Biased Edge Dropout for Enhancing Fairness in Graph Representation Learning

Authors: I. Spinelli, S. Scardapane, A. Hussain, A. Uncini

Presenter: Indro Spinelli

SAPIENZA
UNIVERSITÀ DI ROMA

intelligent signal processing
and multimedia lab

Introduction

Fairness in Machine Learning

Fairness

Law

Sociology

Politics

Economics

Sport

Machine Learning

https://adioma.com/icons/equitable?term=Fairness

A broad definition

Fairness in decision making is broadly defined as the absence of any advantage or discrimination towards an individual or a **group** based on their traits (gender, ethnicity, age, ...).

A sensitive binary attribute **S** partitions the dataset into a protected minority and a majority.

This general definition leaves the door open to several different fairness metrics, each focused on a different type of discrimination.

Group Metrics

Demographic Parity

Satisfied if the likelihood of a positive outcome is the same regardless of the value of the sensitive attribute **S**.

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

Equalized Odds

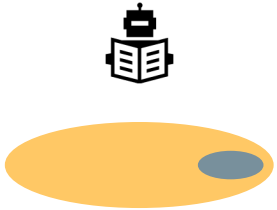
Satisfied if the rates for true positives and false positives between the two groups defined by **S** are the same.

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

For more definitions with examples check: https://developers.google.com/machine-learning/glossary/fairness

From biased data to biased model

When we design a data driven solution we want to minimize its error rate which depends on the number of samples.



Prioritize higher accuracy on the **majority** at the expense of the minority group.

Countermeasures:

Preprocess the data

Include a fairness constraint in the objective function

Postprocess model's predictions

https://adioma.com/icons/machine-learning

Just dont use that S uh?

The information about the sensitive attributes can influence the model even if not explicitly used as input.

Most of the times it is encoded in the combination of other attributes or in the **structure** of the dataset itself.

https://www.nbcnews.com/feature/nbc-out/facebook-can-tell-if-you-re-gay-based-few-likes-n823416

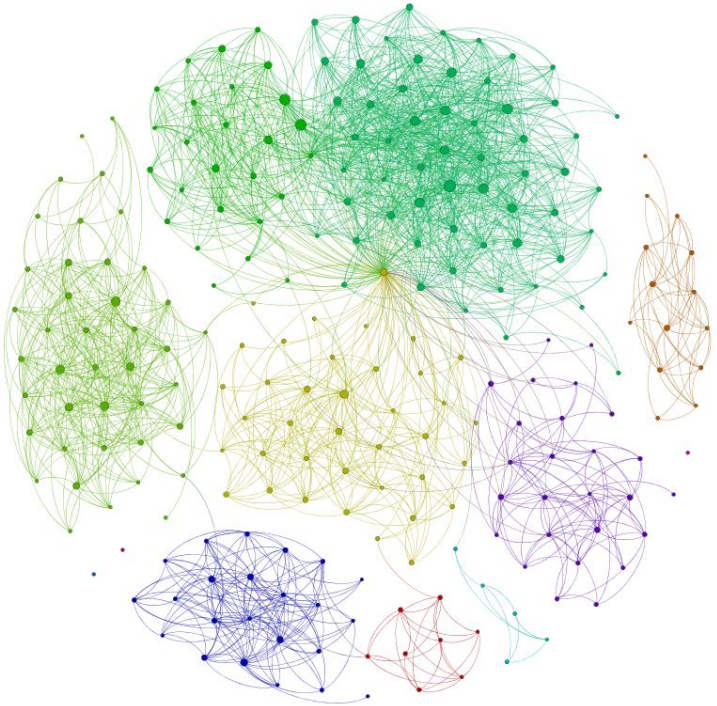
https://www.ox.ac.uk/news/2019-05-16-new-machine-learning-algorithm-can-predict-age-and-gender-just-your-twitter-profile

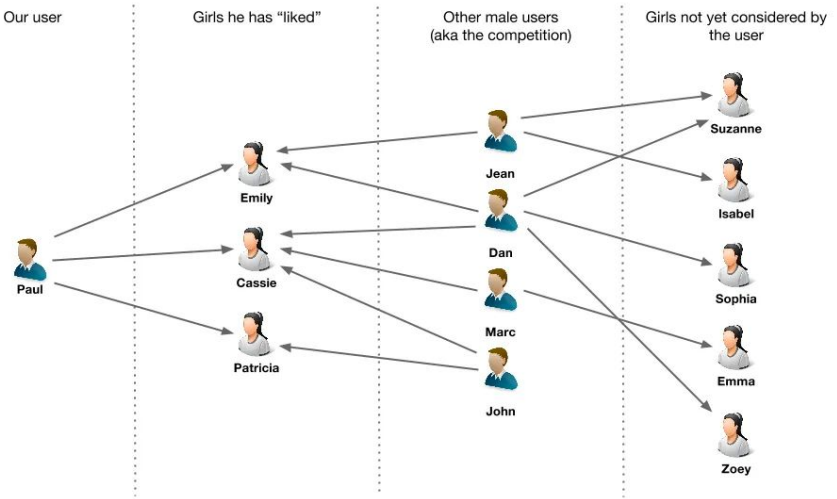
Graphs and Fairness

Social and recommender

Social Networks & Recommender Systems

(I just stole this slide from S.Scardapane)





http://allthingsgraphed.com/2014/08/28/facebook-friends-network/

https://linkurio.us/blog/using-neo4j-to-build-a-recommendation-engine-based-on-collaborative-filtering/

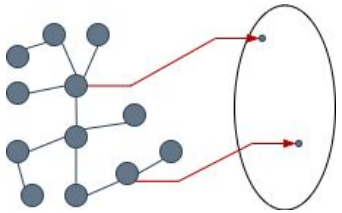
Common tasks

Link Prediction

Supervised

Node Representation Learning

Unsupervised



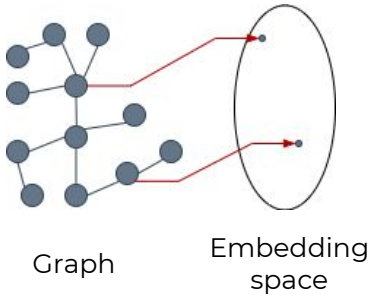
Common tasks

Link Prediction

Supervised

Node Representation Learning

Unsupervised



Homophily

The information of the sensitive attribute can be expressed implicitly by the data.

The tendency of similar nodes to cluster on several real-world graphs (i.e., **homophily**) is a well know source of unfairness. (Birds of a Feather, McPherson et al., 2001)

Homophily is the principle that similar users interact at a higher rate than dissimilar ones.
In a graph, this means that nodes with similar characteristics are more likely to be connected.

Unfair homophily at work

In social networks, the “unfair homophily” of race, gender or nationality, limits the contents accessible by the users, influencing their online behaviour.

For recommender systems and online ad targeting Google has an entire page about….

“ Algorithms don’t remember incidents of unfair bias. But customers do.”

“Once that customer trust is lost, there’s no guarantee brands can get it back.”

https://www.thinkwithgoogle.com/feature/ml-fairness-for-marketers/ https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation

Unfair model makes it worse

A ML model trained on biased data most likely will amplify and consolidate these biases.

An unfair link prediction on a social network increases the segregation of the network isolating the users in their own cultural or ideological bubbles (**filter bubble**).

Similarly, a recommender systems that learns to associate a product to the majority group is subject to a feedback loop. The more users interact with the product, the more the algorithm will reinforce its bias.

