Inequality II.

Networks and Algorithms

Lisette Espín-Noboa | 31.May.2023 | TU Wien

What is inequality?

≠

What is inequality…

…in general?

What is inequality…

…in general?

"A **lack of equality** or fair treatment in the sharing of wealth or opportunities"

AMERICAN (Cambridge dictionary)

What is inequality…

…in general?

“A **lack of equality** or fair treatment in the sharing of wealth or opportunities”

AMERICAN (Cambridge dictionary)

“A situation in which **money or opportunities are not shared equally** between different groups **in society**”

BUSINESS English (Cambridge dictionary)

What is inequality…

…in general?

“A **lack of equality** or fair treatment in the sharing of wealth or opportunities”

AMERICAN (Cambridge dictionary)

“A situation in which **money or opportunities are not shared equally** between different groups **in society**”

BUSINESS English (Cambridge dictionary)

“The relation between two expressions that are **not equal**, employing a sign such as ≠ ‘not equal to’, > ‘greater than’, or < ‘less than’”

Oxford Languages

What is inequality…

…in general?

"A **lack of equality** or fair treatment in the sharing of wealth or opportunities"

AMERICAN (Cambridge dictionary)

"A situation in which **money or opportunities are not shared equally** between different groups **in society**"

BUSINESS English (Cambridge dictionary)

"The relation between two expressions that are **not equal**, employing a sign such as ≠ 'not equal to', > 'greater than', or < 'less than'"

Oxford Languages

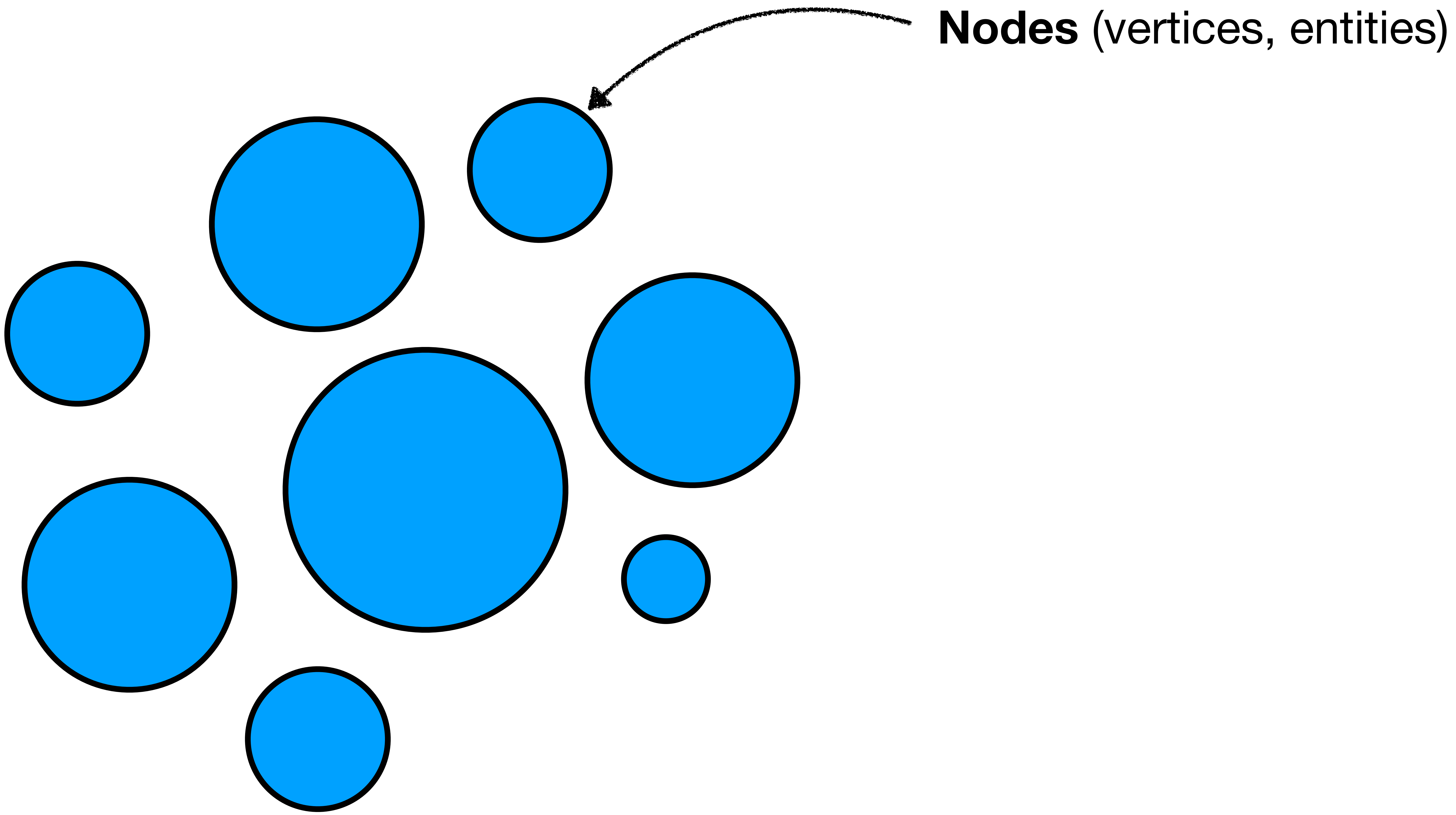
Inequality is also known as: imbalance, disproportion, dissimilarity, difference, <u>bias</u>.

