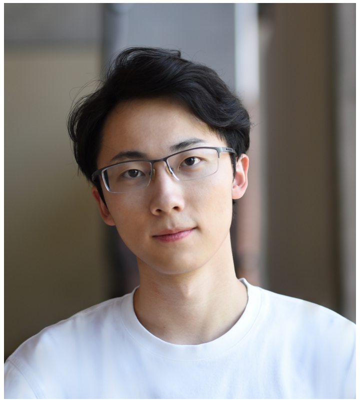
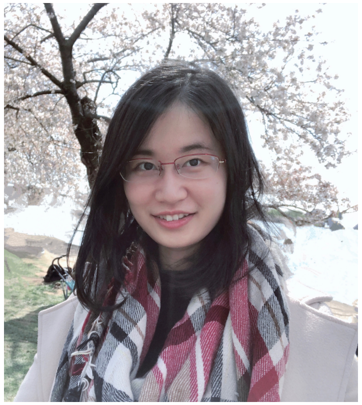
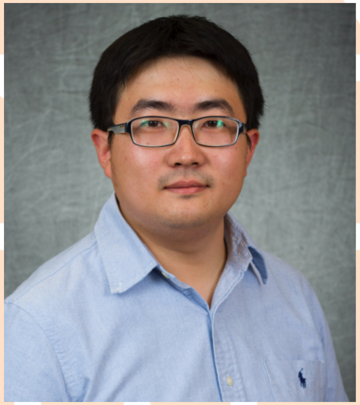
Fairness in Graph Mining: Metrics, Algorithms, and Applications



Yushun Dong **Jing Ma** **Chen Chen** **Jundong Li**

University of Virginia

Related Materials of this Tutorial

Fairness in Graph Mining: A Survey

Yushun Dong, Jing Ma, Chen Chen, and Jundong Li

Abstract—Graph mining algorithms have been playing a significant role in myriad fields over the years. However, despite their promising performance on various graph analytical tasks, most of these algorithms lack fairness considerations. As a consequence, they could lead to discrimination towards certain populations when exploited in human-centered applications. Recently, algorithmic fairness has been extensively studied in graph-based applications. In contrast to algorithmic fairness on independent and identically distributed (i.i.d.) data, fairness in graph mining has exclusive backgrounds, taxonomies, and fulfilling techniques. In this survey, we provide a comprehensive and up-to-date introduction of existing literature under the context of fair graph mining. Specifically, we propose a novel taxonomy of fairness notions on graphs, which sheds light on their connections and differences. We further present an organized summary of existing techniques that promote fairness in graph mining. Finally, we summarize the widely used datasets in this emerging research field and provide insights on current research challenges and open questions, aiming at encouraging cross-breeding ideas and further advances.

Index Terms—Algorithmic Fairness, Graph Mining, Debiasing

1 Introduction

Graph-structured data is pervasive in diverse real-world applications, e.g., E-commerce [94], [112], health care [36], [51], traffic forecasting [66], [92], and drug discovery [15], [162]. In recent years, a number of graph mining algorithms have been proposed to gain a deeper understanding of such data. These algorithms have shown promising performance on various graph analytical tasks such as node classification [56], [79], [119], [152] and link prediction [4], [95], [96], [100], which contribute to great advances in many graph-based applications.

Despite the success of these graph mining algorithms, most of them lack fairness considerations. Consequently, they could yield discriminatory results towards certain populations when such algorithms are exploited in human-centered applications [74]. For example, a social network-based job recommender system may unfavorably recommend fewer job opportunities to individuals of a certain gender [89] or individuals in an underrepresented ethnic group [141]. With the widespread usage of graph mining algorithms, such potential discrimination could also exist in other high-stake applications such as disaster response [150], criminal justice [3], and loan approval [127]. In these applications, critical and life-changing decisions are often made for the individuals involved. Therefore, how to tackle unfairness issues in graph mining algorithms naturally becomes a crucial problem.

Fulfilling fairness in graph mining can be non-trivial due to two main challenges. The first challenge is to formulate proper fairness notions as the criteria to determine the existence of unfairness (i.e., bias). Although a vast amount of traditional algorithmic fairness notions have been proposed in the context of independent and identically distributed (i.i.d.) data [41], [102], they are unable to reflect the relational information (i.e., the topology) in graph data. For example, the same population can be connected with different topologies as in Fig. 1a and 1b, where each node represents an individual, and the color of nodes denotes their demographic subgroup membership, such as different genders. Compared with the graph topology in Fig. 1a, the topology in Fig. 1b has more intra-group edges than inter-group edges. The dominance of intra-group edges in the graph topology is a common type of bias existing in real-world graphs [38], [40], [65], which cannot be captured by traditional algorithmic fairness notions. The second challenge is to prevent the graph mining algorithms from inheriting the bias exhibited in the input graphs [40], [103], [139], [151]. We present a toy example to demonstrate how the information propagation mechanism in Graph Neural Networks (GNNs) [60], [79], [152] induces bias to the output node embeddings from a biased graph topology in Fig. 1c. In the input space, the node features are uniformly distributed. However, when the information propagation is performed on a biased topology as in Fig. 1b, the information received by nodes in different subgroups could be biased [40], leading to a biased embedding distribution in the output space.

There has been emerging research interest in fulfilling algorithmic fairness in graph mining. Nevertheless, the studied fairness notions vary across different works, which can be confusing and impede further progress. Meanwhile, different techniques are developed in achieving various fairness notions. Without a clear understanding of the corresponding mappings, future fair graph mining algorithm design can be difficult. Therefore, a systematic survey of recent advances is needed to shed light on future research. In this survey, we present a comprehensive and up-to-date review of existing works in fair graph mining. The main

- *Y. Dong is with Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, Virginia, US. E-mail: yd6eb@virginia.edu*
- *J. Ma is with Department of Computer Science, University of Virginia, Charlottesville, Virginia, US. E-mail: jm3mr@virginia.edu*
- *C. Chen is with Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, US. E-mail: zrh6du@virginia.edu*
- *J. Li is with Department of Electrical and Computer Engineering, Department of Computer Science, and School of Data Science, University of Virginia, Charlottesville, Virginia, US. E-mail: jundong@virginia.edu*

Our survey paper has been released on arxiv.

Related Materials of this Tutorial

Fairness in Graph Mining: A Survey

Yushun Dong, Jing Ma, Chen Chen, and Jundong Li

Abstract—Graph mining algorithms have been playing a significant role in myriad fields over the years. However, despite their promising performance on various graph analytical tasks, most of these algorithms lack fairness considerations. As a consequence, they could lead to discrimination towards certain populations when exploited in human-centered applications. Recently, algorithmic fairness has been extensively studied in graph-based applications. In contrast to algorithmic fairness on independent and identically distributed (i.i.d.) data, fairness in graph mining has exclusive backgrounds, taxonomies, and fulfilling techniques. In this survey, we provide a comprehensive and up-to-date introduction of existing literature under the context of fair graph mining. Specifically, we propose a novel taxonomy of fairness notions on graphs, which sheds light on their connections and differences. We further present an organized summary of existing techniques that promote fairness in graph mining. Finally, we summarize the widely used datasets in this emerging research field and provide insights on current research challenges and open questions, aiming at encouraging cross-breeding ideas and further advances.

Index Terms—Algorithmic Fairness, Graph Mining, Debiasing

1 Introduction

Graph-structured data is pervasive in diverse real-world applications, e.g., E-commerce [94], [112], health care [36], [51], traffic forecasting [66], [92], and drug discovery [15], [162]. In recent years, a number of graph mining algorithms have been proposed to gain a deeper understanding of such data. These algorithms have shown promising performance on various graph analytical tasks such as node classification [56], [79], [119], [152] and link prediction [4], [95], [96], [100], which contribute to great advances in many graph-based applications.

Despite the success of these graph mining algorithms, most of them lack fairness considerations. Consequently, they could yield discriminatory results towards certain populations when such algorithms are exploited in human-centered applications [74]. For example, a social network-based job recommender system may unfavorably recommend fewer job opportunities to individuals of a certain gender [89] or individuals in an underrepresented ethnic group [141]. With the widespread usage of graph mining algorithms, such potential discrimination could also exist in other high-stake applications such as disaster response [150], criminal justice [3], and loan approval [127]. In these applications, critical and life-changing decisions are often made for the individuals involved. Therefore, how to tackle unfairness issues in graph mining algorithms naturally becomes a crucial problem.

Fulfilling fairness in graph mining can be non-trivial due to two main challenges. The first challenge is to formulate proper fairness notions as the criteria to determine the existence of unfairness (i.e., bias). Although a vast amount of traditional algorithmic fairness notions have been proposed in the context of independent and identically distributed (i.i.d.) data [41], [102], they are unable to reflect the relational information (i.e., the topology) in graph data. For example, the same population can be connected with different topologies as in Fig. 1a and 1b, where each node represents an individual, and the color of nodes denotes their demographic subgroup membership, such as different genders. Compared with the graph topology in Fig. 1a, the topology in Fig. 1b has more intra-group edges than inter-group edges. The dominance of intra-group edges in the graph topology is a common type of bias existing in real-world graphs [38], [40], [65], which cannot be captured by traditional algorithmic fairness notions. The second challenge is to prevent the graph mining algorithms from inheriting the bias exhibited in the input graphs [40], [103], [139], [151]. We present a toy example to demonstrate how the information propagation mechanism in Graph Neural Networks (GNNs) [60], [79], [152] induces bias to the output node embeddings from a biased graph topology in Fig. 1c. In the input space, the node features are uniformly distributed. However, when the information propagation is performed on a biased topology as in Fig. 1b, the information received by nodes in different subgroups could be biased [40], leading to a biased embedding distribution in the output space.

There has been emerging research interest in fulfilling algorithmic fairness in graph mining. Nevertheless, the studied fairness notions vary across different works, which can be confusing and impede further progress. Meanwhile, different techniques are developed in achieving various fairness notions. Without a clear understanding of the corresponding mappings, future fair graph mining algorithm design can be difficult. Therefore, a systematic survey of recent advances is needed to shed light on future research. In this survey, we present a comprehensive and up-to-date review of existing works in fair graph mining. The main

- *Y. Dong is with Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, Virginia, US. E-mail: yd6eb@virginia.edu*
- *J. Ma is with Department of Computer Science, University of Virginia, Charlottesville, Virginia, US. E-mail: jm3mr@virginia.edu*
- *C. Chen is with Biocomplexity Institute, University of Virginia, Charlottesville, Virginia, US. E-mail: zrh6du@virginia.edu*
- *J. Li is with Department of Electrical and Computer Engineering, Department of Computer Science, and School of Data Science, University of Virginia, Charlottesville, Virginia, US. E-mail: jundong@virginia.edu*

Our survey paper has been released on arxiv.

We also released an open-source library PyGDebias, including 13 build-in methods and 26 build-in popular graph datasets.

Collected Algorithms

13 different methods in total are implemented in this library. We provide an overview of their characteristics as follows.

Methods	Debiasing Technique	Fairness Notions	Paper & Code
FairGNN [2]	Adversarial Learning	Group Fairness	[Paper] [Code]
EDITS [3]	Edge Rewiring	Group Fairness	[Paper] [Code]
FairWalk [4]	Rebalancing	Group Fairness	[Paper] [Code]
CrossWalk [5]	Rebalancing	Group Fairness	[Paper] [Code]
UGE [6]	Edge Rewiring	Group Fairness	[Paper] [Code]
FairVGNN [7]	Adversarial Learning	Group Fairness	[Paper] [Code]
FairEdit [8]	Edge Rewiring	Group Fairness	[Paper] [Code]
NIFTY [9]	Optimization with Regularization	Group/Counterfactual Fairness	[Paper] [Code]
GEAR [10]	Edge Rewiring	Group/Counterfactual Fairness	[Paper] [Code]
InFoRM [11]	Optimization with Regularization	Individual Fairness	[Paper] [Code]
REDRESS [12]	Optimization with Regularization	Individual Fairness	[Paper] [Code]
GUIDE [13]	Optimization with Regularization	Individual Fairness	[Paper] [Code]
RawlsGCN [14]	Rebalancing	Degree-Related Fairness	[Paper] [Code]

Outline

- Background Information
- Fairness Notions and Metrics
- Methodologies to Mitigate Bias
- Real-World Applications
- Summary & Existing Challenges

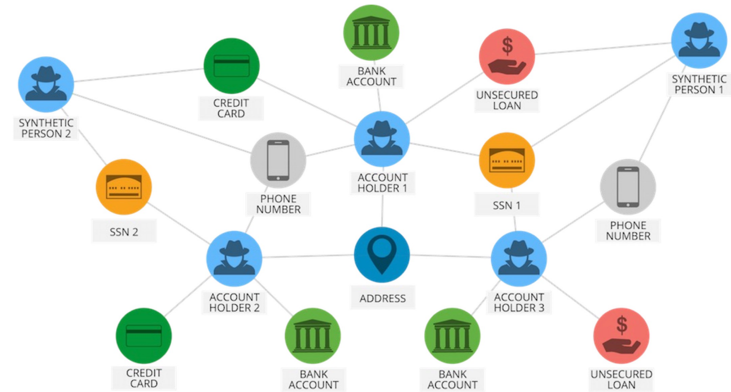
Outline

- Background Information
- Fairness Notions and Metrics
- Methodologies to Mitigate Bias
- Real-World Applications
- Summary & Existing Challenges

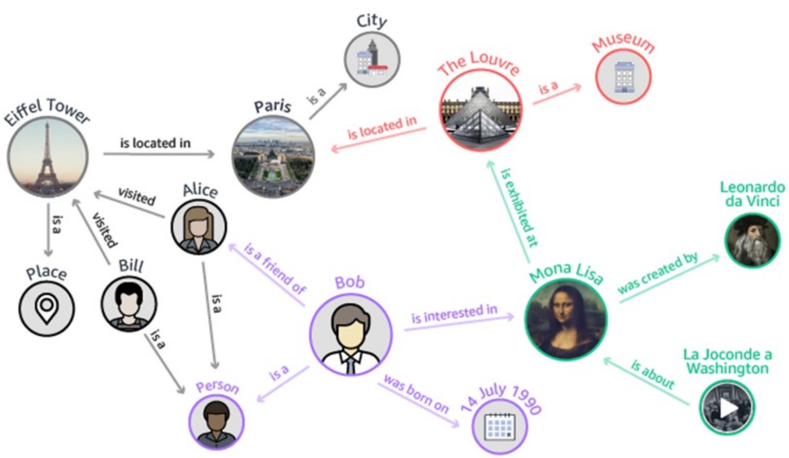
Graph Mining Algorithms

What are graph mining algorithms?

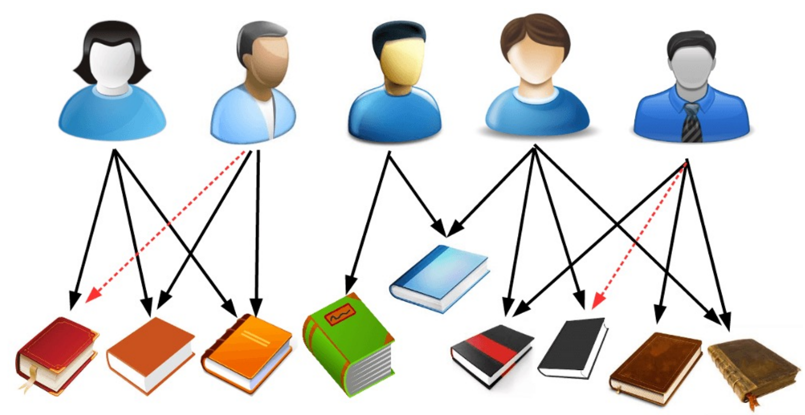
Social networks



Financial Networks



Knowledge Graphs

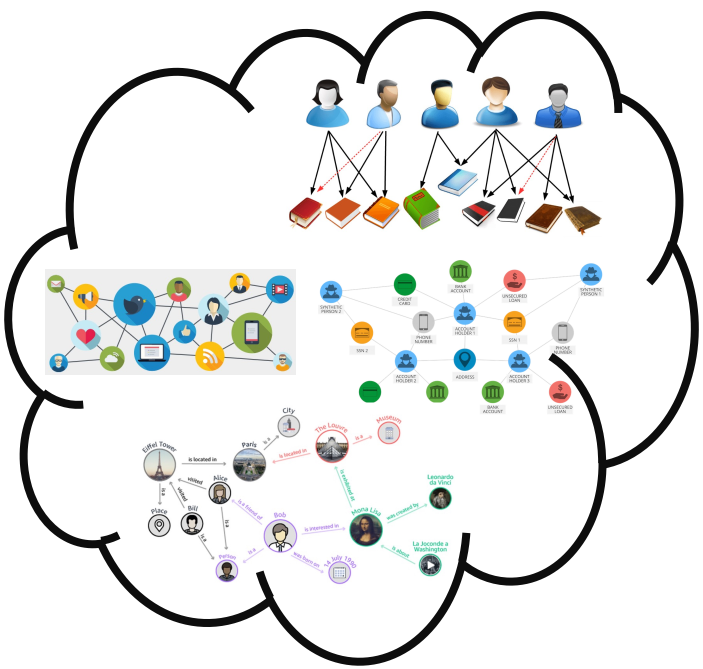


E-Commerce Networks

Graph Mining Algorithms (Cont.)

What are graph mining algorithms?

In general, graph mining algorithms **extracts information encoded in the graph data** to facilitate our understanding (on these graphs) and gain benefit on various predictive tasks.

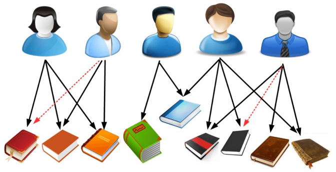


Graph Mining Algorithms

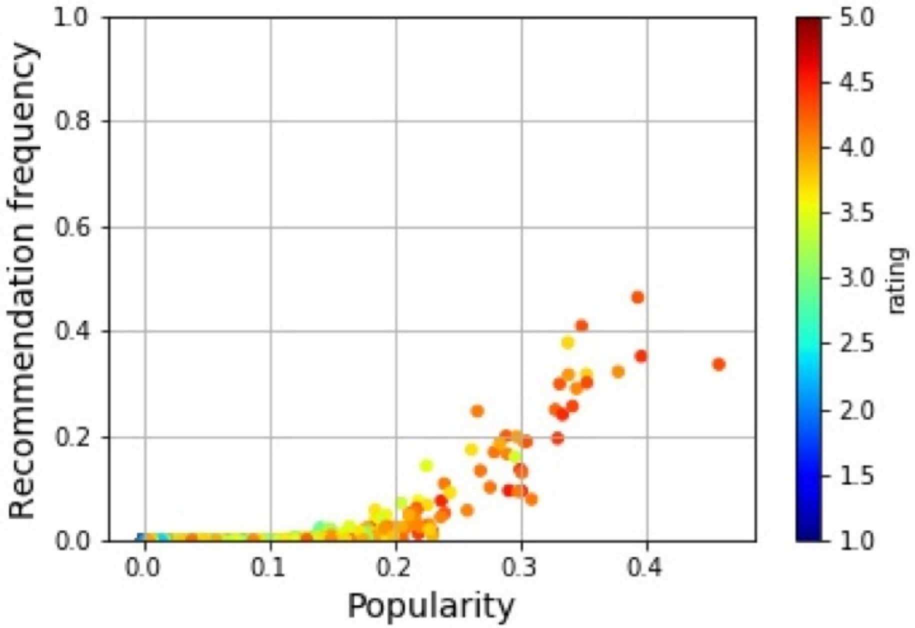
- Who are potential friends?
- Which item will this customer buy?
- Which loan applicant is with the lowest risk of debt default?

...

The Risk of Bias in Graph Mining



Potential discrimination in **recommender systems**.

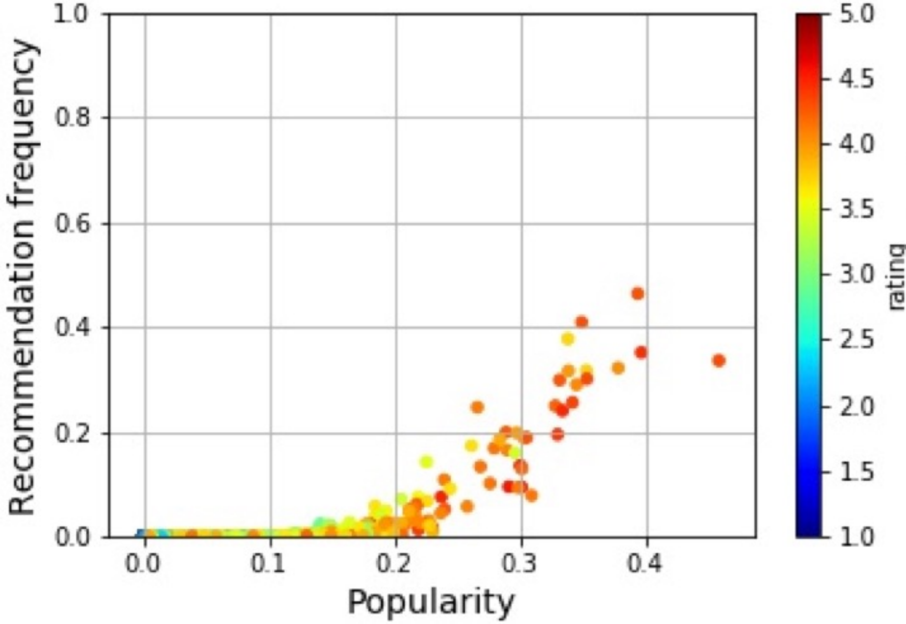


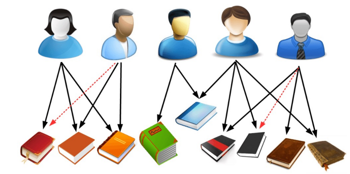
Popular items are often over-emphasized in recommendations, while less popular ones get less exposure [1].

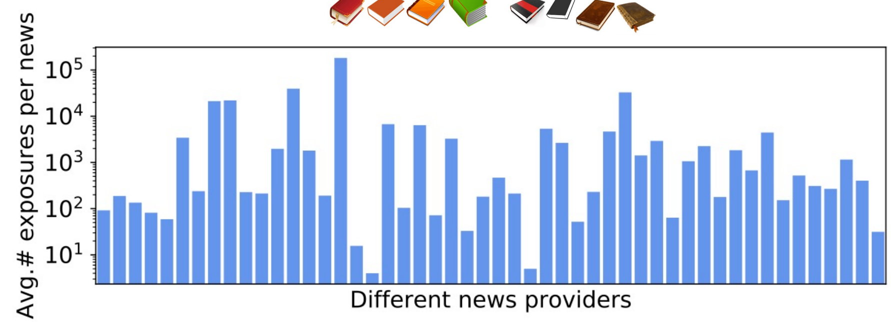
[1] Abdollahpouri H, et al. The impact of popularity bias on fairness and calibration in recommendation[J]. arXiv preprint arXiv:1910.05755, 2019.

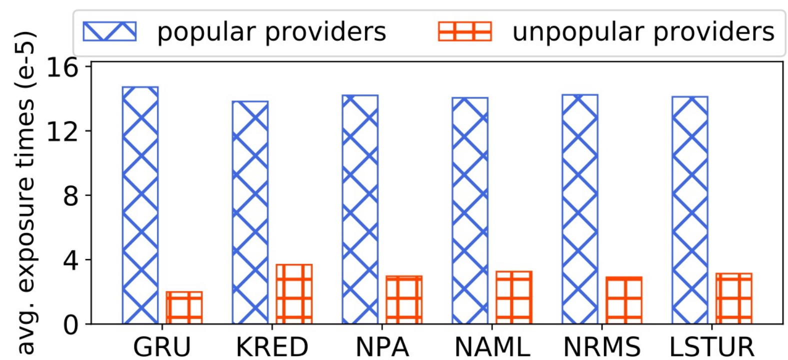
The Risk of Bias in Graph Mining

Potential discrimination in **recommender systems**.









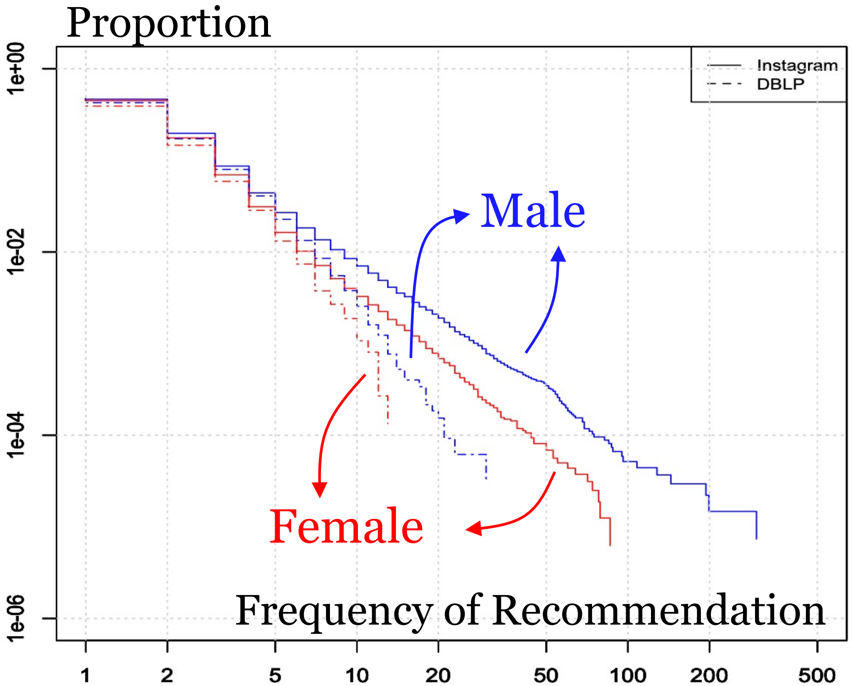
Popular items are often over-emphasized in recommendations, while less popular ones get less exposure [1].

Unpopular providers always bear much less exposure rate across different recommendation models [2].

[1] Abdollahpouri H, et al. The impact of popularity bias on fairness and calibration in recommendation[J]. arXiv preprint arXiv:1910.05755, 2019.
[2] Tao Qi, et al. ProFairRec: Provider Fairness-aware News Recommendation[J]. arXiv preprint arXiv:2204.04724, 2022.

The Risk of Bias in Graph Mining (Cont.)

Potential discrimination in **social networks**.

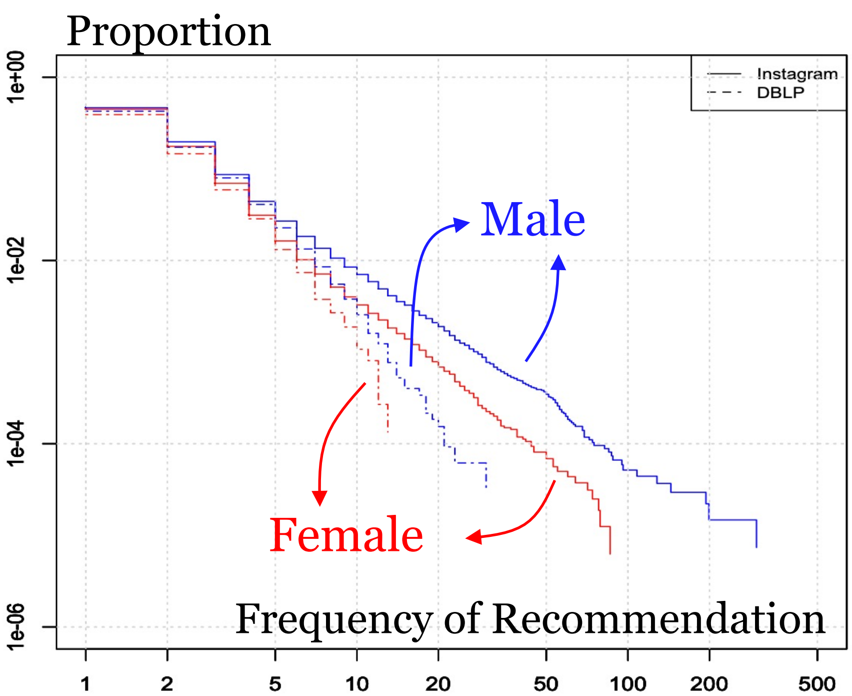


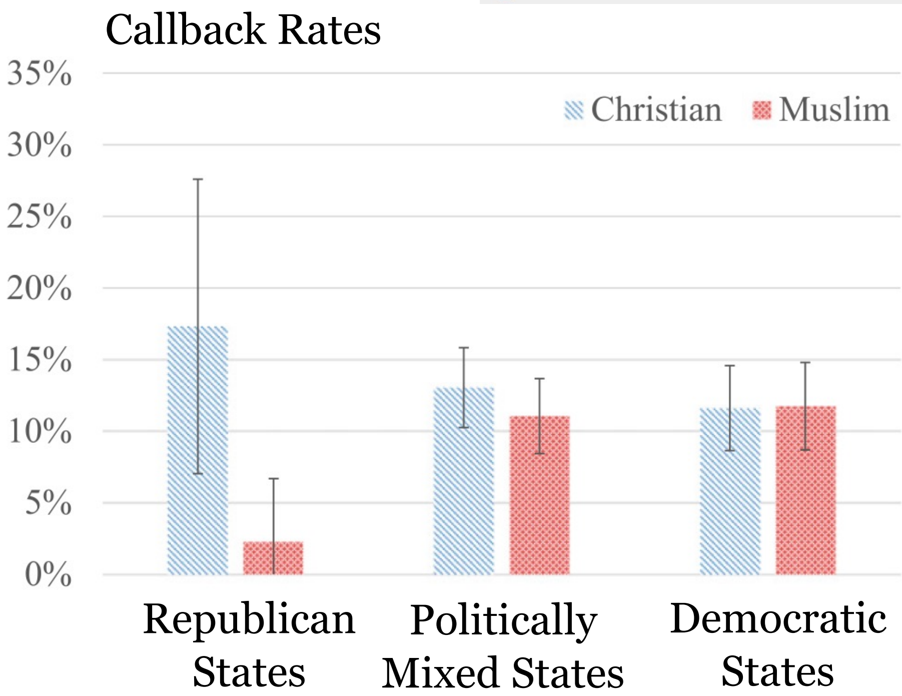
Users who get recommended to connect exhibit divergence between males and females [1].

[1] Stoica A A, et al. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In WWW 2018.

The Risk of Bias in Graph Mining (Cont.)

Potential discrimination in **social networks**.





Users who get recommended to connect exhibit divergence between males and females [1].

Users' religion could also be a source of hiring discrimination in social networks [2].

[1] Stoica A A, et al. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In WWW 2018.
[2] Acquisti, Alessandro, et al. "An experiment in hiring discrimination via online social networks." Management Science 66.3 (2020): 1005-1024.

Algorithmic Fairness

Then how to define fairness?

Algorithmic Fairness

Then how to define fairness?

Fairness can be defined in different ways [1]: different real-world applications show biases from various perspectives [2].

[1] Mengnan Du, Fan Yang, Na Zou, and Xia Hu. Fairness in deep learning: A computational perspective. IEEE Intelligent Systems, 2020.
[2] Yushun Dong, Jing Ma, Chen Chen, and Jundong Li. Fairness in Graph Mining: A Survey. arXiv preprint arXiv:2204.09888, 2022.

Algorithmic Fairness

Then how to define fairness?

Fairness can be defined in different ways [1]: different real-world applications show biases from various perspectives [2].

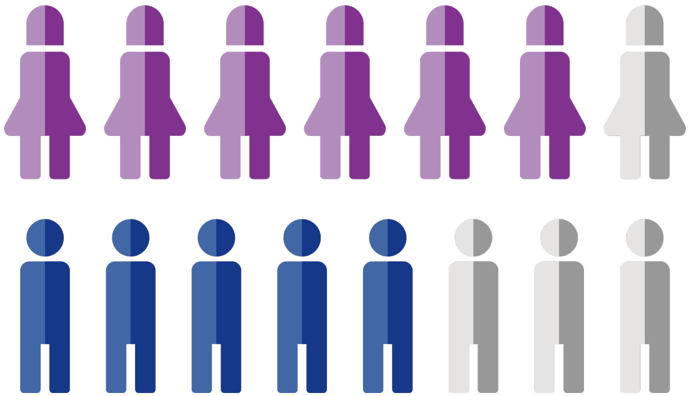
For example, it **depends on the specific studied problem** to determine which case should be considered as fair.

[1] Mengnan Du, Fan Yang, Na Zou, and Xia Hu. Fairness in deep learning: A computational perspective. IEEE Intelligent Systems, 2020.
[2] Yushun Dong, Jing Ma, Chen Chen, and Jundong Li. Fairness in Graph Mining: A Survey. arXiv preprint arXiv:2204.09888, 2022.

Algorithmic Fairness (Cont.)

Then how to define fairness?

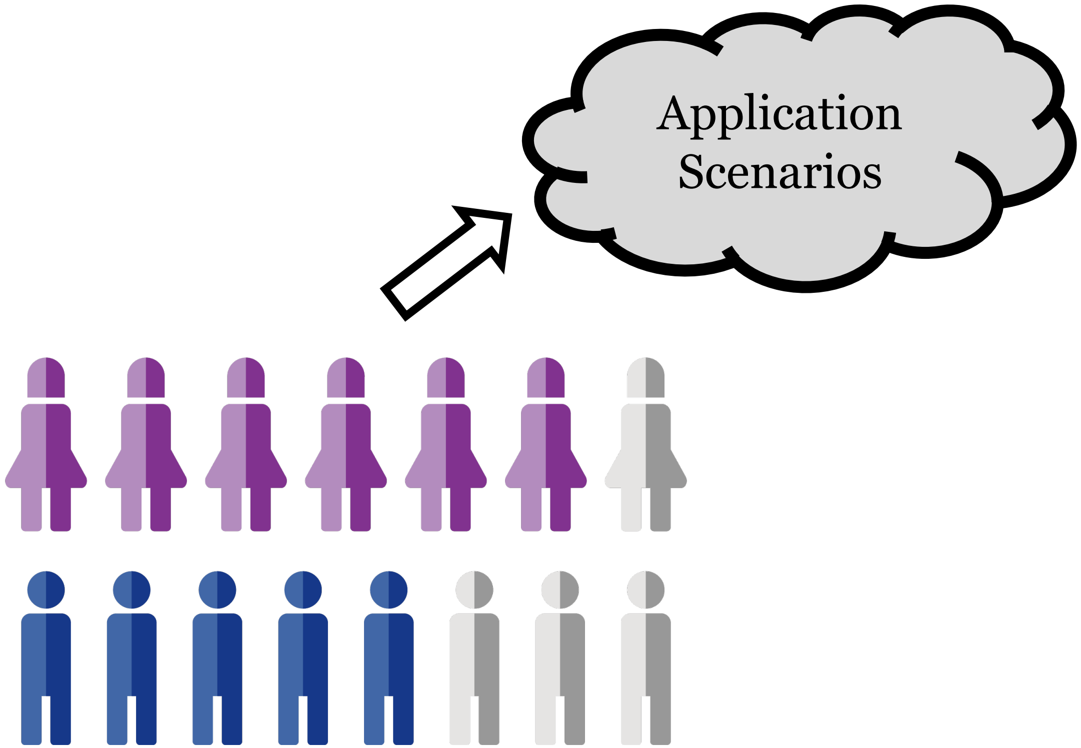
Despite the lack of a **universal criterion** for fairness, we could still study fairness in algorithms: there are **various existing fairness notions** based on people's awareness.



Algorithmic Fairness (Cont.)

Then how to define fairness?

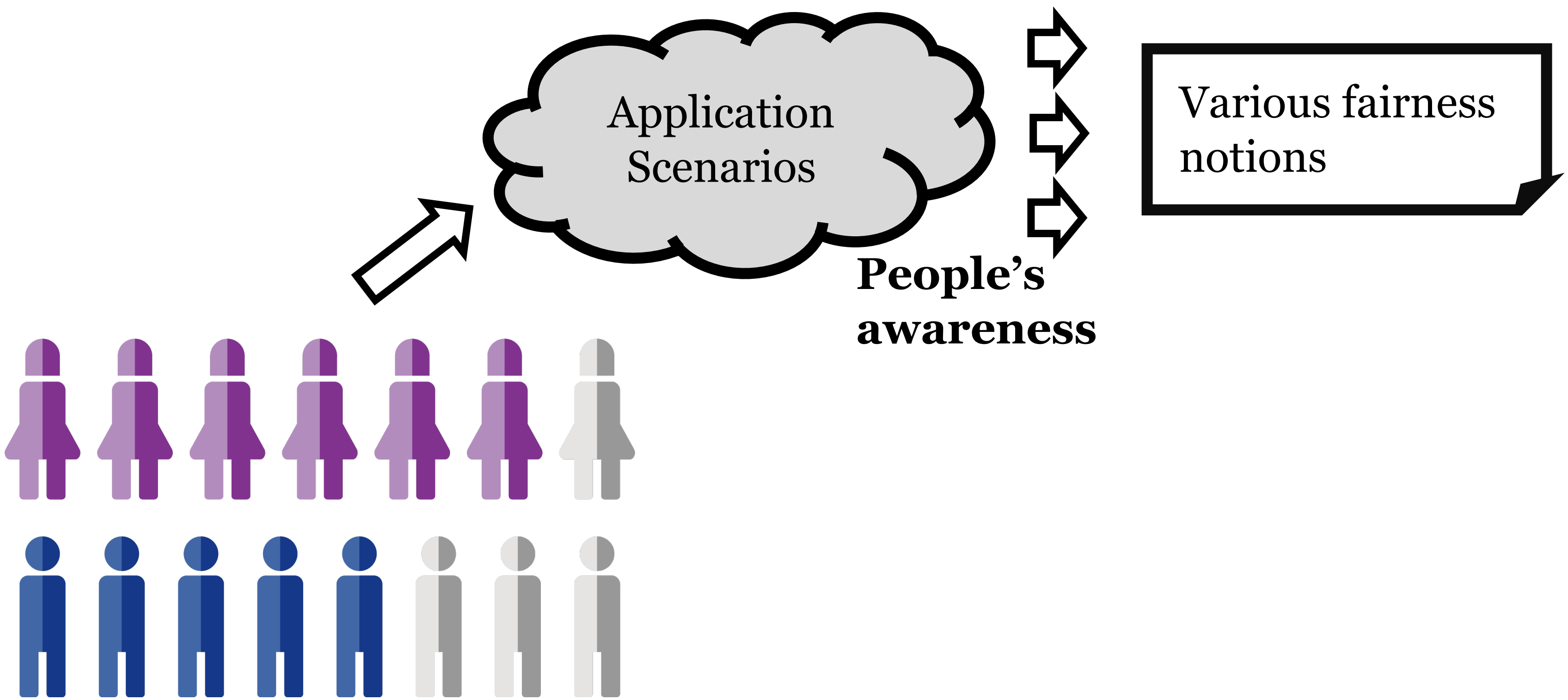
Despite the lack of a **universal criterion** for fairness, we could still study fairness in algorithms: there are **various existing fairness notions** based on people's awareness.



Algorithmic Fairness (Cont.)

Then how to define fairness?

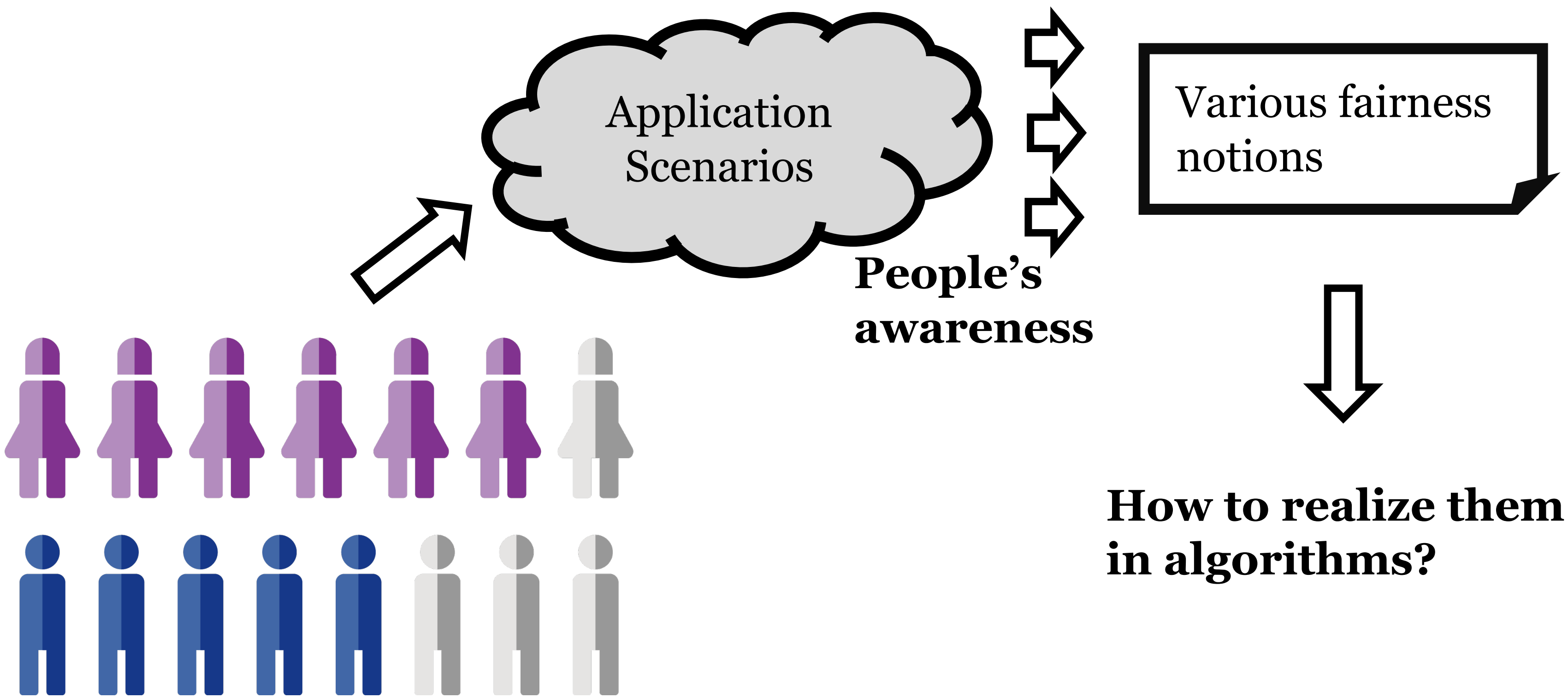
Despite the lack of a **universal criterion** for fairness, we could still study fairness in algorithms: there are **various existing fairness notions** based on people's awareness.



Algorithmic Fairness (Cont.)

Then how to define fairness?

Despite the lack of a **universal criterion** for fairness, we could still study fairness in algorithms: there are **various existing fairness notions** based on people's awareness.