FairDrop

A biased edge dropout

FairDrop

We propose a biased edge dropout algorithm (**FairDrop**) to counter-act homophily and improve fairness in graph representation learning.

- Easy to integrate into existing algorithms.
- It’s flexible, yet it has just 1 hyperparameter.
- It’s efficient.

FairDrop

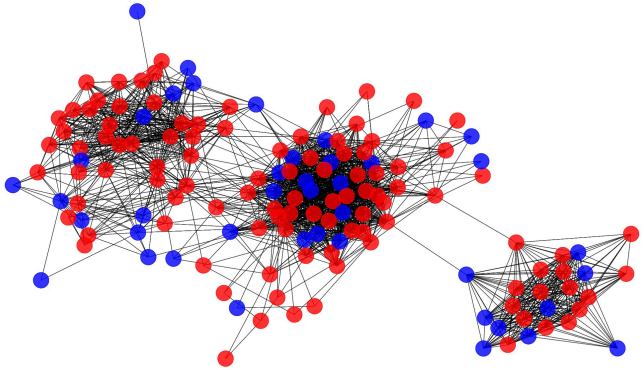
We propose a biased edge dropout algorithm (**FairDrop**) to counter-act homophily and improve fairness in graph representation learning.

- Easy to integrate into existing algorithms.
- It’s flexible, yet it has just 1 hyperparameter.
- It’s efficient.

Tested: Link Prediction

Node Representation Learning

1° Step



Small ego-network of the Facebook dataset.

RED: majority gender

BLUE: minority gender

We build an adjacency matrix $\mathbf{M}$ encoding the heterogeneous connections between the sensitive attributes.

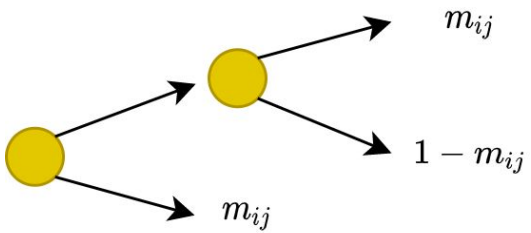
$$m_{ij} = \begin{cases} 1 & \text{if } s_i \neq s_j \\ 0 & \text{otherwise} \end{cases}$$

https://snap.stanford.edu/data/egonets-Facebook.html

2° Step

Randomized response used to **inject randomness** and to regulate the bias imposed by the constraint.

Randomized response adds a layer of “plausible deniability” for structured interviews. With this protection, respondents are able to answer sensitive issues while maintaining their privacy.



2° Step

Randomized response used to inject randomness and to **regulate the bias** imposed by the constraint.

$$rr(m_{ij}) = \begin{cases} m_{ij} & \text{with probability} : \frac{1}{2} + \delta \\ 1 - m_{ij} & \text{with probability} : \frac{1}{2} - \delta \end{cases}$$

2° Step

Randomized response used to inject randomness and to **regulate the bias** imposed by the constraint.

$$rr(m_{ij}) = \begin{cases} m_{ij} & \text{with probability : } \frac{1}{2} + \delta \\ 1 - m_{ij} & \text{with probability : } \frac{1}{2} - \delta \end{cases}$$

$\delta = 0$

Randomized response will always give random answers, maximizing the privacy protection but making the acquired data useless.

2° Step

Randomized response used to inject randomness and to **regulate the bias** imposed by the constraint.

$$rr(m_{ij}) = \begin{cases} m_{ij} & \text{with probability : } \frac{1}{2} + \delta \\ 1 - m_{ij} & \text{with probability : } \frac{1}{2} - \delta \end{cases}$$

$\delta = 0$

Randomized response will always give random answers, maximizing the privacy protection but making the acquired data useless.

$\delta = 1/2$

Randomized response will always give the true answers, minimizing the privacy protection but making the most of the acquired data.

3° Step

Finally, we are ready to drop the unfair connection from the original adjacency matrix.

$$\mathbf{A}_{fair} = \mathbf{A} \circ rr(\mathbf{M})$$

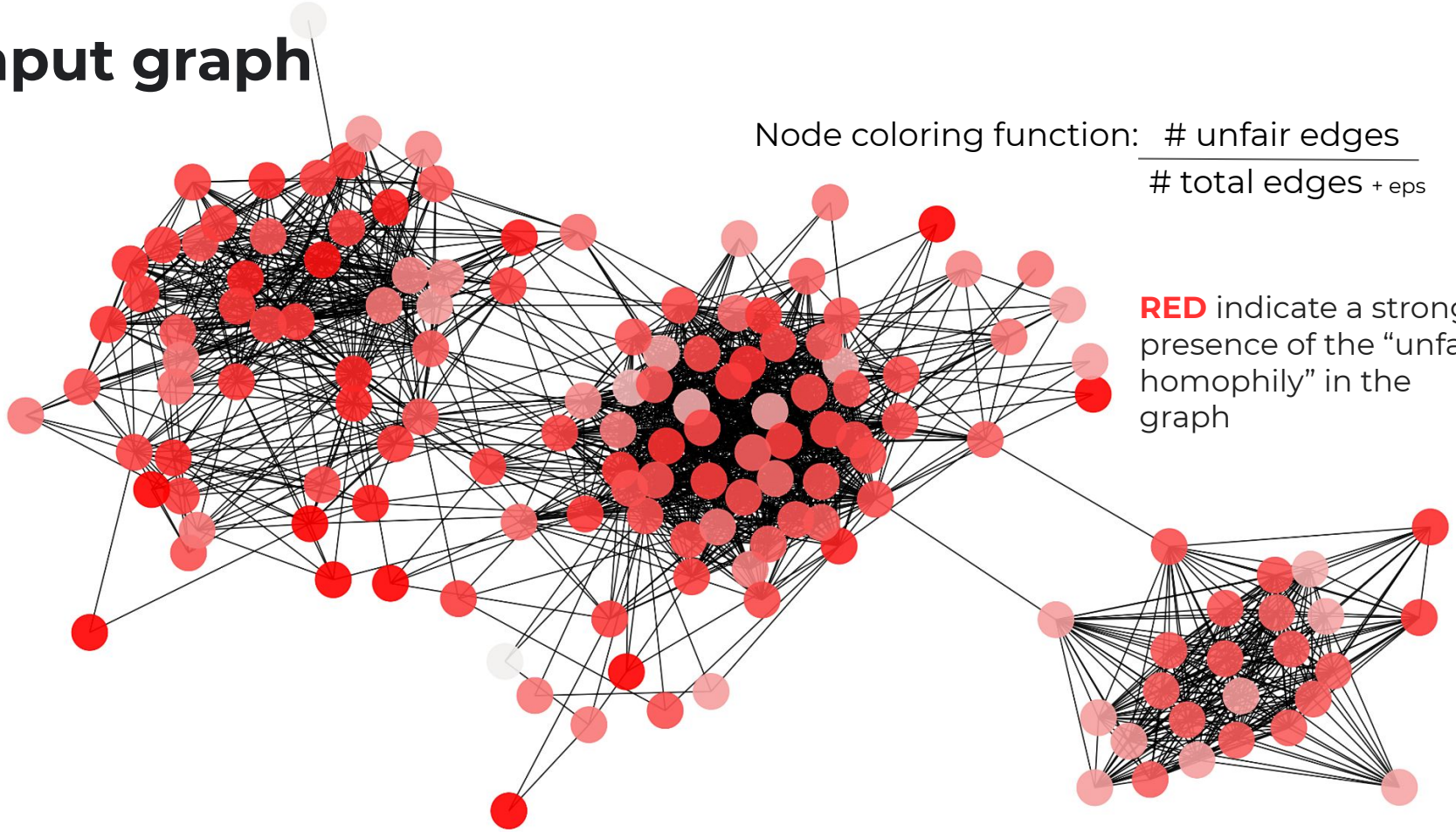
$\delta = 0$

Corresponds to an unbiased edge dropout.