What is inequality (bias)…

…in (social) networks?

What is inequality (bias)…

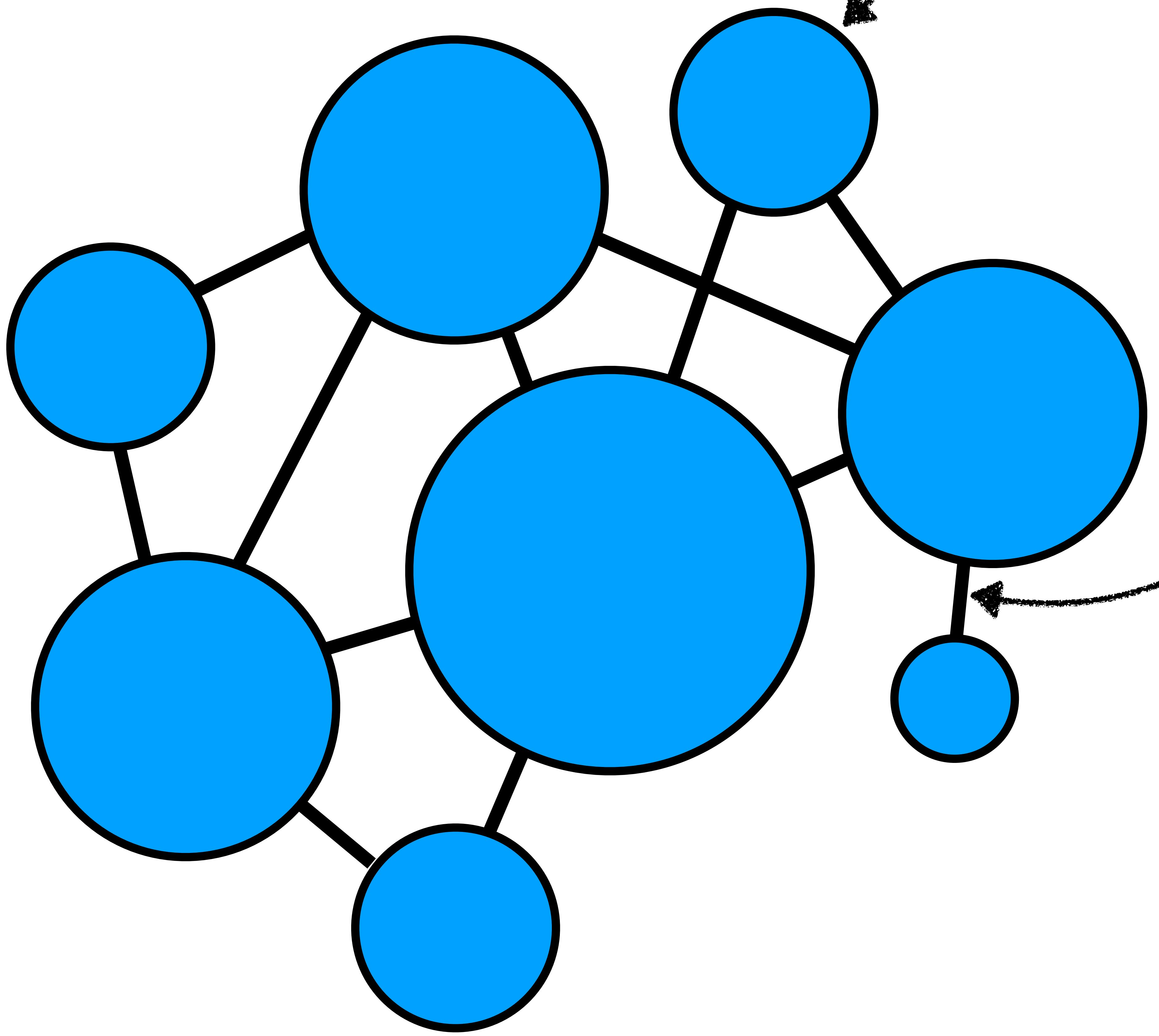
…in (social) networks?