Algorithmic Fairness in Graph Mining

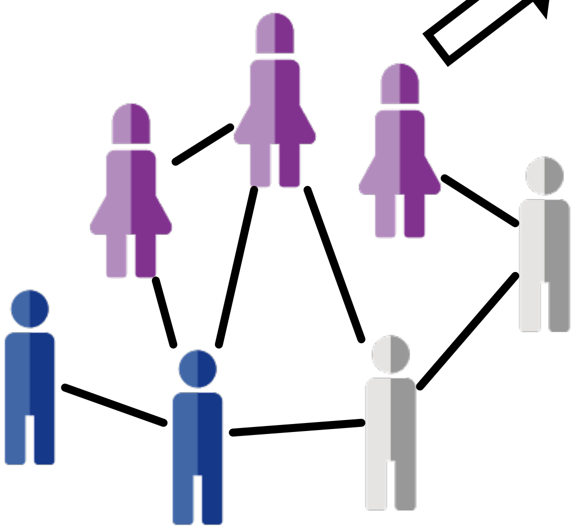
Then how to define fairness?

In the realm of **graph mining...**

Graph-based application scenarios

People's awareness

Traditional & **Novel** fairness notions



How to realize them in graph mining algorithms?

Fulfilling Fairness in Graph Mining

Unique Challenges of fulfilling fairness in graph mining.

Fulfilling Fairness in Graph Mining

Unique Challenges of fulfilling fairness in graph mining.

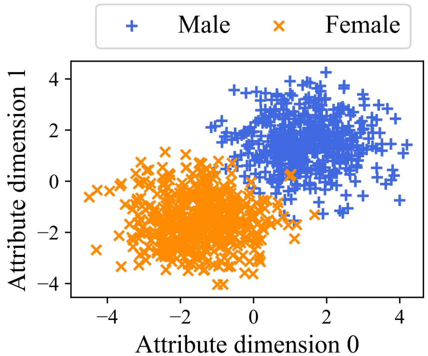
(1) Formulating **proper fairness notions** as the criteria to determine the existence of unfairness (i.e., bias).

Fulfilling Fairness in Graph Mining

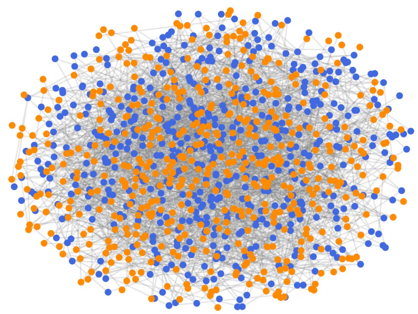
Unique Challenges of fulfilling fairness in graph mining.

(1) Formulating **proper fairness notions** as the criteria to determine the existence of unfairness (i.e., bias).

Attributed Graph 1

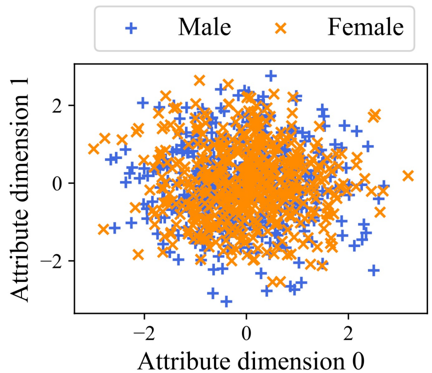


+

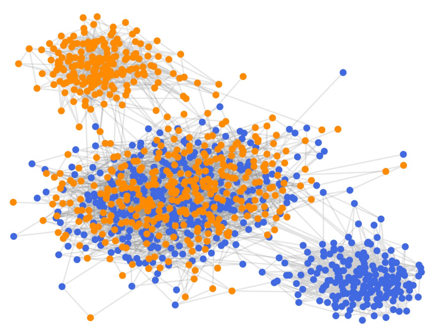


Fair or biased?

Attributed Graph 2



+



Fair or biased?

Fulfilling Fairness in Graph Mining

Unique Challenges of fulfilling fairness in graph mining.

(1) Formulating **proper fairness notions** as the criteria to determine the existence of unfairness (i.e., bias).

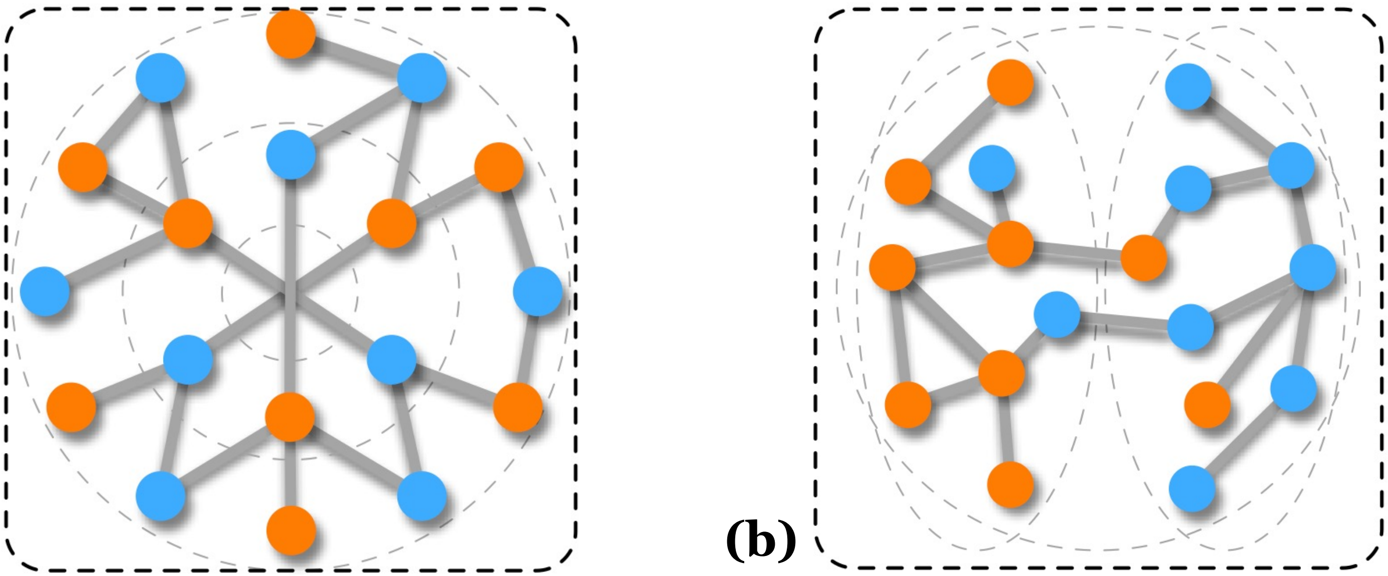
(2) Preventing the graph mining algorithms from **inheriting the bias** exhibited in the input graphs.

Fulfilling Fairness in Graph Mining

Unique Challenges of fulfilling fairness in graph mining.

(1) Formulating **proper fairness notions** as the criteria to determine the existence of unfairness (i.e., bias).

(2) Preventing the graph mining algorithms from **inheriting the bias** exhibited in the input graphs.



(a) (b)

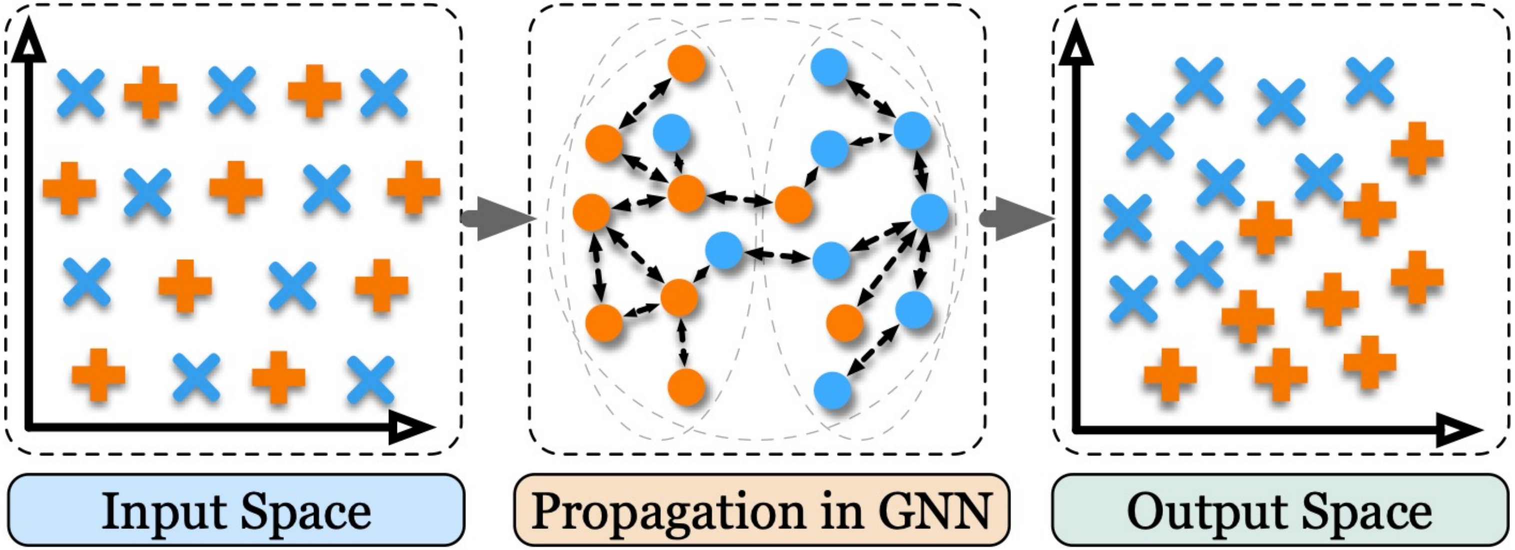
Compared with the structure in (a), the bias in the graph structure of (b) could lead to biased embedding in Graph Neural Networks (GNNs).

Fulfilling Fairness in Graph Mining

Unique Challenges of fulfilling fairness in graph mining.

(1) Formulating **proper fairness notions** as the criteria to determine the existence of unfairness (i.e., bias).

(2) Preventing the graph mining algorithms from **inheriting the bias** exhibited in the input graphs.



An example in Graph Neural Networks (GNN): the unbalance between intra-group and inter-group edges could easily induce bias in the outcome space [1].

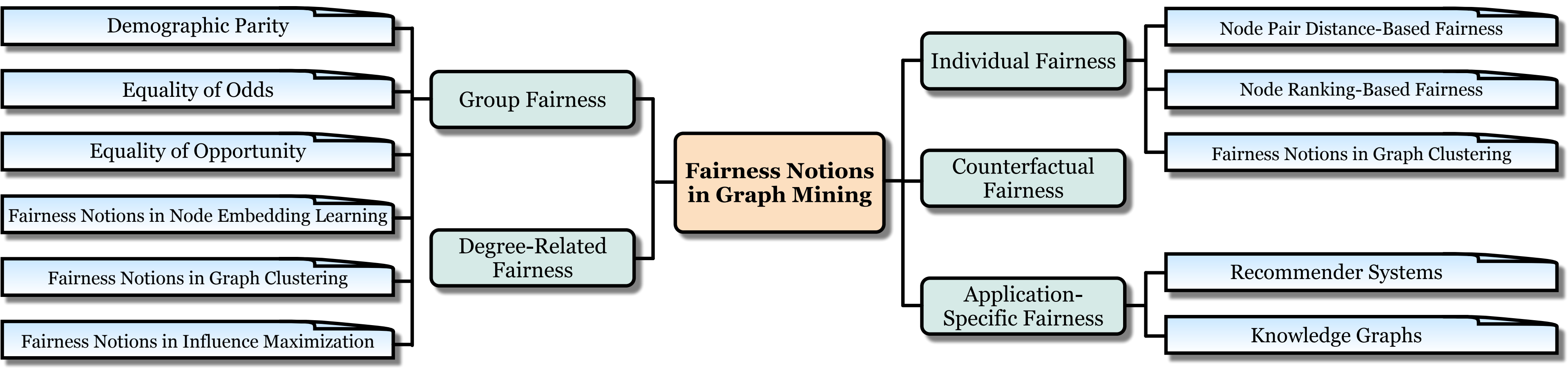
[1] Yushun Dong, Jing Ma, Chen Chen, and Jundong Li. Fairness in Graph Mining: A Survey. arXiv preprint arXiv:2204.09888, 2022.

Outline

- Background Information
- Fairness Notions and Metrics
- Methodologies to Mitigate Bias
- Real-World Applications
- Summary & Existing Challenges

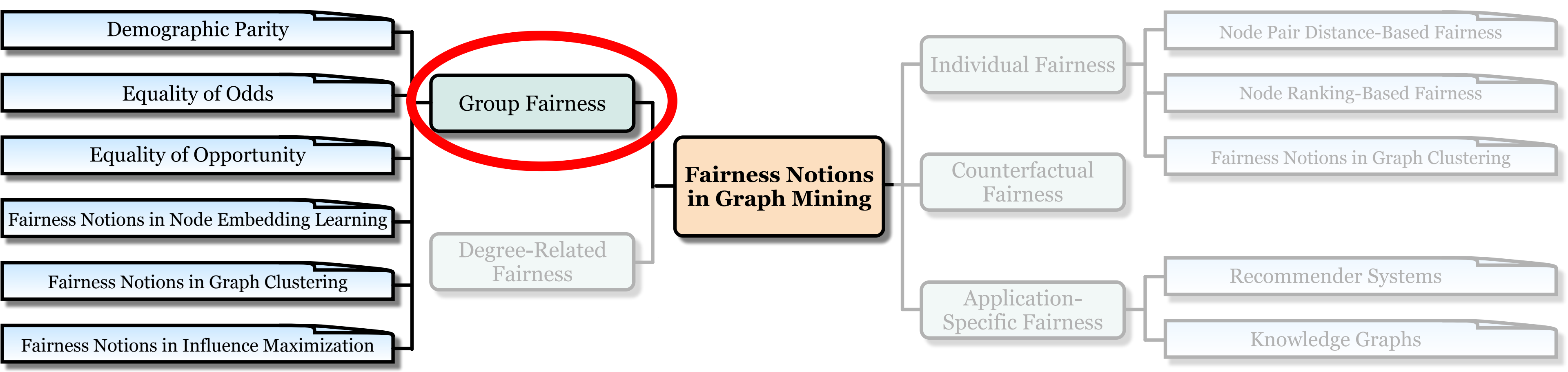
Taxonomy of Fairness Notions

A taxonomy of commonly used algorithmic fairness notions in graph mining algorithms.



Taxonomy of Fairness Notions (Cont.)

A taxonomy of commonly used algorithmic fairness notions in graph mining algorithms.



A general idea of group fairness: categorical **sensitive attributes** (e.g., gender, race) divide the whole population into different sensitive subgroups, and each group should gain **their fair share of interest** [1].

[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

Demographic Parity

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

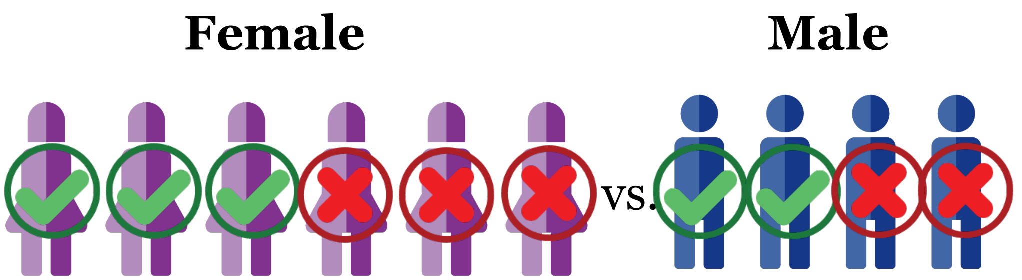
[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

Demographic Parity

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.



[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

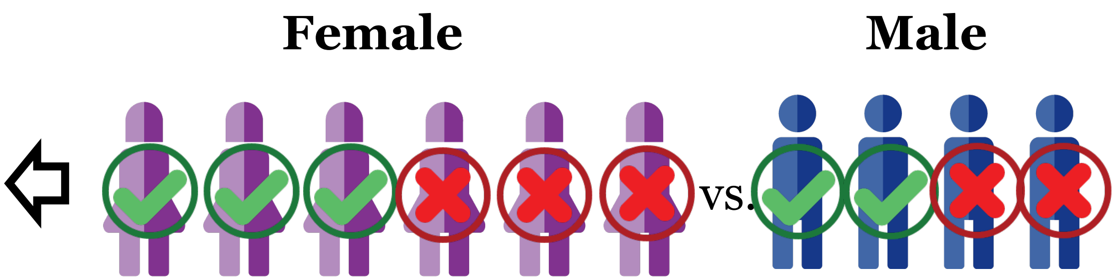
Demographic Parity

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

Fair in perspective of Demographic Parity.



[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

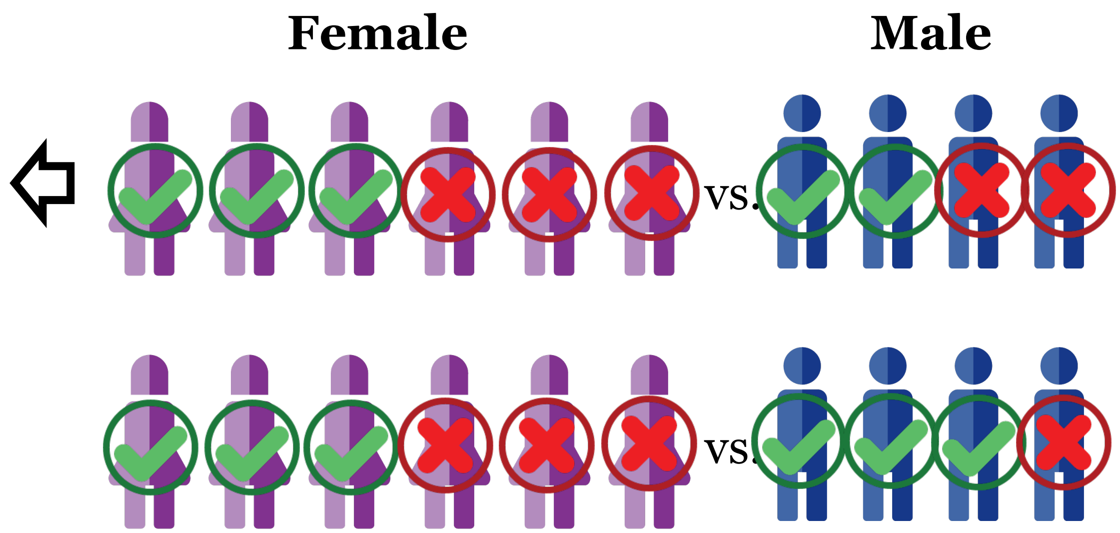
Demographic Parity

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

Fair in perspective of Demographic Parity.



[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

Demographic Parity

Group Fairness: Demographic Parity

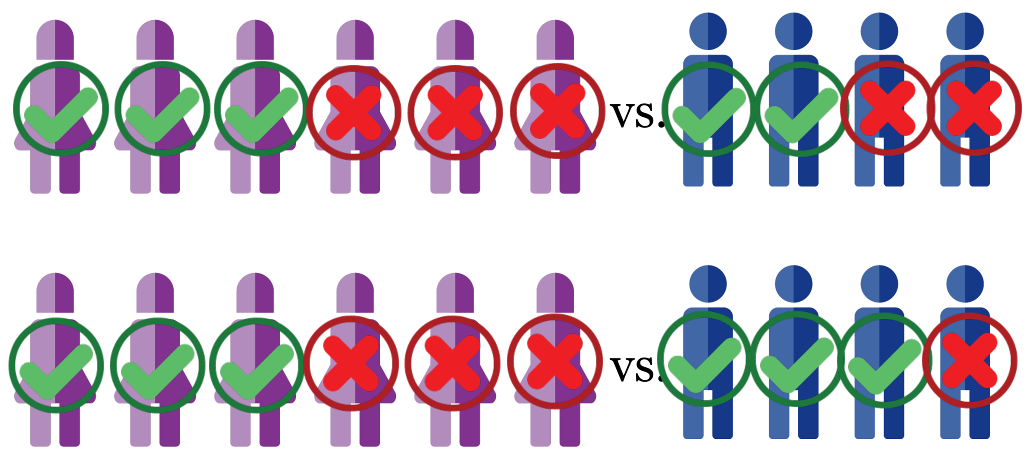
Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

Female **Male**

Fair in perspective of Demographic Parity.

Unfair in perspective of Demographic Parity.



[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

Demographic Parity (Cont.)

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

Criterion: $\mathrm{P}(\hat{Y}=1 \mid S=0) = \mathrm{P}(\hat{Y}=1 \mid S=1)$

Metric: $\Delta_{DP} = |\mathrm{P}(\hat{Y}=1 \mid S=0) - \mathrm{P}(\hat{Y}=1 \mid S=1)|$

[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.

Demographic Parity (Cont.)

Group Fairness: Demographic Parity

Demographic Parity is first proposed in **binary classification task** for tabular data [1].

Demographic Parity is considered as achieved if the model yields the **same positive rate** for individuals in both **sensitive subgroups**.

Criterion: $\mathrm{P}(\hat{Y}=1|S=0) = \mathrm{P}(\hat{Y}=1|S=1)$

Metric: $\Delta_{DP} = |\mathrm{P}(\hat{Y}=1|S=0) - \mathrm{P}(\hat{Y}=1|S=1)|$

Recent works on fairness have **extended this notion to other settings**, including link prediction [2, 3] and scenarios with continuous sensitive feature(s) values [4];

[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science, 2012.
[2] Acquisti, Alessandro, and Christina Fong. "An experiment in hiring discrimination via online social networks." Management Science 66.3 (2020): 1005-1024.
[3] Mengnan Du, Fan Yang, Na Zou, and Xia Hu. Fairness in deep learning: A computational perspective. IEEE Intelligent Systems, 2020.
[4] Yushun Dong, Jing Ma, Chen Chen, and Jundong Li. Fairness in Graph Mining: A Survey. arXiv preprint arXiv:2204.09888, 2022.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

Equality of Odds: the **positive rate** are enforced to be the same between sensitive subgroups conditional on the **ground truth class labels**.

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

Equality of Odds: the **positive rate** are enforced to be the same between sensitive subgroups conditional on the **ground truth class labels**.

Criterion: $\mathrm{P}(\hat{Y}=1 \mid S=0, Y=y) = \mathrm{P}(\hat{Y}=1 \mid S=1, Y=y)$

Metric:

$$\Delta_{EOD} = \left|P(\hat{Y}=1 \mid S=0, Y=1) - P(\hat{Y}=1 \mid S=1, Y=1)\right| + \left|P(\hat{Y}=1 \mid S=0, Y=0) - P(\hat{Y}=1 \mid S=1, Y=0)\right|$$

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

The intuition of Equality of Odds: to enforce the true positive rate (right and positive results) and false positive rate (wrong but positive results) to be the same across sensitive subgroups;

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

The intuition of Equality of Odds: to enforce the true positive rate (right and positive results) and false positive rate (wrong but positive results) to be the same across sensitive subgroups;

Equality of Opportunity: the **positive rate** is enforced to be the same between sensitive subgroups conditional on the **positive ground truth class labels**.

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

The intuition of Equality of Odds: to enforce the true positive rate (right and positive results) and false positive rate (wrong but positive results) to be the same across sensitive subgroups;

Equality of Opportunity: the **positive rate** is enforced to be the same between sensitive subgroups conditional on the **positive ground truth class labels**.

Criterion: $\mathrm{P}\left(\hat{Y}=1 \middle| S=0, Y=1\right) = \mathrm{P}\left(\hat{Y}=1 \middle| S=1, Y=1\right)$

Metric: $\Delta_{EO} = \left|P\left(\hat{Y}=1 \middle| S=0, Y=1\right) - P(\hat{Y}=1 | S=1, Y=1)\right|$

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

The intuition of Equality of Odds: to enforce the true positive rate (right and positive results) and false positive rate (wrong but positive results) to be the same across sensitive subgroups;

The intuition of Equality of Opportunity: to enforce the true positive rate (right and positive results) to be the same across sensitive subgroups;

[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.

Equality of Odds/Opportunity

Group Fairness:

Equality of Odds [1] vs. **Equality of Opportunity** [1]

The intuition of Equality of Odds: to enforce the true positive rate (right and positive results) and false positive rate (wrong but positive results) to be the same across sensitive subgroups;

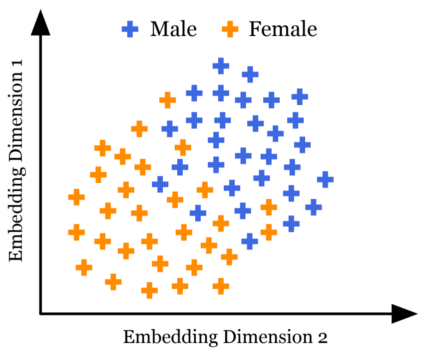
The intuition of Equality of Opportunity: to enforce the true positive rate (right and positive results) to be the same across groups;

Extension to tasks other than node classification, e.g., edge prediction [1, 2].

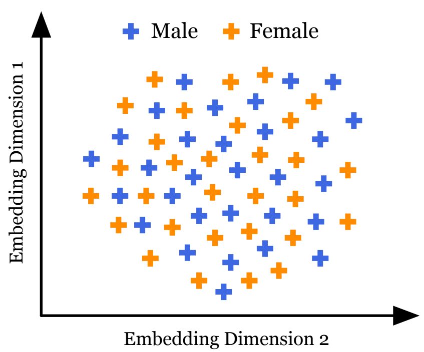
[1] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In NeurIPS, 2016.
[2] Acquisti, Alessandro, and Christina Fong. "An experiment in hiring discrimination via online social networks." Management Science 66.3 (2020): 1005-1024.

Fairness in Node Embedding Learning

(1) Distribution-Based Fairness.



Unfair node embeddings



Fair node embeddings

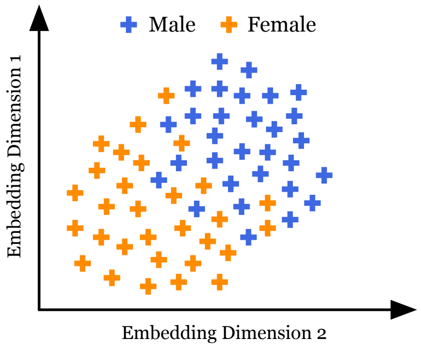
Criterion: Learned node embedding distributions across sensitive subgroups should be **similar**.

Metric: Measures of distance between distributions, e.g., Wasserstein distance [1, 2].

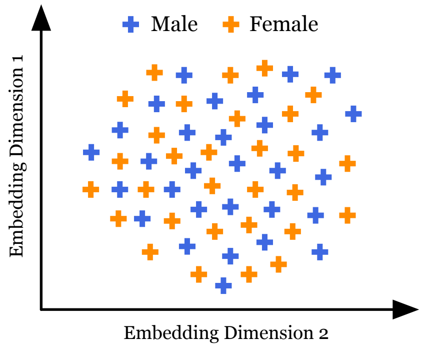
[1] Yushun Dong, Ninghao Liu, Brian Jalaian, and Jundong Li. EDITS: modeling and mitigating data bias for graph neural networks. In WWW, 2022.
[2] Wei Fan, Kunpeng Liu, Rui Xie, Hao Liu, Hui Xiong, and Yanjie Fu. Fair graph auto-encoder for unbiased graph representations with Wasserstein distance. In ICDM, 2021.

Fairness in Node Embedding Learning

(1) Distribution-Based Fairness.



Unfair node embeddings



Fair node embeddings

Criterion: Learned node embedding distributions across sensitive subgroups should be **similar**.

Metric: Measures of distance between distributions, e.g., Wasserstein distance [1, 2].

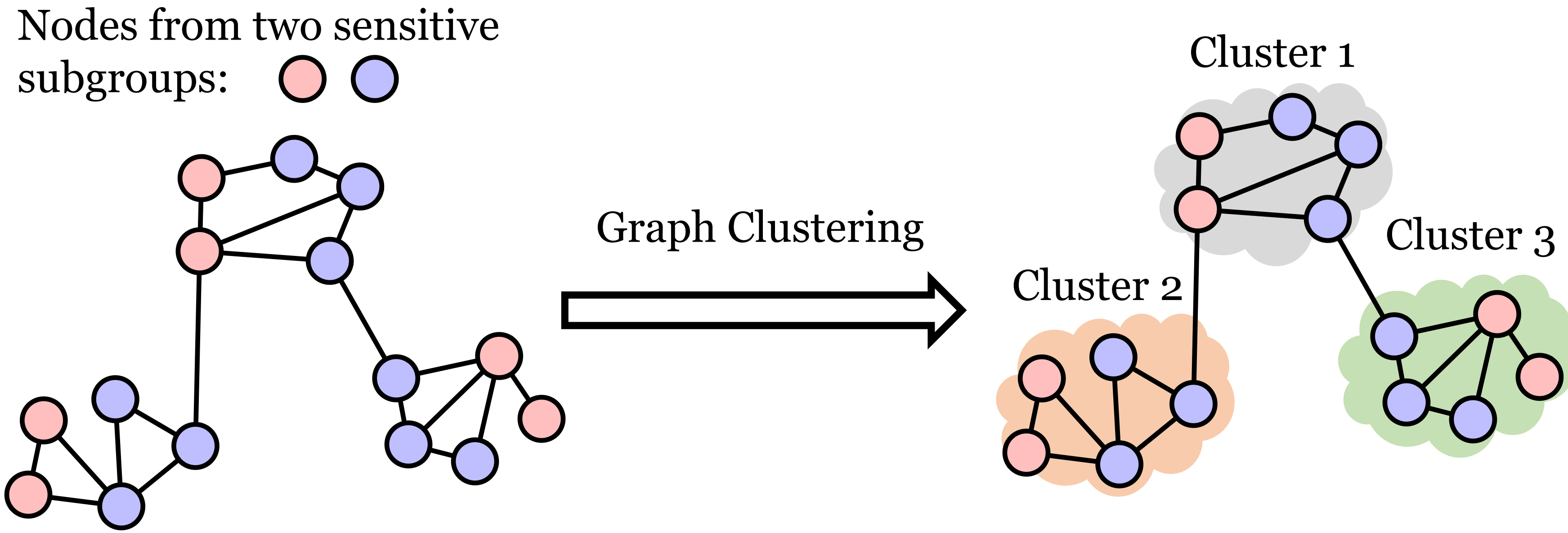
(2) Model-Based Fairness.

Criterion: There should be no information about sensitive attributes encoded in the learned node embeddings.

Metric: **Prediction accuracy** on the sensitive attributes with another model (the lower, the better) [3].

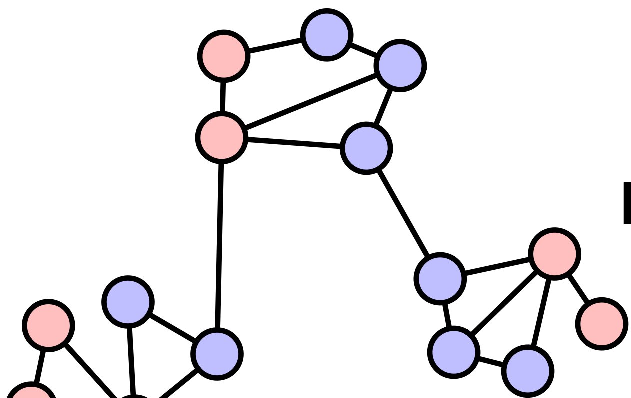
[1] Yushun Dong, Ninghao Liu, Brian Jalaian, and Jundong Li. EDITS: modeling and mitigating data bias for graph neural networks. In WWW, 2022.
[2] Wei Fan, Kunpeng Liu, Rui Xie, Hao Liu, Hui Xiong, and Yanjie Fu. Fair graph auto-encoder for unbiased graph representations with Wasserstein distance. In ICDM, 2021.
[3] Le Wu, Lei Chen, Pengyang Shao, Richang Hong, Xiting Wang, and Meng Wang. Learning fair representations for recommendation: A graph-based perspective. In WWW, 2021.

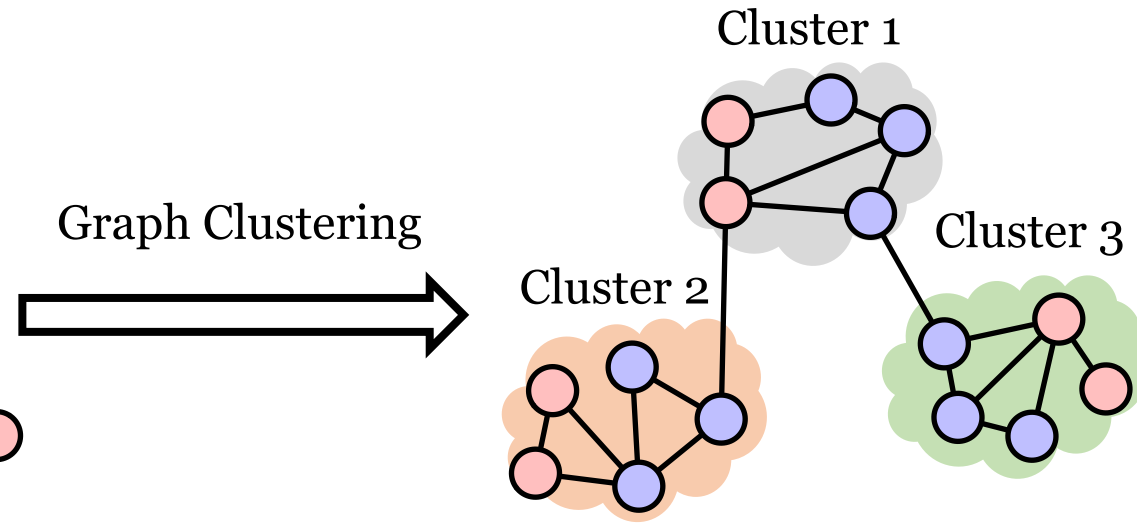
Fairness in Graph Clustering

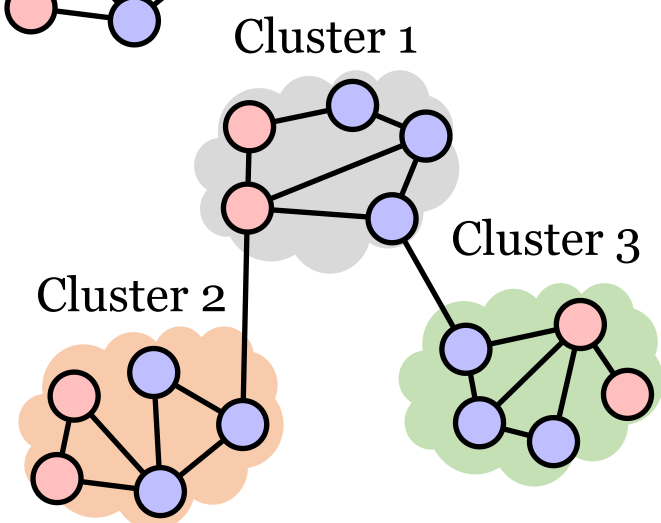


Fairness in Graph Clustering

Nodes from two sensitive subgroups:







Criterion: Nodes from different sensitive subgroups should be **proportionally represented** in each cluster [1].

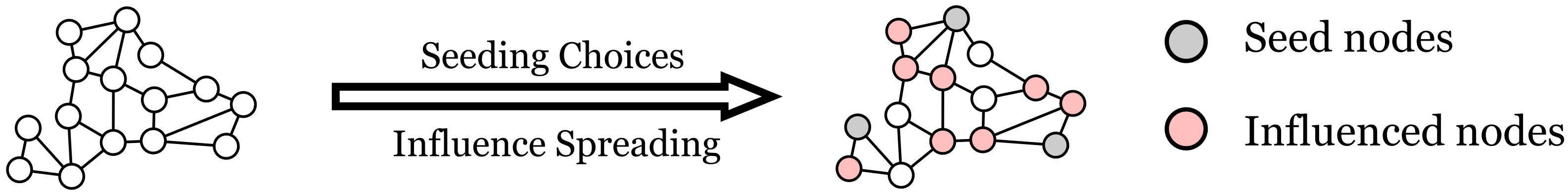
Metric: *Balance score*, which measures fairness with the minimum number ratio [1].

$$Balance(C_k) = \min_{i \neq i', i, i' \in \{1,\ldots,H\}} \frac{|V_i \cap C_k|}{|V_{i'} \cap C_k|}$$

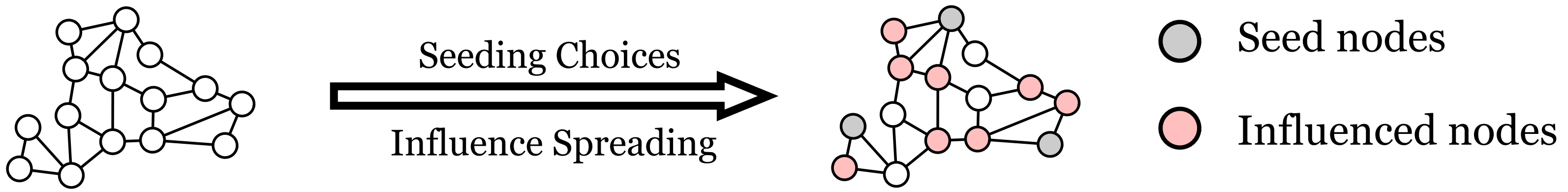
V_i: node set of sensitive subgroup i;
C_l: node set of cluster l;

[1] Matthaus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In ICML, 2019.

Fairness in Influence Maximization



Fairness in Influence Maximization



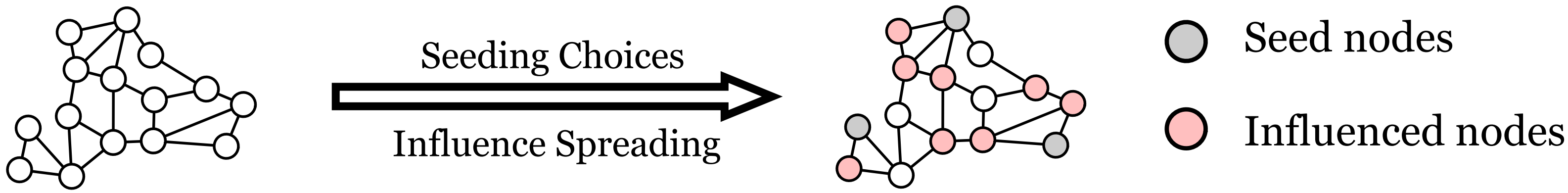
(1) Maxmin Fairness [1].

Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.

[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.

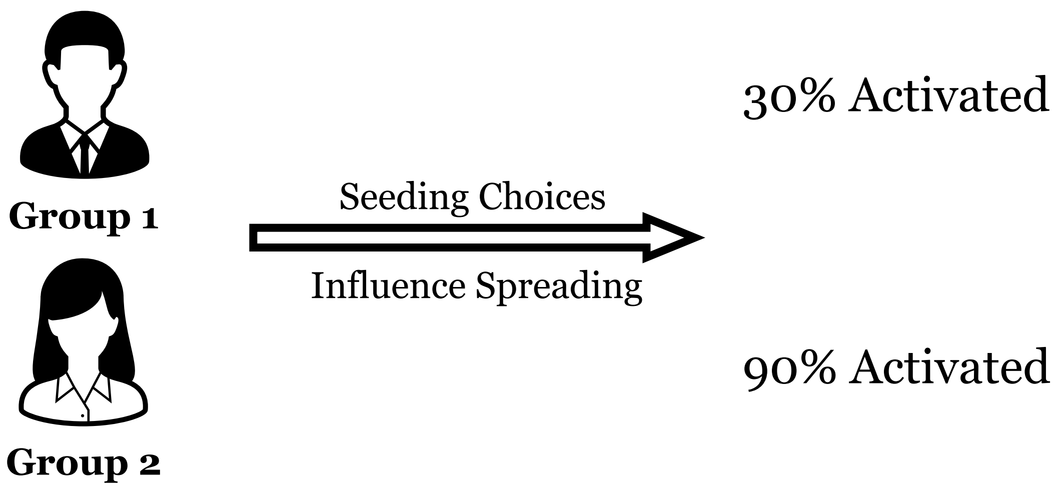
Fairness in Influence Maximization



(1) Maxmin Fairness [1].

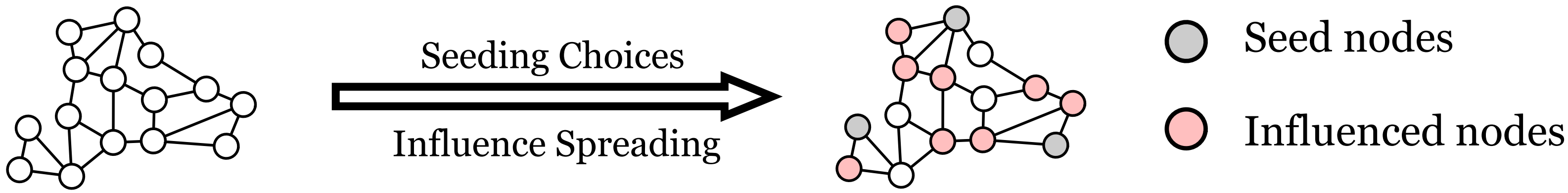
Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.



[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.

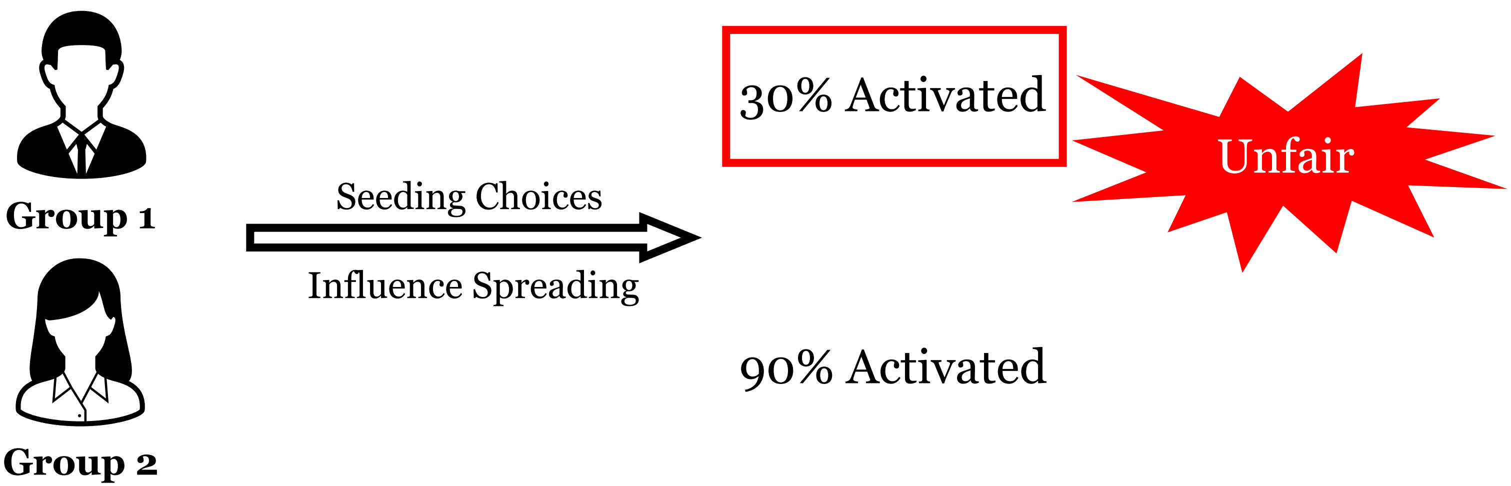
Fairness in Influence Maximization



(1) Maxmin Fairness [1].