$\delta = 1 / 2$

Removes all the edges representing the “unfair homophily” and keeps all the fair ones.

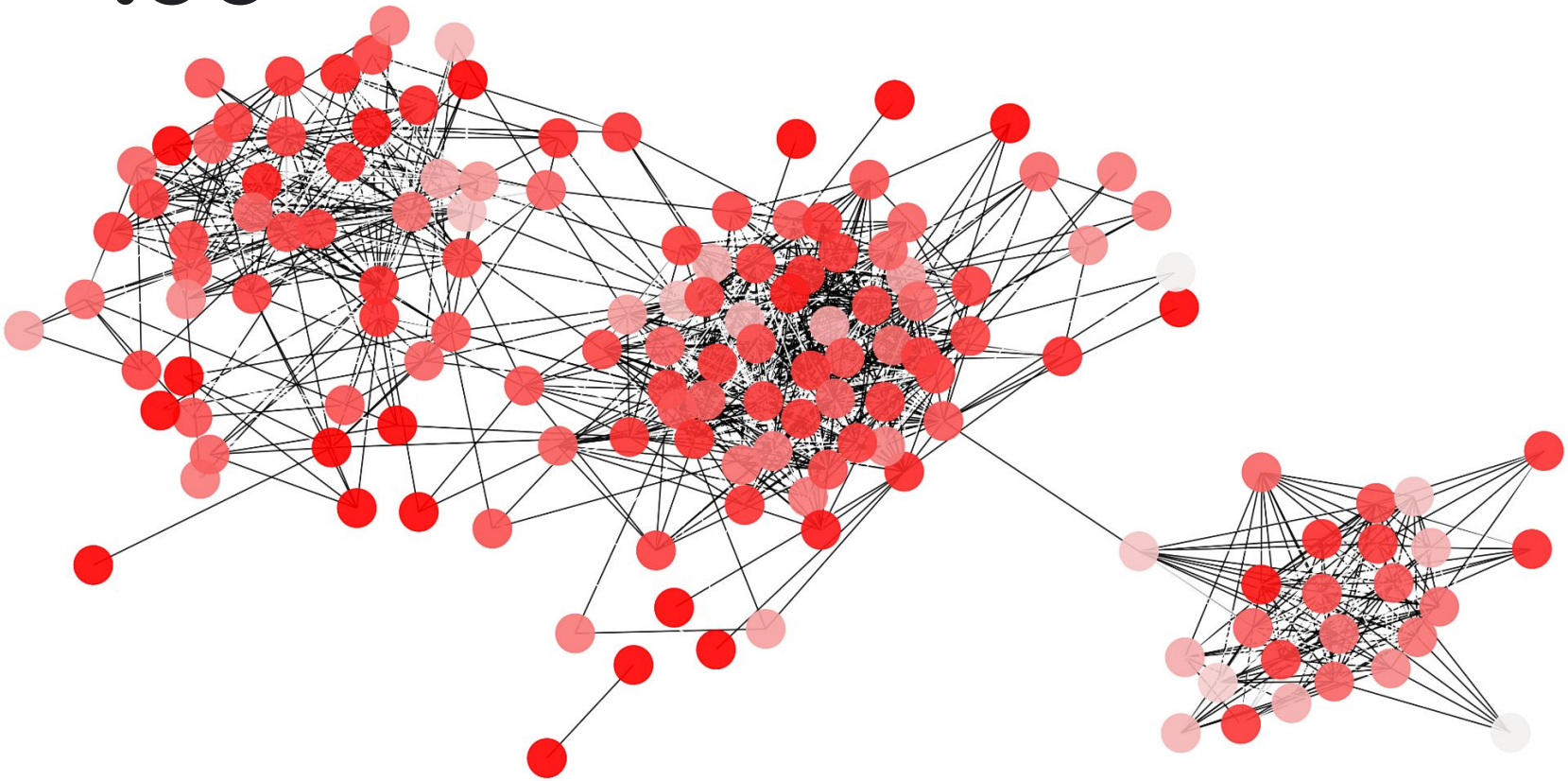
Input graph



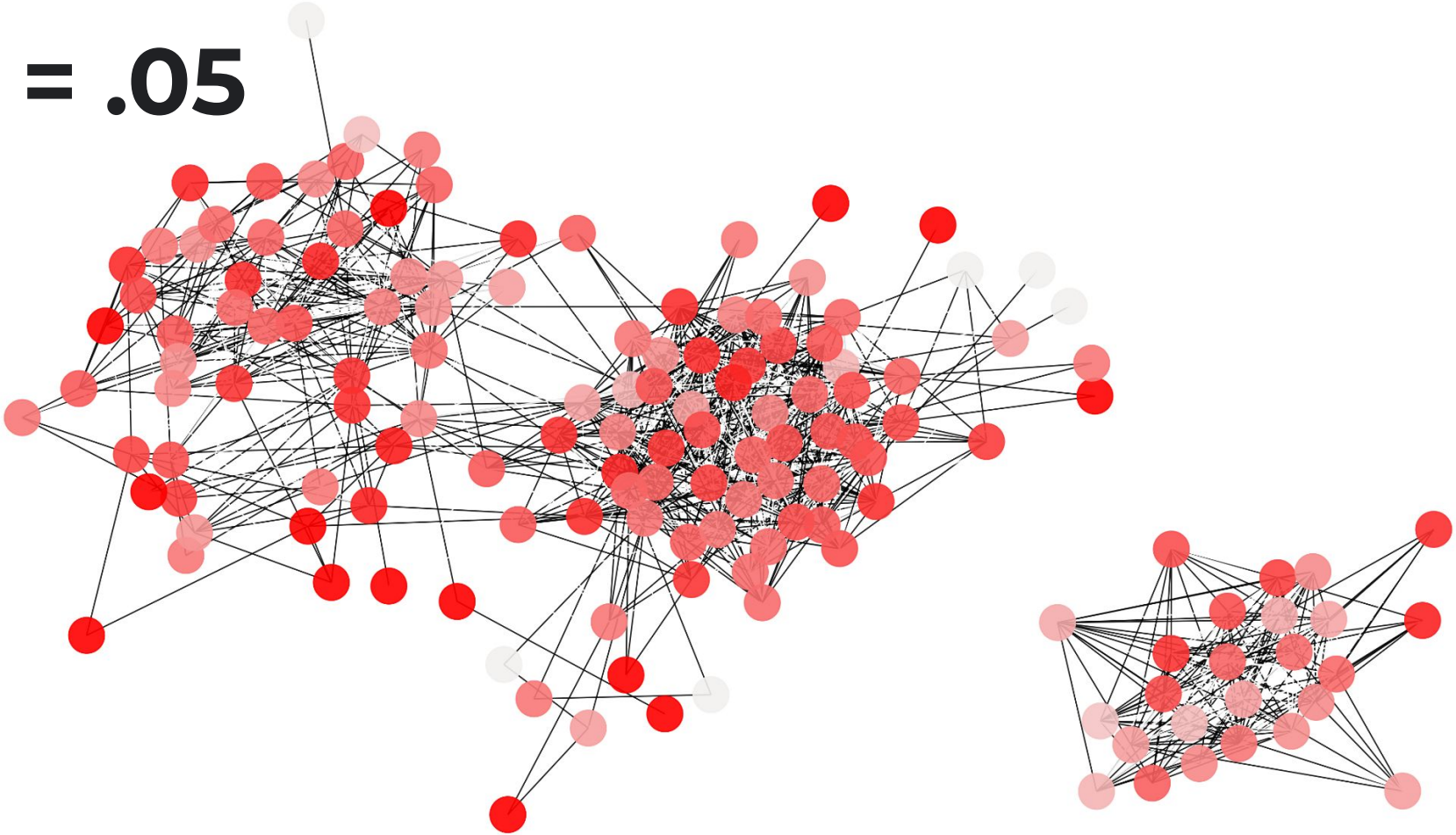
Node coloring function: $\frac{\text{\# unfair edges}}{\text{\# total edges} + \text{eps}}$