What is inequality (bias)…

…in (social) networks?

Nodes (vertices, entities)

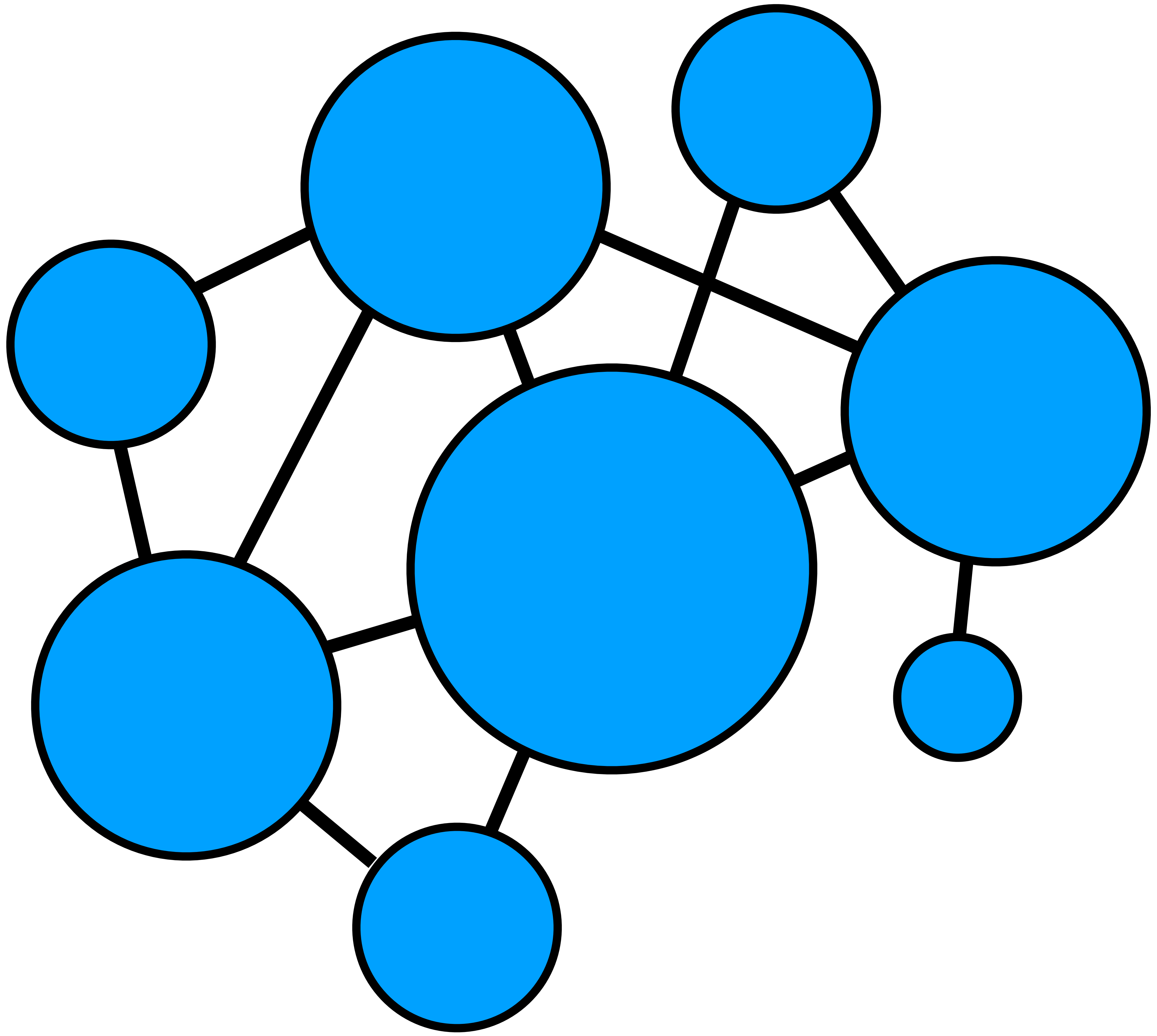