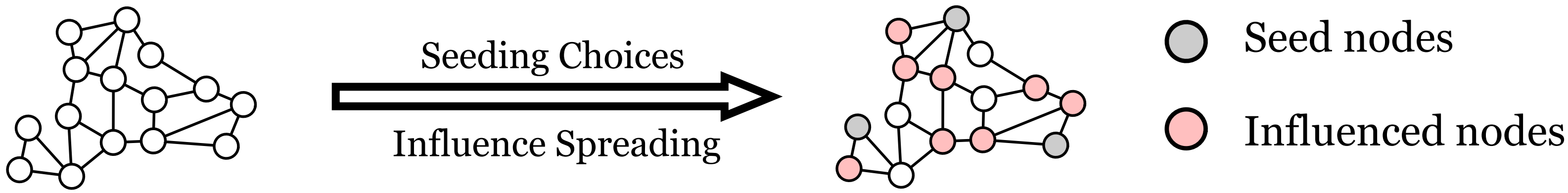
Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.



[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.

Fairness in Influence Maximization



(1) Maxmin Fairness [1].

Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.

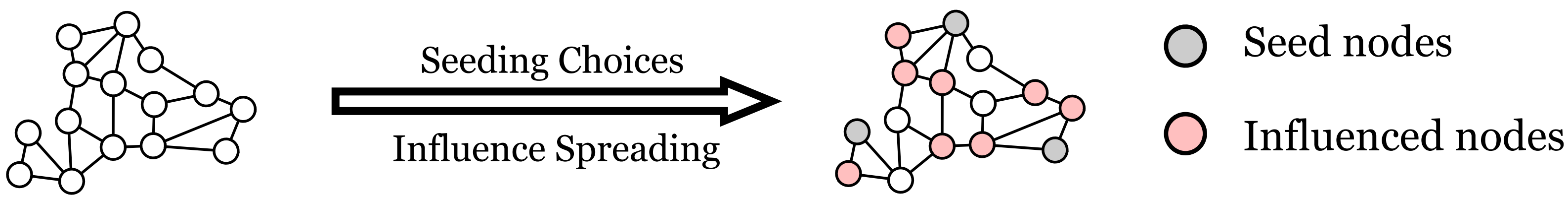
(2) Diversity [1].

Criterion: The influence rate in each sensitive subgroup should be larger than (or equal to) the rate when this subgroup is given a proportional seeding budget.

Metric: The percentage of sensitive subgroups that violates such criterion.

[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.

Fairness in Influence Maximization



(1) Maxmin Fairness [1].

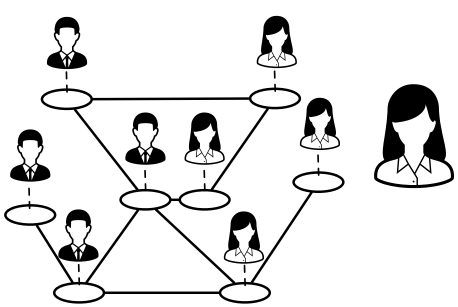
Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.

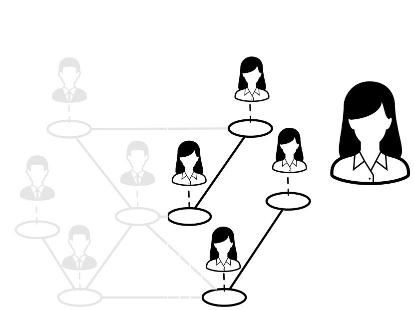
(2) Diversity [1].

Criterion: The influence rate in each sensitive subgroup should be larger than (or equal to) the rate when this subgroup is given a proportional seeding budget.

Metric: The percentage of sensitive subgroups that violates such criterion.



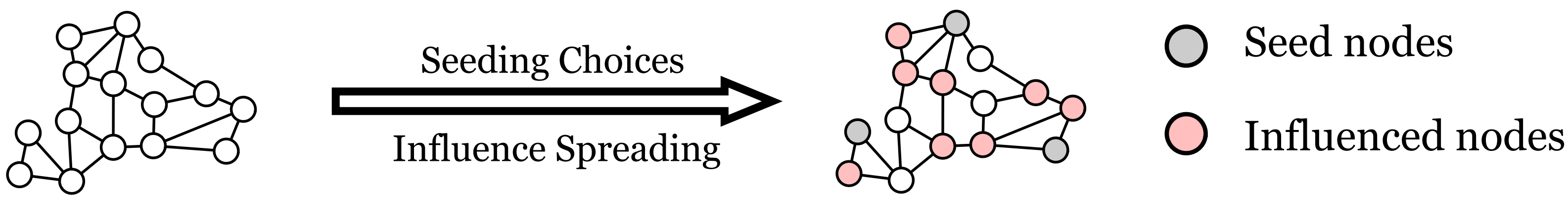
:25% activated with seeding budget 4



:100% activated with seeding budget 2

[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.

Fairness in Influence Maximization



(1) Maxmin Fairness [1].

Criterion: The lowest influence rate among sensitive subgroups should be maximized.

Metric: The lowest influence rate among all sensitive subgroups.

(2) Diversity [1].

Criterion: The influence rate in each sensitive subgroup should be larger than (or equal to) the rate when this subgroup is given a proportional seeding budget.

Metric: The percentage of sensitive subgroups that violates such criterion.

(3) Utility Difference-Based Fairness [2].

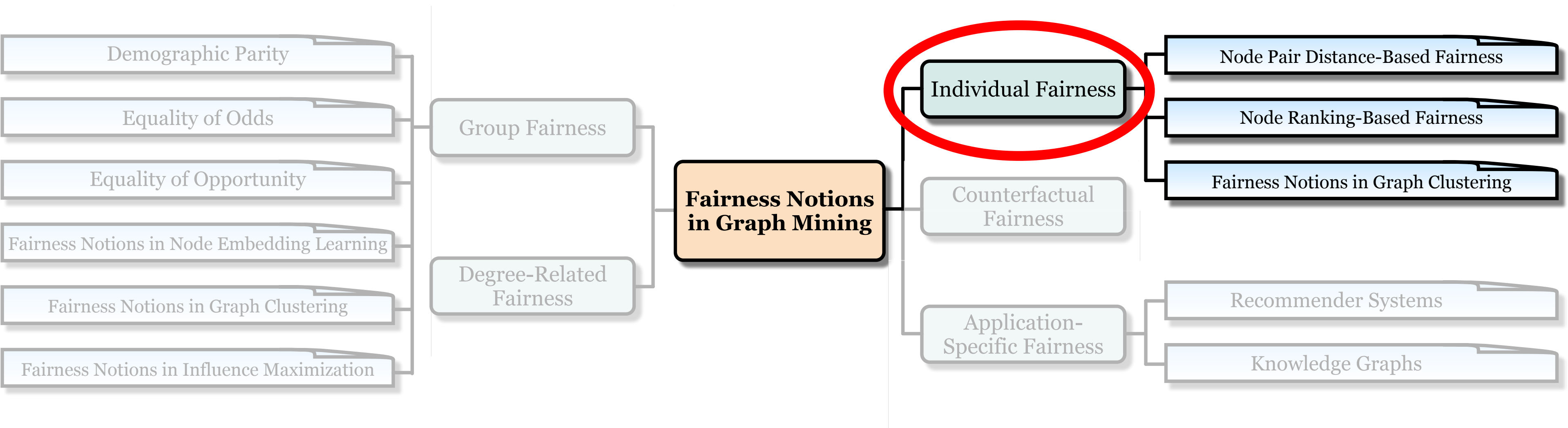
Criterion: The influence rate should be the same across different sensitive subgroups.

Metric: The maximum influence rate difference among all sensitive subgroup pairs.

[1] Alan Tsang, et al. Group fairness in influence maximization. In IJCAI, 2019.
[2] Junaid Ali, et al. On the fairness of time-critical influence maximization in social networks. In NeurIPS, 2019.

Taxonomy of Fairness Notions

Another critical fairness notion in graph mining: Individual Fairness.

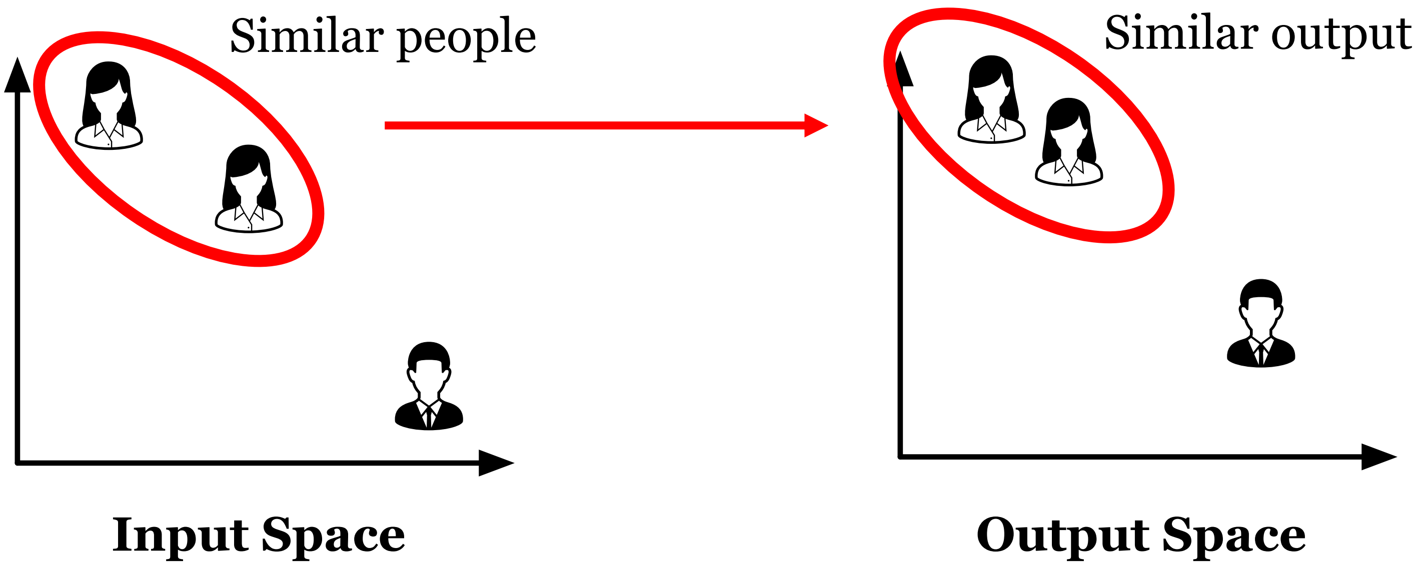


A general idea of individual fairness: **similar individuals should receive similar outputs** from the graph mining algorithms [1].

[1] Ziqian Zeng, Rashidul Islam, et al. Fair representation learning for heterogeneous information networks. In AAAI, 2021.

Node Pair Distance-Based Fairness

For any pair of node, this fairness notion enforces **the output distance to be smaller than a scaled input distance** - which is consistent with the general idea of "similar individual should receive similar output" [1].



[1] Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong. Inform: Individual fairness on graph mining. In SIGKDD, 2020.

Node Pair Distance-Based Fairness

For any pair of node, this fairness notion enforces **the output distance to be smaller than a scaled input distance** - which is consistent with the general idea of "similar individual should receive similar output" [1].

Mathematically, we have

$$\boxed{D_1(f(x), f(y))} \leq L\boxed{D_2(x, y)} \quad \forall\, (x, y) \qquad L\text{: Lipschitz Constant}$$

Output distance   Input distance

In practice, we enforce the following inequality

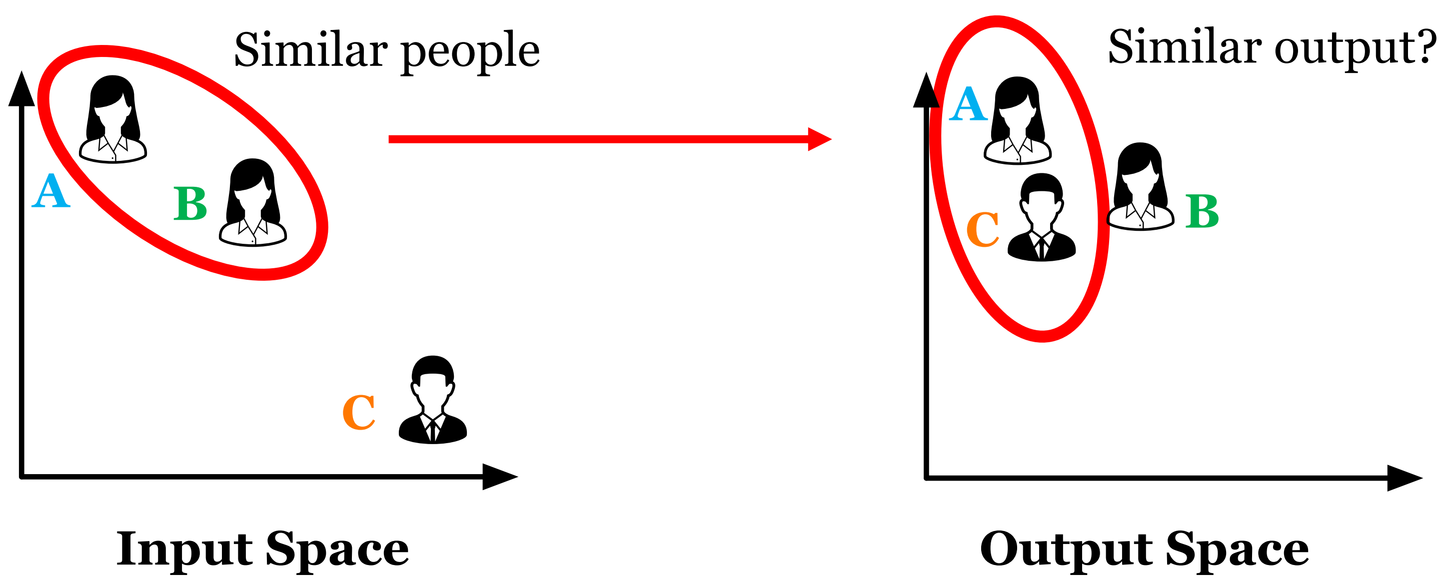
$$\|\mathbf{Y}[i,:] - \mathbf{Y}[j,:]\|_F^2 \leq \frac{\epsilon}{\mathbf{S}[i,j]} \quad \forall\, i, j = 1, \ldots, n$$

Y: Output matrix to compute D_1; **S**: Similarity matrix according to $D_2(x, y)$

[1] Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong. Inform: Individual fairness on graph mining. In SIGKDD, 2020.

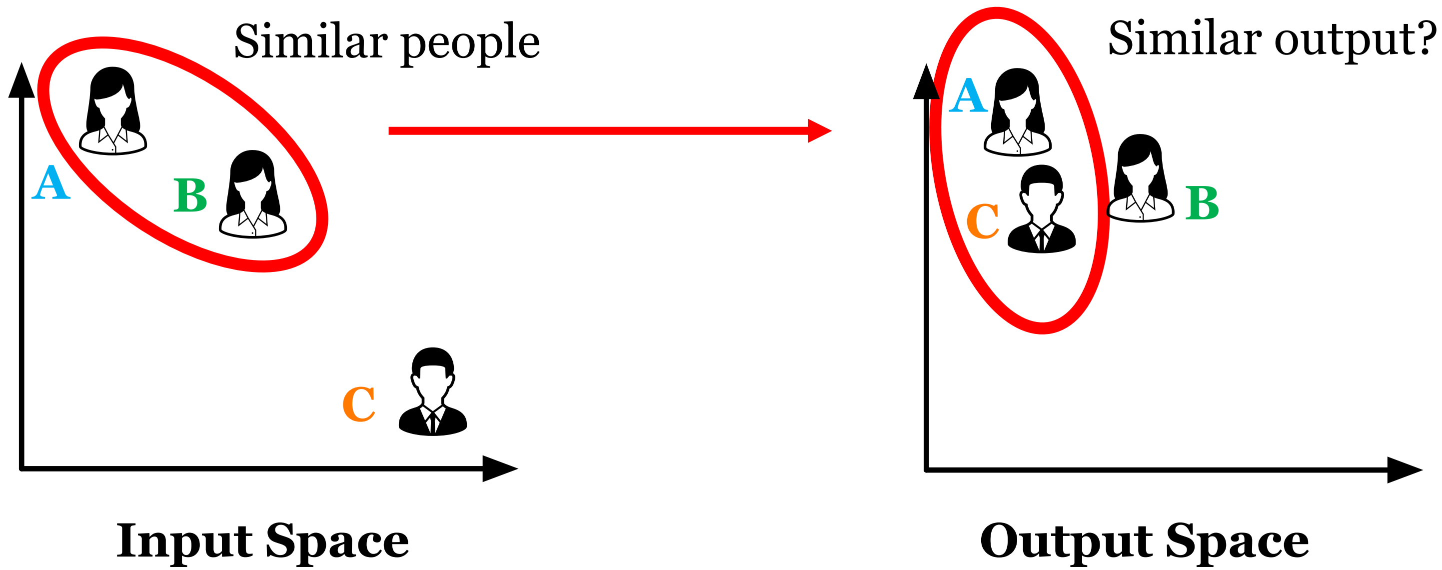
Node Ranking-Based Fairness

Node Pair Distance-Based Fairness can lead to unfairness in a relative perspective: **B is closer to A compared with C in the input space, but A and C is closer in the output space.**



Node Ranking-Based Fairness

Node Pair Distance-Based Fairness can lead to unfairness in a relative perspective: **B is closer to A compared with C in the input space, but A and C is closer in the output space.**



This could lead to a **sense of unfairness** for involved individuals.

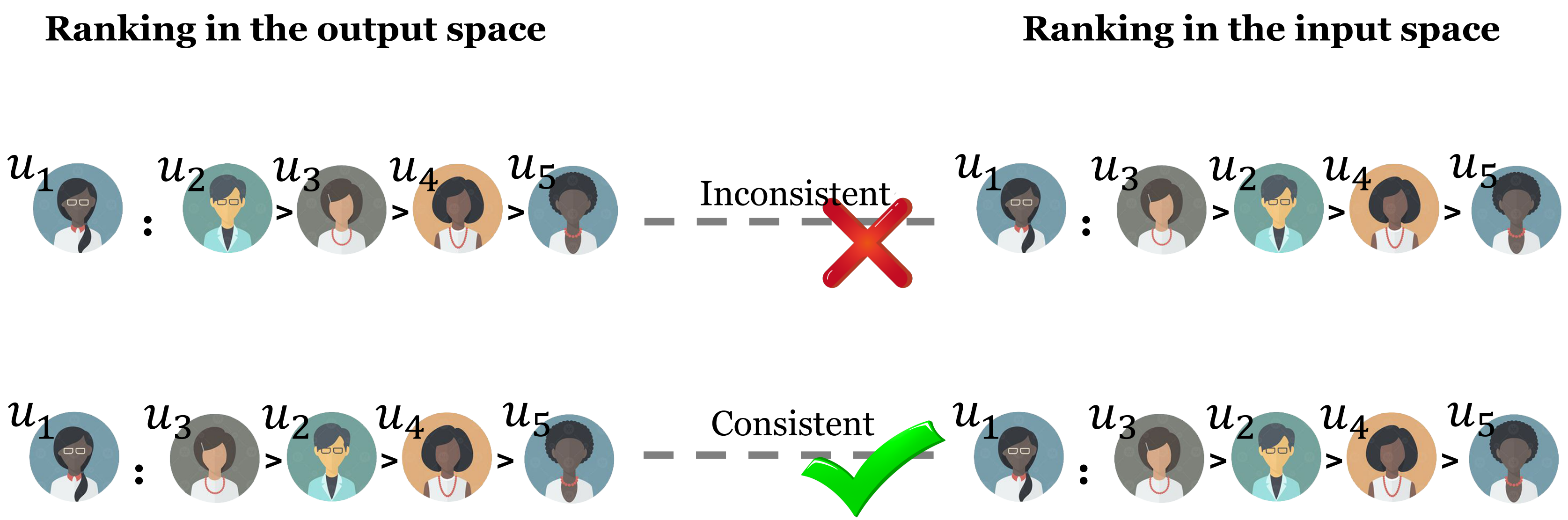
Node Ranking-Based Fairness

Criterion: for each individual, the similarity rankings (between itself and all other people) in both input and output space should be the same [1].

[1] Yushun Dong, Jian Kang, Hanghang Tong, and Jundong Li. Individual fairness for graph neural networks: A ranking based approach. In SIGKDD, 2021.

Node Ranking-Based Fairness

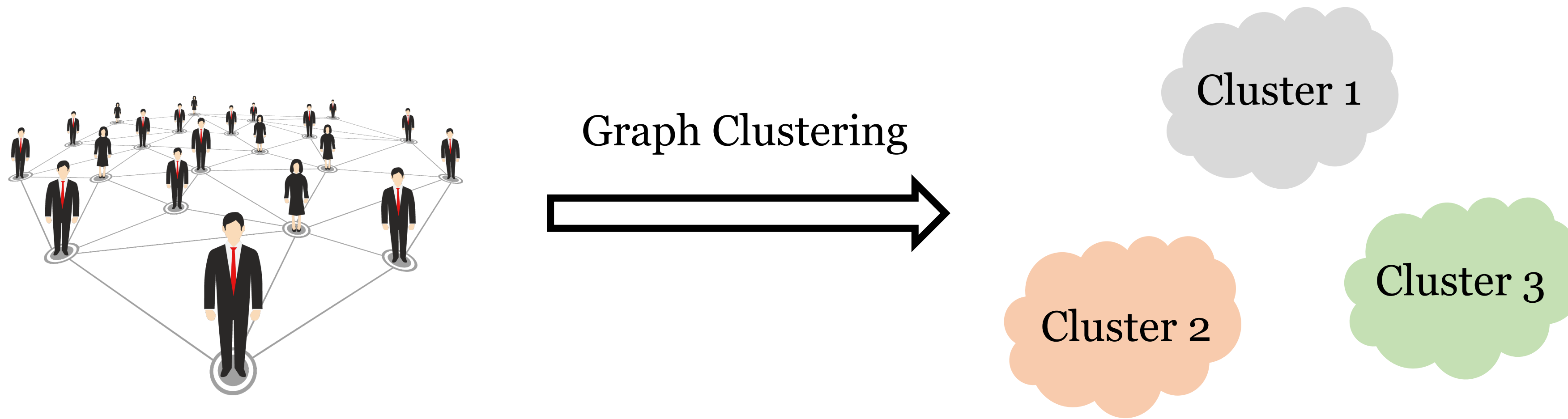
Criterion: for each individual, the similarity rankings (between itself and all other people) in both input and output space should be the same [1].



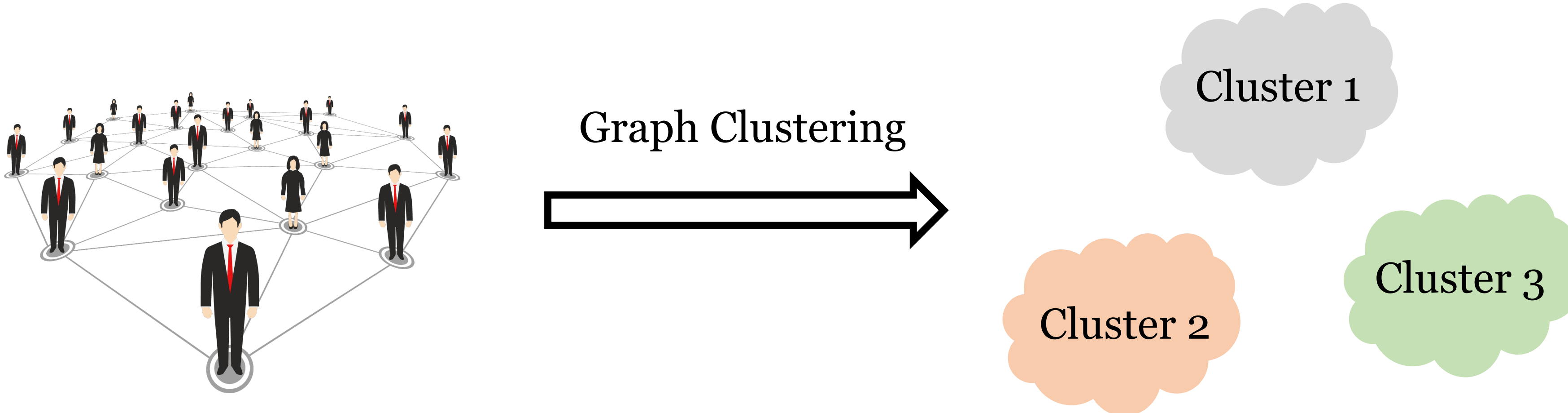
Metrics: average ranking similarity across all individuals, e.g., average NDCG@k [2].

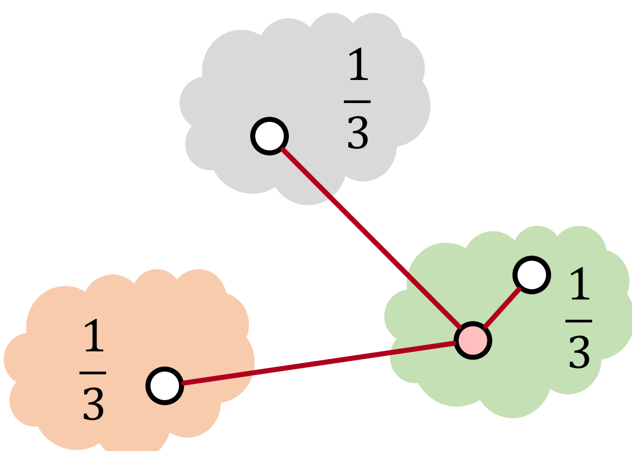
[1] Yushun Dong, Jian Kang, Hanghang Tong, and Jundong Li. Individual fairness for graph neural networks: A ranking based approach. In SIGKDD, 2021.
[2] Mattha¨us Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In ICML, 2019.

Individual Fairness in Graph Clustering



Individual Fairness in Graph Clustering

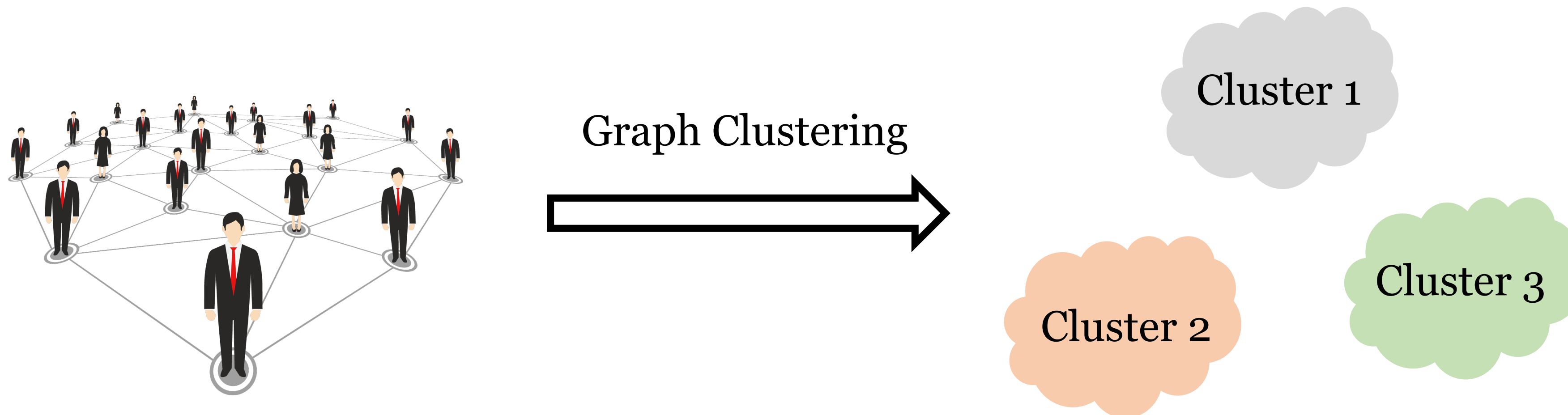


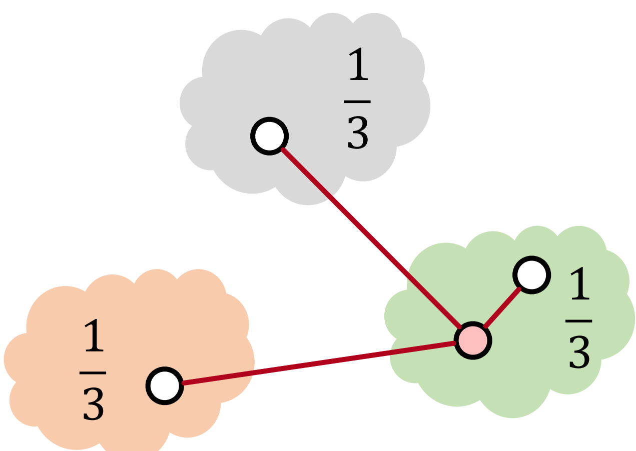


Criterion: For every node , its neighbors should be proportionally represented by each cluster [1].

[1] Shubham Gupta and Ambedkar Dukkipati. Protecting individual interests across clusters: Spectral clustering with guarantees. arXiv preprint arXiv:2105.03714, 2021.

Individual Fairness in Graph Clustering





Criterion: For every node ○, its neighbors should be proportionally represented by each cluster [1].

Metric: how disproportionately neighbors of a node are assigned in different clusters (node-level) [1].

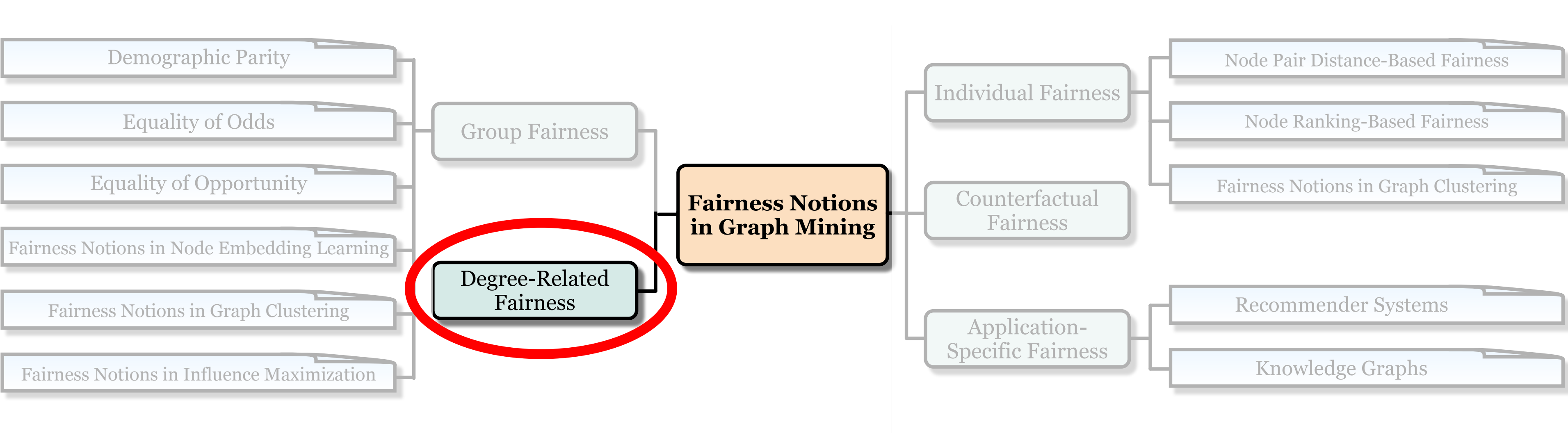
$$\rho_i = \min_{k,l\in\{1,\dots,K\}} \frac{|C_k \cap N_{v_i}|}{|C_l \cap N_{v_i}|}$$

C_k: node set of cluster k;
C_l: node set in cluster l;
N_{v_i}: Neighbor set of node v_i;

[1] Shubham Gupta and Ambedkar Dukkipati. Protecting individual interests across clusters: Spectral clustering with guarantees. arXiv preprint arXiv:2105.03714, 2021.

Taxonomy of Fairness Notions

- A fairness notion **tailored with graph structure**: Degree-Related Fairness.

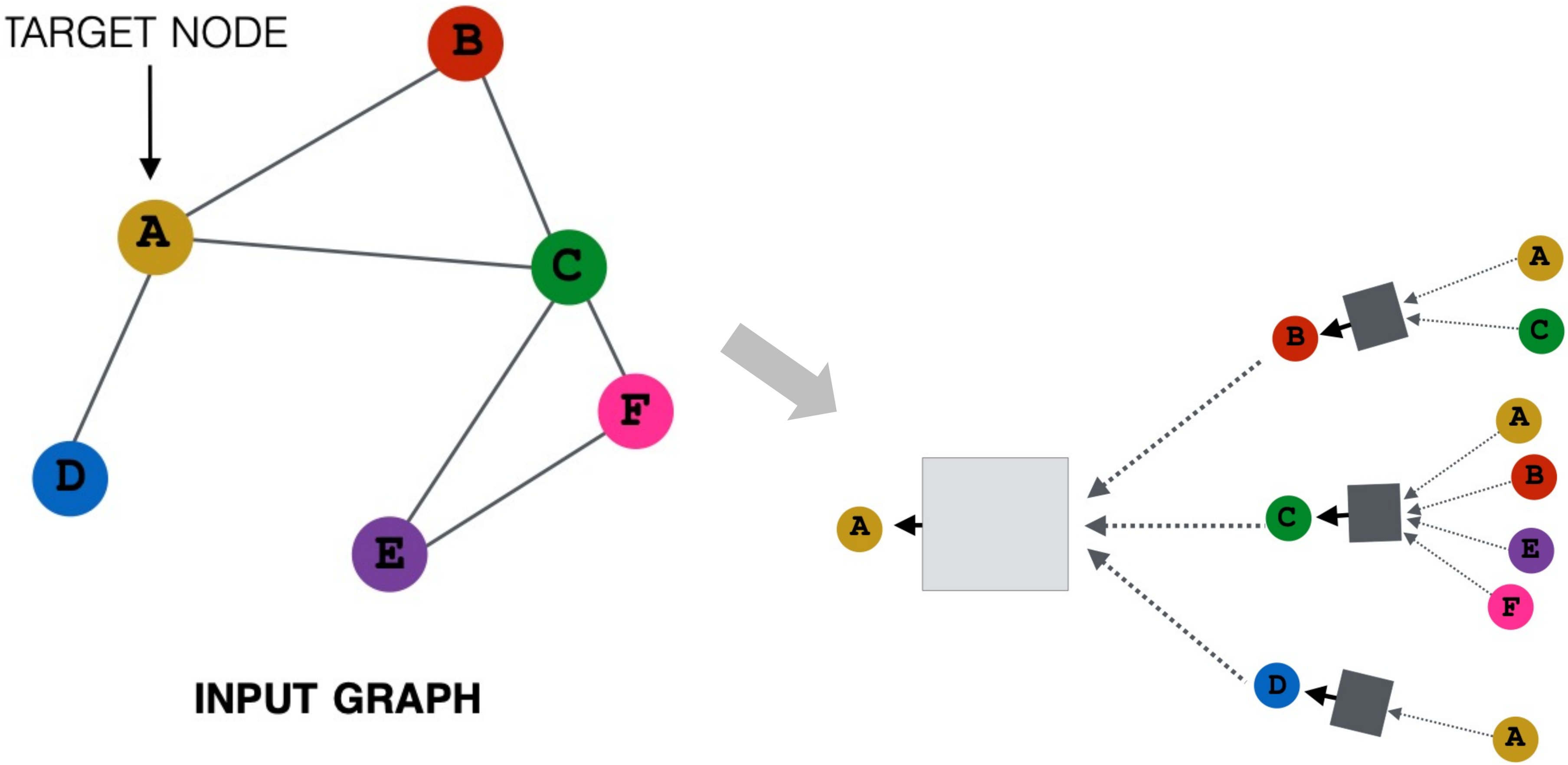


A general idea of degree-related fairness: the degree of nodes should be independent from the quality of their corresponding predictions [1, 2].

[1] Xianfeng Tang, et al. Investigating and mitigating degree-related biases in graph convoltuional networks. In CIKM, 2020
[2] Jian Kang, et al. Rawlsgcn: Towards Rawlsian difference principle on graph convolutional network. In WWW, 2022.

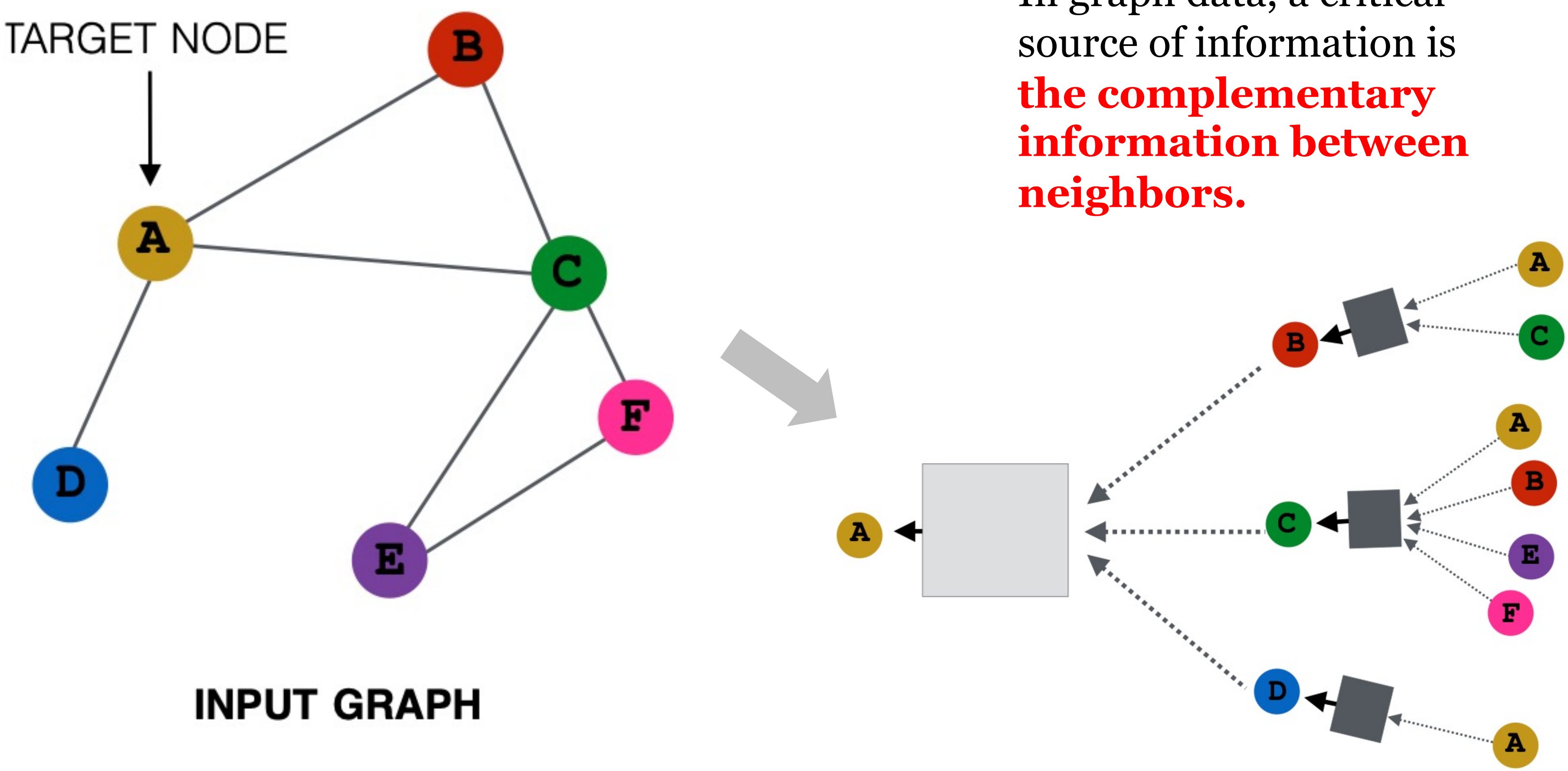
Degree-Related Fairness

A typical **information aggregation** in Graph Neural Networks:



Degree-Related Fairness

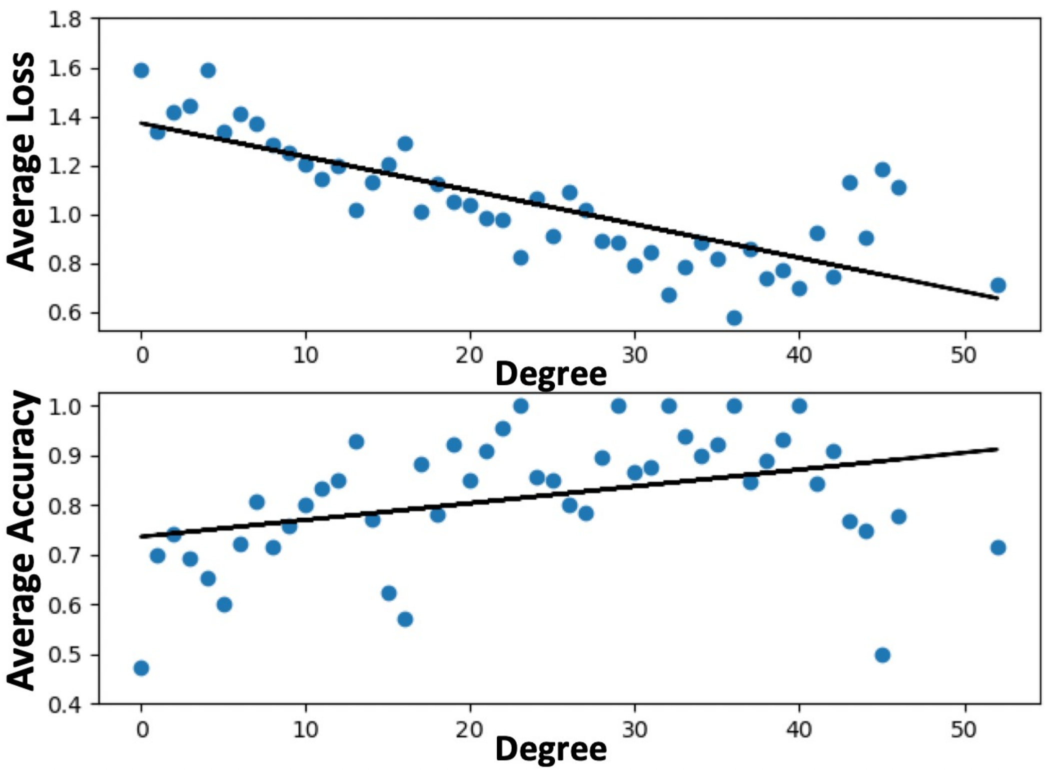
A typical **information aggregation** in Graph Neural Networks:



In graph data, a critical source of information is **the complementary information between neighbors.**

Degree-Related Fairness (Cont.)