RED indicate a strong presence of the “unfair homophily” in the graph

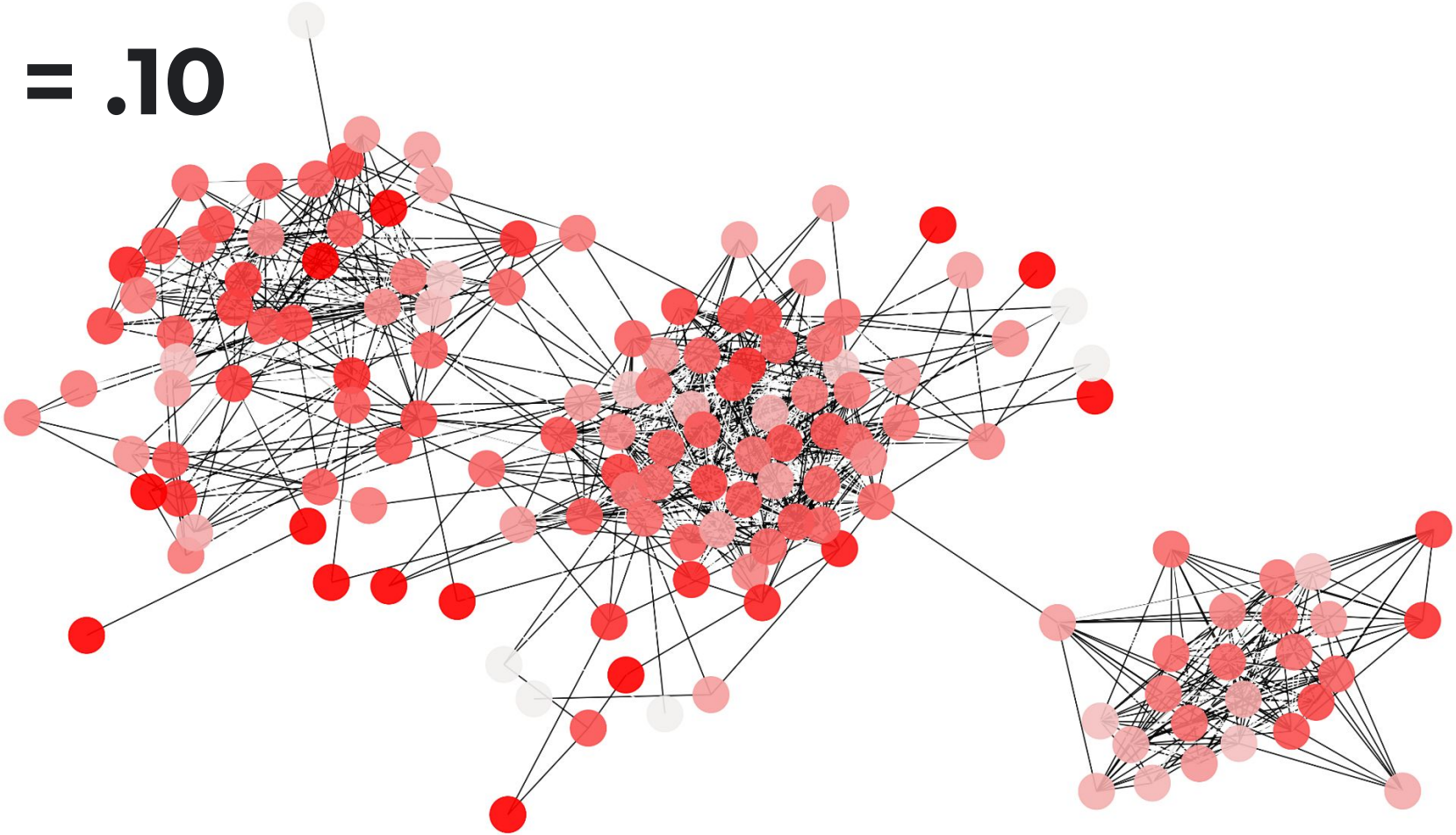
δ = .00



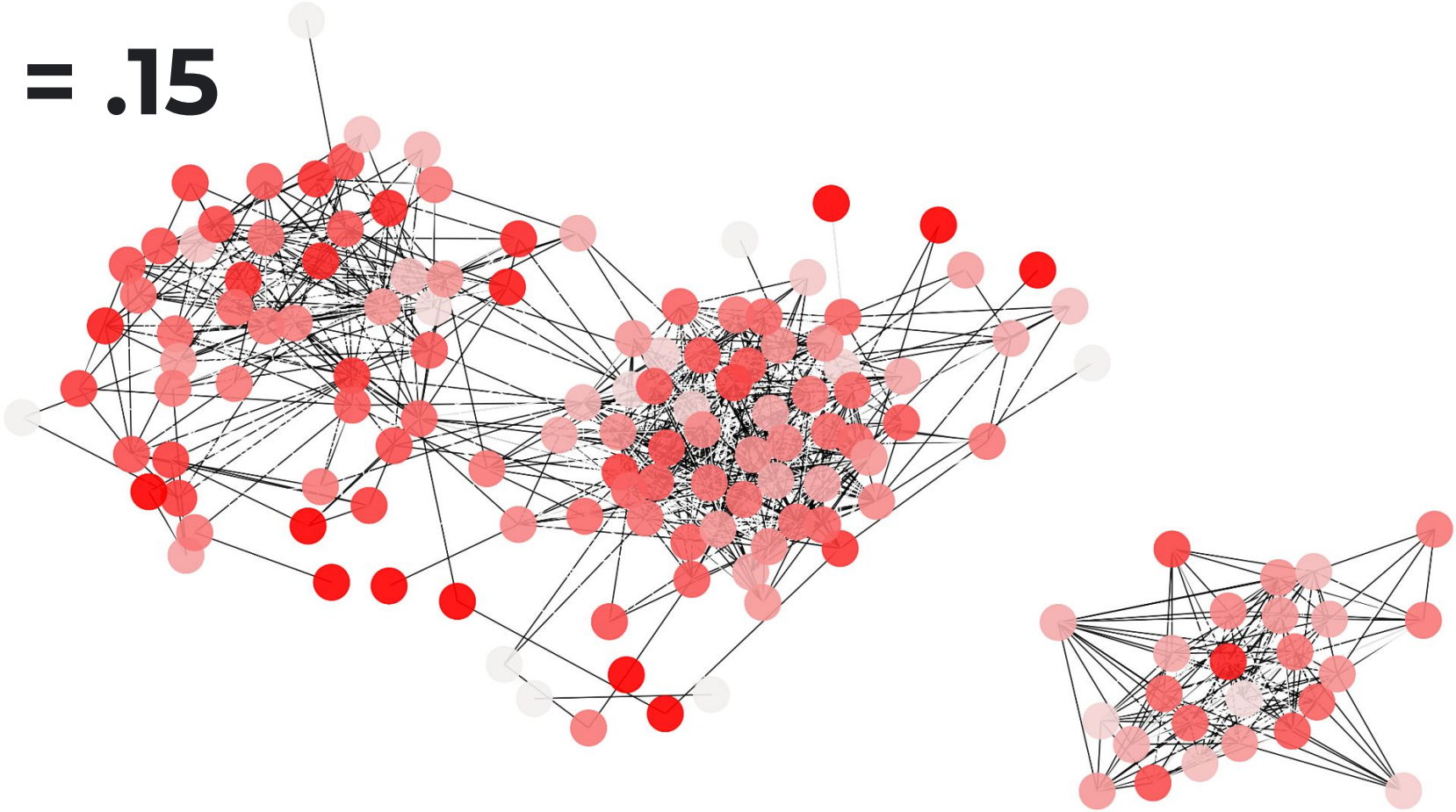
δ = .05



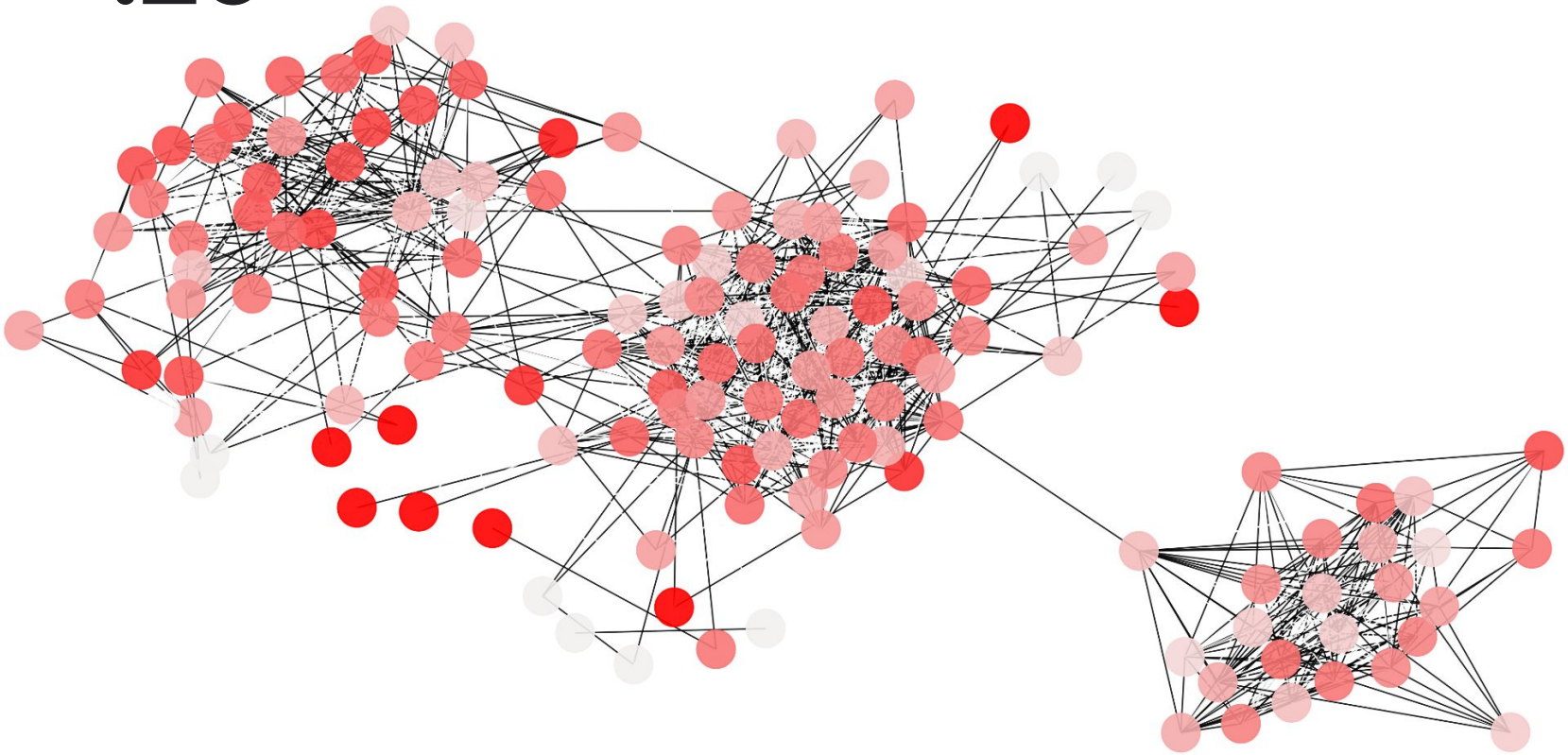
δ = .10



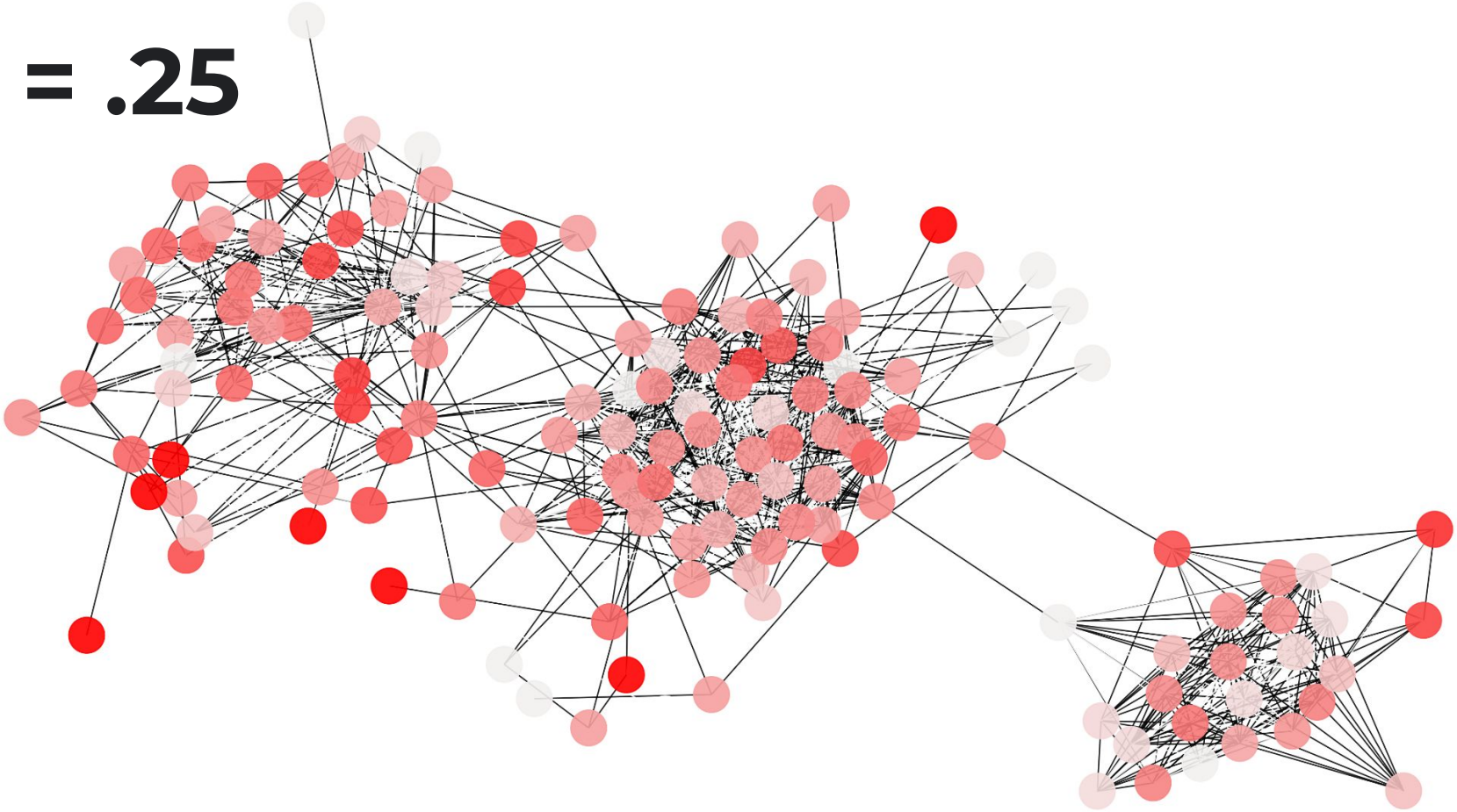
$\delta = .15$



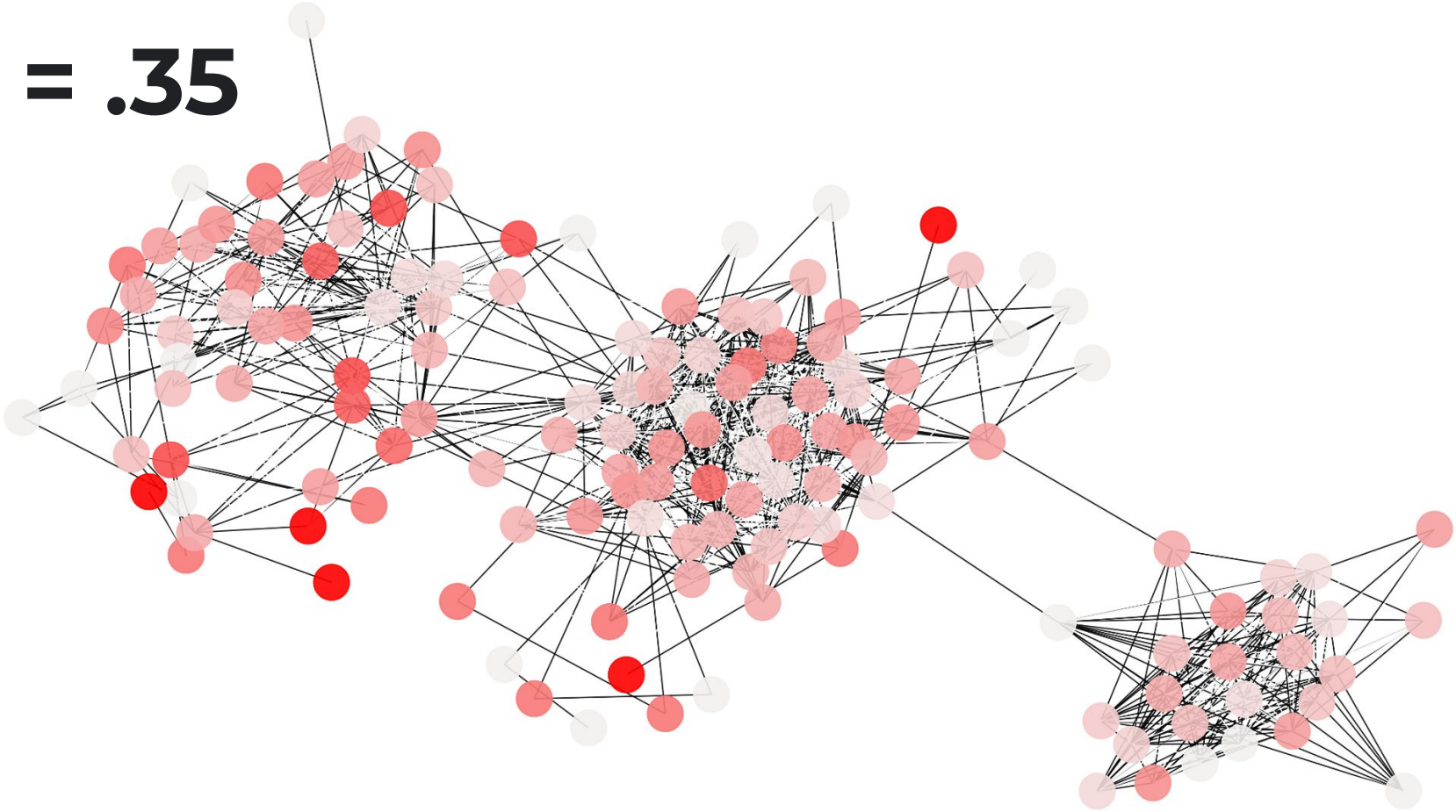