Links (edges, ties, relations)

What is inequality (bias)…

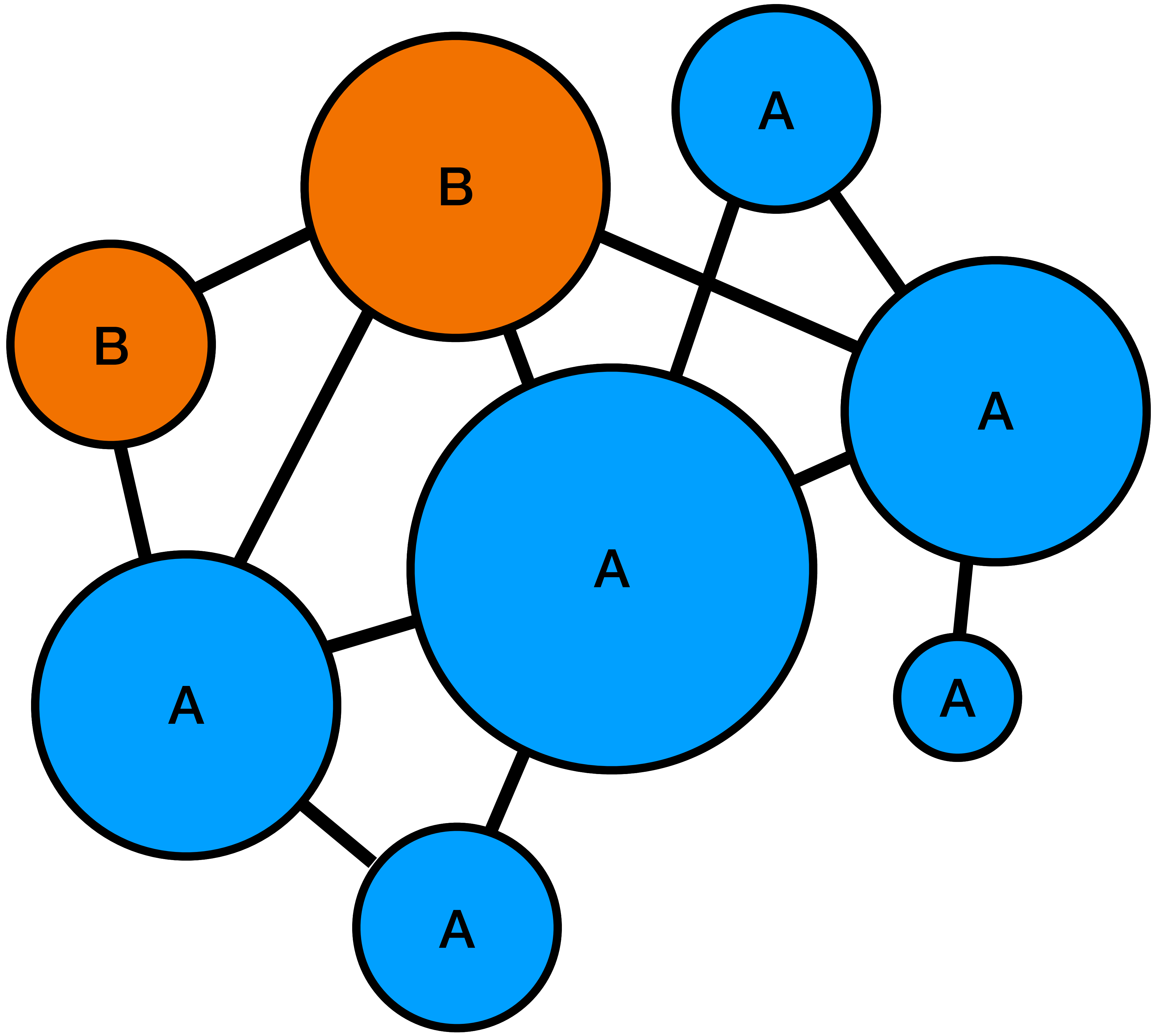
…in (social) networks?

