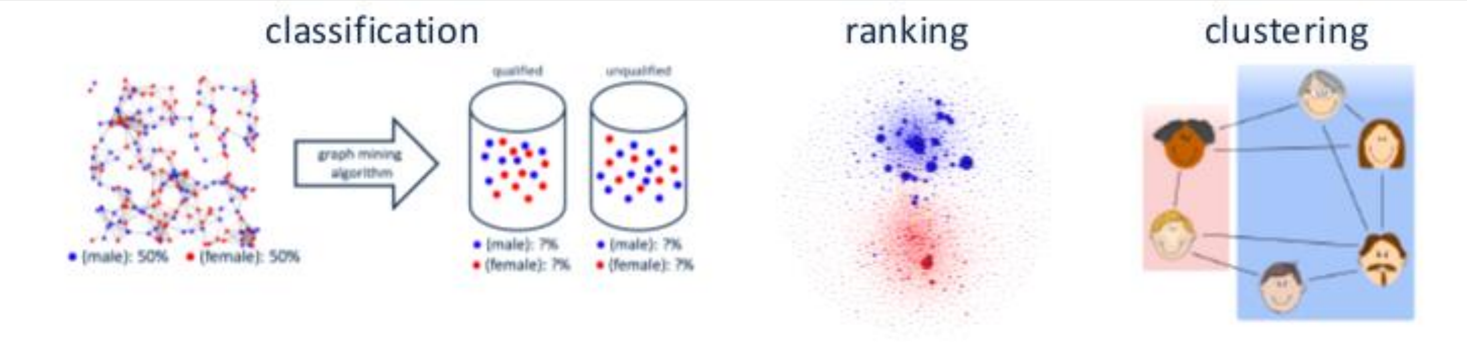


[1] Perozzi, B., Al-Rfou, R., & Skiena, S.. DeepWalk: Online Learning of Social Representations. KDD 2014.
[2] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q.. LINE: Large-scale Information Network Embedding. WWW 2015.
[3] Tang, J., Liu, J., Zhang, M., & Mei, Q.. Visualizing Large-scale and High-dimensional Data. WWW 2016.

Fairness on Graph Embedding

– Why fairness for embedding?

– Allow multiple tasks that consider fairness (e.g., classification, ranking, clustering)

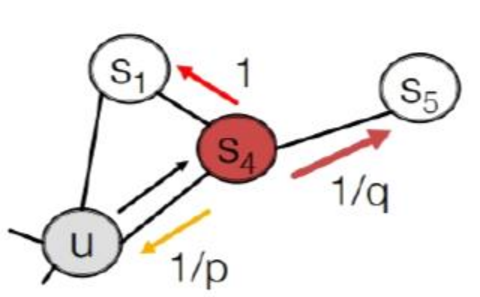


[1] Bose, A., & Hamilton, W.. Compositional Fairness Constraints for Graph Embeddings. ICML 2019.

Fairness on Graph Embedding: Random Walk-based method – node2vec

- Goal: Learn node embeddings that are predictive of nodes in its neighborhood

- Key idea: Skip-gram model with biased random walk



- Example:
  - Return parameter p: How fast the walk explores the neighborhood of the starting node
  - In-out parameter q: How fast the walk leaves the neighborhood of the starting node

Fairness Measures – node2vec

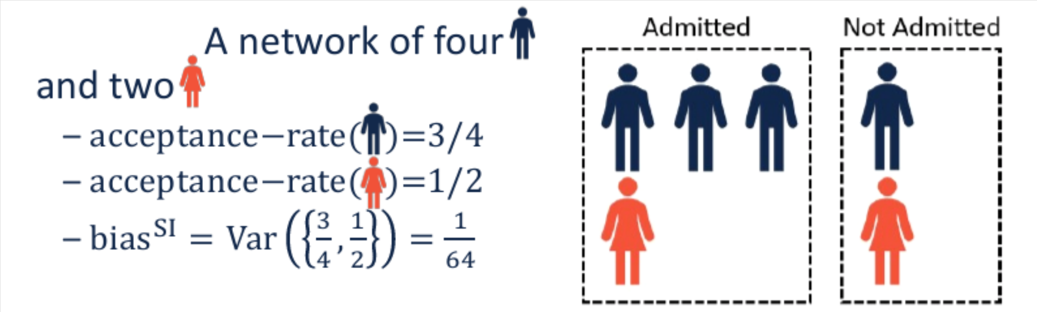
– Statistical parity

  – Given: (1) Sensitive attribute $\mathfrak{S}$; (2) multiple demographics group $\mathfrak{G}^{\mathfrak{S}}$ partitioned by $\mathfrak{S}$

    ❑ Case of college admission – Bias is defined by variance among the acceptance rates of each group in $\mathfrak{G}^{\mathfrak{S}}$

$$bias^{SI}\left(\mathfrak{G}^{\mathfrak{S}}\right) = Var(\{acceptance\ rate\ (G^S)|G^S \in \mathfrak{G}^{\mathfrak{S}})$$

– Example:



[1] Rahman, T., Surma, B., Backes, M., & Zhang, Y.. FairWalk: Towards Fair Graph Embedding. IJCAI 2019.