A typical **average loss distribution** across node degrees in Graph Neural Networks:

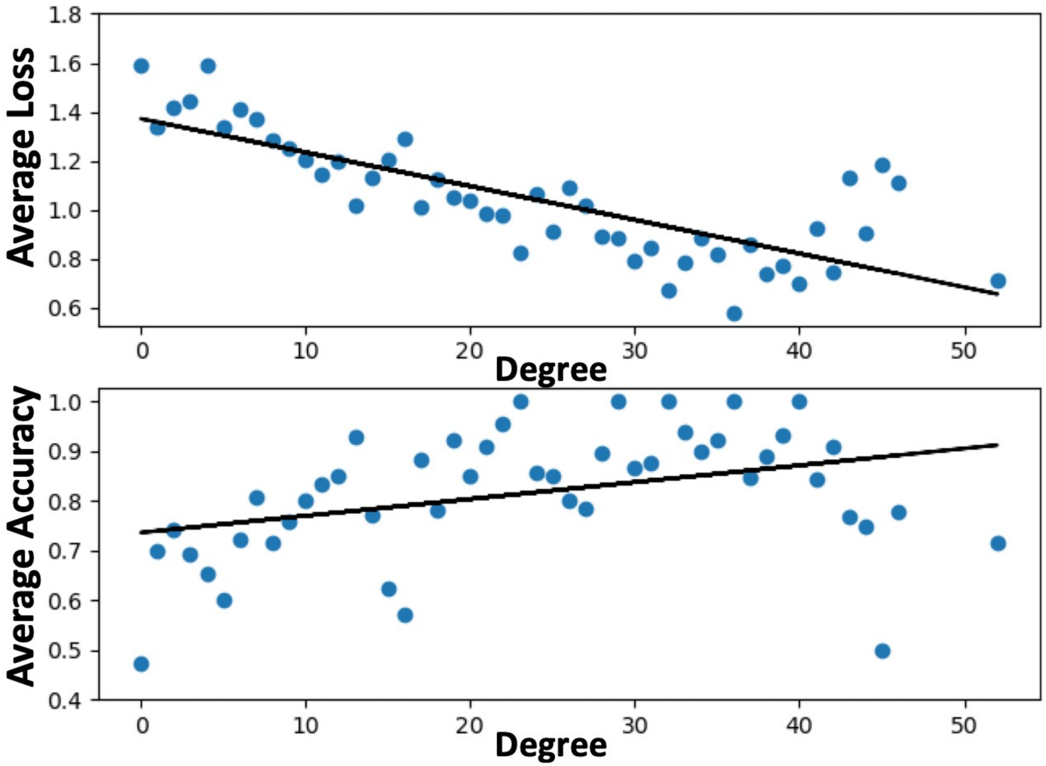


In graph data, a critical source of information is **the complementary information between neighbors.**

However, graph mining algorithms rely on such information tend to yield predictions with **much worse quality** for low-degree nodes, as they have **fewer neighbors**.

Degree-Related Fairness (Cont.)

A typical **average loss distribution** across node degrees in Graph Neural Networks:



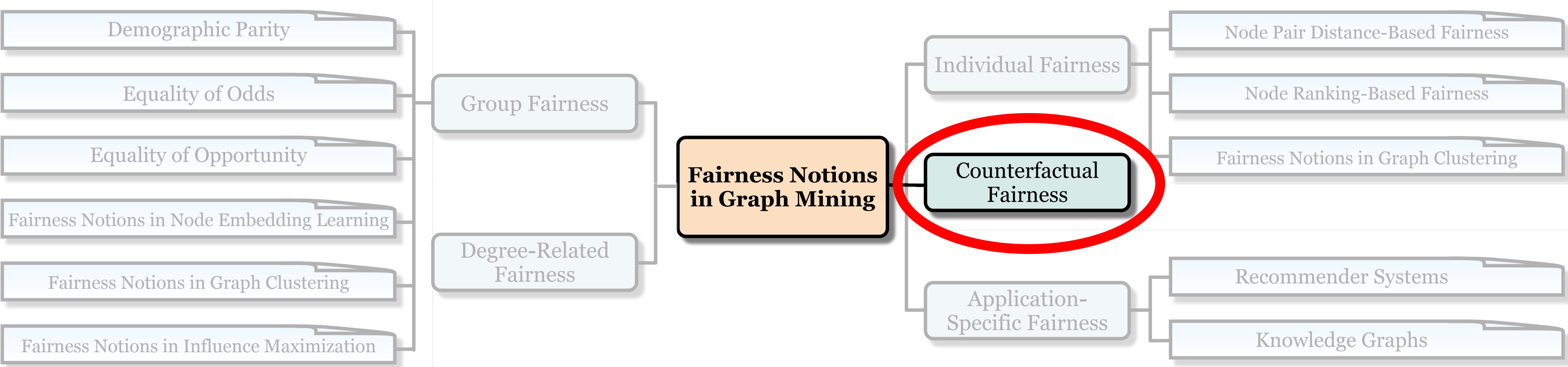
In graph data, a critical source of information is **the complementary information between neighbors.**

However, graph mining algorithms rely on such information tend to yield predictions with **much worse quality** for low-degree nodes, as they have **fewer neighbors**.

Degree-Related Fairness requires that nodes should bear similar utility (e.g., node classification accuracy) in the graph mining algorithms **regardless of their degrees**.

Taxonomy of Fairness Notions

A fairness notion **from the causal perspective**: counterfactual fairness.

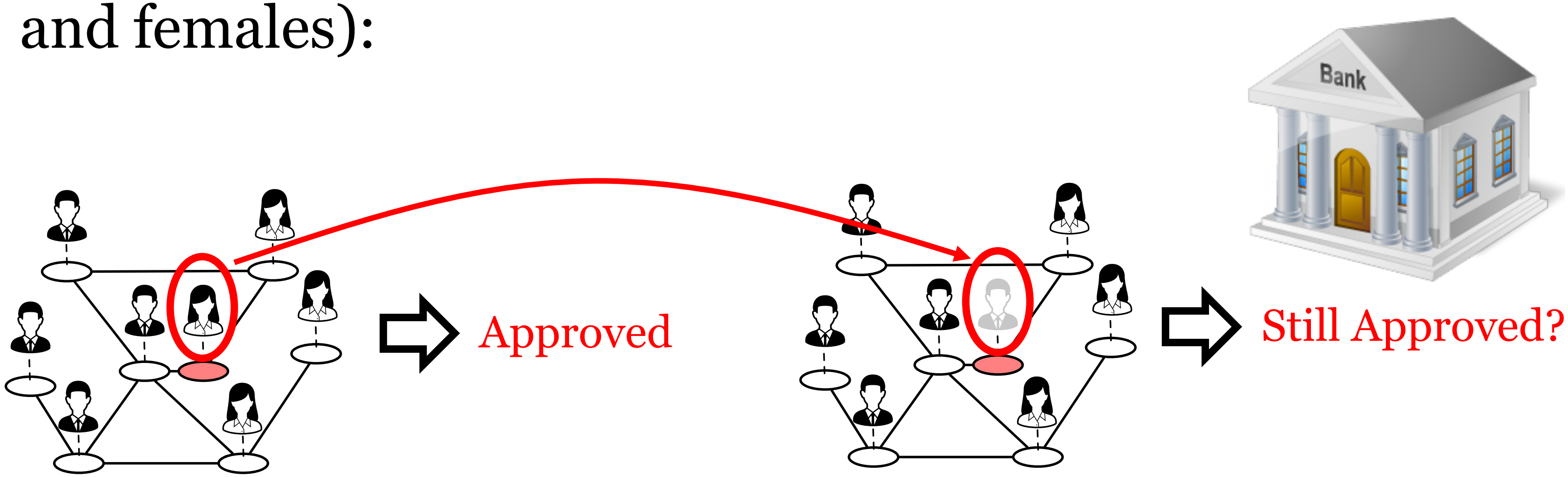


A general idea of counterfactual fairness: the sensitive information of any individual **should not causally influence** the corresponding output [1].

[1] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. NeurIPS, 2017.

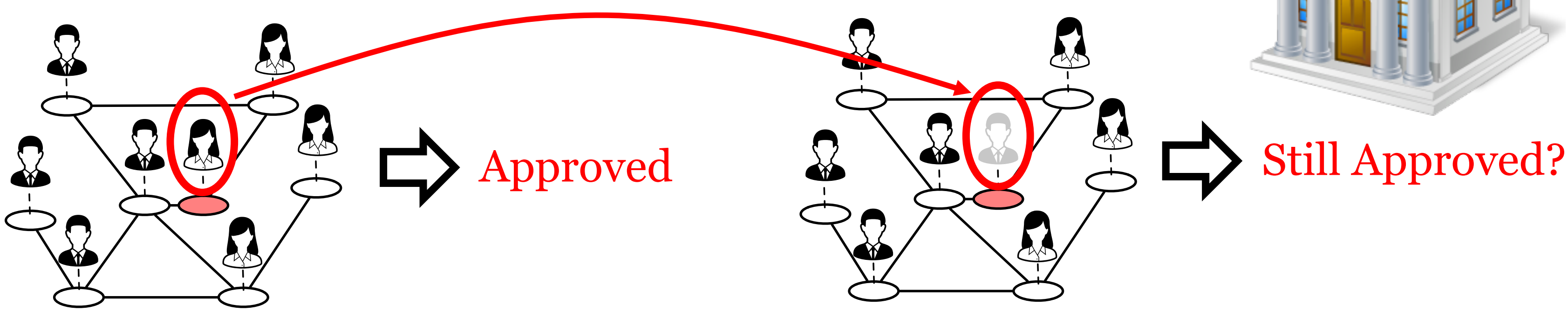
Counterfactual Fairness

Consider a network of loan applicants (including males and females):



Counterfactual Fairness

Consider a network of loan applicants (including males and females):



Criterion: If the sensitive feature of an individual is changed into a different value (e.g., from s to s'), the output should still be maintained the same [1].

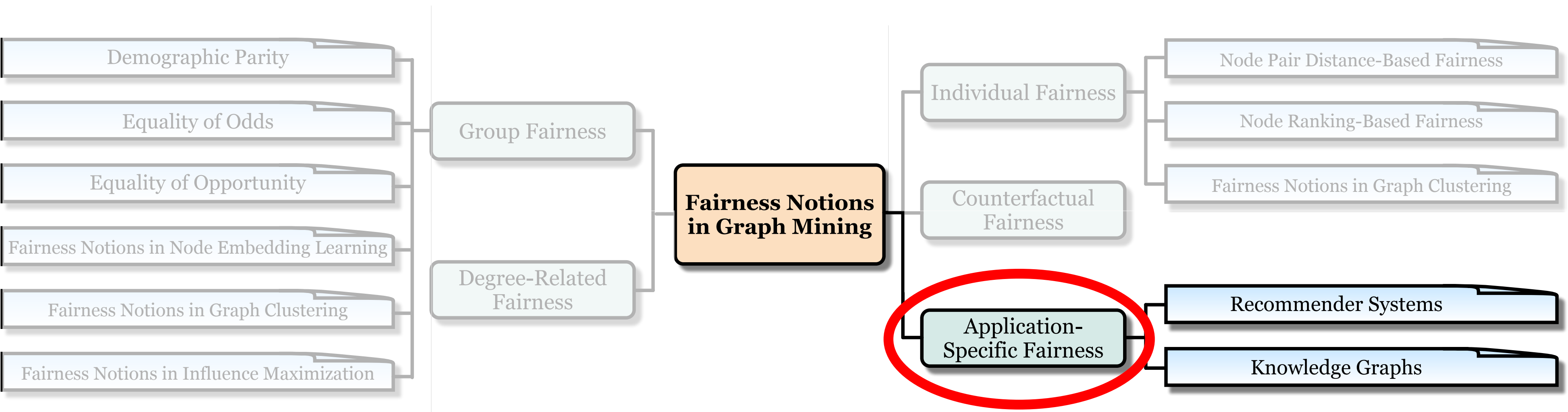
$$P\left(\hat{Y}_{S\leftarrow s} = y | X = x, S = s\right) = P\left(\hat{Y}_{S\leftarrow s'} = y | X = x, S = s\right)$$

Metric: the percentage of nodes whose predicted label changes when their sensitive feature values are changed.

[1] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. NeurIPS, 2017.

Taxonomy of Fairness Notions

Fairness notions **in real-world applications:** application-specific fairness.



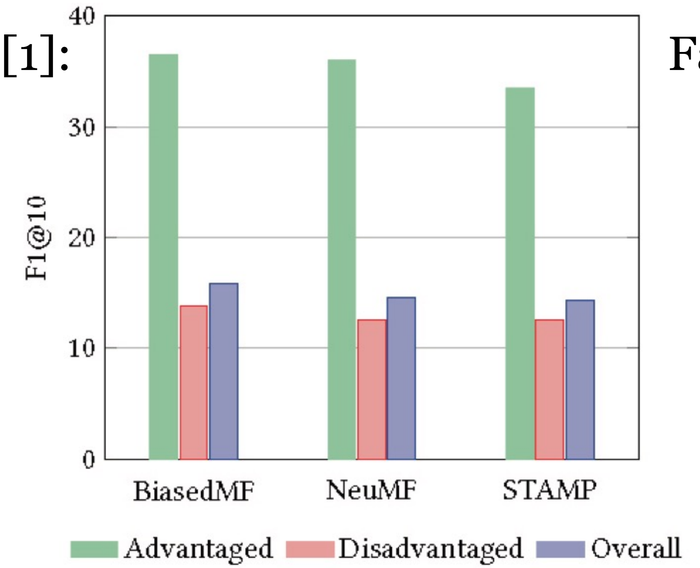
In real-world applications, certain scenarios could bring a sense of unfairness, which requires defining **application-specific fairness** to depict if there is any exhibited bias.

User Fairness in Recommendation

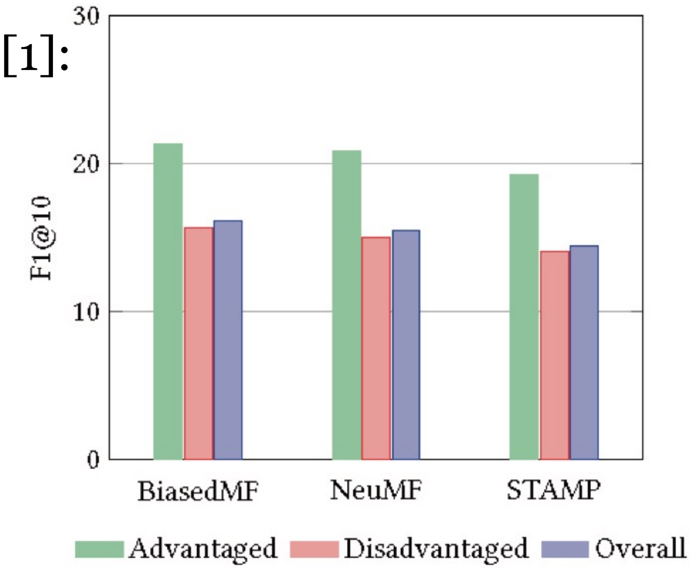
Application-specific fairness in recommender systems.

(1) User Fairness. **Quantitative recommendation utility** for different groups.

More biased [1]:



Fairer [1]:



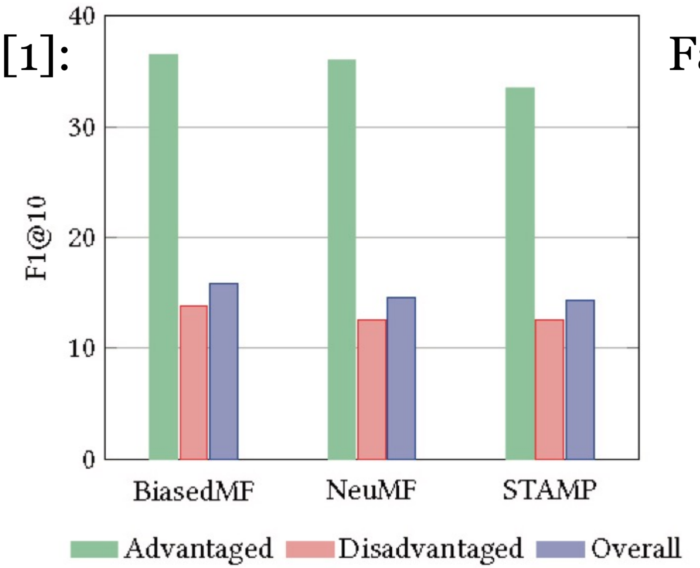
[1] Yunqi Li, et al. User-oriented fairness in recommendation. In WWW, 2021.

User Fairness in Recommendation

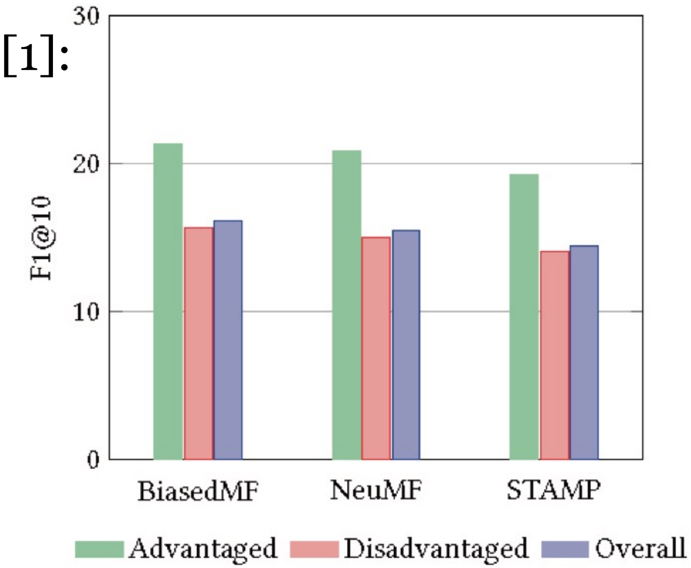
Application-specific fairness in recommender systems.

(1) User Fairness. **Quantitative recommendation utility** for different groups.

More biased [1]:



Fairer [1]:



Criterion: User fairness requires that the **recommendation quality** for different users should be similar [1, 2].

Metric: Measured with the recommendation quality discrepancy between different groups of users (e.g., active users vs. inactive users) [1, 2].

[1] Yunqi Li, et al. User-oriented fairness in recommendation. In WWW, 2021.
[2] Zuohui Fu, et al. Fairness-aware explainable recommenda- tion over knowledge graphs. In SIGIR, 2020.

Popularity Fairness in Recommendation

Application-specific fairness in recommender systems.

(2) Popularity Fairness.

The filter bubble phenomenon: sometimes users are isolated from less popular items or information.

Popularity Fairness in Recommendation

Application-specific fairness in recommender systems.

(2) Popularity Fairness.

The filter bubble phenomenon: sometimes users are isolated from less popular items or information.

Criterion: Popular instances **should not be over-emphasized** compared with other instances [1].

Metric: Measured with the average recommendation rate of less popular instances.

[1] Joseph Fisher, Dave Palfrey, Christos Christodoulopoulos, and Arpit Mittal. Measuring social bias in knowledge graph embeddings. In workshop of AKBC, 2020.

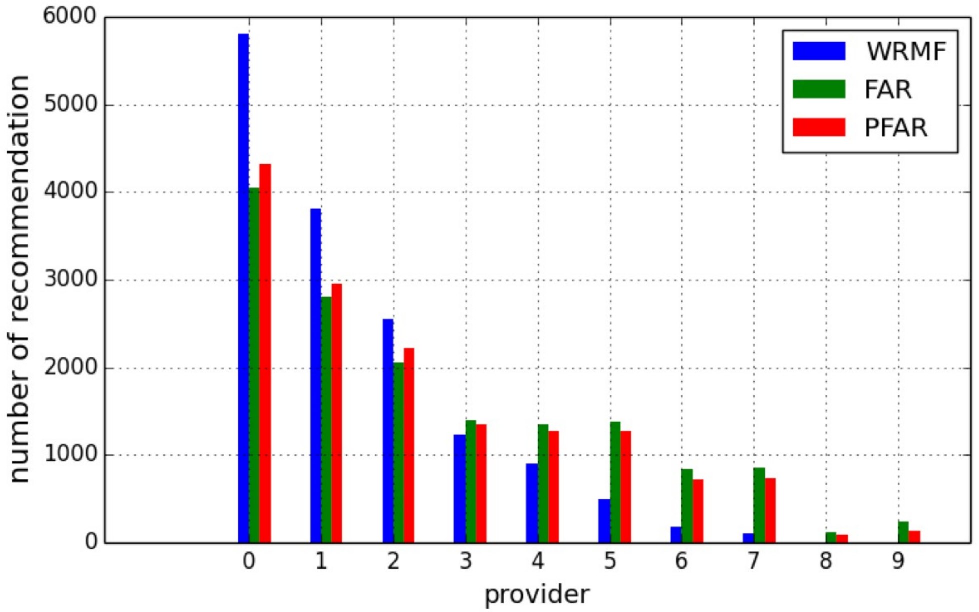
Provider Fairness in Recommendation

Application-specific fairness in recommender systems.

(3) Provider Fairness.

In a recommender system:

there could be significant differences in **the exposure rate** of items from different providers in a recommendation system [1].



[1] Weiwen Liu et al. Personalizing fairness-aware re-ranking. arXiv preprint arXiv:1809.02921, 2018.

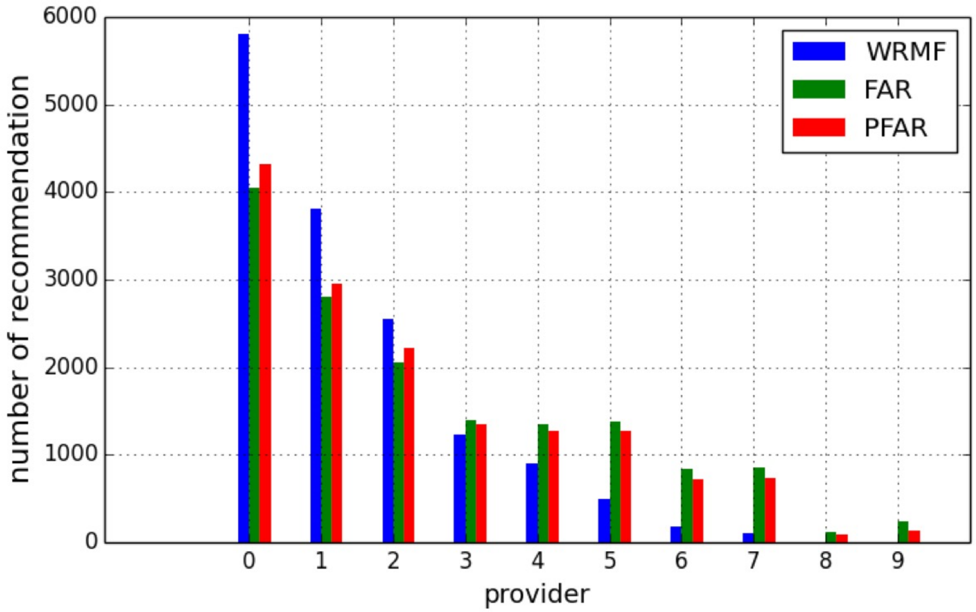
Provider Fairness in Recommendation

Application-specific fairness in recommender systems.

(3) Provider Fairness.

In a recommender system:

there could be significant differences in **the exposure rate** of items from different providers in a recommendation system [2].



Criterion: Items from different providers should receive **the same exposure rate** to the customers [1, 2, 3].

Metrics: (1) number of providers whose corresponding exposure rates are lower than a **threshold** exposure rate [1]; (2) **diversity** of providers for recommended items [2]; (3) item **exposure rate difference** between different providers [3];

[1] Ludovico Boratto, et al. Interplay between upsampling and regularization for provider fairness in recommender systems. In UMUAI, 2020.
[2] Weiwen Liu et al. Personalizing fairness-aware re-ranking. arXiv preprint arXiv:1809.02921, 2018.
[3] Gourab K. Patro, et al. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In WWW, 2020.

Marketing Fairness in Recommendation

Application-specific fairness in recommender systems.

(4) Marketing Fairness. Users' interactions are biased according to the marketing strategies: under certain marketing strategy, **identity-consistent users** interact more with this item [1].

[1] Mengting Wan,Jianmo Ni,Rishabh Misra,and Julian J. McAuley. Addressing marketing bias in product recommendations. In WSDM, 2020.

Marketing Fairness in Recommendation

Application-specific fairness in recommender systems.

(4) Marketing Fairness. Users' interactions are biased according to the marketing strategies: under certain marketing strategy, **identity-consistent users** interact more with this item [1].

Identity Consistent Users: Females

Identity Consistent Users: Males

[1] Mengting Wan,Jianmo Ni,Rishabh Misra,and Julian J. McAuley. Addressing marketing bias in product recommendations. In WSDM, 2020.

Marketing Fairness in Recommendation

Application-specific fairness in recommender systems.

(4) Marketing Fairness. Users' interactions are biased according to the marketing strategies: under certain marketing strategy, **identity-consistent users** interact more with this item [1].

Identity Consistent Users: Females

Identity Consistent Users: Males

Criterion: Recommender systems should not inherit such bias from data and yield biased recommendations [1].

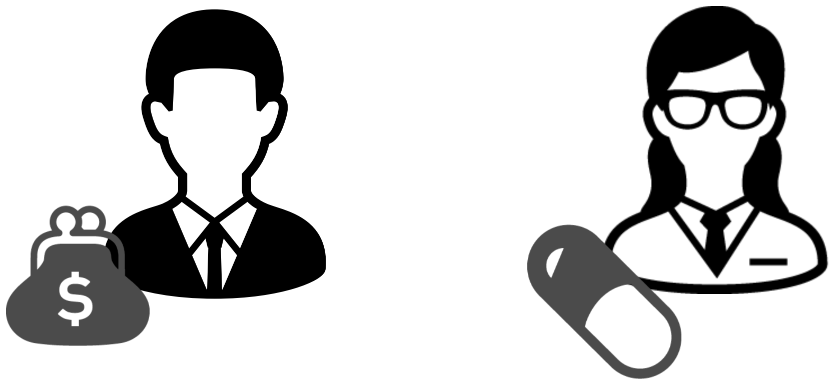
Metric: The difference of the recommendation error variance between identity-consistent and identity-inconsistent users [1].

[1] Mengting Wan,Jianmo Ni,Rishabh Misra,and Julian J. McAuley. Addressing marketing bias in product recommendations. In WSDM, 2020.

Social Fairness in Knowledge Graphs

Application-specific fairness in knowledge graphs.

(1) Social Fairness.



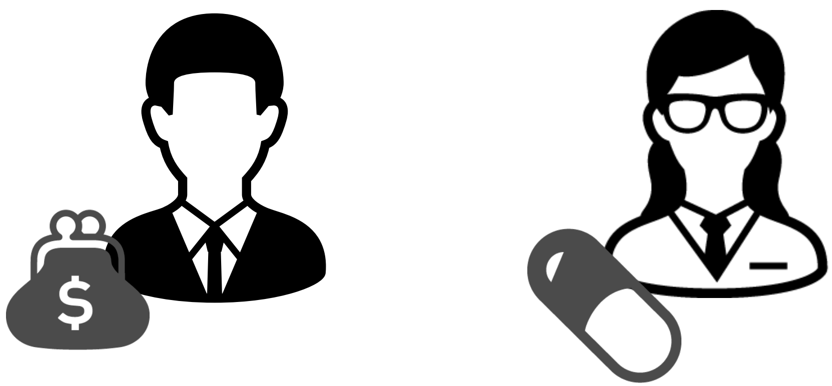
A traditional stereotype: bankers are males, while nurses are females [1].

[1] Ziqian Zeng, Rashidul Islam, et al. Fair representation learning for heterogeneous information networks. In AAAI, 2021.

Social Fairness in Knowledge Graphs

Application-specific fairness in knowledge graphs.

(1) Social Fairness.



A traditional stereotype: bankers are males, while nurses are females [1].

Criterion: The **historical biases** should not be encoded in the learned entity embeddings in knowledge graphs [1].
Metric: Measured with the predicted probability change on stereotype-related labels when the predicted probability on a certain gender changes under perturbation [1].

[1] Ziqian Zeng, Rashidul Islam, et al. Fair representation learning for heterogeneous information networks. In AAAI, 2021.

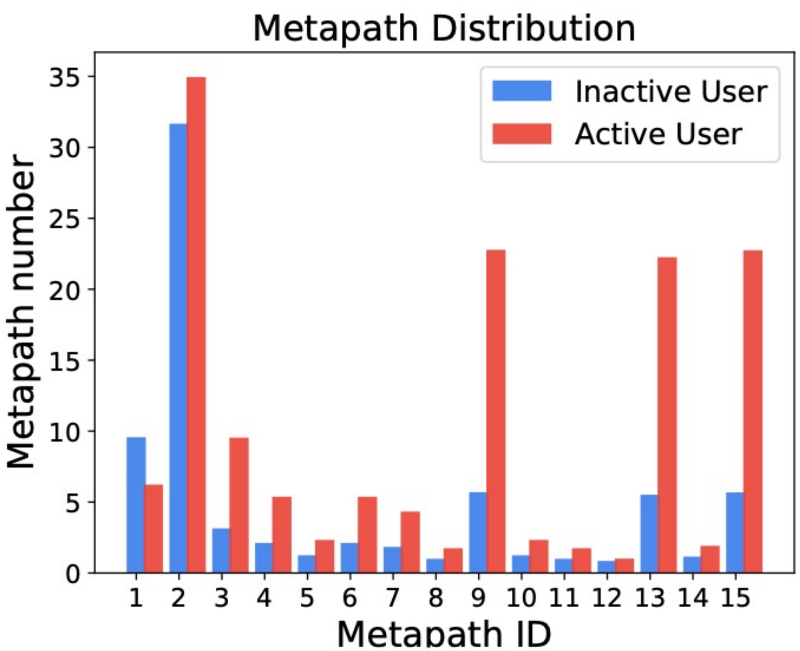
Path Diversity Fairness in Knowledge Graphs

Application-specific fairness in knowledge graphs.

(2) Path Diversity Fairness.

On a **user-item knowledge graph**:

Meta-path distributions over their types can be different across different person entity groups [1].



[1] Zuohui Fu, et al. Fairness-aware explainable recommendation over knowledge graphs. In SIGIR, 2020.

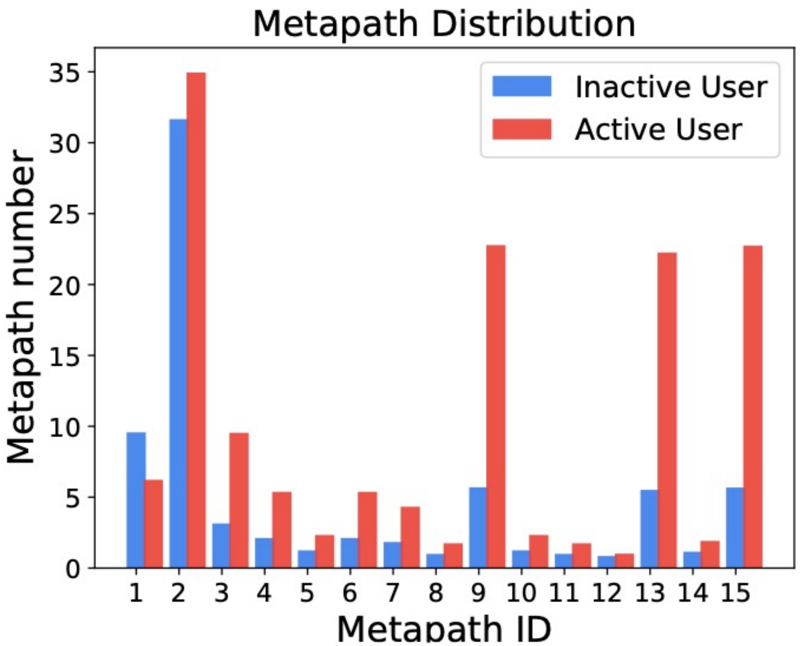
Path Diversity Fairness in Knowledge Graphs

Application-specific fairness in knowledge graphs.

(2) Path Diversity Fairness.

On a **user-item knowledge graph**:

Meta-path distributions over their types can be different across different person entity groups [1].



Criterion: The distributions of meta-paths (over their types) should be similar across different demographic subgroups in the knowledge graph [1].
Metric: The difference of Simpson's Index of Diversity (SID) between the meta-path distributions of different demographic subgroups [1].

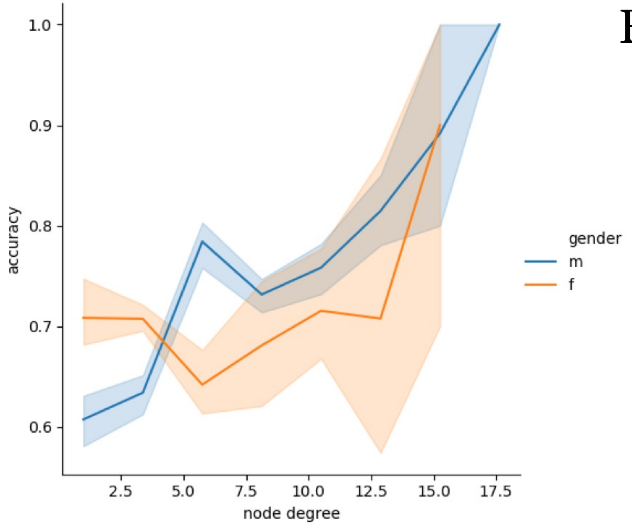
[1] Zuohui Fu, et al. Fairness-aware explainable recommendation over knowledge graphs. In SIGIR, 2020.

Popularity Fairness in Knowledge Graphs

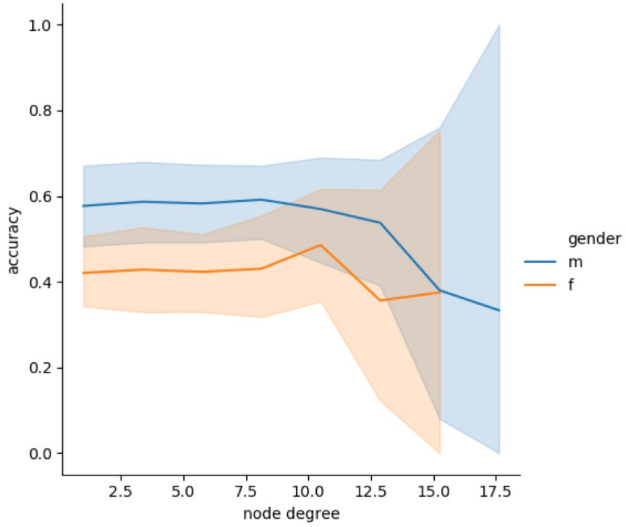
Application-specific fairness in knowledge graphs.

(3) Popularity Fairness. **Prediction for person entities** based on DBpedia.

More biased [1]:



Fairer [1]:

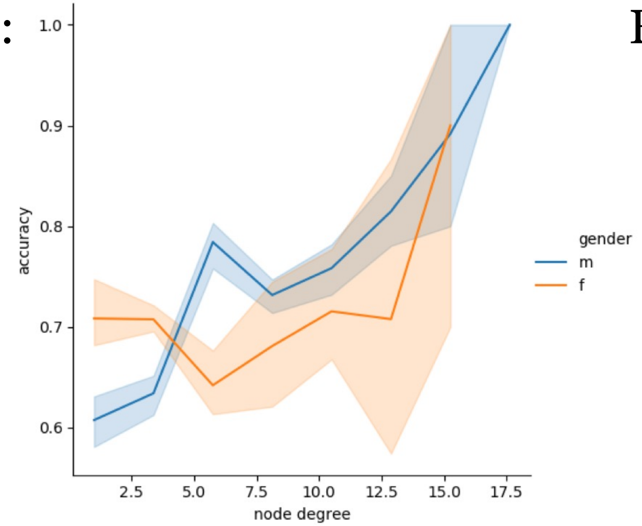


Popularity Fairness in Knowledge Graphs

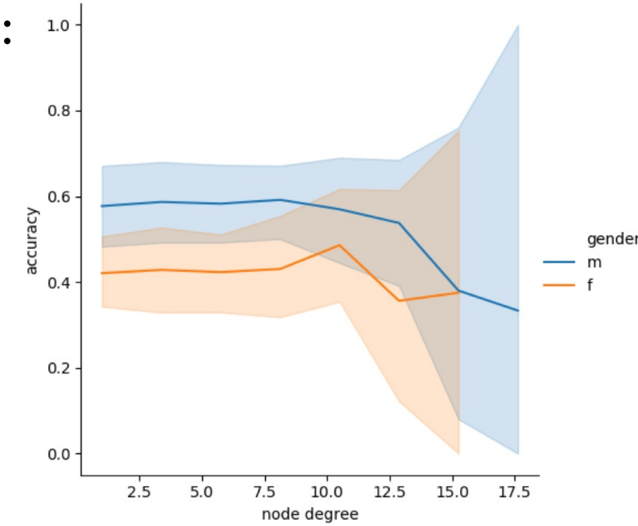
Application-specific fairness in knowledge graphs.

(3) Popularity Fairness. **Prediction for person entities** based on DBpedia.

More biased [1]:



Fairer [1]:



Criterion: The prediction accuracy under certain tasks should be uniformly distributed w.r.t. entity node popularity (e.g., defined as the entity node degree) in the knowledge graph [1].
Metric: Difference between the output distribution of accuracy w.r.t. entity popularity and a uniform distribution [1].

[1] Mario Arduini, et al. Adversarial learning for debiasing knowledge graph embeddings. In SIGKDD, 2020.

Coffee Break

VIRGINIA DATA SCIENCE AND ANALYTICS (VAST) LAB

We will be back in 10 mins.

Outline

Background Information

Fairness Notions and Metrics

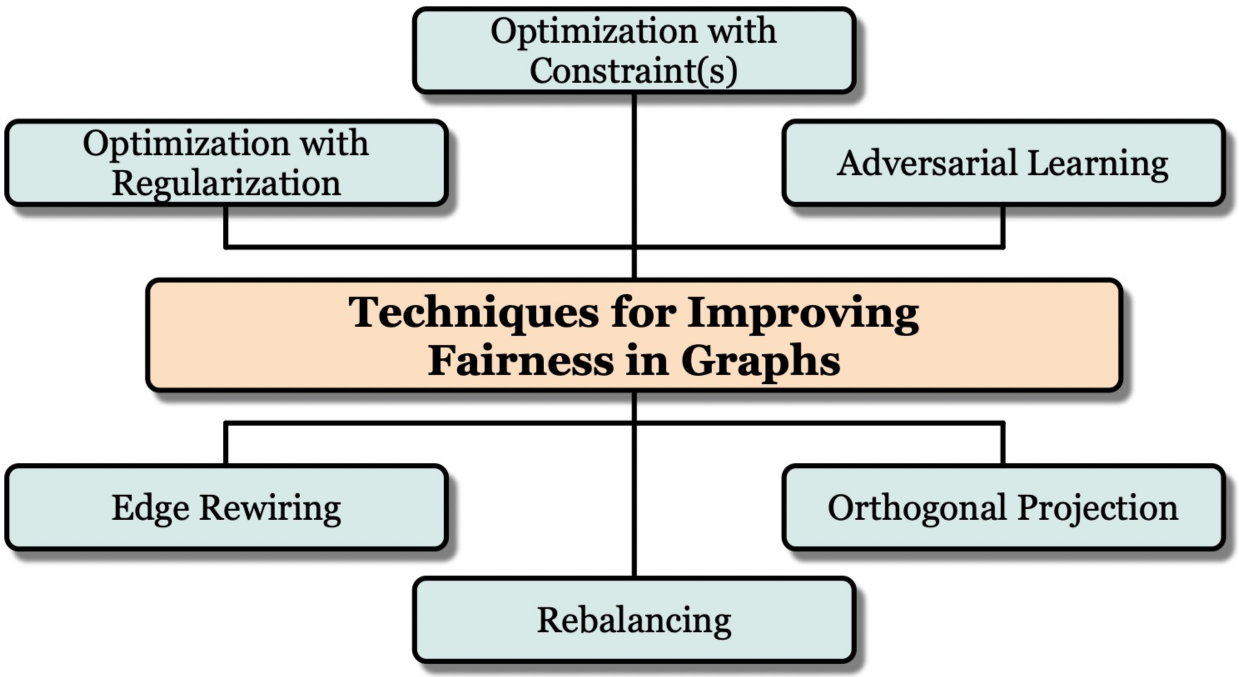
Methodologies to Mitigate Bias

Real-World Applications

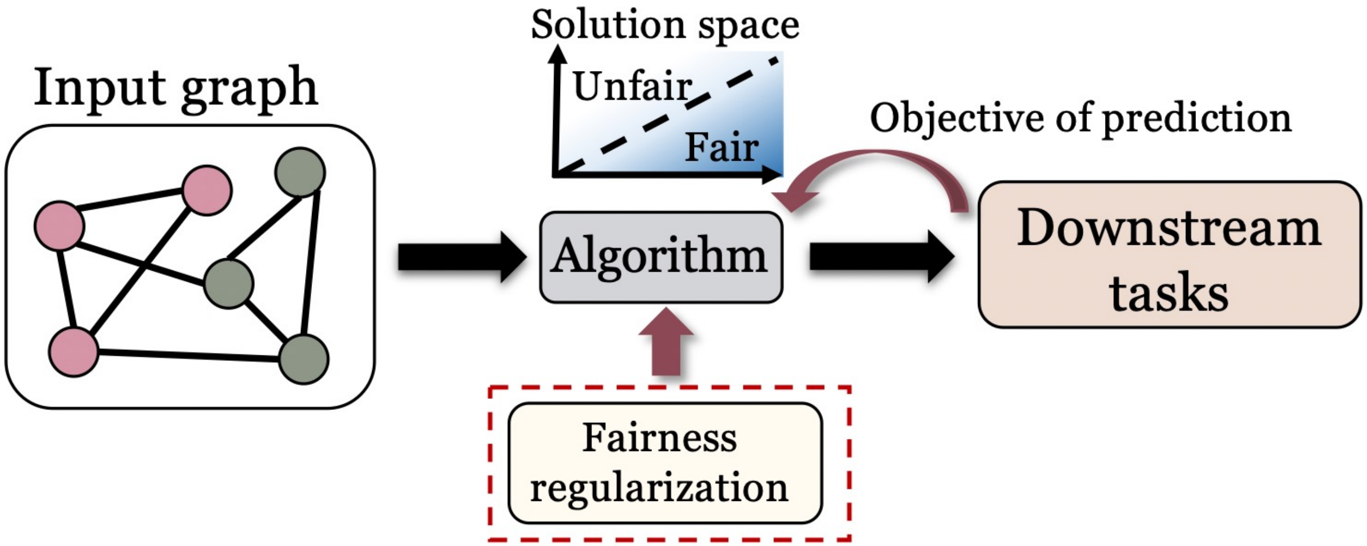
Summary & Existing Challenges

Methodologies to Mitigate Bias

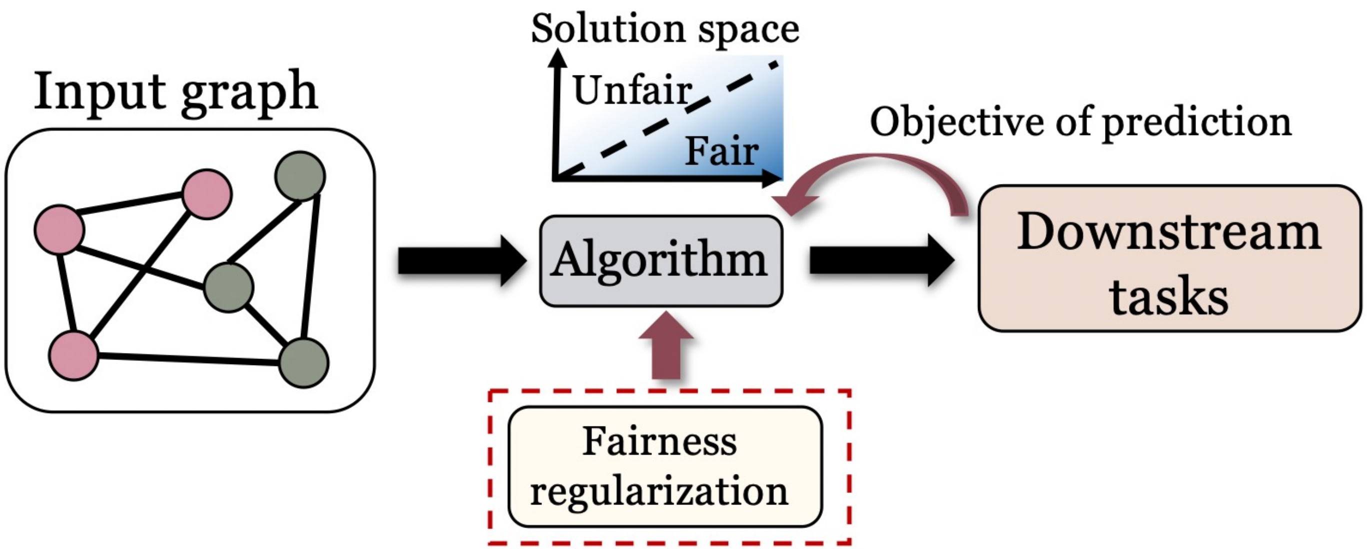
- In general, there are six main categories of commonly used techniques to improve fairness in graph mining.