Biases at a node level:



What is inequality (bias)…

…in (social) networks?

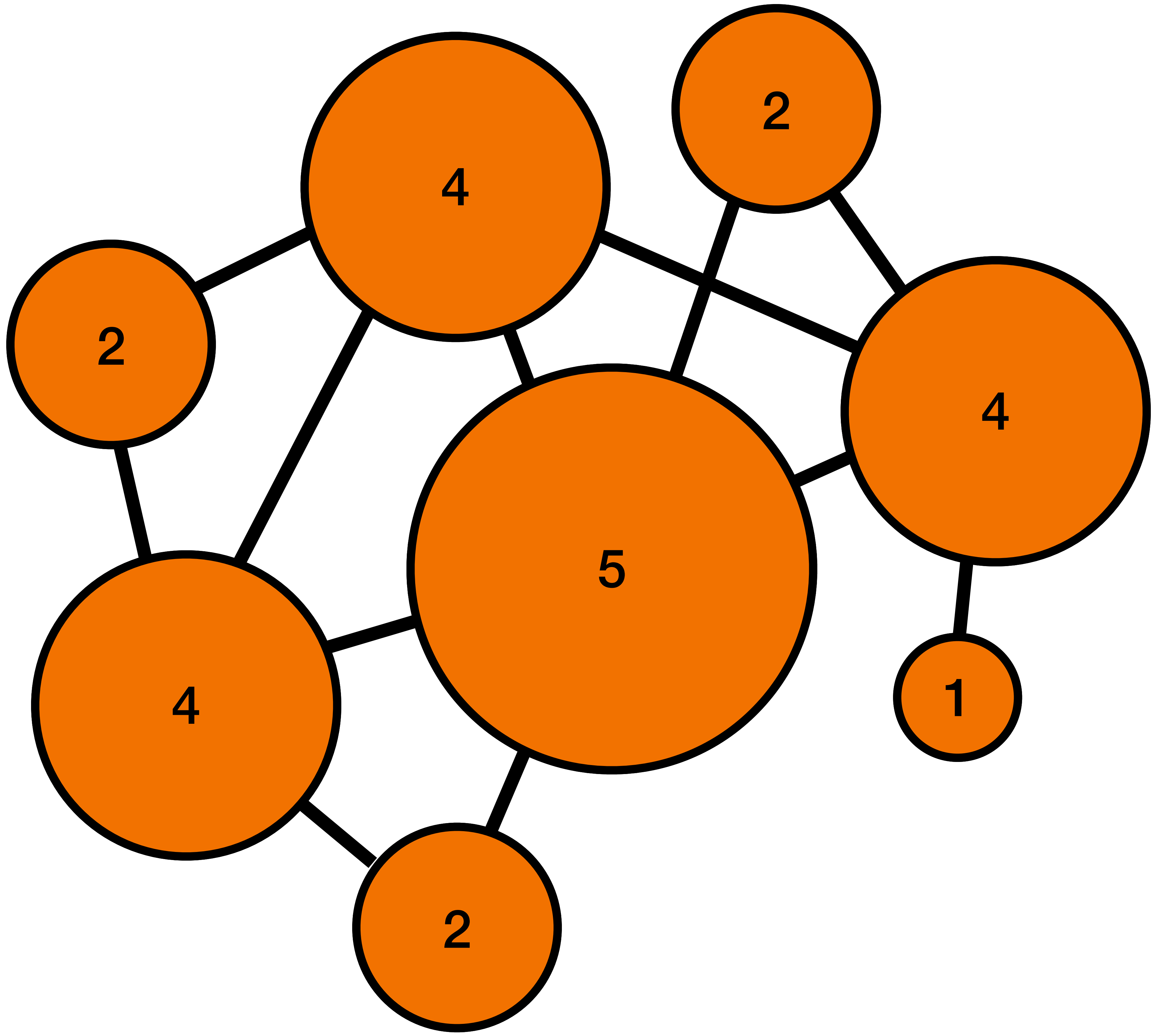


Biases at a node level:

Class imbalance (e.g., majority vs. minority)

What is inequality (bias)…

…in (social) networks?