δ = .20



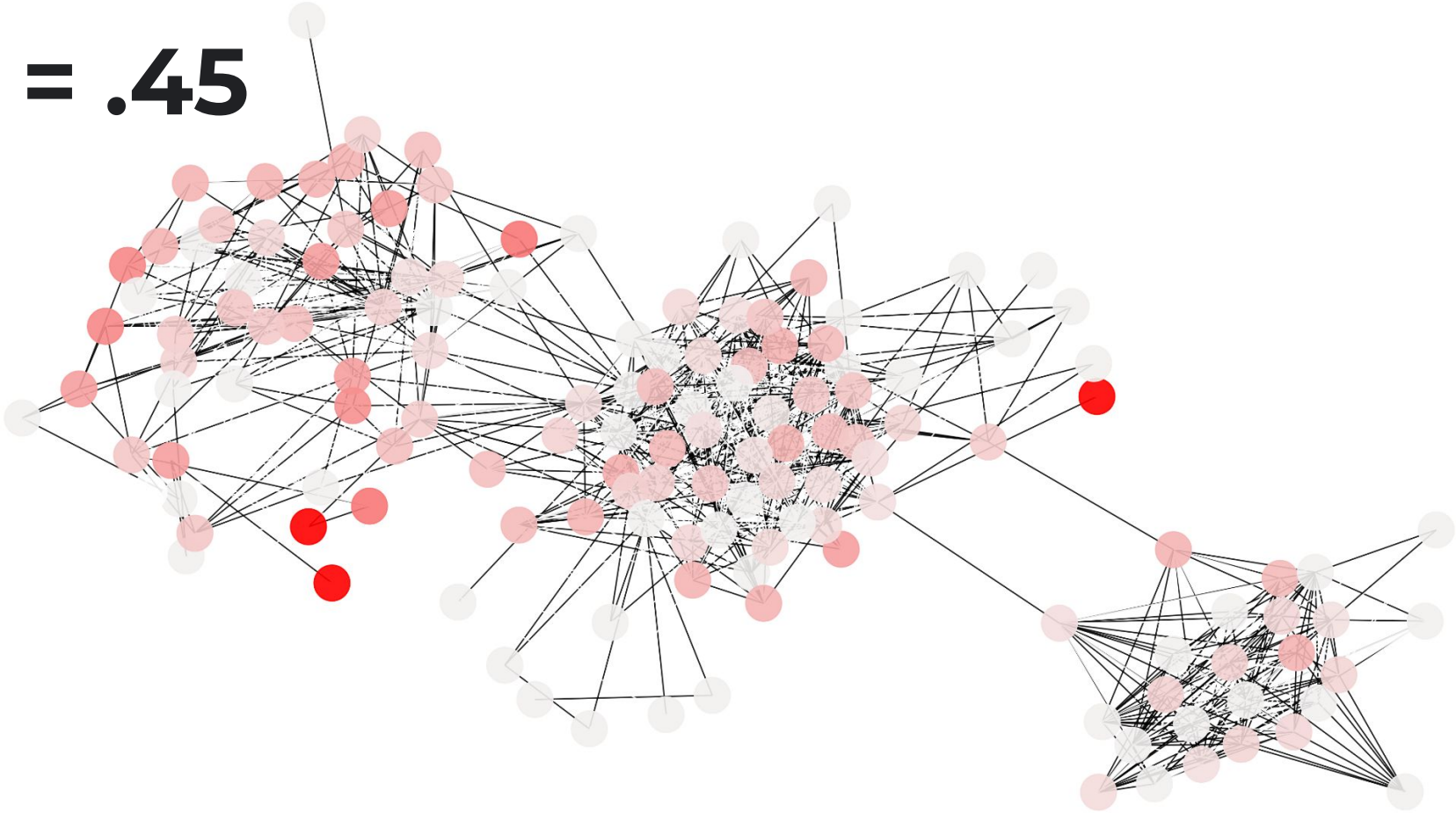
δ = .25



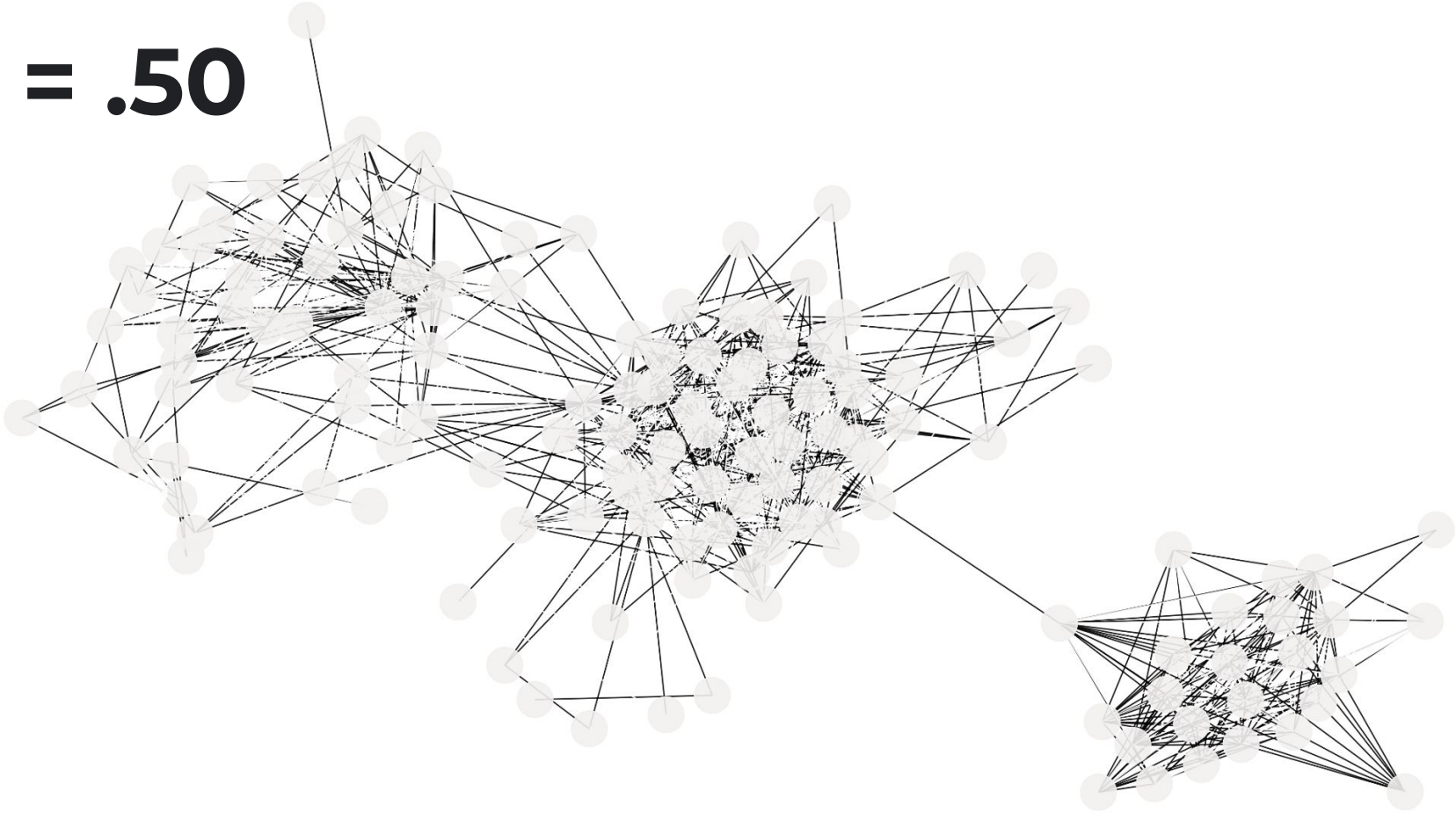
$\delta = .35$



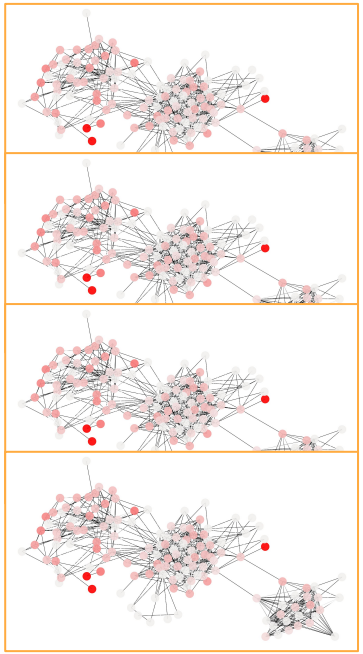
$\delta = .45$



δ = .50



Deployment



Fair node embeddings

This may create a small or negligible drop in accuracy

Fair link predictions