Fairness Measures – node2vec

- Equality of representation – User level
- Z-share: Among recommendations $\rho(u)$ given to a specific user u, measure the fraction of users having sensitive value z

$$z - share(u) = \frac{|\rho_Z(u)|}{|\rho(u)|}$$

- Intuition: Measure the bias as the difference between a fair fraction and the average z-share over all users U

$$bias^{ER_{user}}(z) = \frac{1}{|Z^S|} - \frac{\sum_{u \in U} z - share(u)}{|U|}$$

[1] Rahman, T., Surma, B., Backes, M., & Zhang, Y.. FairWalk: Towards Fair Graph Embedding. IJCAI 2019.

Fairness Measures – node2vec

– Equality of representation – User level

- **Example:** For *any* user u in the social network of ten and ten
  - $|Z^{\mathcal{S}}| = |\{ , \}| = 2$ and fair fraction $\frac{1}{|Z^{\mathcal{S}}|} = \frac{1}{2}$
  - The recommendations w.r.t. any user u are constant:
  - Let $z =$, we know $\rho_z(u) = 1$ and $\rho(u) = 3$
  - $\text{z–share}(u) = \frac{|\rho_z(u)|}{|\rho(u)|} = \frac{1}{3}$
  - $\text{bias}^{\text{ER}_{\text{user}}}(\ ) = \frac{1}{2} - \frac{\sum_u 1/3}{20} = \frac{1}{6}$

[1] Rahman, T., Surma, B., Backes, M., & Zhang, Y.. FairWalk: Towards Fair Graph Embedding. IJCAI 2019.

Fairwalk: Solution

- Key idea: Modify the random walk procedure in node2vec

- Steps of Fairwalk
  - Partition neighbors into demographic groups
  - Assign equal probability to each demographic group
  - Select a demographic group to walk to
  - Randomly select a node within the chosen demographic group

Fairwalk: Statistical Parity

– Fairwalk achieves a more balanced acceptance rates among groups

[1] Rahman, T., Surma, B., Backes, M., & Zhang, Y.. FairWalk: Towards Fair Graph Embedding. IJCAI 2019.

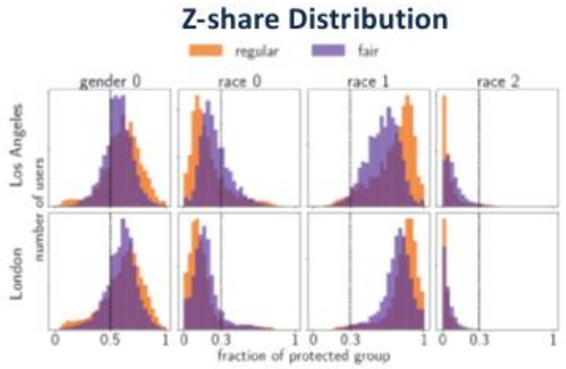
Fairwalk: User-level Equality of Representation

– Fairwalk decreases the user-level bias

– Z-share distribution of Fairwalk leans towards the fair fraction

Bias for User-level Equality of Representation

		gender		race		
		0	1	0	1	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



[1] Rahman, T., Surma, B., Backes, M., & Zhang, Y.. FairWalk: Towards Fair Graph Embedding. IJCAI 2019.

Remarks and Summary

- Fairness-Aware PageRank vs Fairwalk
  - Fairness-aware PageRank: The minority group should have a certain proportion of PageRank probability mass
  - Fairwalk: All demographic group have the same random walk transition probability mass

- Node2vec vs Fairwalk
  - Node2vec: skip-gram model + walk sequences by original random walk
  - Fairwalk: skip-gram model + walk sequences by fair random walk

Further Topics of Fairness on Graph

- Individual fairness [1]
- Counterfactual fairness [2]
- Degree-related fairness [3]
- Rawlsian fairness [4]

[1] Kang, J., He, J., Maciejewski, R., & Tong, H.. InFoRM: Individual Fairness on Graph Mining. KDD 2020.

[2] Agarwal, C., Lakkaraju, H., & Zitnik, M.. Towards a Unified Framework for Fair and Stable Graph Representation Learning. UAI 2021.

[3] Tang, X., Yao, H., Sun, Y., Wang, Y., Tang, J., Aggarwal, C., ... & Wang, S.. Investigating and Mitigating Degree-Related Biases in Graph Convolutional Networks. CIKM 2020.

[4] Rahmattalabi, A., Vayanos, P., Fulginiti, A., Rice, E., Wilder, B., Yadav, A., & Tambe, M.. Exploring Algorithmic Fairness in Robust Graph Covering Problems. NeurIPS 2019.