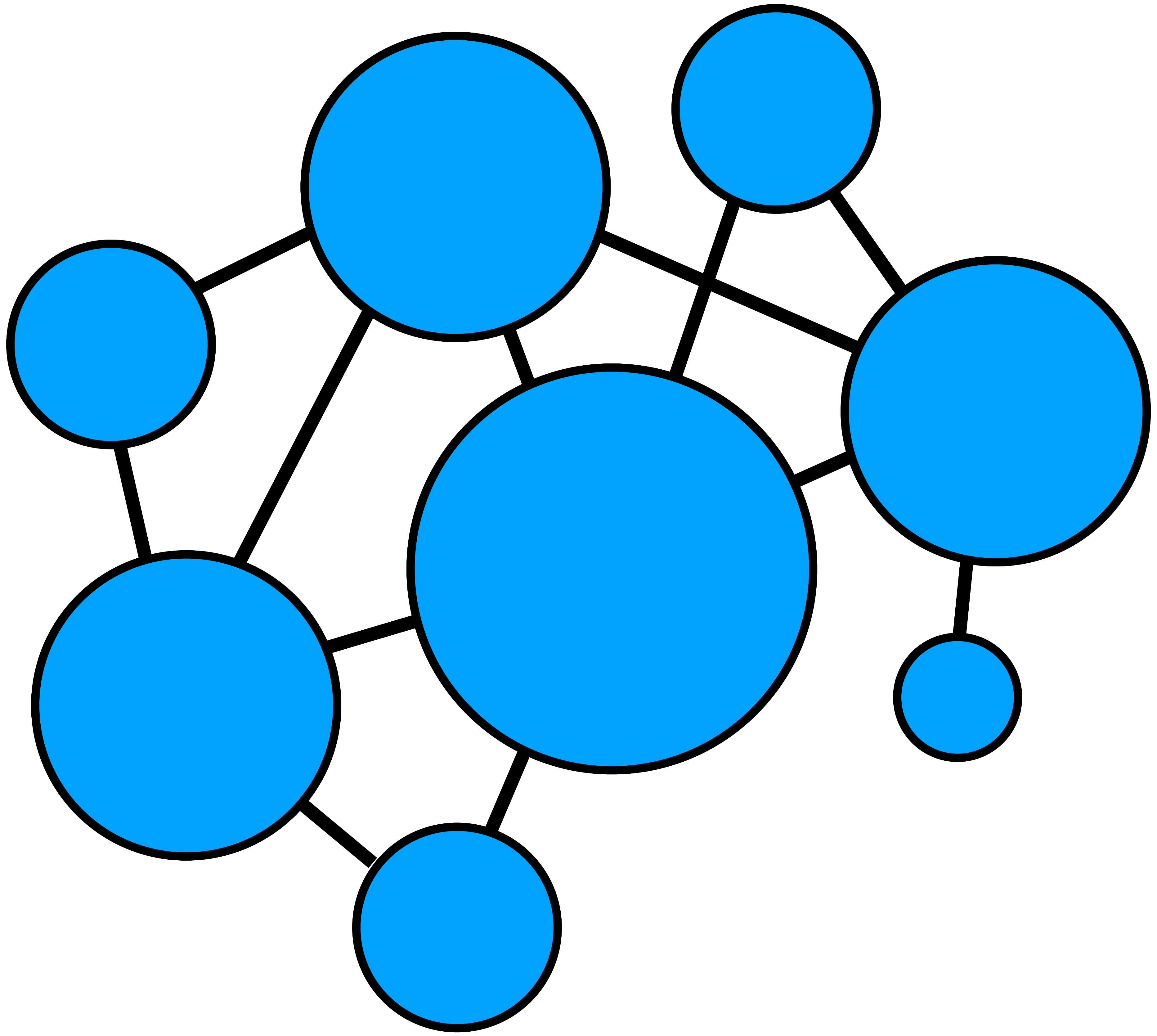
Biases at a node level:

Class imbalance (e.g., majority vs. minority)

Non-uniform degree distribution
e.g., skewed or Power law (high vs low degree)

What is inequality (bias)…

…in (social) networks?