https://adioma.com/icons/machine-learning

Fairness Measures

For graphs

Two tasks, more than two measures

Node representation learning

Is fair if it obfuscates as much as possible the unfair presence of sensitive attributes in the resulting embeddings.

Use a classifier to predict the sensitive attribute. If it yields the same accuracy as a random classifier the embeddings are fair and the downstream task built upon them should too.

Link prediction

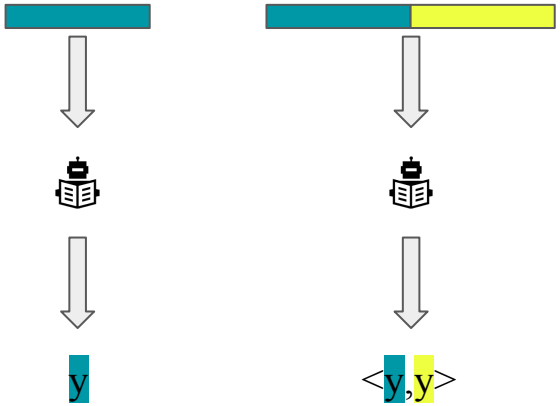
Issue: an edge connects two nodes and thus two sensitive attributes. It is impossible to apply the group fairness used for i.i.d. data because of this **dyadic** relation.