Optimization with Regularization



Optimization with Regularization

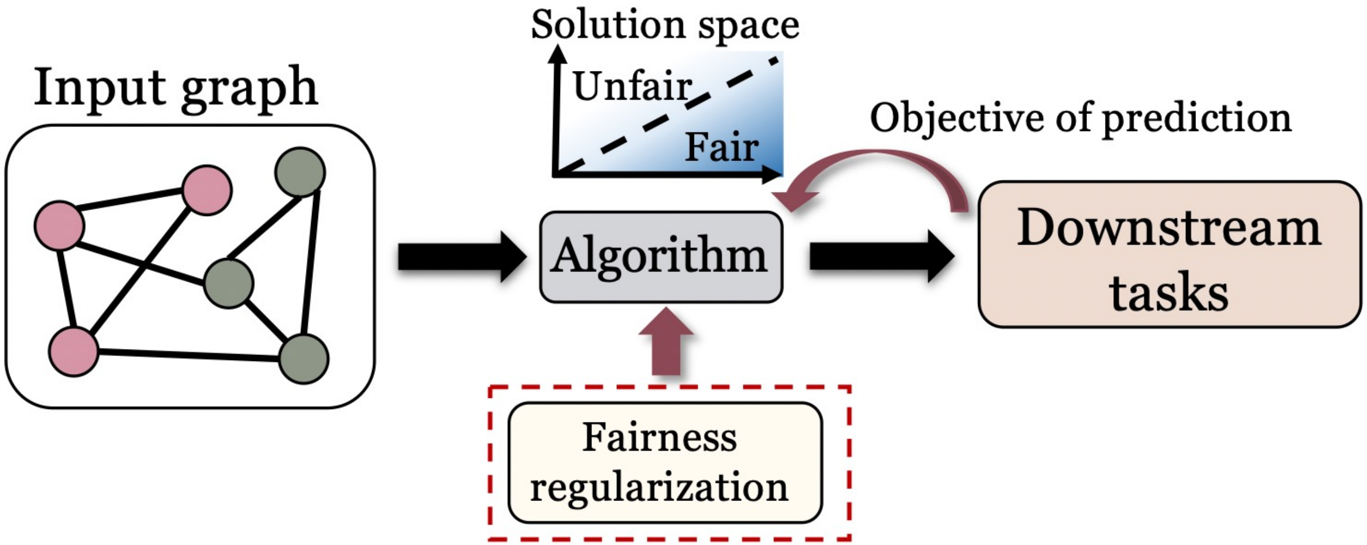


- Improving Group Fairness

$$\mathcal{L} = \mathcal{L}_{\text{utility}} + \lambda \mathcal{L}_{\text{fair}}$$

- Algorithm Output-Based Regularization.
- Network Topology-Based Regularization.
- Node Embedding-Based Regularization.

Optimization with Regularization



- Improving Group Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

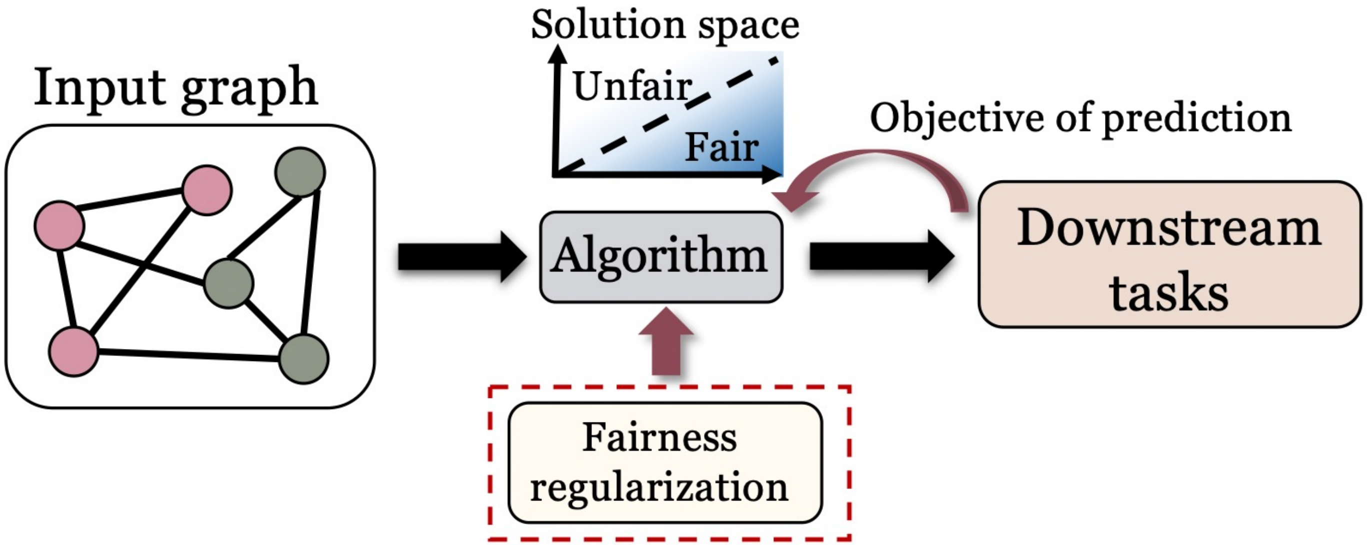
- Algorithm Output-Based Regularization.
- Network Topology-Based Regularization.
- Node Embedding-Based Regularization.

An example [1]:

$$\mathscr{L}_{sp} = \sum_{j=1}^{c} \left(\frac{\sum_{v_i \in \mathcal{V}_0} P(\hat{Y} = j \mid v_i)}{|\mathcal{V}_0|} - \frac{\sum_{v_i \in \mathcal{V}_1} P(\hat{Y} = j \mid v_i)}{|\mathcal{V}_1|} \right)^2$$

[1] Ziqian Zeng, et al. Fair representation learning for heterogeneous information networks. In AAAI, 2021.

Optimization with Regularization



- Improving Group Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

Algorithm Output-Based Regularization.
Network Topology-Based Regularization.
Node Embedding-Based Regularization.

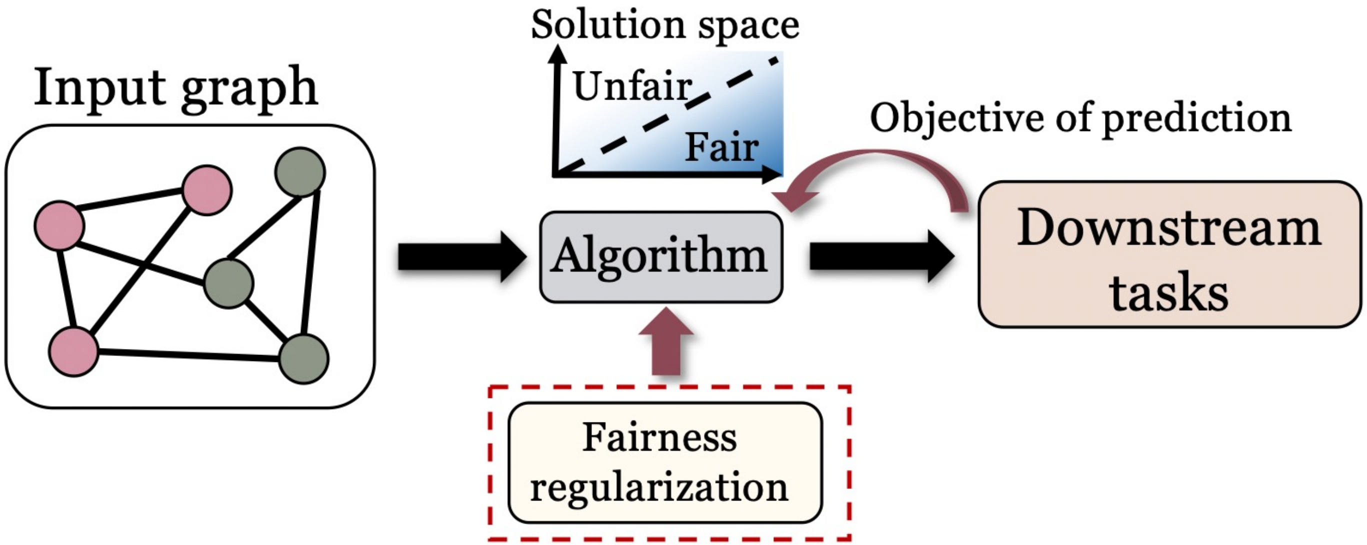
An example [1]:

$$\mathscr{L}_{sp} = \sum_{j=1}^{c} \left(\frac{\sum_{v_i \in \mathcal{V}_0} P(\hat{Y} = j \mid v_i)}{\boxed{|\mathcal{V}_0|}} - \frac{\sum_{v_i \in \mathcal{V}_1} P(\hat{Y} = j \mid v_i)}{\boxed{|\mathcal{V}_1|}} \right)^2$$

node sets for the two sensitive subgroup (S=0 and S=1)

[1] Ziqian Zeng, et al. Fair representation learning for heterogeneous information networks. In AAAI, 2021.

Optimization with Regularization



- Improving Group Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

Algorithm Output-Based Regularization.

Network Topology-Based Regularization.

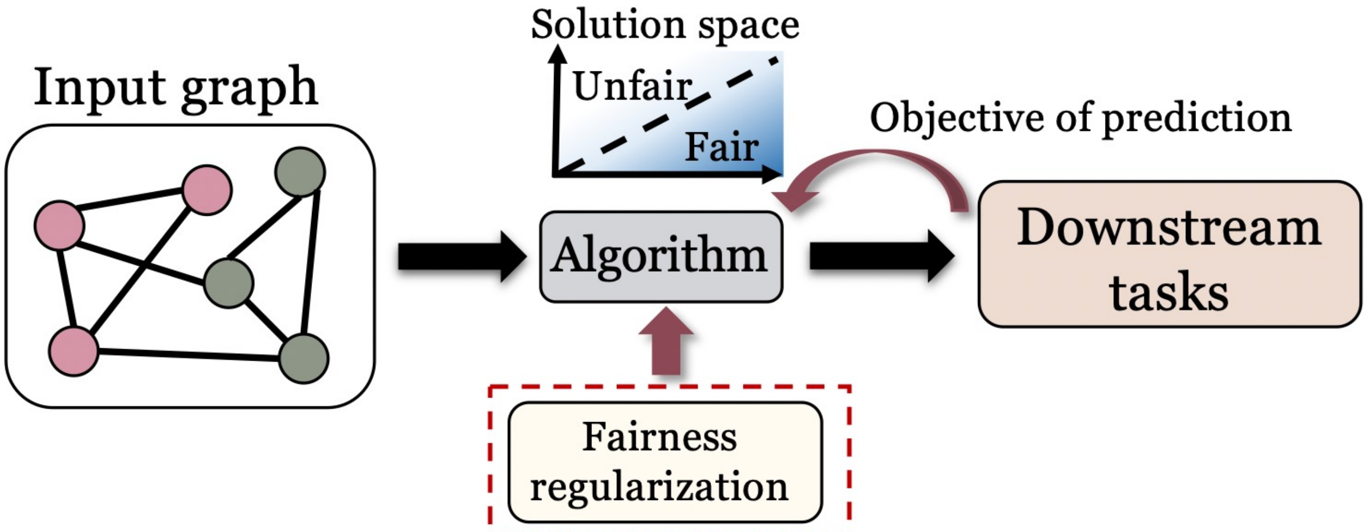
Node Embedding-Based Regularization.

An example [1]:

$$\mathscr{L}_{\text{fair}} = \|\boldsymbol{\Delta}_s \text{softmax}(\hat{\mathbf{X}})\|_1 \qquad \boldsymbol{\Delta}_s = \frac{\mathbb{1}_{=1}(\mathbf{s})}{\|\mathbb{1}_{=1}(\mathbf{s})\|_1} - \frac{\mathbb{1}_{=0}(\mathbf{s})}{\|\mathbb{1}_{=0}(\mathbf{s})\|_1}$$

[1] Zhimeng Jiang, et al. Fmp: Toward fair graph message passing against topology bias. arXiv 2022.

Optimization with Regularization



- Improving Group Fairness

$$\mathcal{L} = \mathcal{L}_{\text{utility}} + \lambda \mathcal{L}_{\text{fair}}$$

- Algorithm Output-Based Regularization.
- Network Topology-Based Regularization.
- Node Embedding-Based Regularization.

An example [1]:

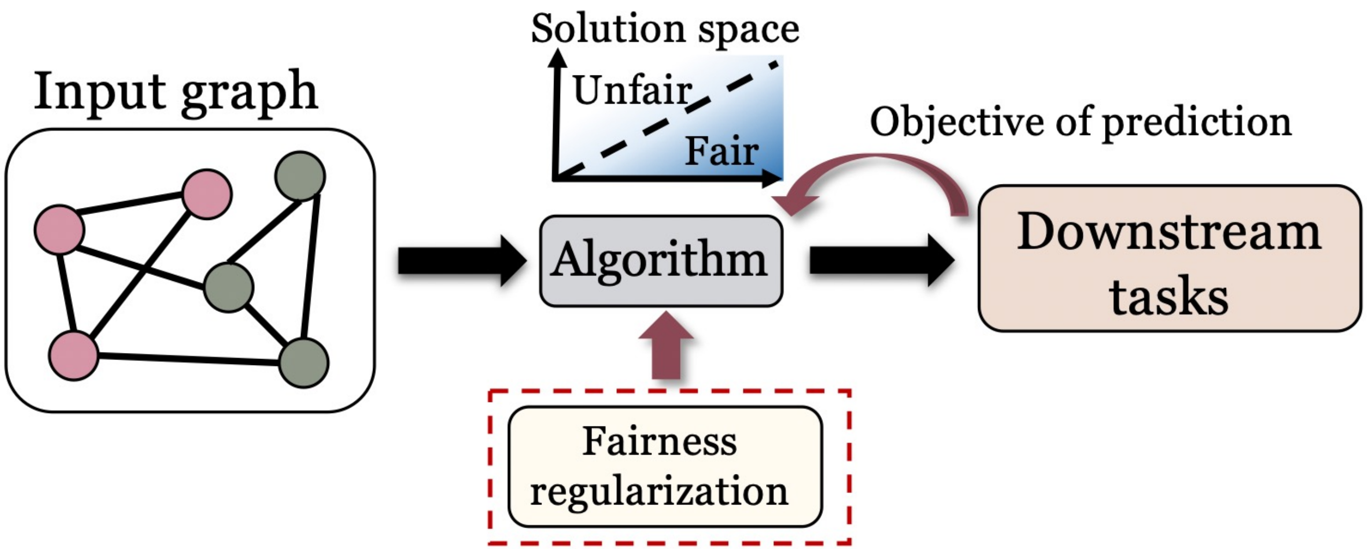
node feature matrix after propagation

$$\mathcal{L}_{\text{fair}} = \| \boxed{\boldsymbol{\Delta}_s} \text{softmax}\boxed{(\hat{\mathbf{X}})} \|_1 \qquad \boldsymbol{\Delta}_s = \frac{\mathbb{1}_{=1}(\mathbf{s})}{\|\mathbb{1}_{=1}(\mathbf{s})\|_1} - \frac{\mathbb{1}_{=0}(\mathbf{s})}{\|\mathbb{1}_{=0}(\mathbf{s})\|_1}$$

$\{1, -1\}^{1\times n}$ sensitive subgroup membership normalized by subgroup size

[1] Zhimeng Jiang, et al. Fmp: Toward fair graph message passing against topology bias. arXiv 2022.

Optimization with Regularization



- Improving Group Fairness

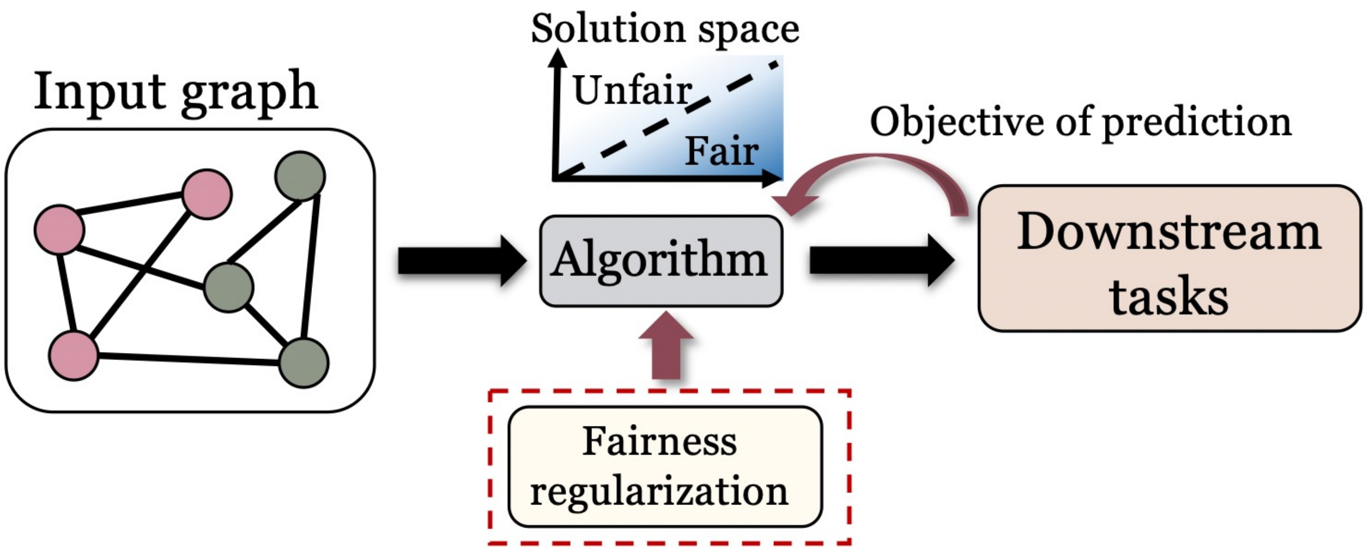
$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

- Algorithm Output-Based Regularization.
- Network Topology-Based Regularization.
- Node Embedding-Based Regularization.

E.g., The total Euclidean distance of all embedding pairs spanning across different sensitive subgroups [1].

[1] Preethi Lahoti, et al. Operationalizing Individual fairness with pairwise fair representations. VLDB, 2019

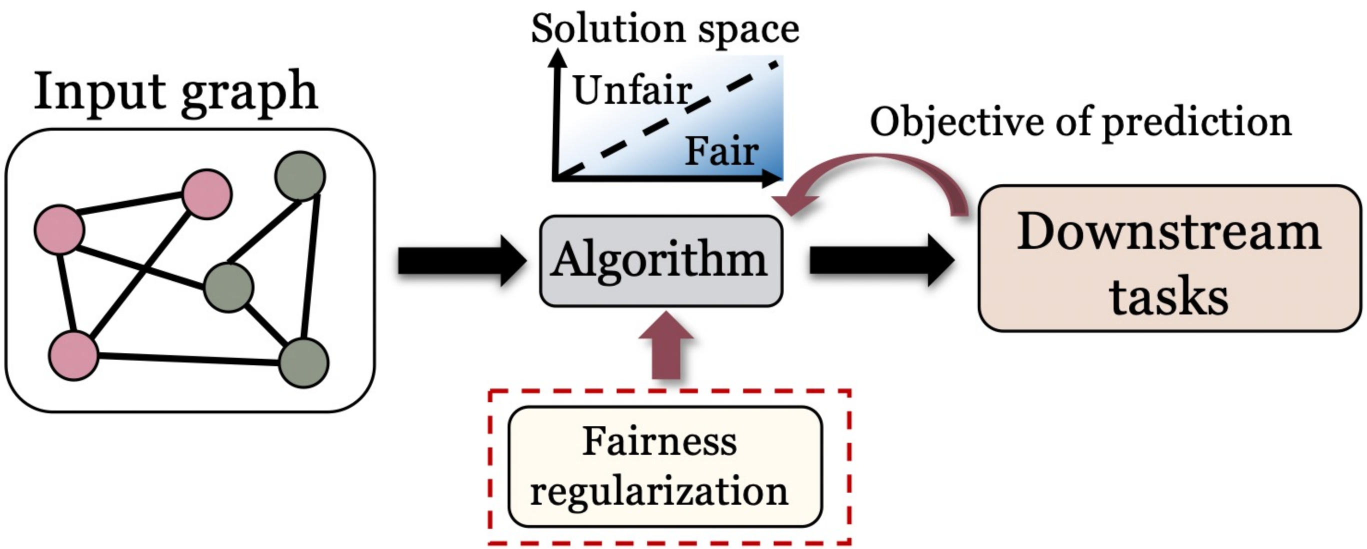
Optimization with Regularization



- Improving Individual Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}} \begin{cases} \text{Output Logits-Based Regularization.} \\ \text{Node Embedding-Based Regularization.} \end{cases}$$

Optimization with Regularization



- Improving Individual Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}} \begin{cases} \text{Output Logits-Based Regularization.} \\ \text{Node Embedding-Based Regularization.} \end{cases}$$

An example [1]:

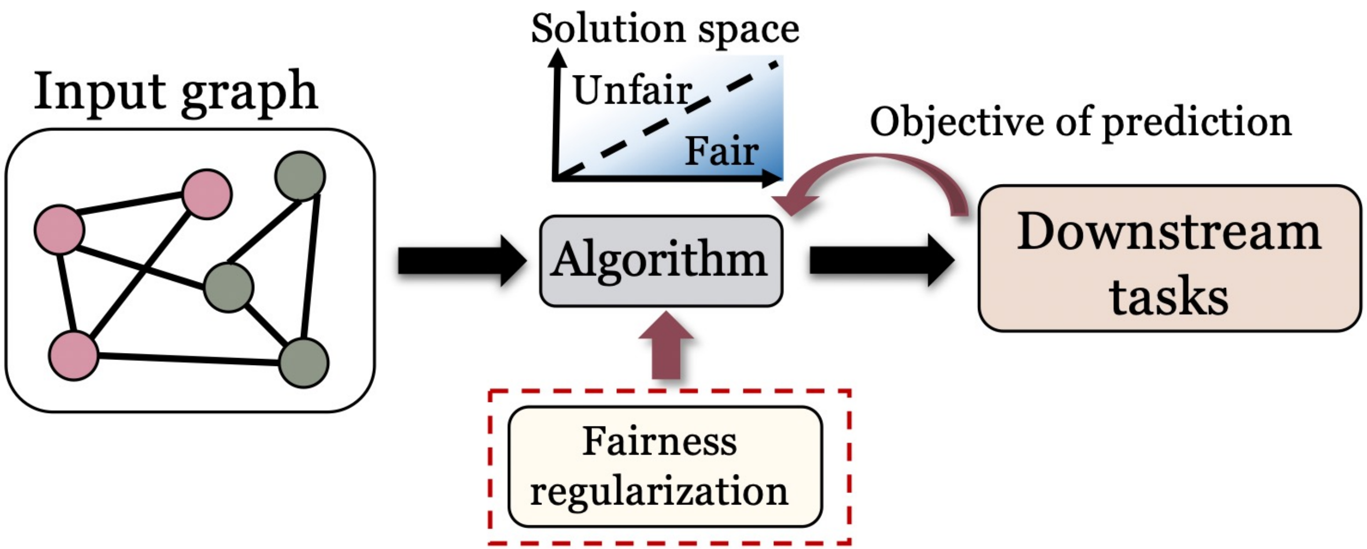
$$\mathrm{Tr}(\hat{\mathbf{Y}}^{\top} \boxed{\mathbf{L_S}} \hat{\mathbf{Y}})$$

Laplacian matrix for the similarity matrix

With such a regularization, the algorithm yields similar outputs for similar nodes

[1] Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong. Inform: Individual fairness on graph mining. In SIGKDD, 2020.

Optimization with Regularization

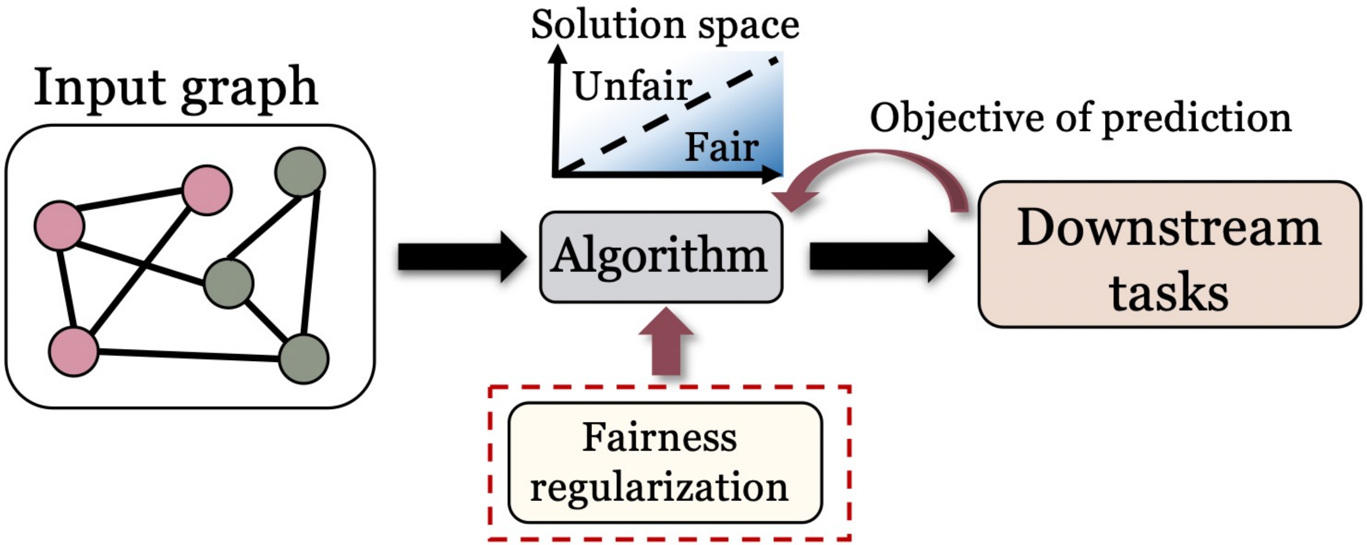


- Improving Individual Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}} \begin{cases} \text{Output Logits-Based Regularization.} \\ \text{Node Embedding-Based Regularization.} \end{cases}$$

Promoting the level of group fairness based on node embedding distributions also helps to impose individual fairness

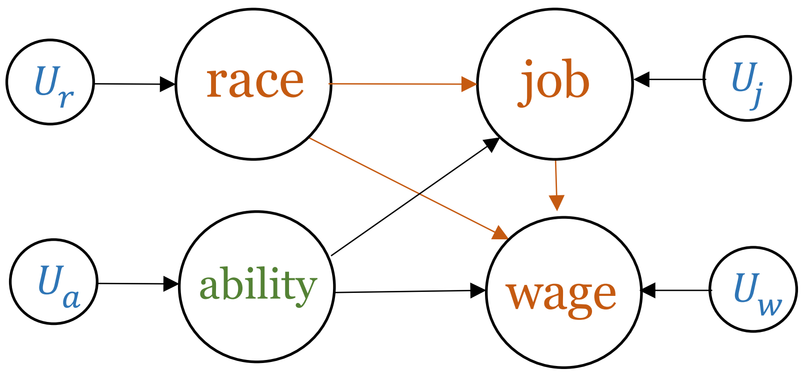
Optimization with Regularization



- Improving Counterfactual Fairness

Background: Causal Model

- Structural causal model [1]
  - Independent exogenous variables (U)
  - Endogenous variables
  - Causal graph (a Directed Acyclic Graph) & structural equations (functions which describe the relations between variables)

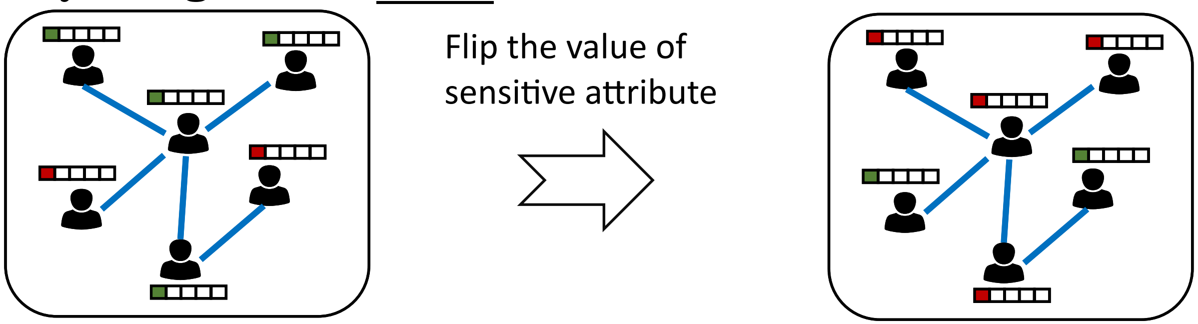


Biased information

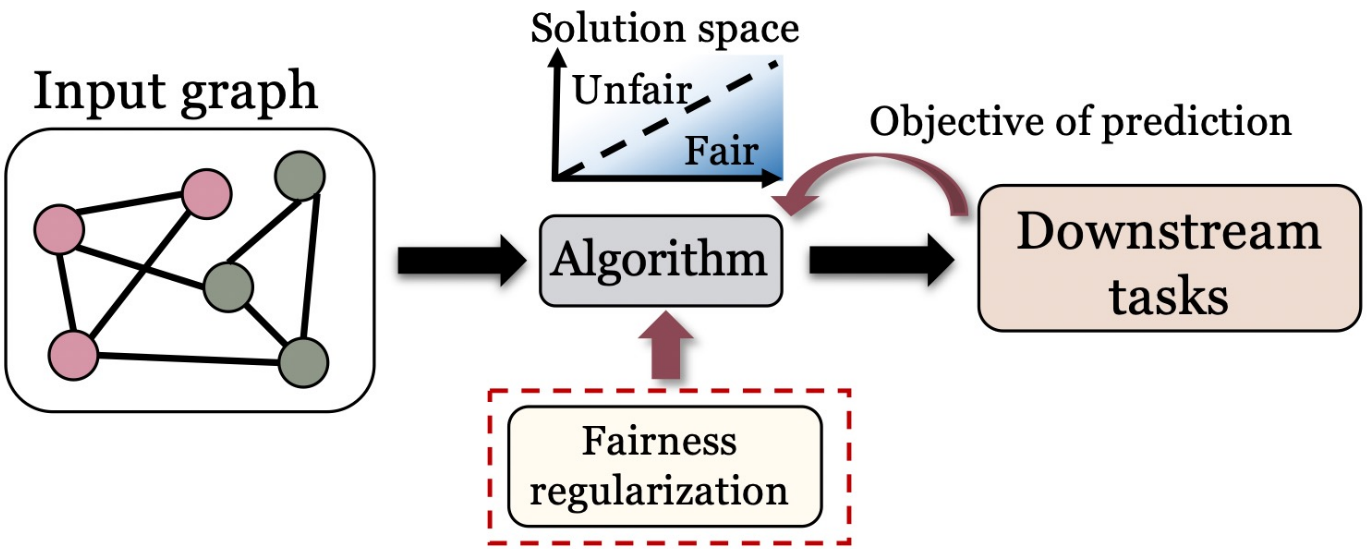
[1] Pearl J. Causality[M]. Cambridge university press, 2009.

Counterfactual Fairness on Graphs

- A few works extend counterfactual fairness on graphs:
  - The node representations for each node i should be the same after setting sensitive attribute S_i as different values, while everything else is fixed.



Optimization with Regularization



- Improving Counterfactual Fairness

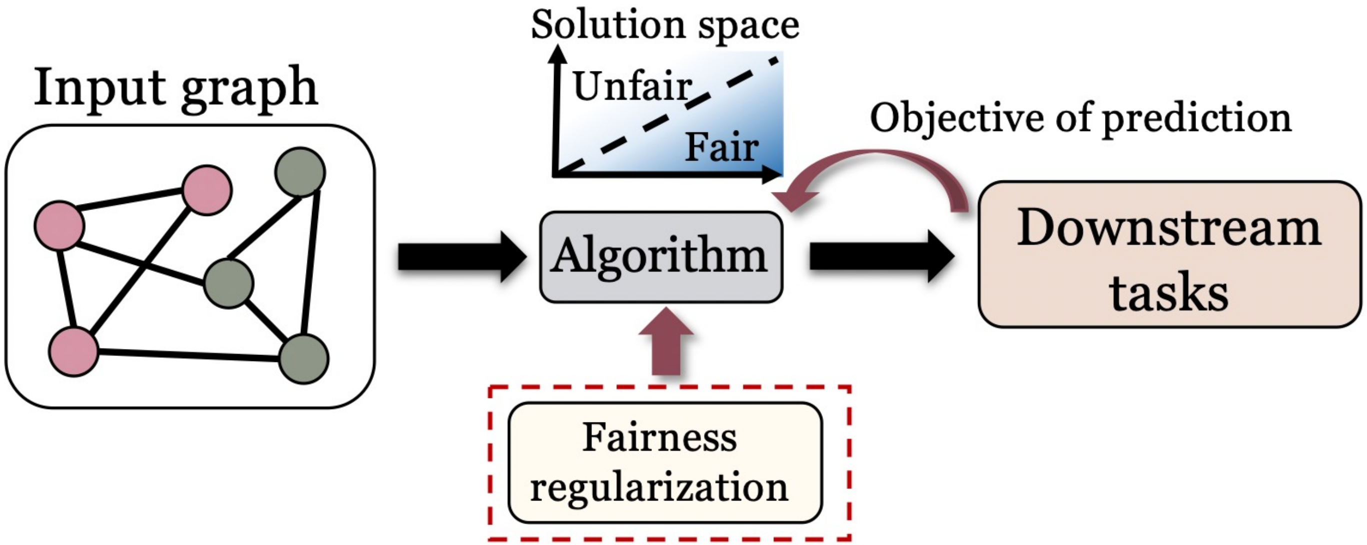
$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

Node Embedding-Based Regularization [1].

$$E[D(\mathbf{z}_i, \mathbf{z}_i')]$$

[1] Chirag Agarwal, et al. Towards a unified framework for fair and stable graph representation learning. UAI, 2021.

Optimization with Regularization



- Improving Counterfactual Fairness

$$\mathscr{L} = \mathscr{L}_{\text{utility}} + \lambda \mathscr{L}_{\text{fair}}$$

Node Embedding-Based Regularization [1].

$$E[D(\boxed{\mathbf{z}_i}, \boxed{\mathbf{z}_i'})]$$

embeddings of node i learned based on the **factual** graph

embeddings of node i learned based on the **counterfactual** graph

[1] Chirag Agarwal, et al. Towards a unified framework for fair and stable graph representation learning. UAI, 2021.

Counterfactual Fairness

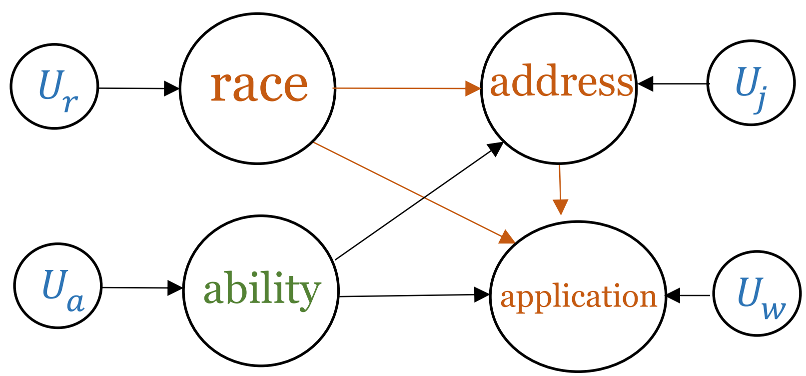
- Prediction $\hat{Y}$ is **counterfactually fair** if under any features $X = x$ and sensitive attribute $S = s$:

$$P(\boxed{\widehat{Y}_{S \leftarrow s}} = y | X = x, S = s) = P(\boxed{\widehat{Y}_{S \leftarrow s'}} = y | \boxed{X = x}, \boxed{S = s})$$

The value of the prediction if S had been set to s (s')
Notice: other features may change correspondingly.

Features

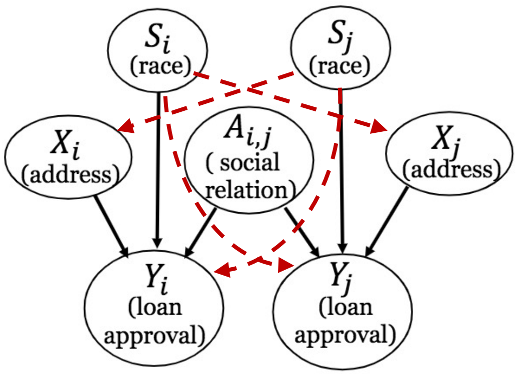
Sensitive attribute

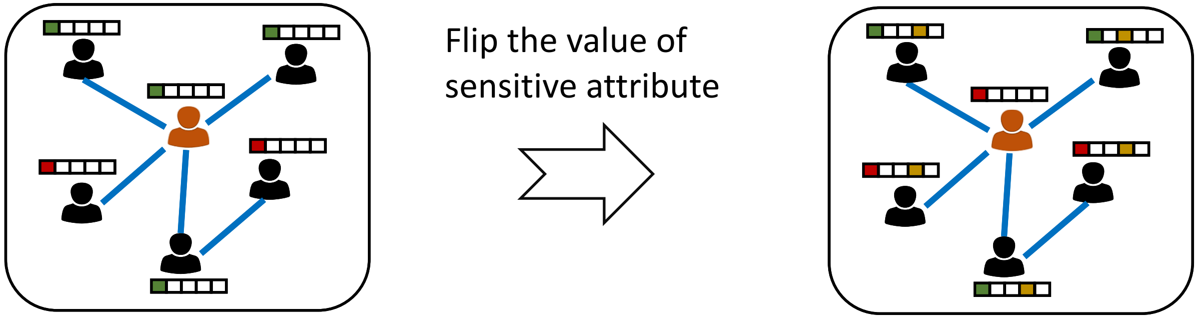


Descendants of the sensitive attribute will be also changed after intervention

Counterfactual Fairness on Graphs

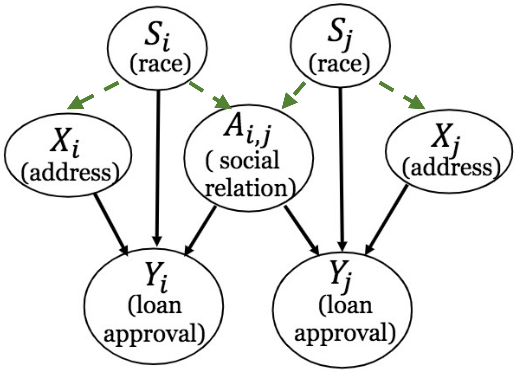
- Limitations of the above fairness notion:
  - In graphs, the sensitive attributes of each node's neighbors may causally affect the prediction w.r.t. this node (red dashed edges);

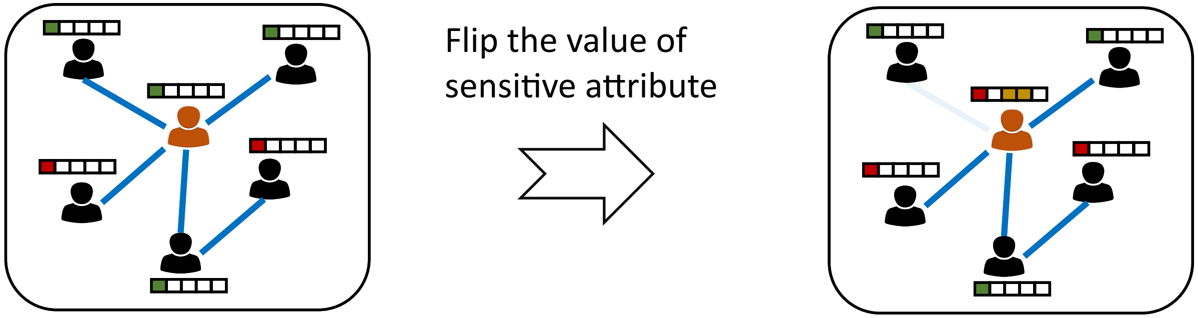




Counterfactual Fairness on Graphs

- Limitations of the above fairness notion:
  - The sensitive attributes may causally affect other features and the graph structure (green dashed edges).