Define dyadic groups that maps the sensitive attributes from the node to the edges. Evaluate the fairness of the link prediction by using common metrics for i.i.d data.

Our contributions

Node representation learning

We proposed a modification to the current Representation Bias (RB) measure to take into account the additional information provided by the presence of a link



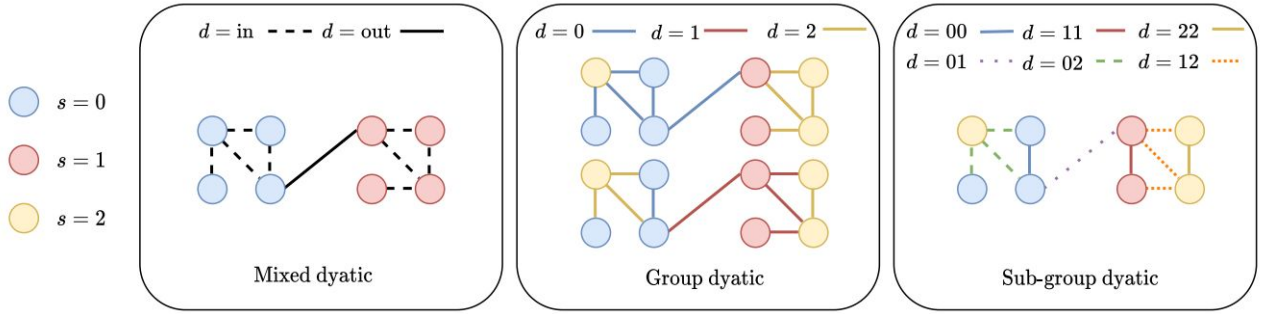
https://adioma.com/icons/machine-learning

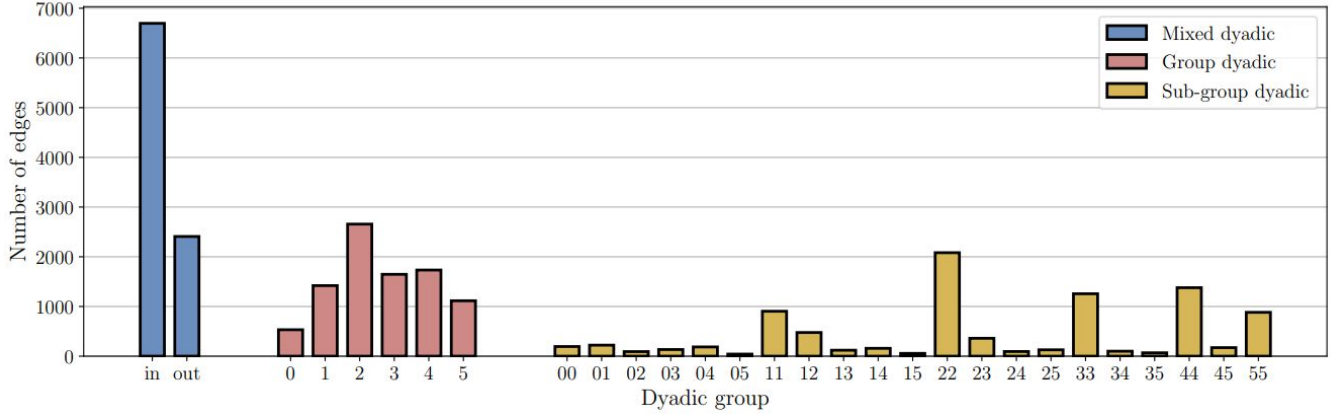
Link prediction

How we define the dyadic groups strongly affects our measures.
Two dyadics groups are already been discussed in the literature **mixed** dyadic and **sub-group** dyadic.

We propose a third dyadic group definition called **group** dyadic that aims to be closer to the original groups defined by the sensitive attribute.

Dyadic groups





Two tasks, more than two measures

Node representation learning on FB

Method	Classifier	Node RB ↓	Link RB ↓	Accuracy ↑
DeBayes	LR	52.0 ± 0.5	**55.7** ± 0.5	96.7 ± 0.1
	NN	58.9 ± 1.3	**91.6** ± 0.5	**98.5** ± 0.2
	RF	53.4 ± 0.6	62.7 ± 0.6	**97.9** ± 0.2
FairWalk	LR	53.1 ± 0.6	63.3 ± 1.0	**97.4** ± 0.2
	NN	58.9 ± 1.3	99.5 ± 0.2	97.6 ± 0.3
	RF	53.5 ± 0.6	62.3 ± 0.8	97.0 ± 0.2
Node2Vec+FairDrop	LR	**51.3** ± 1.0	62.3 ± 0.8	96.0 ± 0.2
	NN	**49.9** ± 1.3	99.0 ± 0.1	96.7 ± 0.5
	RF	**50.0** ± 0.0	**61.1** ± 0.4	96.4 ± 0.2

https://snap.stanford.edu/data/egonets-Facebook.html

Link prediction on Pubmed

Method	Accuracy ↑	AUC ↑
GCN+EdgeDrop	88.0 ± 0.5	94.6 ± 0.3
GAT+EdgeDrop	80.6 ± 0.9	88.8 ± 0.7
GCN+FairDrop	**88.4** ± 0.4	**94.8** ± 0.2
GAT+FairDrop	79.0 ± 0.8	87.6 ± 0.7

Mixed

$\Delta DP_m \downarrow$	$\Delta EO_m \downarrow$
43.7 ± 1.0	12.8 ± 0.8
43.5 ± 1.1	24.5 ± 1.9
42.5 ± 0.5	**12.2** ± 0.7
37.4 ± 0.9	19.7 ± 1.1

Group

$\Delta DP_g \downarrow$	$\Delta EO_g \downarrow$
6.3 ± 0.7	6.0 ± 1.1
4.8 ± 1.6	7.5 ± 1.5
5.6 ± 1.8	**5.1** ± 0.9
2.0 ± 1.0	6.4 ± 1.4

Sub-group

$\Delta DP_s \downarrow$	$\Delta EO_s \downarrow$
57.5 ± 1.4	**26.3** ± 2.3
60.1 ± 1.9	49.3 ± 3.6
55.7 ± 1.5	26.6 ± 2.6
56.8 ± 2.1	47.3 ± 4.1

https://pytorch-geometric.readthedocs.io/en/latest/modules/datasets.html

Conclusions

Conclusion

Pros:

Easy to deploy

Flexible

Efficient

Cons:

When we have **N** sensitive attributes we have to create **N** “unfair homophily” mask and **alternate** them during the training process.

The concept of homophily is different for **bipartite** graphs and FairDrop cannot be applied as it is.

Pros of the Cons:

We are working on that.

General Con:

Lack of dedicated datasets.

Thanks for listening!

Indro Spinelli
PhD Student

https://spindro.github.io/

https://twitter.com/IndroSpinelli

Simone Scardapane
Assistant Professor

https://www.sscardapane.it/

https://twitter.com/s_scardapane

Amir Hussain
Full Professor

https://www.napier.ac.uk/people/amir-hussain

Aurelio Uncini
Full Professor

http://www.uncini.com/

Many thanks to @thegautamkamath @deliprao