Graph Counterfactual Fairness

- **Graph counterfactual fairness** [1]: An encoder $Z_i = (\Phi(X, A))_i$ satisfies graph counterfactual fairness if for any node i:

$$P(\boxed{(Z_i)_{S \leftarrow \mathbf{s}'}} | X = \mathbf{X}, A = \mathbf{A}) = P(\boxed{(Z_i)_{S \leftarrow \mathbf{s}''}} | \boxed{X = \mathbf{X},} \boxed{A = \mathbf{A}}),$$

The node representation of i when the values of the sensitive attributes of all nodes on the graph are set to s' (s'')

- s' (s'') : a n-dimensional vector for a n-node graph

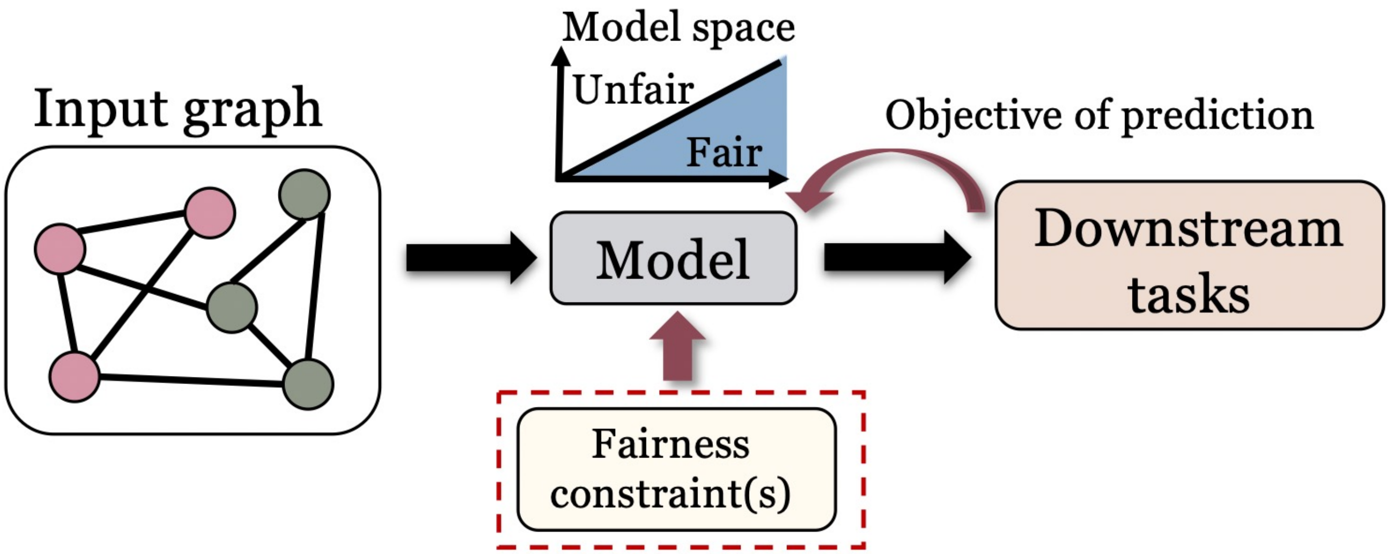
Node features (including sensitive attribute)

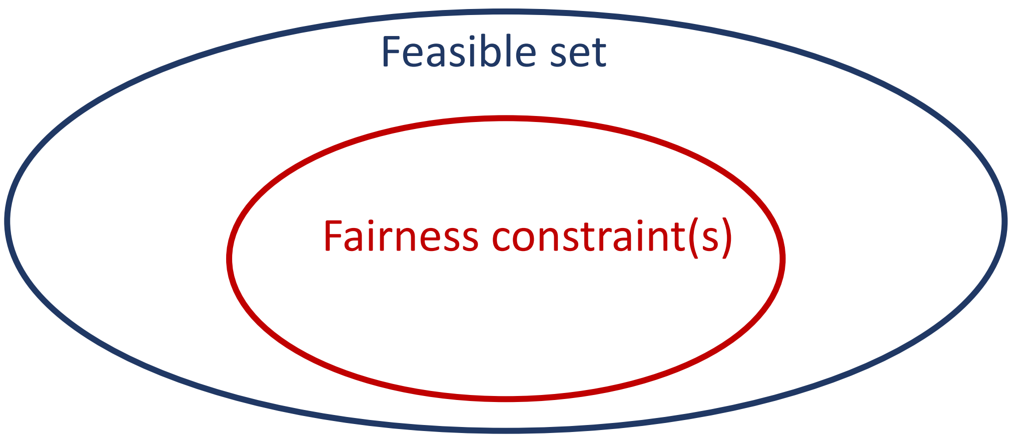
Graph structure

- Example: the prediction for one's loan application being approved should be the same regardless this applicant's <u>and his/her friends' (connected in a social network)</u> race information.

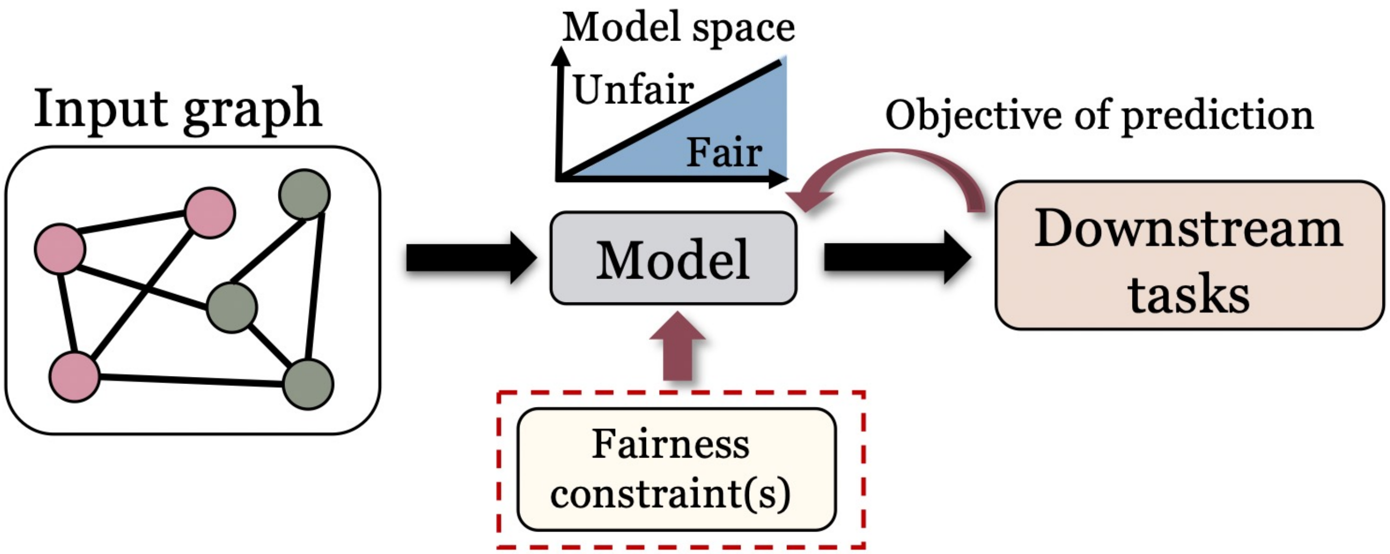
[1] Jing Ma, et al. Learning fair node representations with graph counterfactual fairness. In WSDM,, 2022.

Optimization with Constraints





Optimization with Constraints

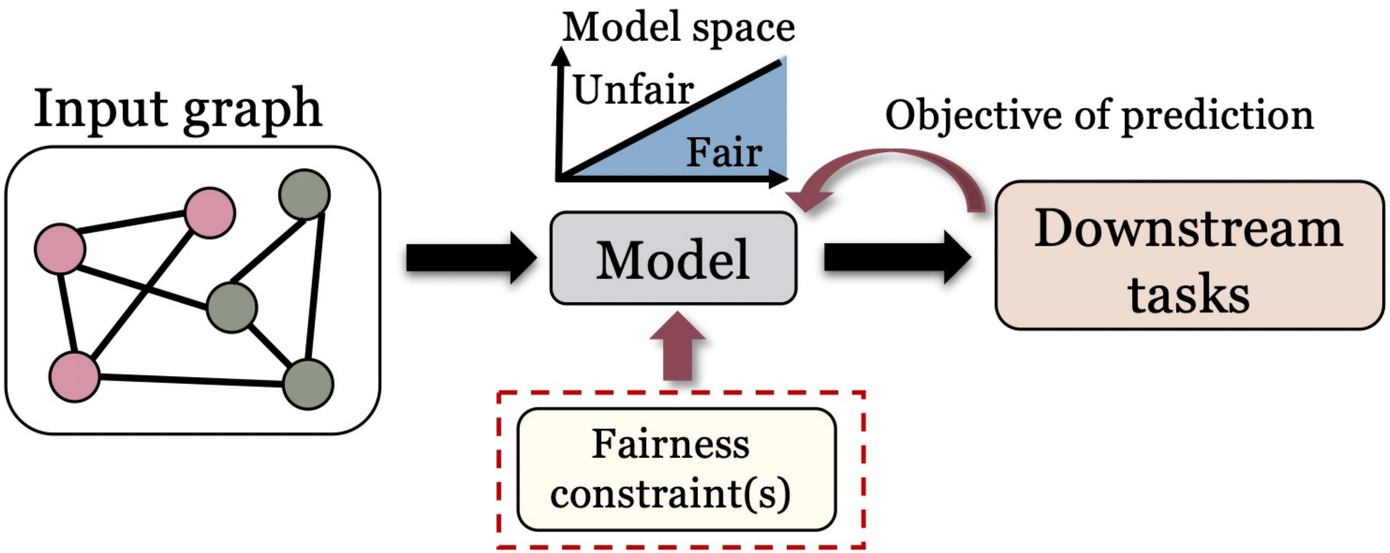


Adding a fairness-aware constraint on the optimization problem.

$$\begin{aligned} \min \quad & \mathscr{L}_{\text{utility}}, \\ \text{subject to} \quad & \text{certain fairness constraint(s)} \end{aligned}$$

Most existing works formulate such a constraint with the **performance difference** on different demographic subgroups.

Optimization with Constraints



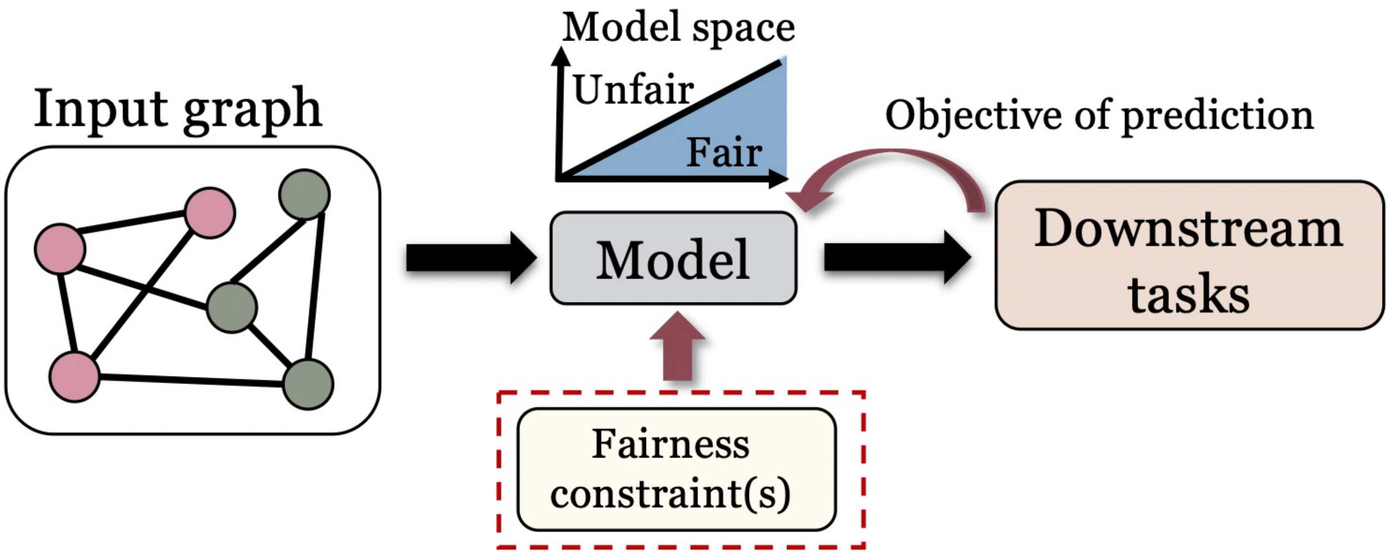
- Improving Group Fairness
  - Influence maximization

$$\max \quad \underbrace{\sum_{i=1}^{H} u_i(\mathcal{A})}_{\text{Expected number of influenced nodes}}$$

$$\text{subject to} \quad \underbrace{|\mathcal{A}| \leq K}_{\text{Bound of seed set size}}$$

$$\text{and} \quad \underbrace{M(\mathcal{A}, u_1, \ldots, u_H, \mathcal{V}_1, \ldots, \mathcal{V}_H) \leq \beta_{\text{unfair}}}_{\text{Fairness constraint}}$$

Optimization with Constraints



- Improving Group Fairness
  - Influence maximization

$$\max \quad \underbrace{\sum_{i=1}^{H} \boxed{u_i(\mathcal{A})}}_{\text{Expected number of influenced nodes}}$$

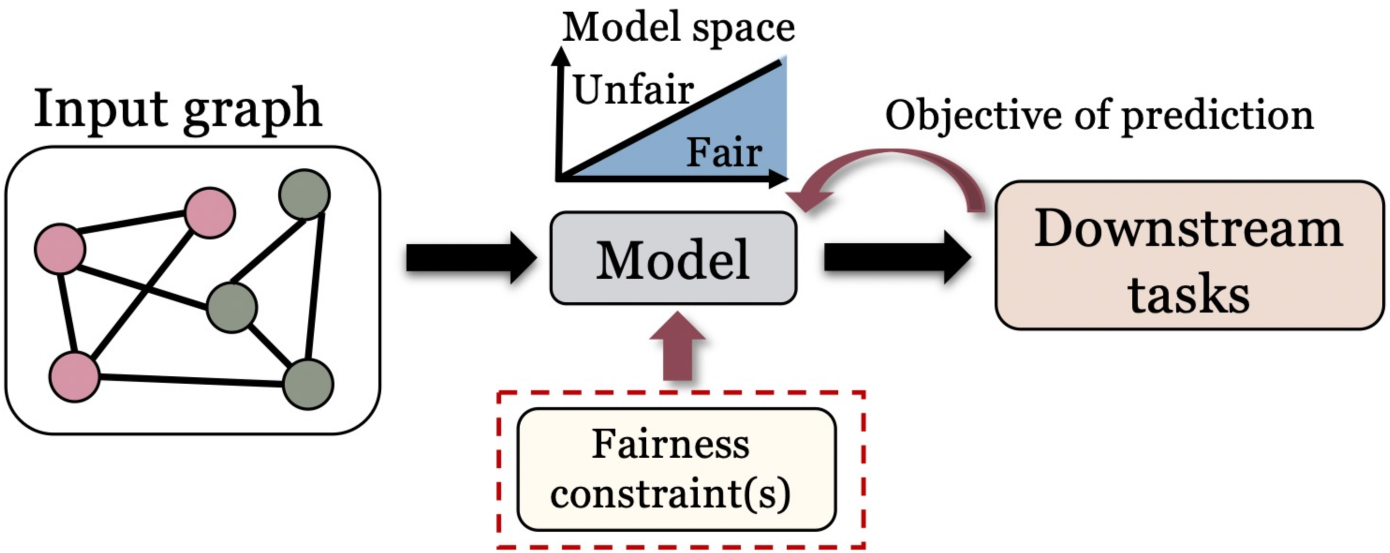
the utility (i.e., the percentage of influenced nodes) of the i-th sensitive subgroup based on the seed node set

$$\text{subject to} \quad \underbrace{|\mathcal{A}| \leq K}_{\text{Bound of seed set size}}$$

$$\text{and} \quad \underbrace{\boxed{M(\mathcal{A}, u_1, \ldots, u_H, \mathcal{V}_1, \ldots, \mathcal{V}_H)} \leq \beta_{\text{unfair}}}_{\text{Fairness constraint}}$$

M: the unfairness level of the influence maximization

Optimization with Constraints



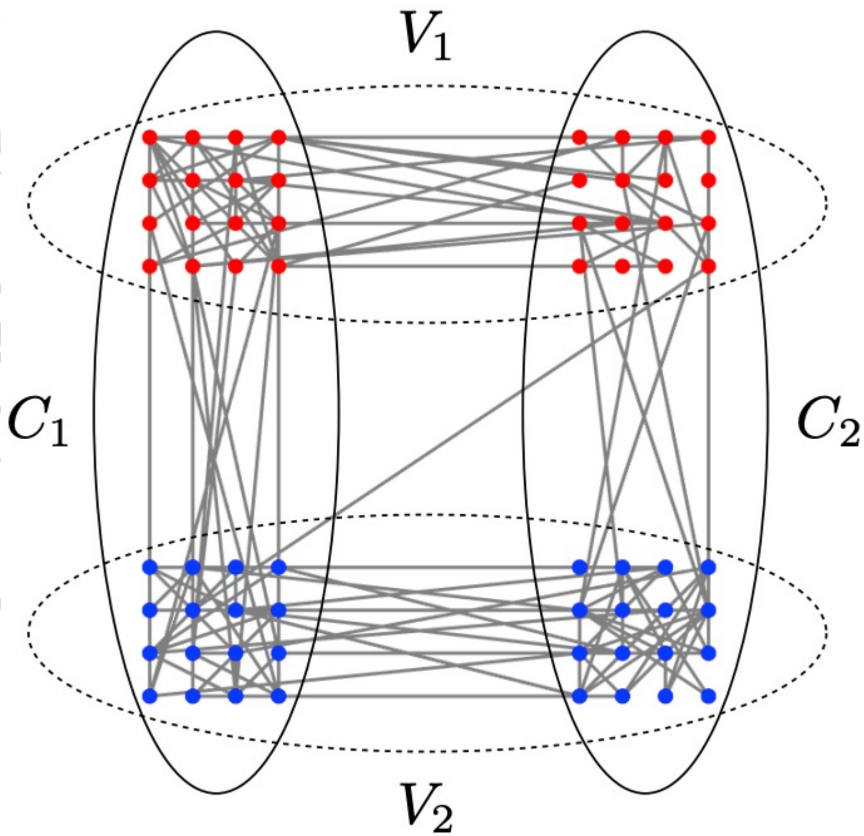
- Improving Group Fairness
    - Graph clustering

Optimization with Constraints

Input gra

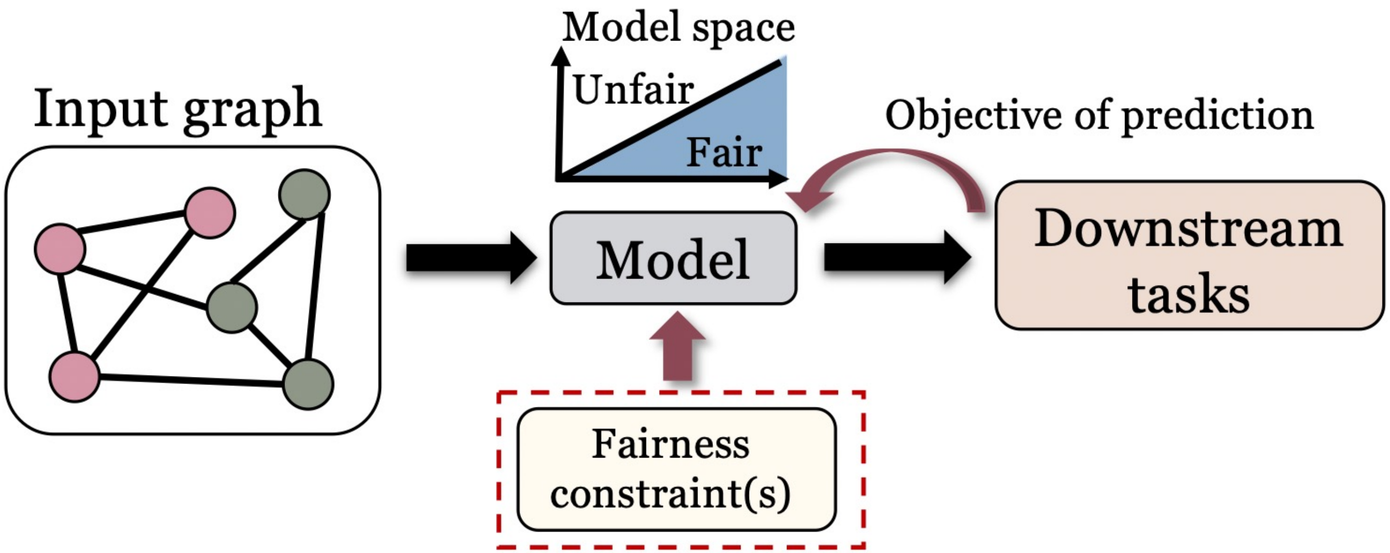
prediction

vnstream tasks



- Improving
  - Graph

Optimization with Constraints

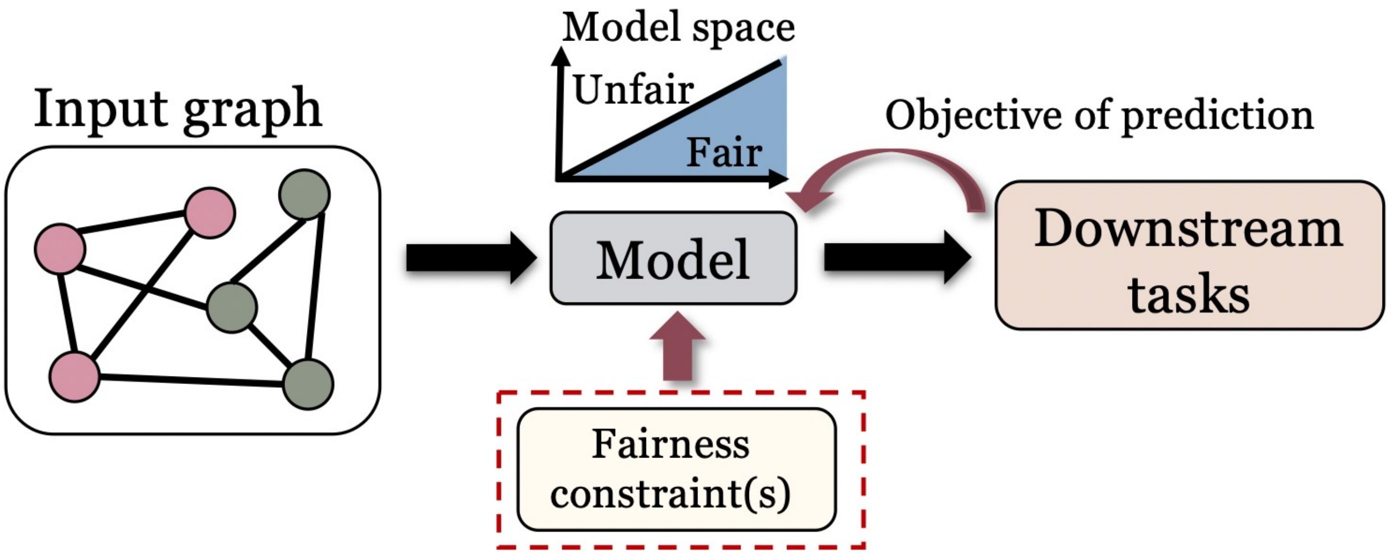


- Improving Group Fairness
  - Graph clustering [1]

$$\forall i \in \{1, \ldots, H\} \text{ and } \forall k \in \{1, \ldots, K\}, \frac{|\mathcal{V}_i \cap \mathcal{C}_k|}{|\mathcal{C}_k|} = \frac{|\mathcal{V}_i|}{|\mathcal{V}|}$$

[1] Matthaus Kleindessner, et al. Guarantees for spectral clustering with fairness constraints. In ICML, 2019

Optimization with Constraints



- Improving Group Fairness
  - Graph clustering [1]

$$\forall i \in \{1, \ldots, H\} \text{ and } \forall k \in \{1, \ldots, K\}, \frac{|\boxed{\mathcal{V}_i} \cap \boxed{\mathcal{C}_k}|}{|\mathcal{C}_k|} = \frac{|\mathcal{V}_i|}{|\mathcal{V}|}$$

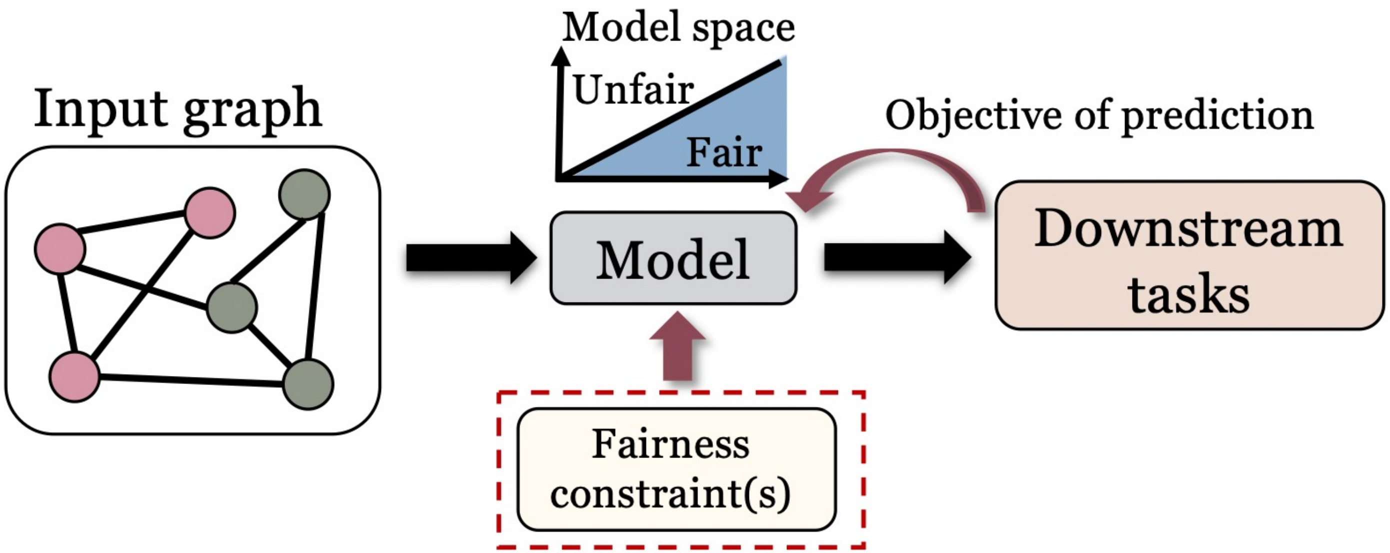
the node set of the i-th subgroup

the node set of the k-th cluster

Each sensitive subgroup is proportionally represented by each cluster.

[1] Matthaus Kleindessner, et al. Guarantees for spectral clustering with fairness constraints. In ICML, 2019

Optimization with Constraints



- Improving Individual Fairness
  - Graph clustering [1]

$$\forall k \in \{1, ..., \boxed{K}\}, \frac{1}{|\mathcal{C}_k|} |\{v_j : \mathbf{A}_{ij} = 1 \wedge v_j \in \boxed{\mathcal{C}_k}\}| = \frac{1}{|\mathcal{V}|} |\{\boxed{v_j} : \mathbf{A}_{ij} = 1\}|$$

Cluster number

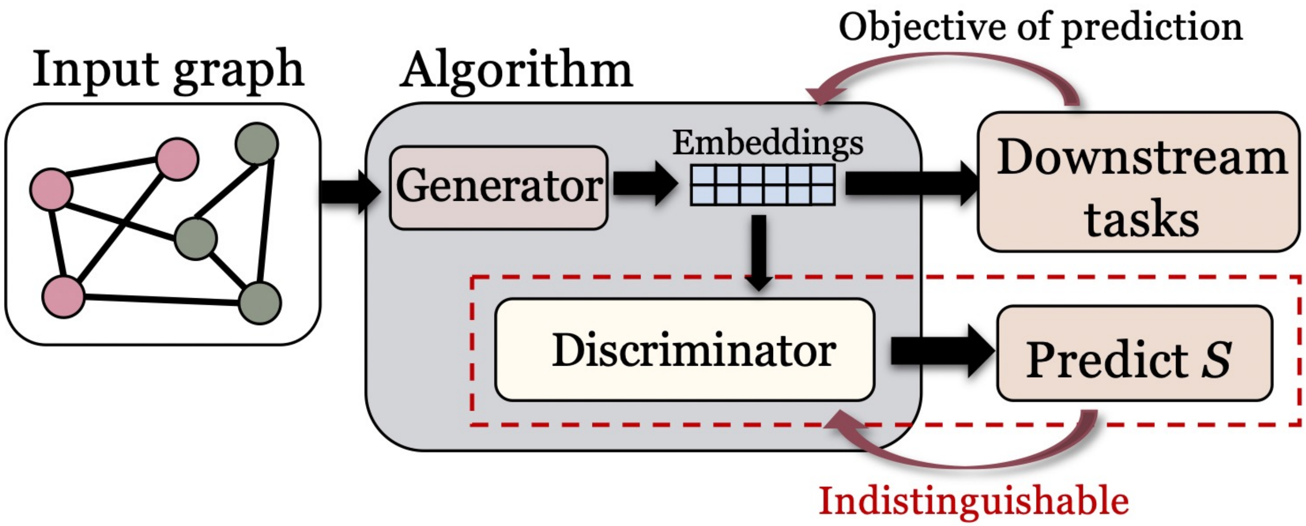
Node set of cluster *k*

Neighbors of node i

for each node in the graph, its neighbors should be proportionally assigned to different clusters

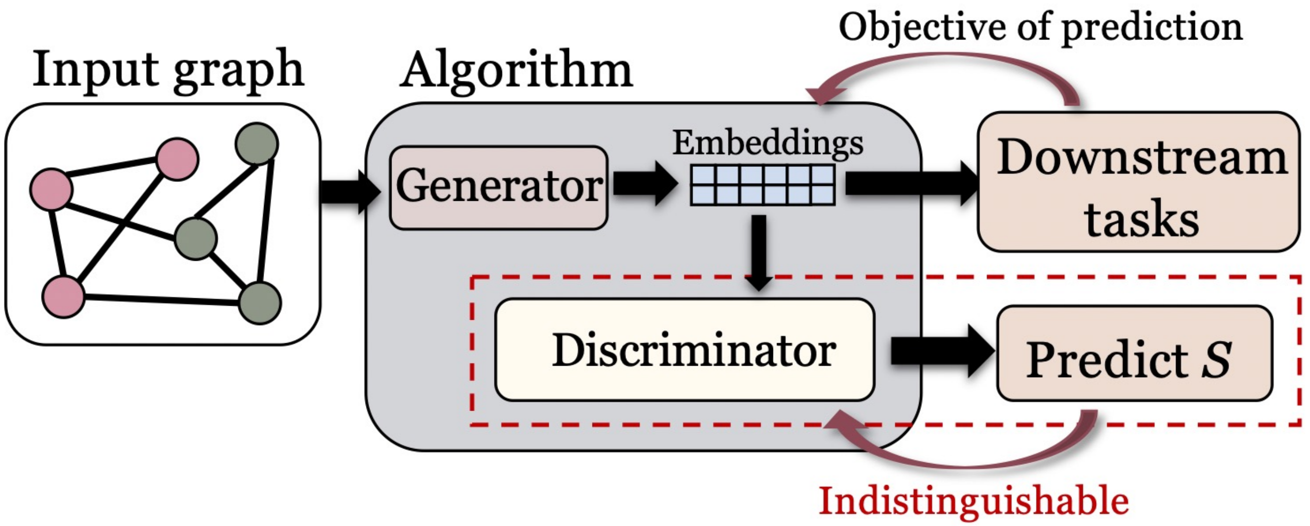
[1] Shubham Gupta and Ambedkar Dukkipati. Protecting individual interests across clusters: Spectral clustering with guarantees. arXiv, 2021

Adversarial Learning



A general formulation of fulfilling fairness with adversarial learning includes a **generator** and a **discriminator**:

Adversarial Learning



A general formulation of fulfilling fairness with adversarial learning includes a **generator** and a **discriminator**:

Generator: generate node embeddings for downstream tasks;

Discriminator: distinguish the embeddings between demographic subgroups;

Adversarial Learning

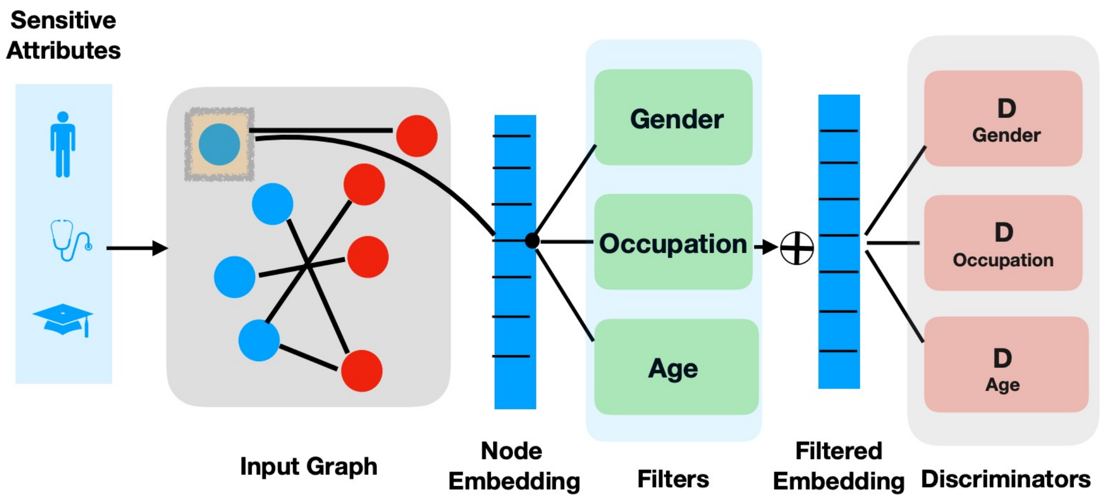


Figure: An example of adversarial learning-based method for fair graph embeddings [1]

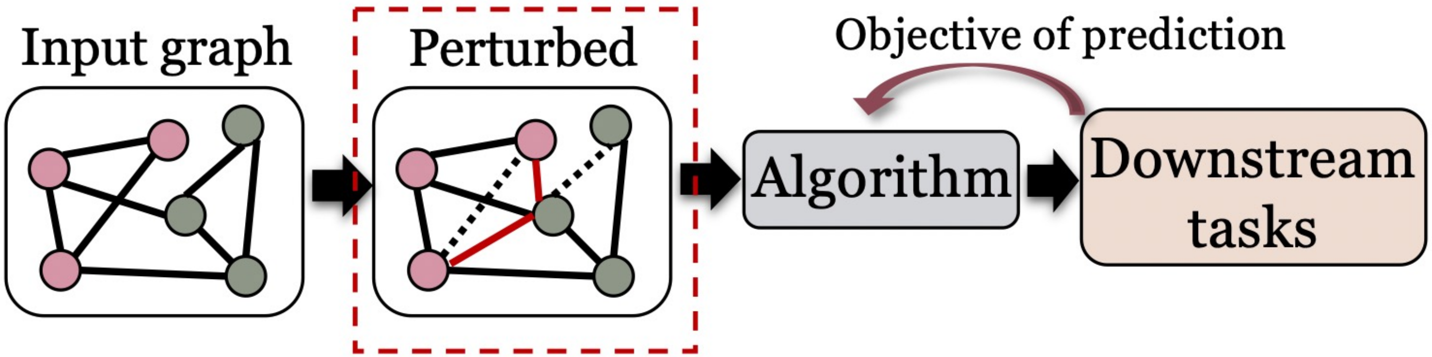
A general formulation of fulfilling fairness with adversarial learning includes a **generator** and a **discriminator**:

Generator: generate node embeddings for downstream tasks;

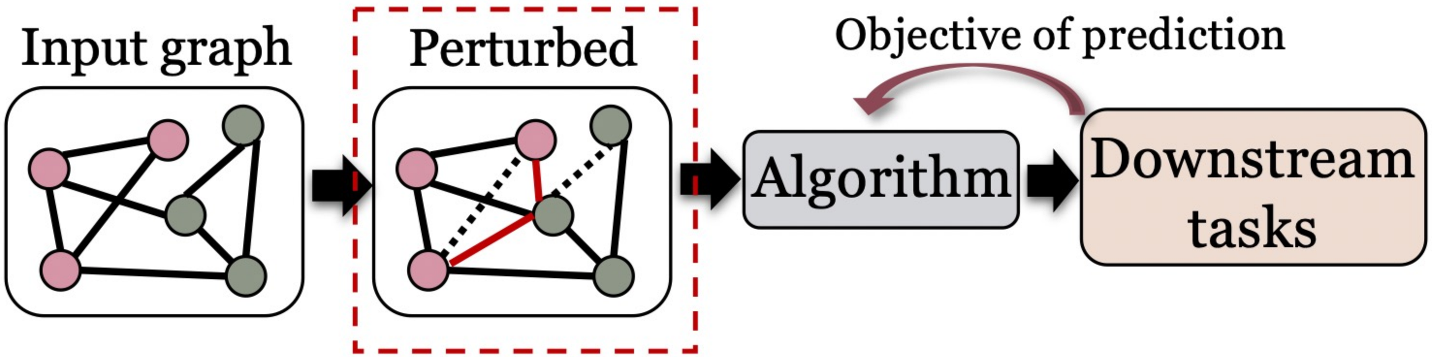
Discriminator: distinguish the embeddings between demographic subgroups;

[1] Avishek Bose and William Hamilton. Compositional fairness constraints for graph embeddings. In ICML, 2019

Edge Rewiring

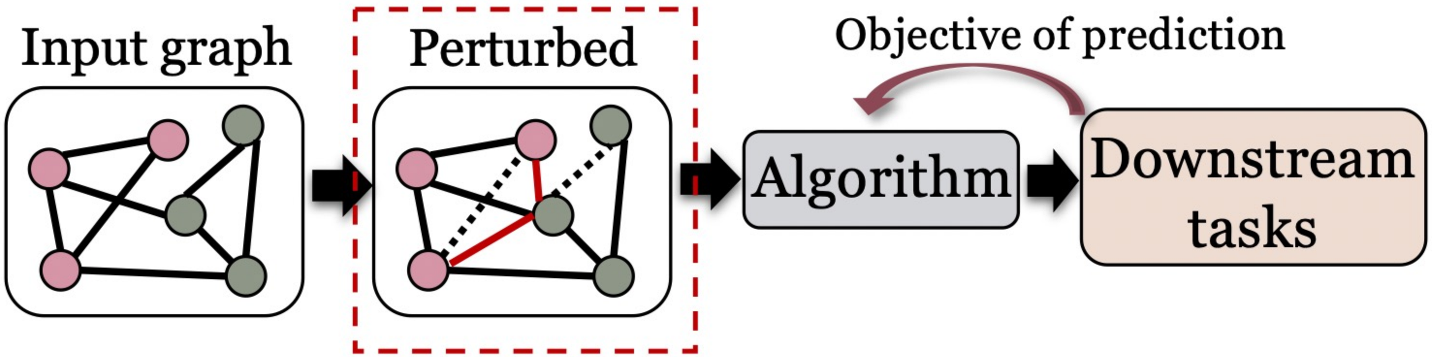


Edge Rewiring

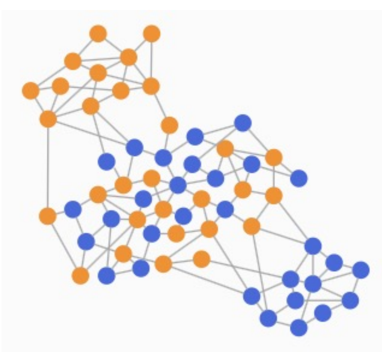


There could be **bias encoded in the network structure**, and edge rewiring aims to achieve a fairer structure for the graph mining algorithm.

Edge Rewiring

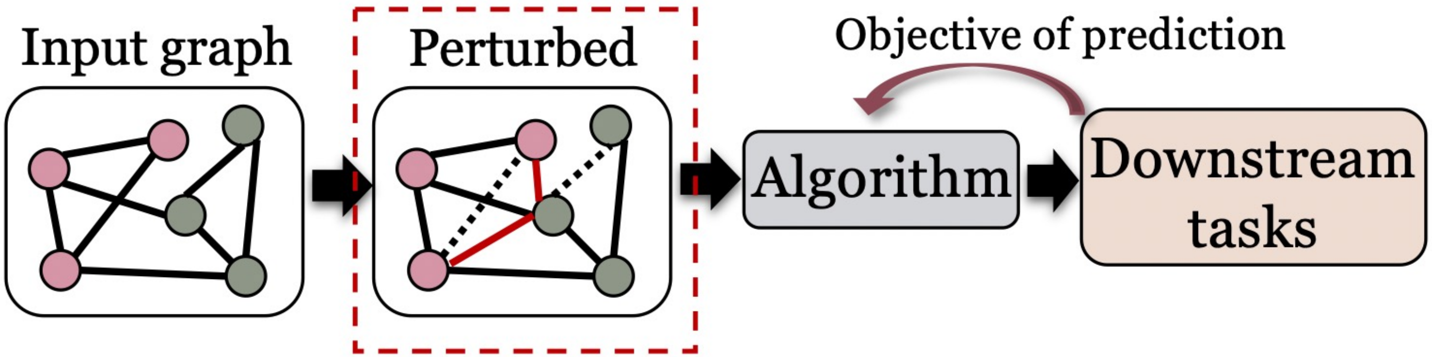


There could be **bias encoded in the network structure**, and edge rewiring aims to achieve a fairer structure for the graph mining algorithm.



An example of **biased graph structure**: clear **community structure** between two groups of nodes, where the membership is dependent on sensitive feature(s).

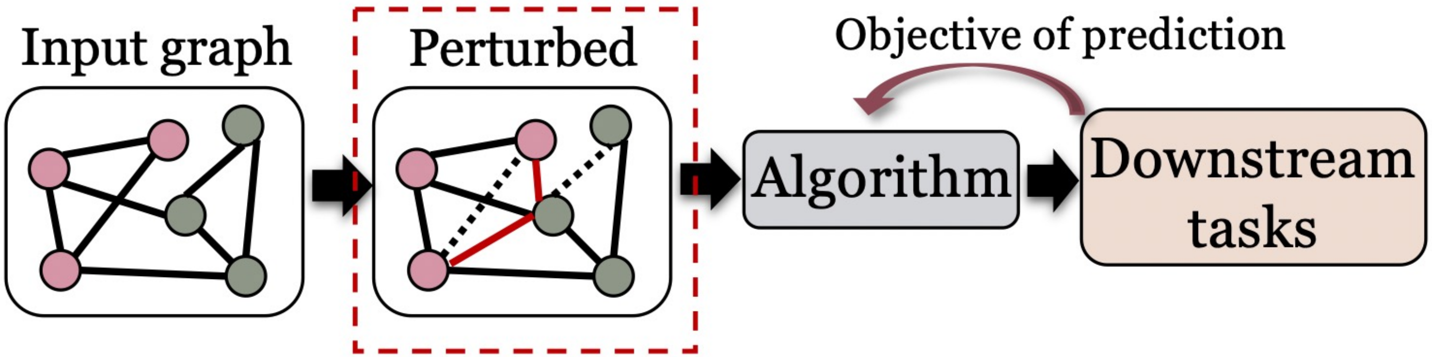
Edge Rewiring



- Improving Group Fairness
  - Information flow-based rewiring

Intuition: modify the graph topology to make information flows as fair as possible.

Edge Rewiring



- Improving Group Fairness
  - Information flow-based rewiring

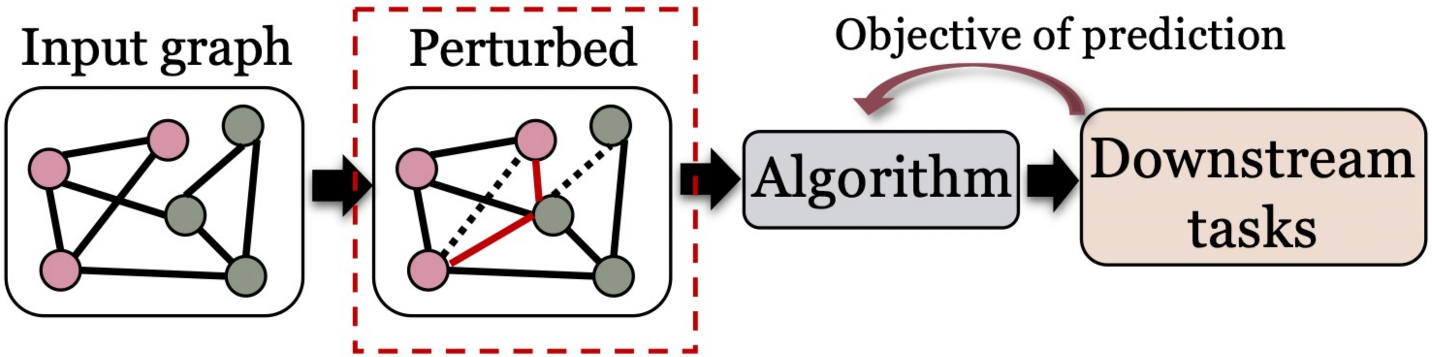
Intuition: modify the graph topology to make information flows as fair as possible.

- **Information unfairness score** [1]: the largest distribution difference of the probabilistic accessibility between two node groups.

- To obtain a fair graph topology, edges are rewired in a greedy manner to maximally reduce the information unfairness score.

[1] Zeinab S Jalali, et al. On the information unfairness of social networks. In SDM, 2020

Edge Rewiring



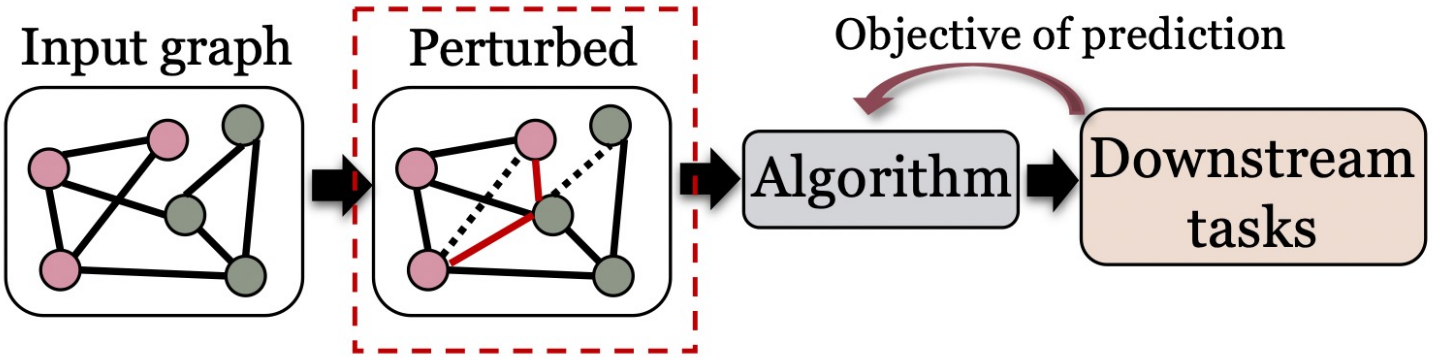
- Improving Group Fairness
  - Information flow-based rewiring

Intuition: modify the graph topology to make information flows as fair as possible.

- The Wasserstein distance between the node embedding distributions from two sensitive subgroups is minimized by learning a less biased (weighted) graph adjacency matrix [1].

[1] Yushun Dong, et al. Individual fairness for graph neural networks: A ranking based approach. In SIGKDD 2021.

Edge Rewiring



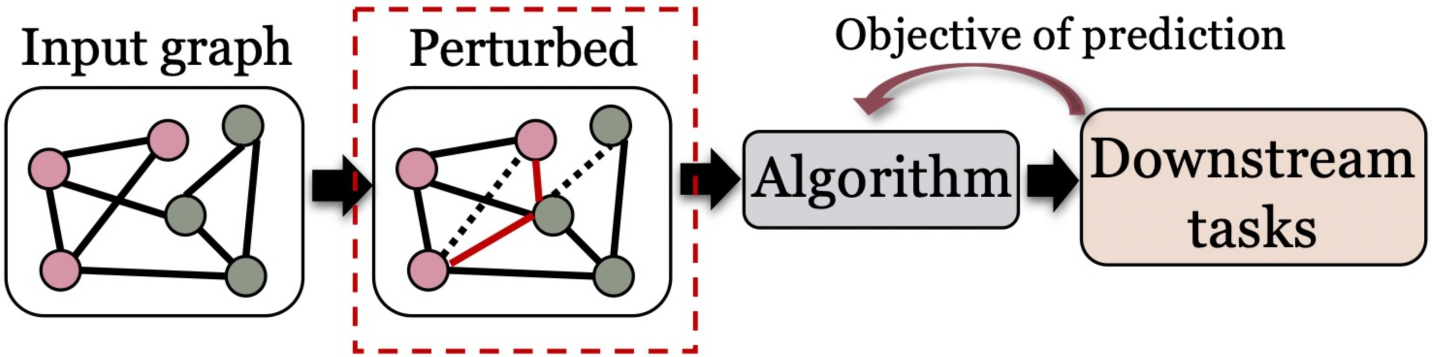
- Improving Group Fairness
  - Edge sampling-based rewiring

Intuition: Edges can be sampled in a probabilistic way to improve group fairness.

- Nodes within the same sensitive subgroup tend to be linked together on homogeneous graphs.
- FairDrop [1] removes more intra-group edges than inter-group edges.

[1] Indro Spinelli, et al. Biased edge dropout for enhancing fairness in graph representation learning. TAI, 2021.

Edge Rewiring



- Improving Individual Fairness

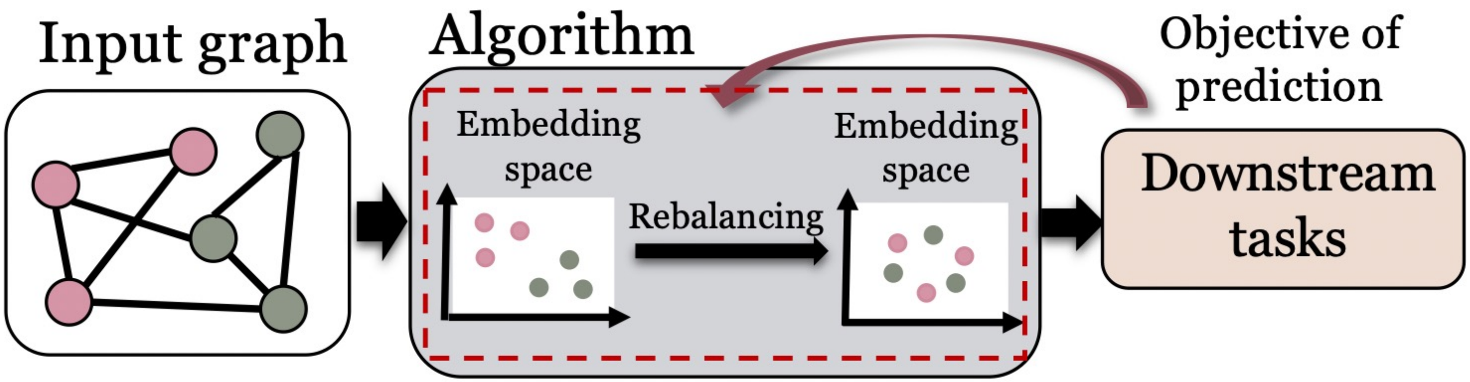
Intuition: encourage similar individuals to share similar topological characteristics.

- An optimization problem [1] is formulated to encourage similar nodes to have highly overlapped neighboring node sets after edge rewiring.

[1] Charlotte Laclau, el al. All of the fairness for edge prediction with optimal transport. In AISTATS, 2021

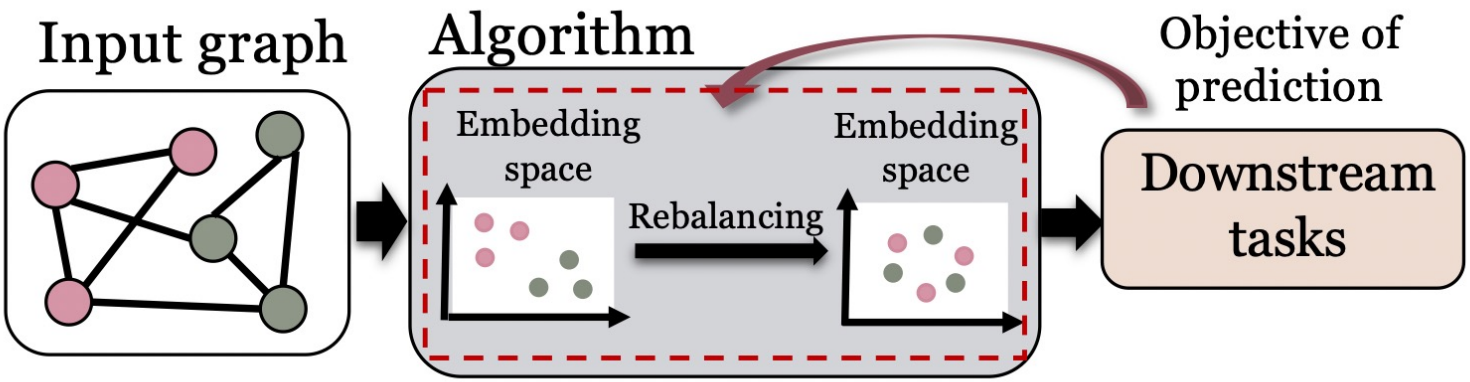
Rebalancing

Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



Rebalancing

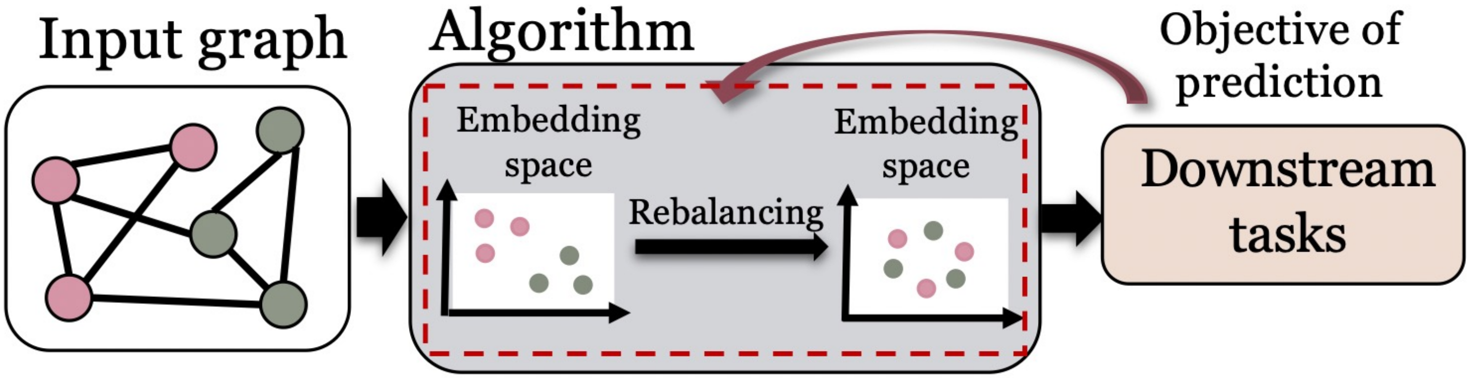
Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



- Improving Group Fairness

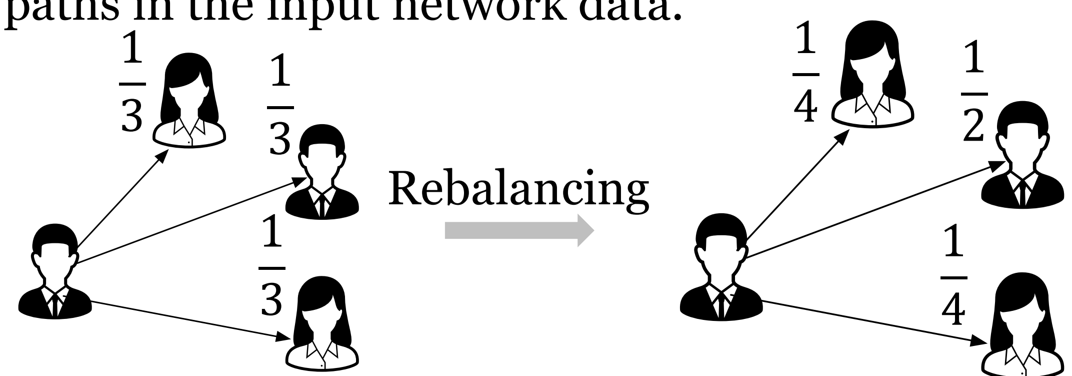
Rebalancing

Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



- Improving Group Fairness
  - Edge/Path-based rebalancing: promote group fairness based on the edges or paths in the input network data.

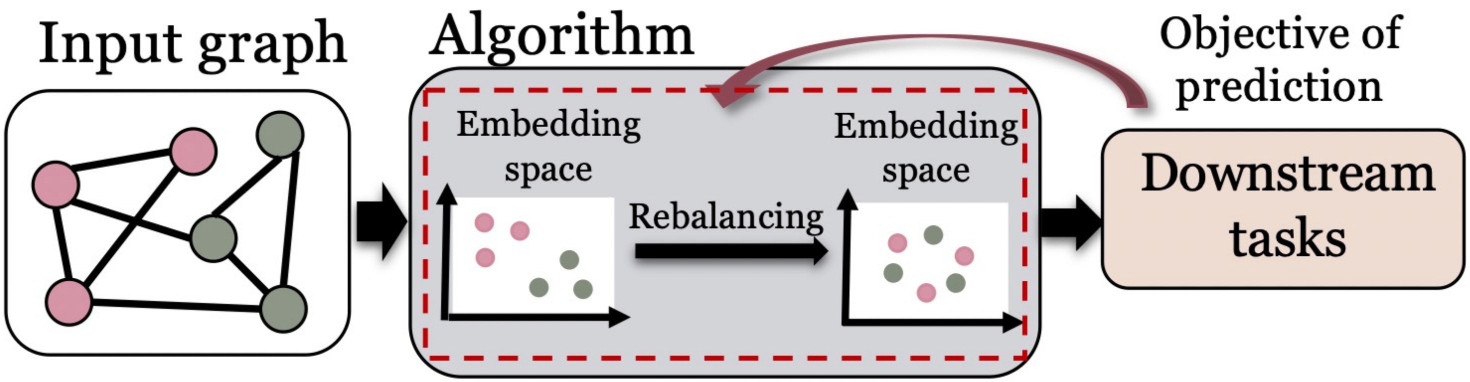
An example of rebalancing in **random walk** on graphs [1]:



[1] Rahman T, Surma B, Backes M, et al. Fairwalk: Towards fair graph embedding[J]. 2019.

Rebalancing

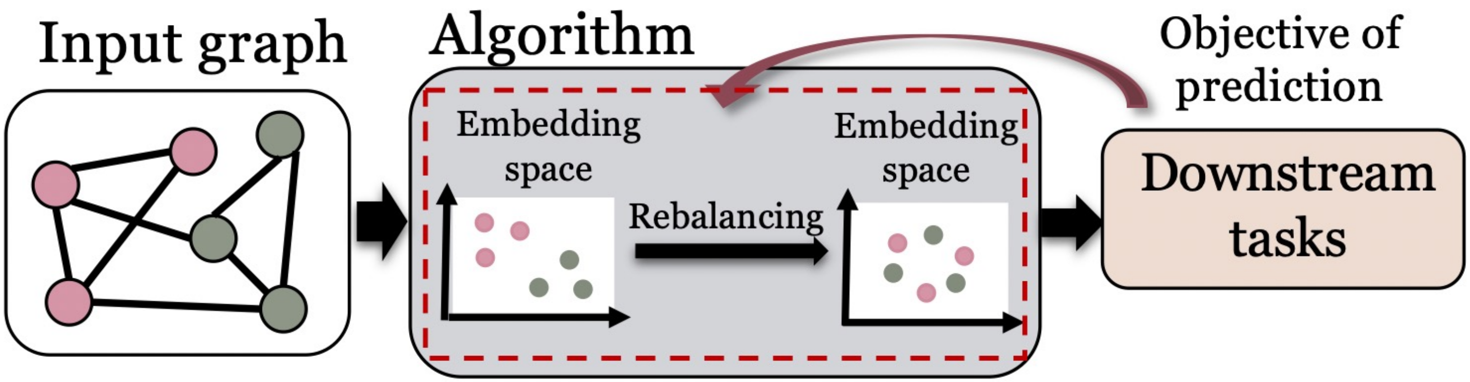
Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



- Improving Group Fairness
  - Node sampling/generation-based rebalancing: rebalance the node number between different sensitive subgroups
    - **Node sampling**: Sample subgraphs with balanced populations from different sensitive subgroups
    - **Node generation**: Generate pseudo nodes and reweight edges to encourage a balanced information propagation.

Rebalancing

Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



- Improving Degree-related Fairness
  - Nodes with low degrees usually benefit less from the information propagation.

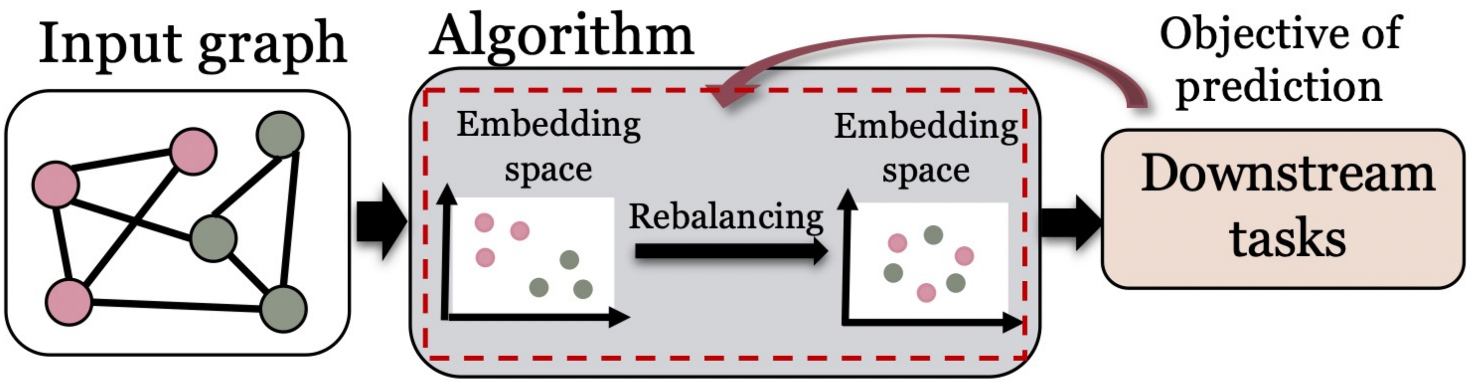
- Generate pseudo labels to improve the probability of labeled nodes appearing in the neighborhood of low-degree nodes [1].

More supervision for low-degree nodes

[1] Xianfeng Tang, et al. Investigating and mitigating degree-related biases in graph convolutional networks. In CIKM 2020

Rebalancing

Rebalancing could be achieved in different ways **depending on the characteristics of the graph mining algorithms & tasks**.



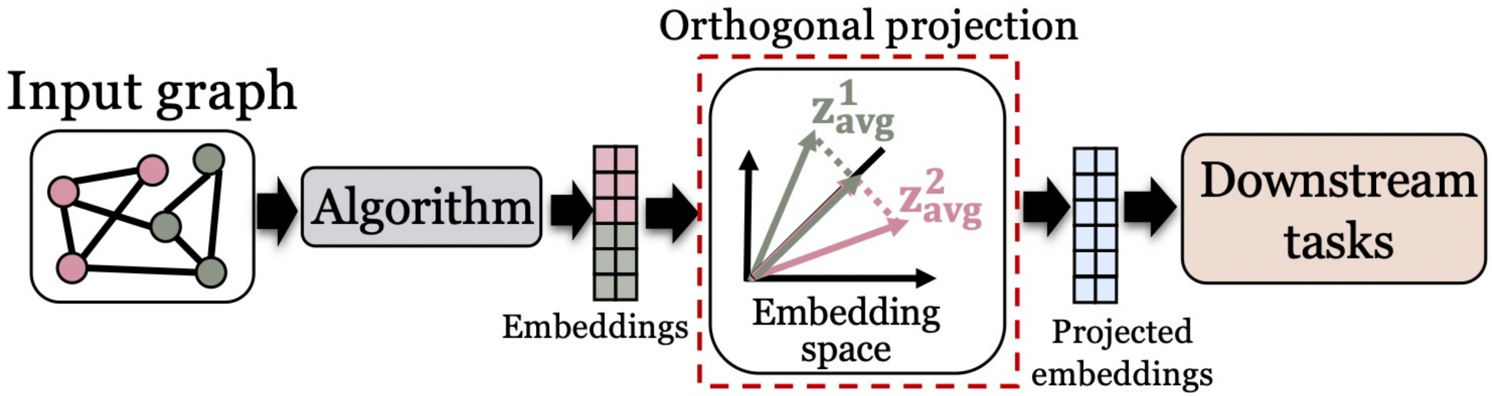
- Improving Degree-related Fairness
  - Nodes with low degrees usually benefit less from the information propagation.

- High-degree nodes often have stronger influence on the gradient of the learnable weights in GNN [1].
- For this problem, a doubly stochastic adjacency matrix (the rows and columns sum up to 1) of the graph is employed as GNN input.

Rebalance the influence of each node in optimization

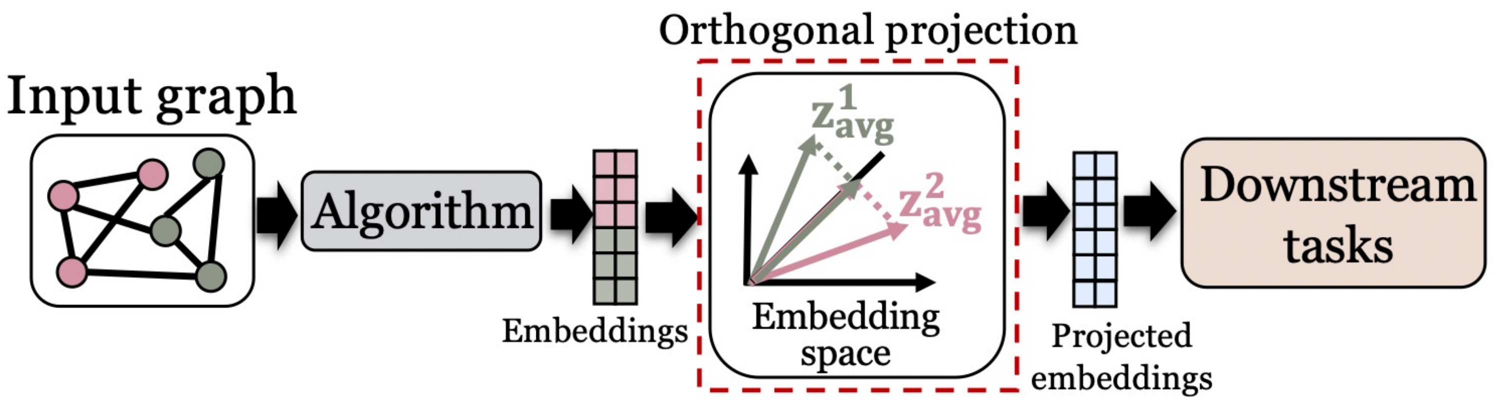
[1] Jian Kang, et al. Rawlsgcn: Towards rawlsian difference principle on graph convolutional network. In WWW 2022

Orthogonal Projection



Intuition: if the node embeddings are projected onto the same hyper-plane (**orthogonal** to the direction of the sensitive features), then there will be **no correlation** between node embeddings and bias.

Orthogonal Projection



Intuition: if the node embeddings are projected onto the same hyper-plane (**orthogonal** to the direction of the sensitive features), then there will be **no correlation** between node embeddings and bias.

Find bias direction.

Project embeddings onto a hyper-plane orthogonal to the bias direction.

$$\mathbf{z}_{\text{avg}}^i = \frac{\mathbf{z}_1 + \mathbf{z}_2 + \ldots + \mathbf{z}_{|\mathcal{V}_i|}}{\|\mathbf{z}_1 + \mathbf{z}_2 + \ldots + \mathbf{z}_{|\mathcal{V}_i|}\|_2} \rightarrow \boxed{\mathbf{z}_{\text{bias}}} = \frac{\mathbf{z}_{\text{avg}}^1 - \mathbf{z}_{\text{avg}}^2}{\|\mathbf{z}_{\text{avg}}^1 - \mathbf{z}_{\text{avg}}^2\|_2} \rightarrow \mathbf{z}_j' = \mathbf{z}_j - < \mathbf{z}_j, \mathbf{z}_{\text{bias}} > \mathbf{z}_{\text{bias}}$$

unit vector in the bias direction

Outline

Background Information

Fairness Notions and Metrics

Methodologies to Mitigate Bias

Real-World Applications

Summary & Existing Challenges

User Fairness in Recommender System

- **User Fairness**: the recommendation quality for different users should be similar.
  - Active/inactive users
  - User in different sensitive subgroups

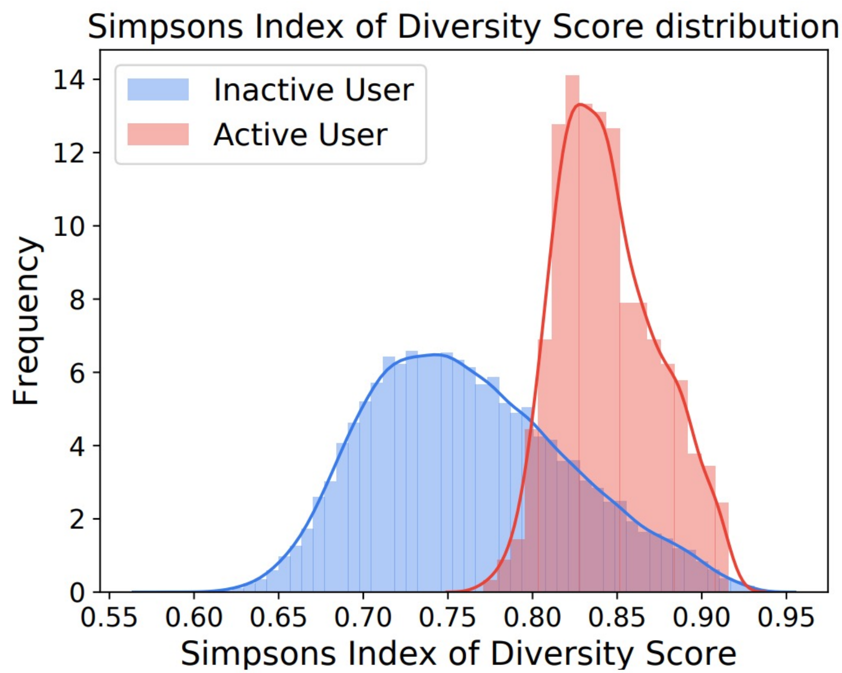


Figure: Statistics of Amazon Beauty dataset [1]

[1] Zuohui Fu, et al. Fairness-aware explainable recommendation over knowledge graphs. In SIGIR, 2020.

User Fairness in Recommender System

- **Rebalancing-based method**: rebalance item ratings given by users from different sensitive subgroups [1].

[1] Golnoosh Farnadi, et al. A fairness-aware hybrid recommender system. In workshop of RecSys, 2018.

User Fairness in Recommender System

- **Adversarial learning-based method**: avoid delivering news with biased content towards certain demographic subgroups.

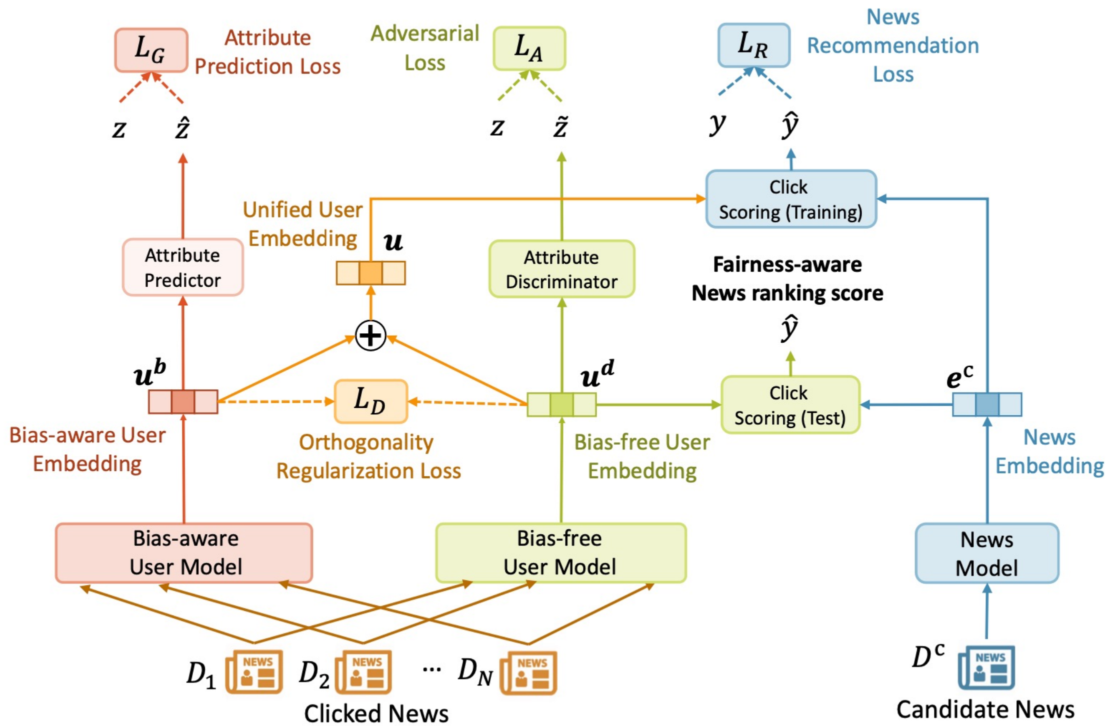


Figure: The architecture of FairRec [1]

[1] Chuhan Wu, et al. Fairness-aware news recommendation with decomposed adversarial learning. In AAAI, 2021

Popularity Fairness in Recommender System

- **Popularity Fairness**: popular items should not be over-emphasized compared with other instances.

Popularity Fairness in Recommender System

- **Popularity Fairness**: popular items should not be over-emphasized compared with other instances.
  - Measurement: the average recommendation rate of less popular instances (e.g., users, items) [1].

$$Q_{\text{fairness}} = \frac{1}{2|\mathcal{E}|} \sum_{i,j} (\mathbf{A}_{i,j} - \frac{d_i d_j}{2|\mathcal{E}|}) \delta(M(v_i), M(v_j))$$

[1] Farzan M. el al. Bursting the filter bubble: Fairness-aware network link prediction. In AAAI, 2020

Popularity Fairness in Recommender System

- **Popularity Fairness**: popular items should not be over-emphasized compared with other instances.
  - Measurement: the average recommendation rate of less popular instances (e.g., users, items) [1].

$$Q_{\text{fairness}} = \frac{1}{2\boxed{|\mathcal{E}|}} \sum_{i,j} \left(\mathbf{A}_{i,j} - \frac{\boxed{d_i} d_j}{2|\mathcal{E}|}\right) \boxed{\delta}\left(\boxed{M(v_i)}, \boxed{M(v_j)}\right)$$

degree of node i

group membership for user i and j

Total edge number

Kronecker delta function

[1] Farzan M. el al. Bursting the filter bubble: Fairness-aware network link prediction. In AAAI, 2020

Popularity Fairness in Recommender System

- **Popularity Fairness**: popular items should not be over-emphasized compared with other instances.
  - Measurement: the average recommendation rate of less popular instances (e.g., users, items) [1].

$$Q_{\text{fairness}} = \frac{1}{2|\mathcal{E}|}\sum_{i,j}\left(\mathbf{A}_{i,j} - \frac{d_i d_j}{2|\mathcal{E}|}\right)\delta\left(M(v_i), M(v_j)\right)$$

A lower value indicates more inter-group edges, which implies that those less-popular groups are encouraged to connect with other groups.

degree of node i

group membership for user i and j

Total edge number

Kronecker delta function

[1] Farzan M. el al. Bursting the filter bubble: Fairness-aware network link prediction. In AAAI, 2020

Popularity Fairness in Recommender System

- **Edge Rewiring-based method:** Based on link prediction result, a proportion of links are rewired in a greedy manner to achieve popularity fairness [1].

[1] Farzan M., Bursting the filter bubble: Fairness-aware network link prediction. In AAAI, 2020.

Popularity Fairness in Recommender System

- **Regularization-based method**
  - commonly used technique to fulfill different fairness notions in recommender system.

An example [1] of regularization for popularity fairness:

$$\mathcal{L}_{fair} = \mathrm{Corr}_{\mathrm{P}}(\boxed{\hat{\mathbf{r}}_{+}}, \boxed{\mathbf{p}_{+}})$$

the vector of predicted **relevance scores** for positive user-item pairs

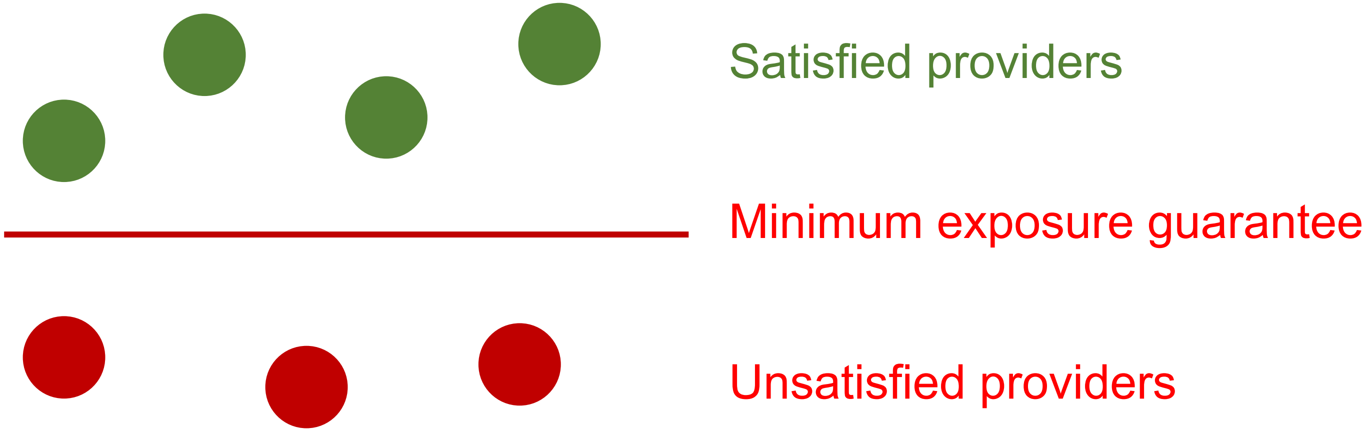
the vector of the feedback number (i.e., **popularity**) received by the items in user-item pairs

This regularization relieves the effect that popular items tend to receive higher relevance scores.

[1] Ziwei Zhu, et al. Popularity-opportunity bias in collaborative filtering. In WSDM, 2021.

Provider Fairness in Recommender System

- **Provider Fairness**: items from different providers should receive the same exposure rate to the customers.
  - **Example of metric 1**: set a minimum exposure guarantee for all providers and used the number of unsatisfied providers to measure provider fairness [1].



[1] Gourab K. Patro, et al. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In WWW, 2020.

Provider Fairness in Recommender System

- **Provider Fairness**: items from different providers should receive the same exposure rate to the customers.
  - **Example of metric 2**: average number of providers appearing in recommendations [1].

[1] Weiwen Liu and Robin Burke. Personalizing fairness-aware reranking. arXiv, 2018.

Provider Fairness in Recommender System

- **Provider Fairness**: items from different providers should receive the same exposure rate to the customers.
  - **Example of metric 2**: average number of providers appearing in recommendations [1].
  - **Example of metric 3**: use both the user-item relevance difference and item exposure rate difference between different providers [2].

[1] Weiwen Liu and Robin Burke. Personalizing fairness-aware reranking. arXiv, 2018.
[2] Ludovico Boratto, et al. Interplay between upsampling and regularization for provider fairness in recommender systems. In UMUAI, 2020.

Provider Fairness in Recommender System

- **Provider Fairness**: items from different providers should receive the same exposure rate to the customers.
  - **Example of metric 2**: average number of providers appearing in recommendations [1].
  - **Example of metric 3**: use both the user-item relevance difference and item exposure rate difference between different providers [2].

Provider Fairness in Recommender System

- **Rebalancing-based method:** upsample interactions between users and items from minority providers.

Marketing Fairness in Recommender System

- **Marketing Fairness**: users are less likely to interact with items whose marketing strategy is not consistent with their identity.
  - E.g., some gender-neutral items (e.g., armband) could be marketed only with images of males.
  - Measurement: variance of recommendation errors for identity-consistent and identity-inconsistent users [1].

[1] Mengting Wan, et al. Addressing marketing bias in product recommendations. In WSDM, 2020.

Marketing Fairness in Recommender System

- **Regularization-based method** [1]: add an additional term to regularize the correlation between prediction errors and the distribution of market segments.

$$\mathcal{L}^* = \sum (s_{u,i} - r_{u,i})^2 + \alpha \mathcal{L}_{corr.},$$

$$\mathcal{L}_{corr.} = \kappa^{(u.)} \overbrace{\frac{V^{(u.)}}{U^{(u.)}}}^{\text{error parity on user identity}} + \kappa^{(p.)} \underbrace{\frac{V^{(p.)}}{U^{(p.)}}}_{\text{error parity on product image}} + \kappa^{(market)} \overbrace{\frac{V^{(market)}}{U^{(market)}}}^{\text{error parity on market segments}},$$

[1] Mengting Wan, et al. Addressing marketing bias in product recommendations. In WSDM, 2020.

Social Fairness in Knowledge Graph

- **Social Fairness**: knowledge graph embeddings could encode historical social biases.
  - E.g., bankers are males and nurses are female.
  - Example of measurement: tail prediction (e.g., female/male) based on sensitive relations (e.g., gender) + head entity (e.g., human)

Social Fairness in Knowledge Graph

- **Regularization-based method**
  - Example of regularization [1]: the KL-divergence between the prediction distribution and uniform distribution over all possible sensitive feature values.

[1] Joseph Fisher, et al. Measuring social bias in knowledge graph embeddings. In workshop of AKBC, 2020.

Social Fairness in Knowledge Graph

- **Adversarial Learning-based method**
  - Use a sensitive information filter to remove social bias from the embeddings of human entities with a min-max game [1].

[1] Mario Arduini, et al. Adversarial learning for debiasing knowledge graph embeddings. In SIGKDD, 2020.

Path Diversity Fairness in Knowledge Graph

- **Path Diversity Fairness**: the distributions of meta-paths should be similar across different demographic subgroups.
- **Constraint-based method**: use a fairness constrained approach [1] via heuristic re-ranking to mitigate unfairness in recommendation over knowledge graphs.

[1] Zuohui Fu, et al. Fairness-aware explainable recommendation over knowledge graphs. In SIGIR, 2020.

Fairness in Other Applications

- **Criminal justice**: predict whether a defendant deserves bail over a similarity network between defendants.

"The United States inarguably has a mass-incarceration crisis, but it is poor people and minorities who bear its brunt. Punishment profiling will exacerbate these disparities—including racial disparities. It also confirms the widespread impression that the criminal justice system is rigged against the poor."

[1] Starr, S. B. 2014a. "Sentencing by the Numbers." New York Times op-ed, August 10, 2014.

Fairness in Other Applications

- **Economics:** default and credit risk prediction over the network between bank clients.

Fairness in Other Applications

- **Social network:**
  - Information diffusion over social networks.
  - Gender gap on social media.
  - Fair influence on social networks.

Fairness in Other Applications

- **Health:** prevent homeless youth from HIV over real-world social connections.
  - E.g., in the HIV prevention domain, we may wish to ensure that members of racial minorities or of LGBTQ identity are not disproportionately excluded [1].

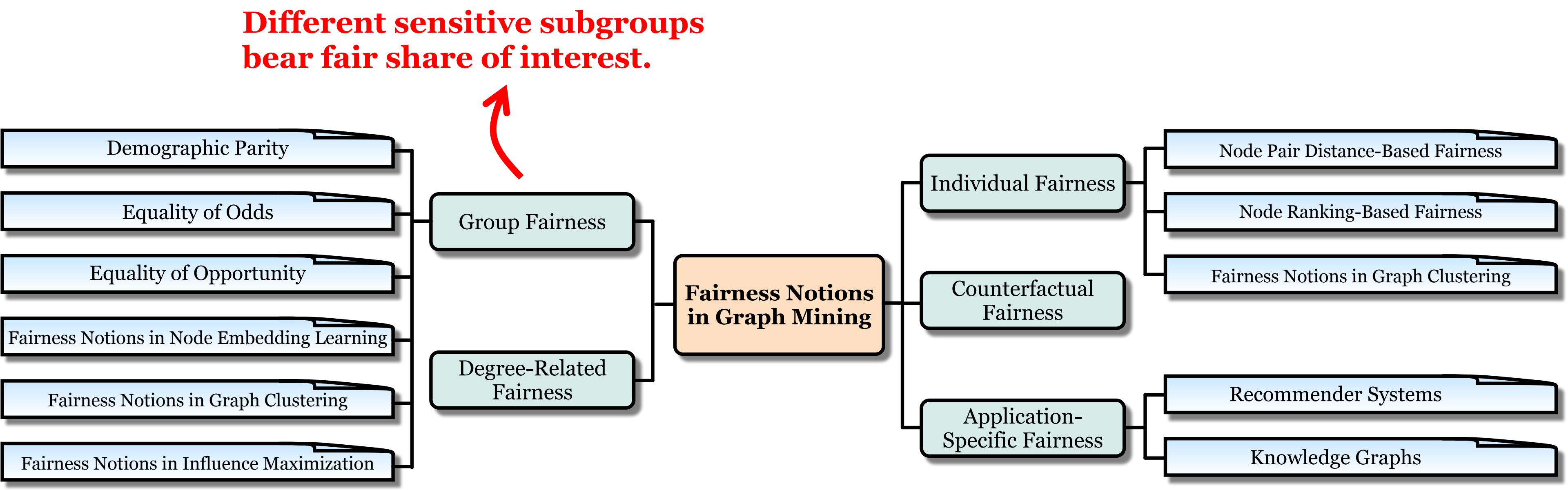
[1] Alan Tsang, et al. Group-fairness in influence maximization. In IJCAI, 2019.

Outline

- Background Information
- Fairness Notions and Metrics
- Methodologies to Mitigate Bias
- Real-World Applications
- Summary & Existing Challenges

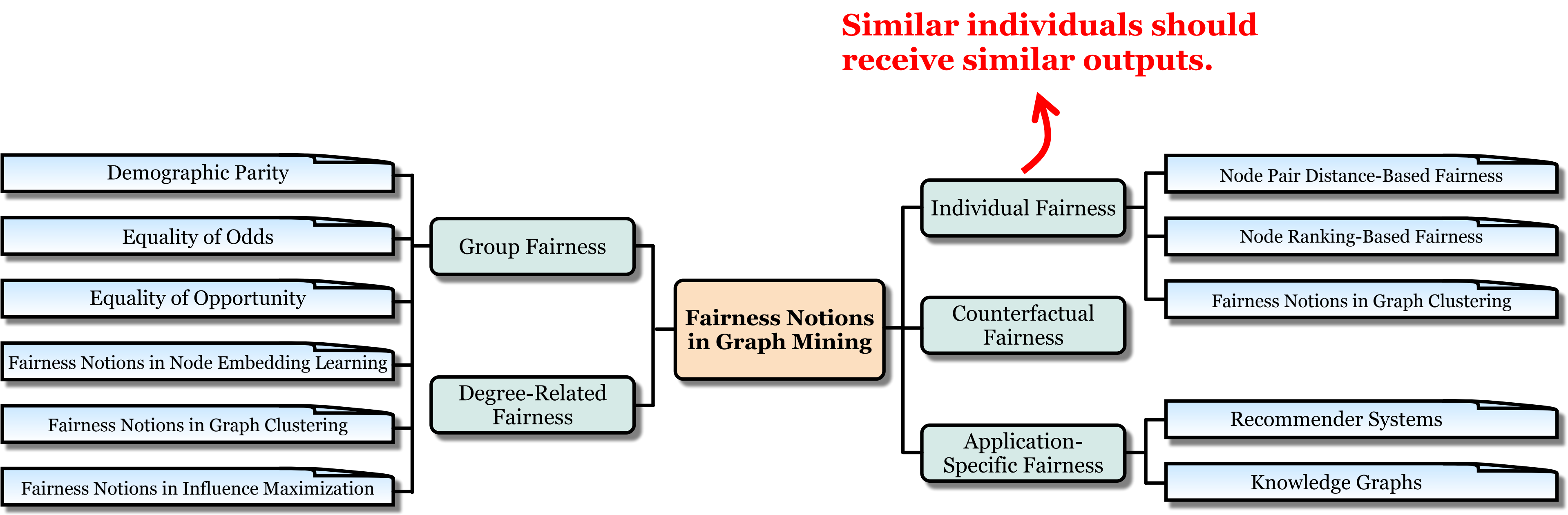
Summary on Fairness Notions

The taxonomy of fairness notions:



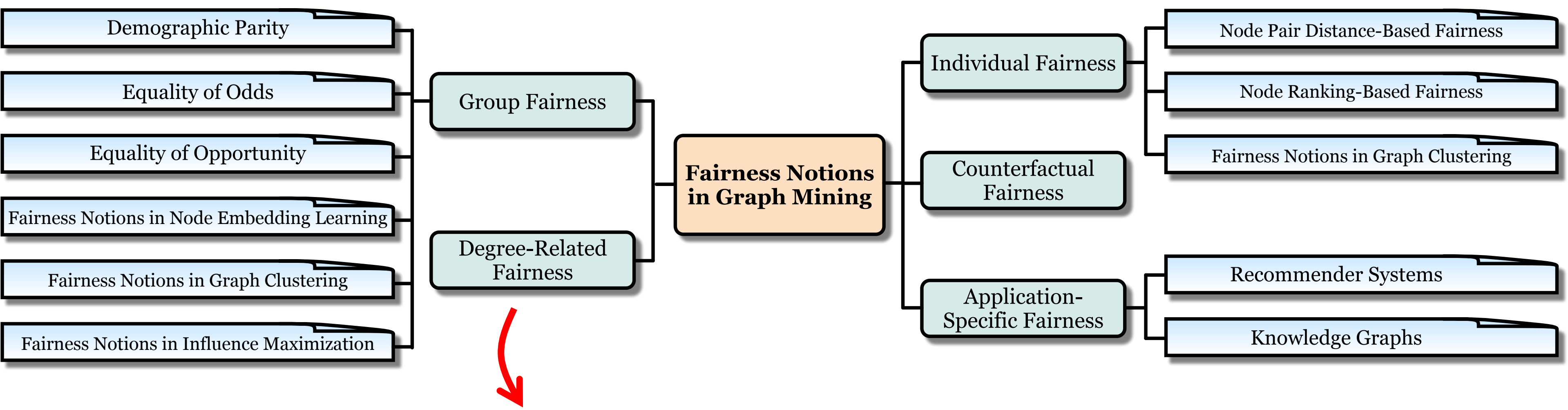
Summary on Fairness Notions

The taxonomy of fairness notions:



Summary on Fairness Notions

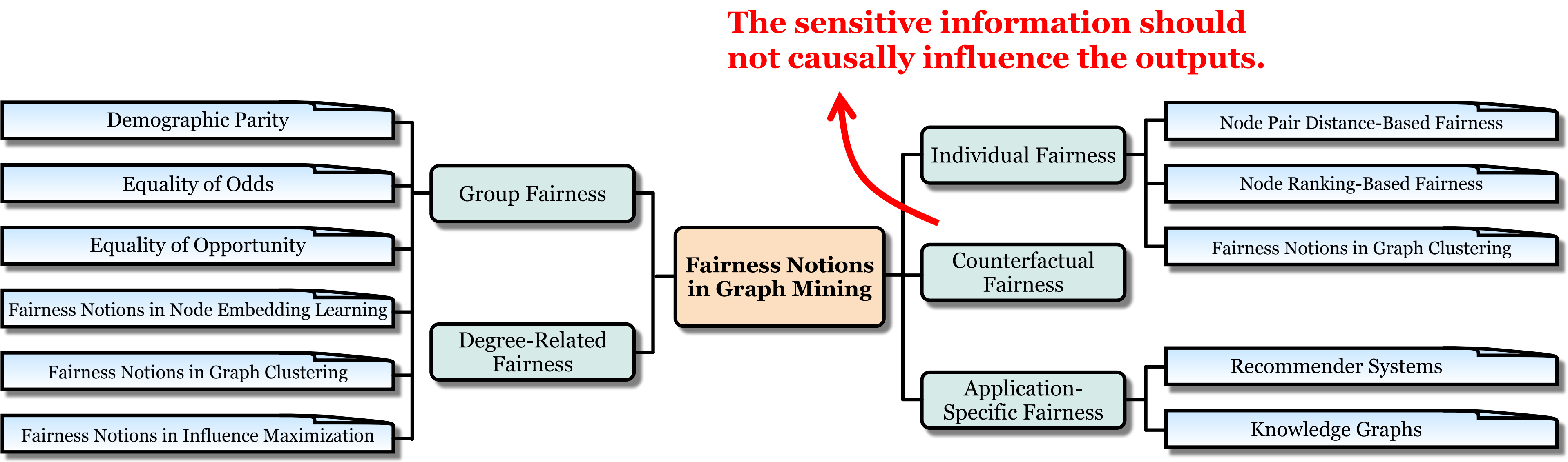
The taxonomy of fairness notions:



Nodes with different degrees should bear similar level of utility from the graph mining model.

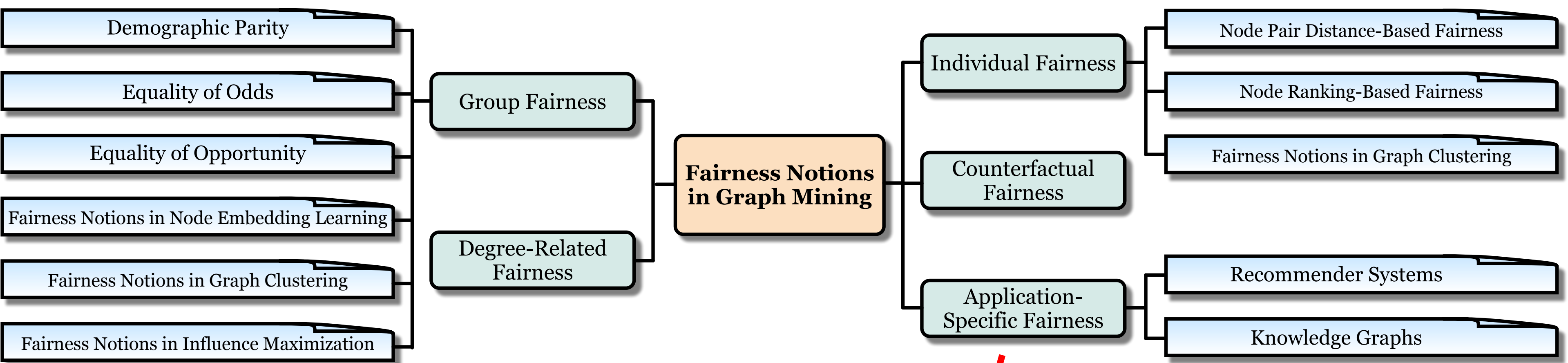
Summary on Fairness Notions

The taxonomy of fairness notions:



Summary on Fairness Notions

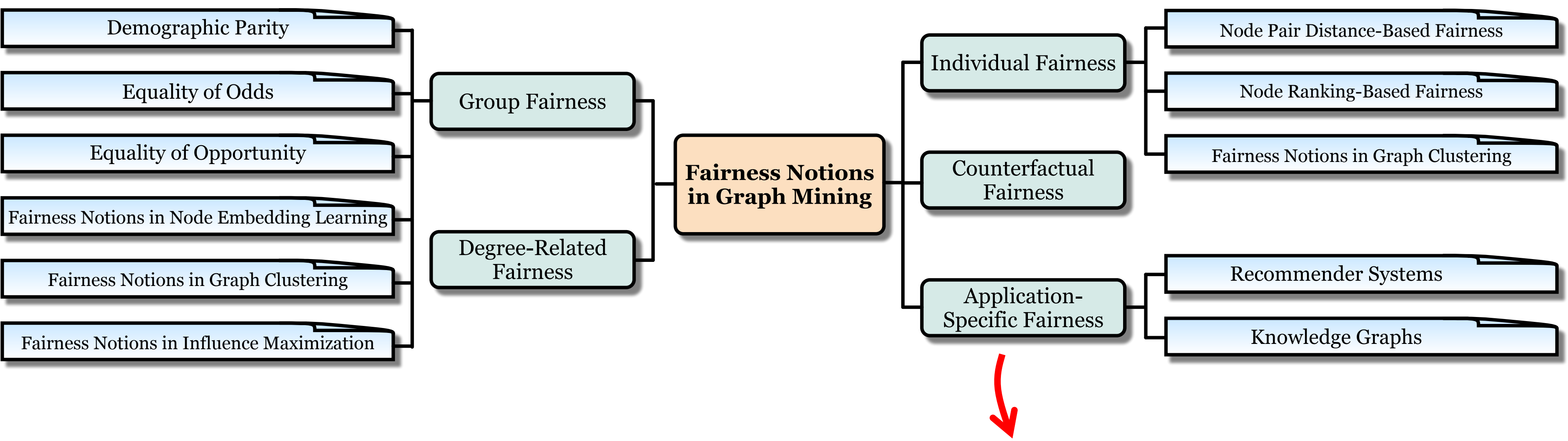
The taxonomy of fairness notions:



Recommender Systems:
(1) User Fairness
(2) Popularity Fairness
(3) Provider Fairness
(4) Marketing Fairness

Summary on Fairness Notions

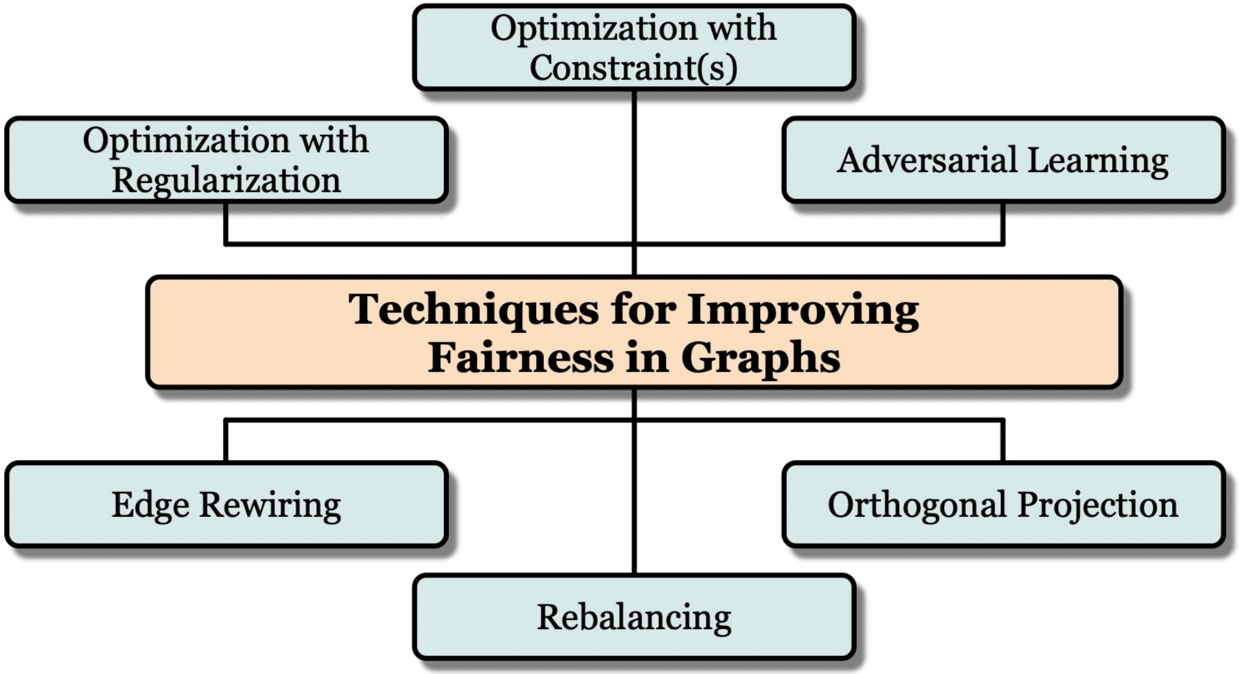
The taxonomy of fairness notions:



Knowledge Graphs:
(1) Social Fairness
(2) Path Diversity Fairness
(3) Popularity Fairness

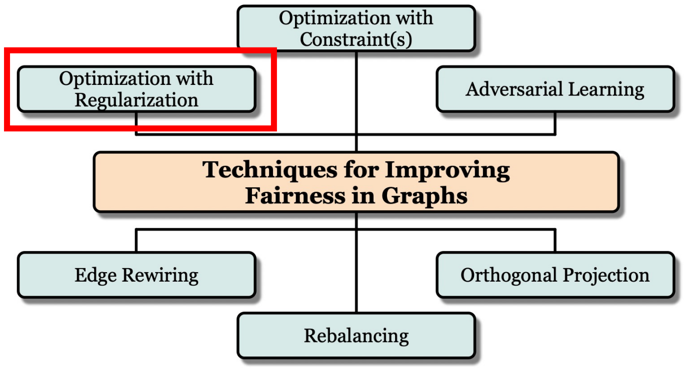
Summary on Techniques Fulfilling Fairness

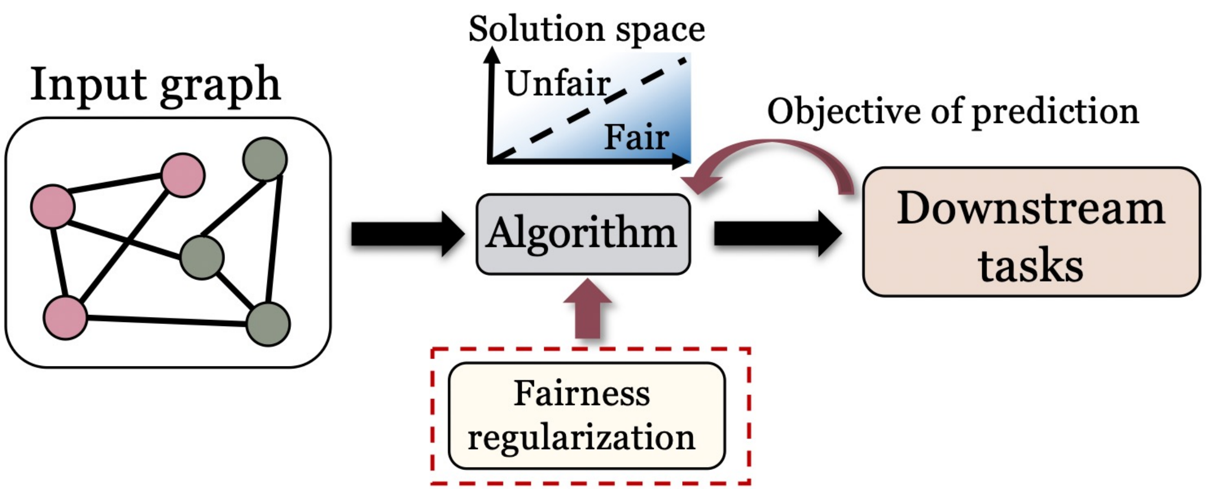
The taxonomy of techniques fulfilling fairness:



Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:

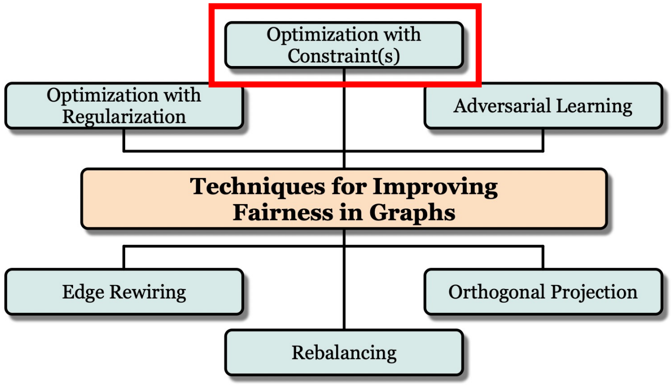


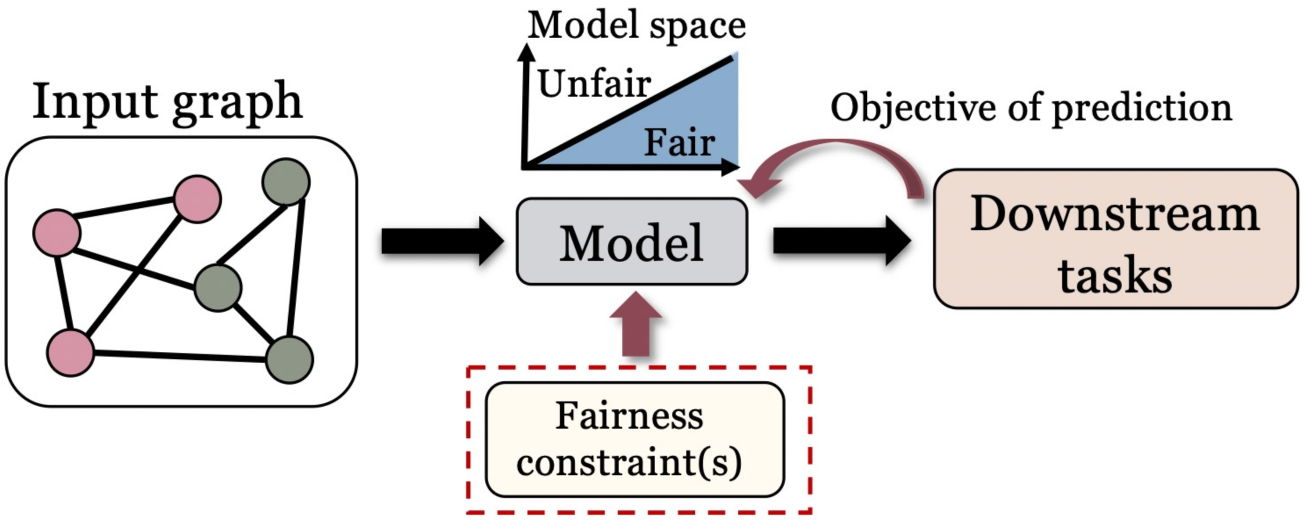


Formulating fairness-aware regularizations to achieve as fair solutions as possible.

Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:

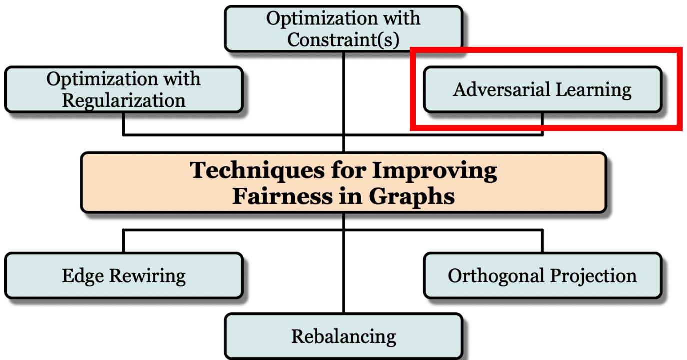


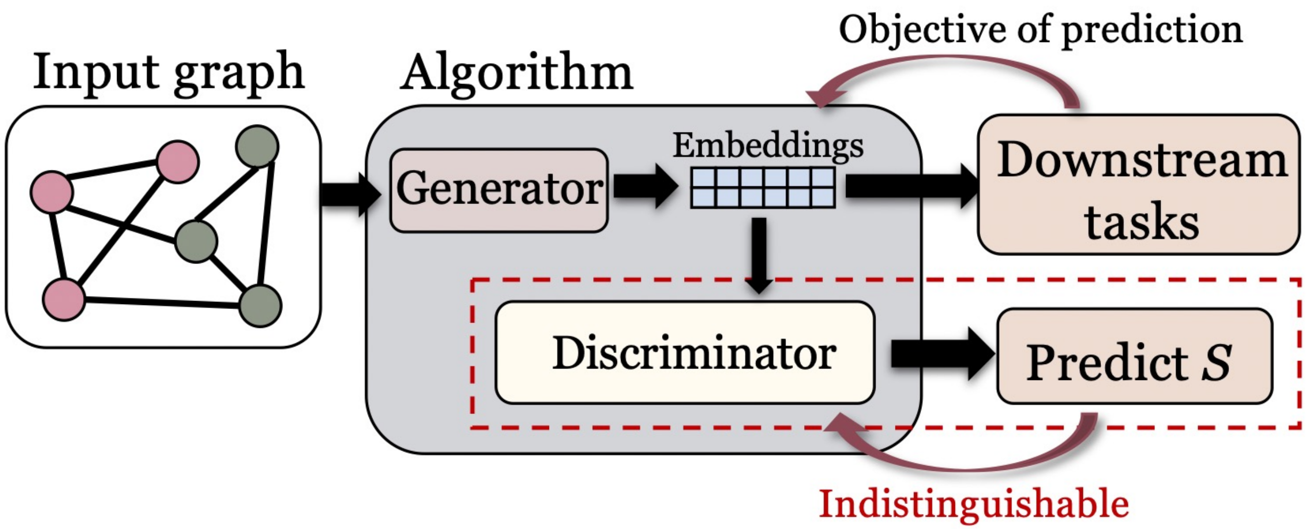


Formulating fairness-aware constraints to define fair area in the model space.

Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:

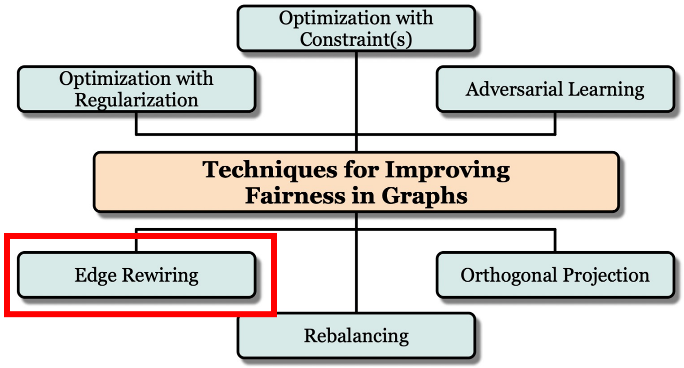


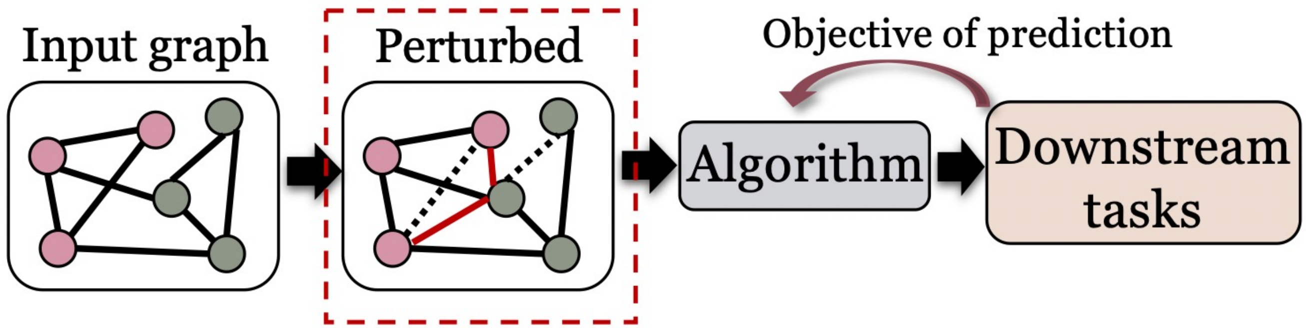


Learn embeddings that fools the discriminator to exclude sensitive information.

Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:

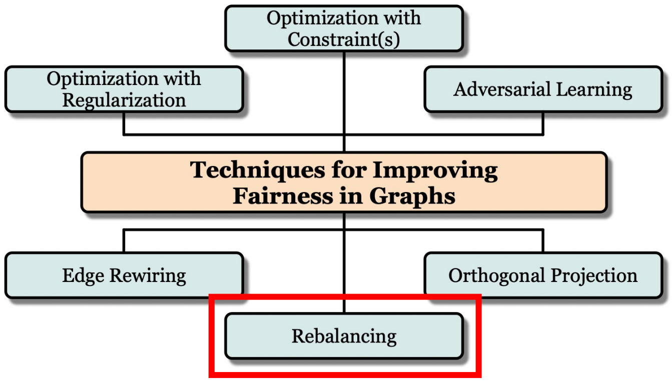


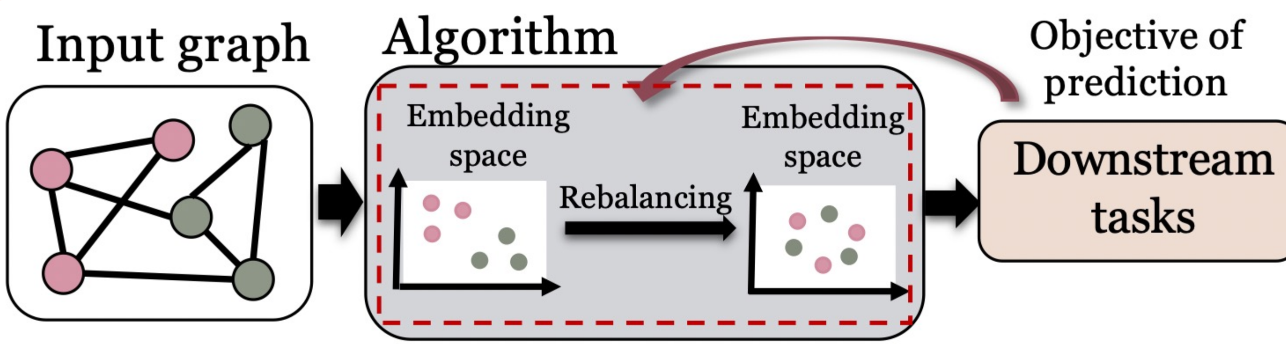


Edit the graph topology to achieve fairness-aware objectives in downstream tasks.

Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:

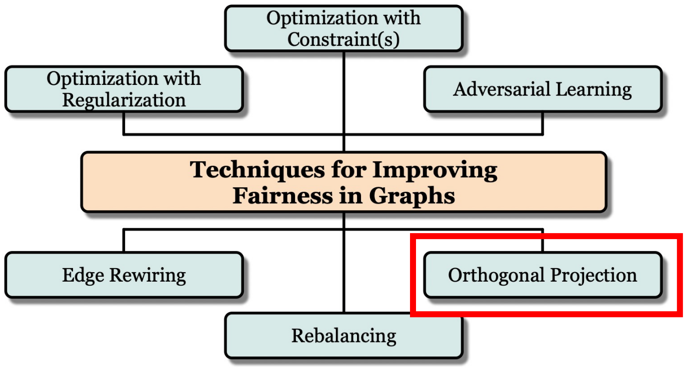


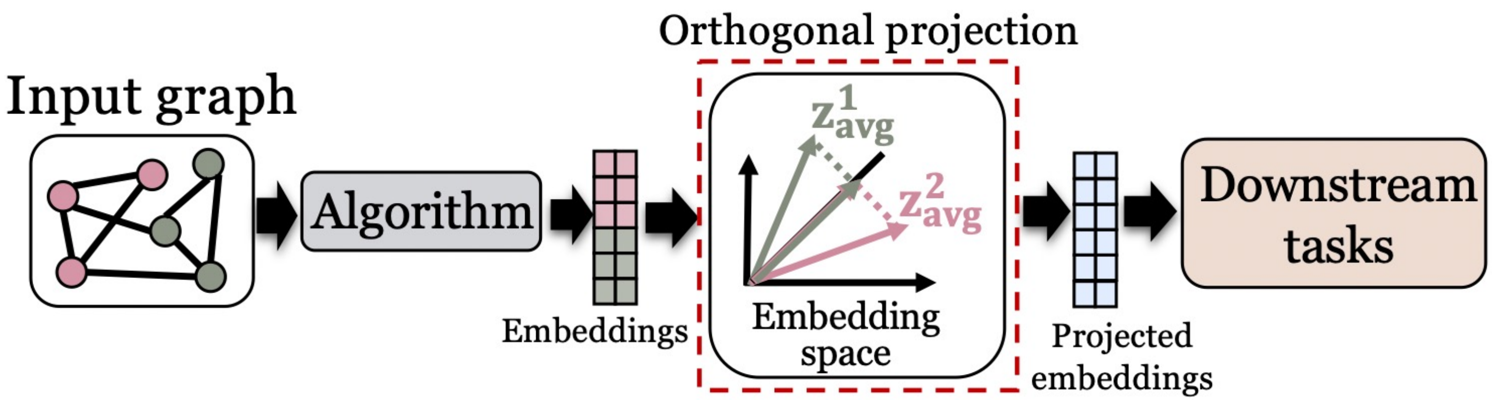


Rebalance certain statistics between different demographic subgroups to reduce their output difference from certain perspectives to achieve fairness.

Summary on Techniques Fulfilling Fairness

The taxonomy of techniques fulfilling fairness:



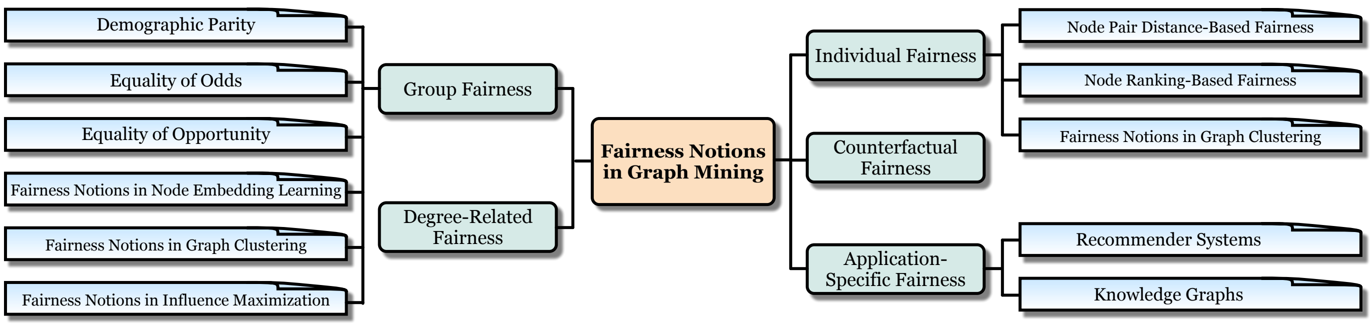


Project the learned embedding onto a hyper-plane that is orthogonal to the exhibited bias.

Problem 1: Insufficient Fairness Notions

- (1) The Insufficiency of Fairness Notions.

Can existing fairness notions help to avoid **all cases where people may feel unfair**?



Problem 2: Multiple Types of Fairness

- (1) The Insufficiency of Fairness Notions.
- (2) Fulfilling Multiple Types of Fairness.

How to achieve **multiple types of fairness**?

Problem 2: Multiple Types of Fairness

- (1) The Insufficiency of Fairness Notions.

- (2) Fulfilling Multiple Types of Fairness.

How to achieve **multiple types of fairness**?

Are some of the existing fairness notions in **conflict** with each other?

Problem 2: Multiple Types of Fairness

- (1) The Insufficiency of Fairness Notions.

- (2) Fulfilling Multiple Types of Fairness.

How to achieve **multiple types of fairness**?

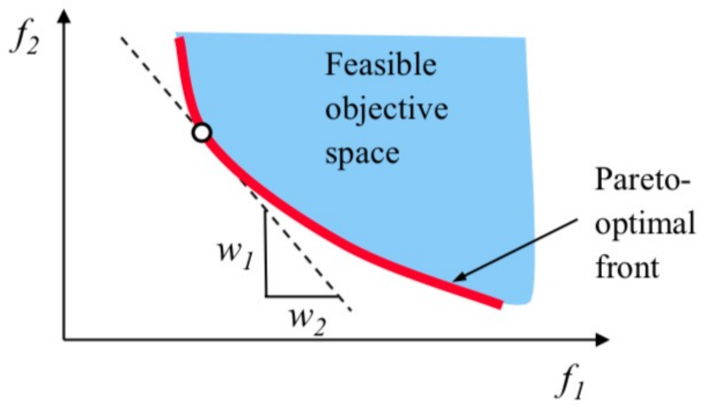
Are some of the existing fairness notions in **conflict** with each other?

If we could achieve multiple types of fairness, will people get a **stronger sense of fairness**? If not, what will be beneficial for social good?

Problem 3: Balance Utility and Fairness

- (1) The Insufficiency of Fairness Notions.
- (2) Fulfilling Multiple Types of Fairness.
- (3) Balancing Model Utility and Algorithmic Fairness.

How to achieve fairness **at low or no cost of utility**?

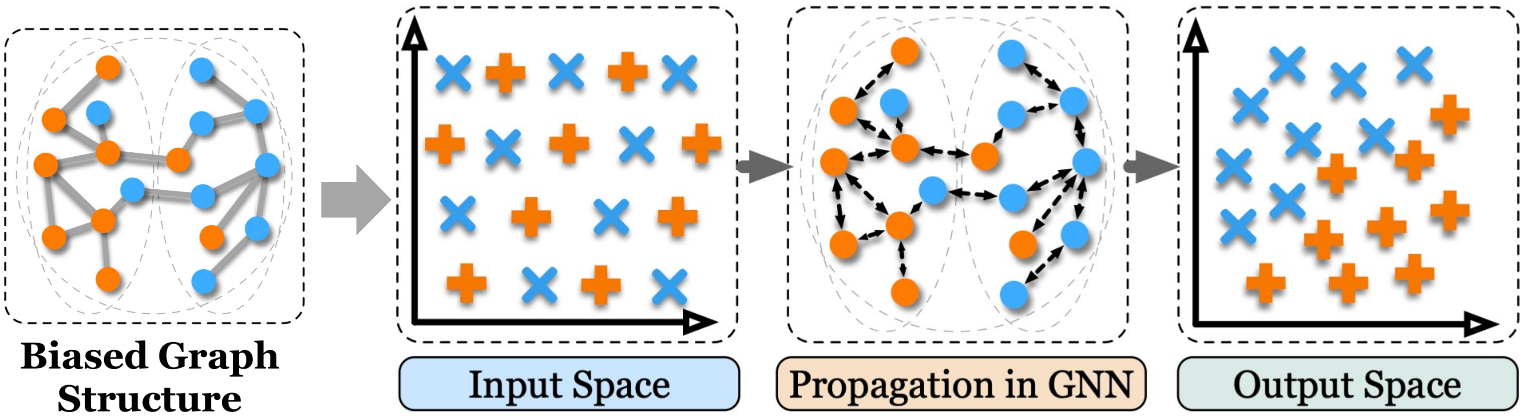


Problem 4: Explainability of Unfairness

- (1) The Insufficiency of Fairness Notions.
- (2) Fulfilling Multiple Types of Fairness.
- (3) Balancing Model Utility and Algorithmic Fairness.
- (4) Explaining How Unfairness Arises.

How to **interpret why unfairness arises** in graph mining algorithms?

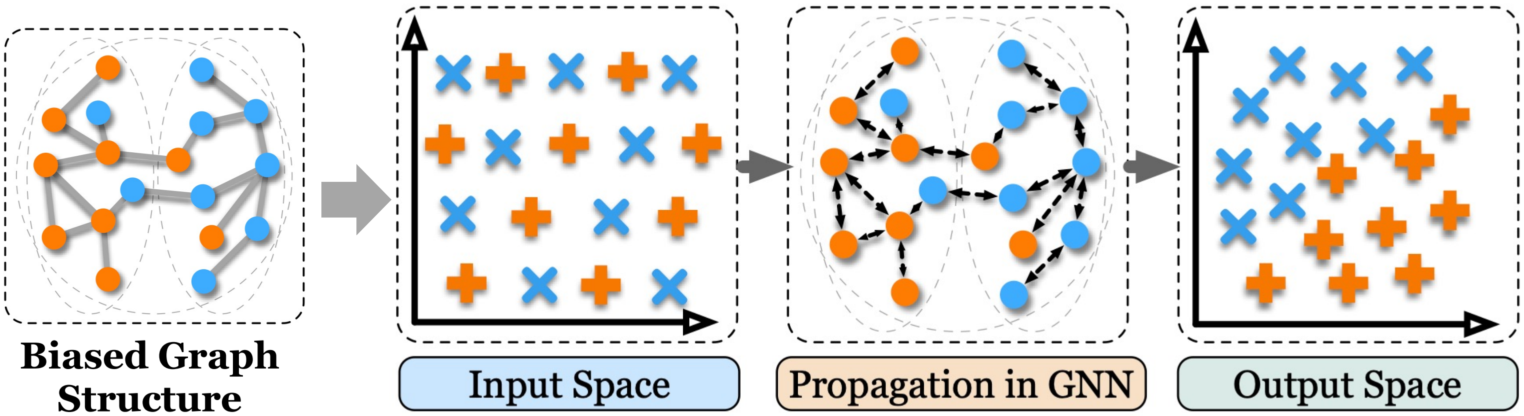
Problem 4: Explainability of Unfairness



- (4) Explaining How Unfairness Arises.

How to **interpret why unfairness arises** in graph mining algorithms?

Problem 4: Explainability of Unfairness



- (4) Explaining How Unfairness Arises.

How to **interpret why unfairness arises** in graph mining algorithms?

Is the **graph data** biased?

Is the **model** biased naturally?

Problem 5: Robustness on Fairness

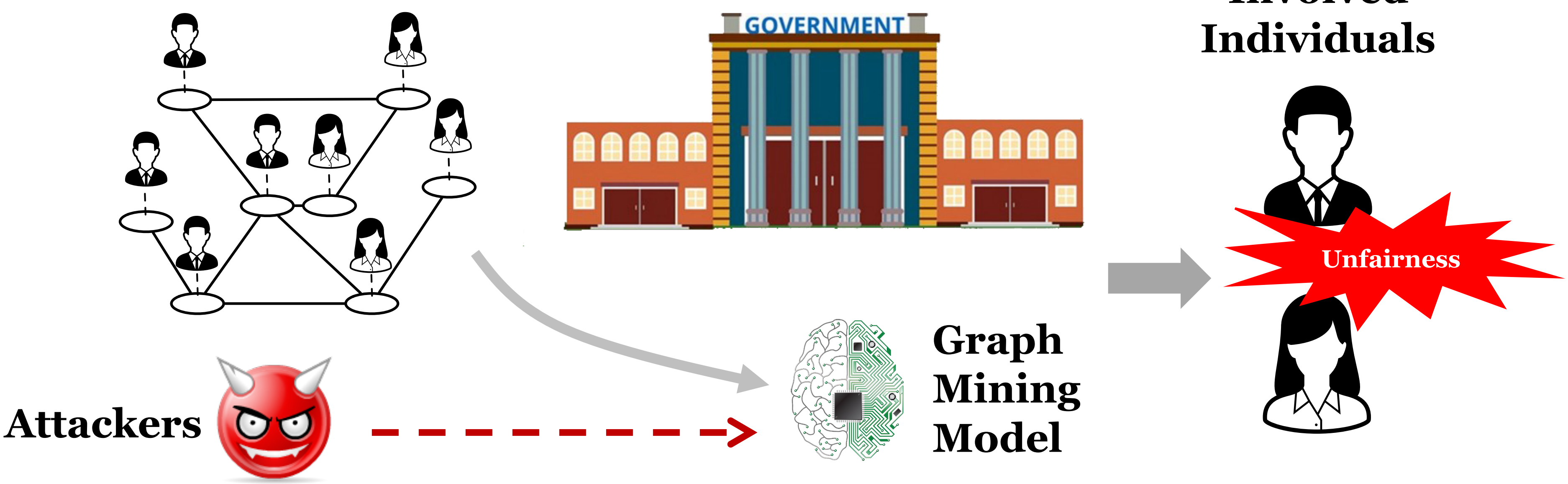
- (1) The Insufficiency of Fairness Notions.
- (2) Fulfilling Multiple Types of Fairness.
- (3) Balancing Model Utility and Algorithmic Fairness.
- (4) Explaining How Unfairness Arises.
- (5) Enhancing Robustness of Algorithms on Fairness.

How would existing graph mining algorithms perform in perspective of fairness **under malicious attack**?

How to achieve **better robustness** in perspective of fairness?

Problem 5: Robustness on Fairness

Consider the case where there are malicious attackers **whose goal is to induce bias** in the decisions made by the government.



- (5) Enhancing Robustness of Algorithms on Fairness.

How would existing graph mining algorithms perform in perspective of fairness **under malicious attack**?

How to achieve **better robustness** in perspective of fairness?

Q&A

Thanks for listening!