Biases at a node level:

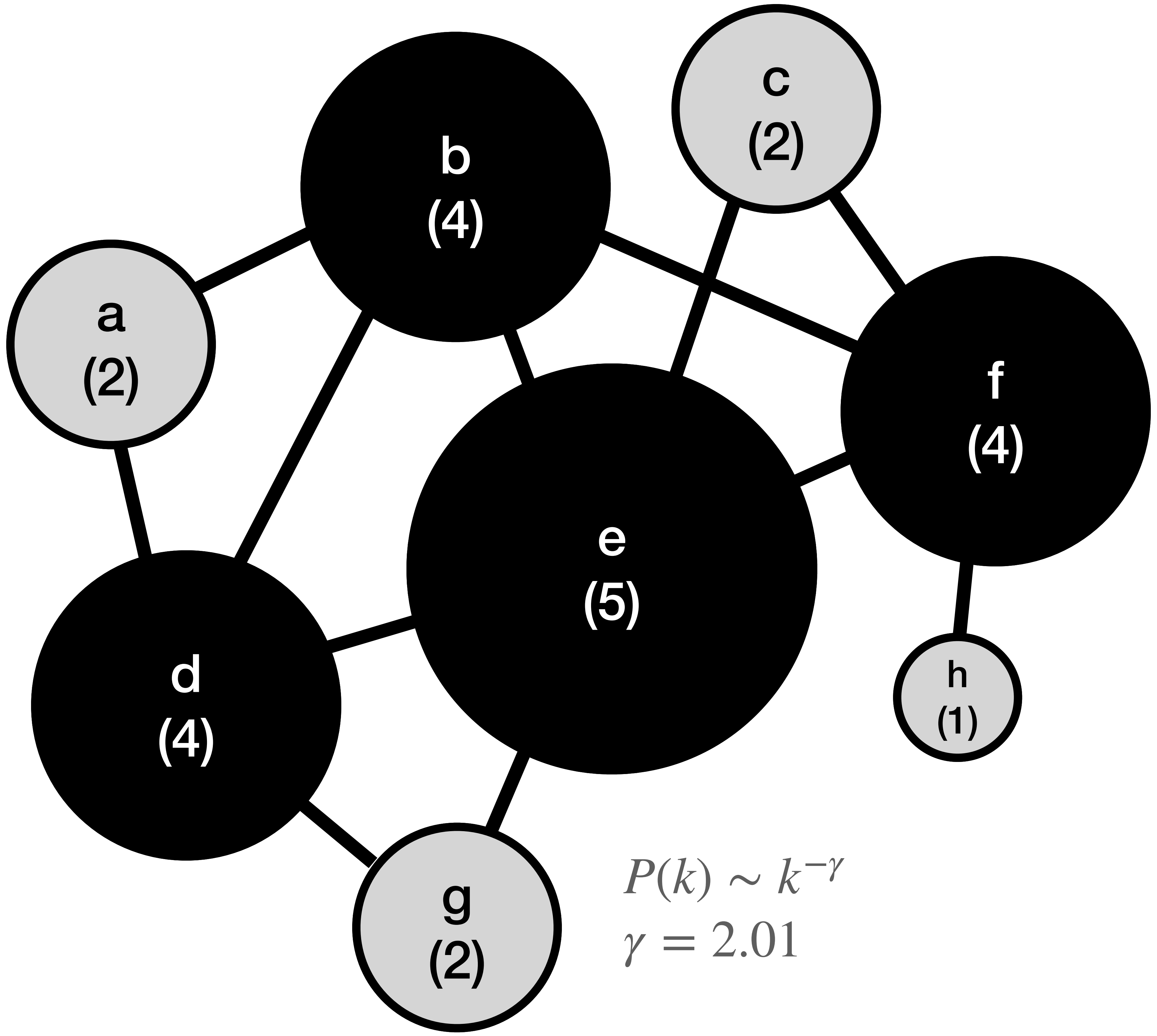
Class imbalance (e.g., majority vs. minority)

Non-uniform degree distribution
e.g., skewed or Power law (high vs low degree)

Biases at a link (formation) level:

What is inequality (bias)…

…in (social) networks?



$P(k) \sim k^{-\gamma}$
$\gamma = 2.01$

Biases at a node level:

Class imbalance (e.g., majority vs. minority)

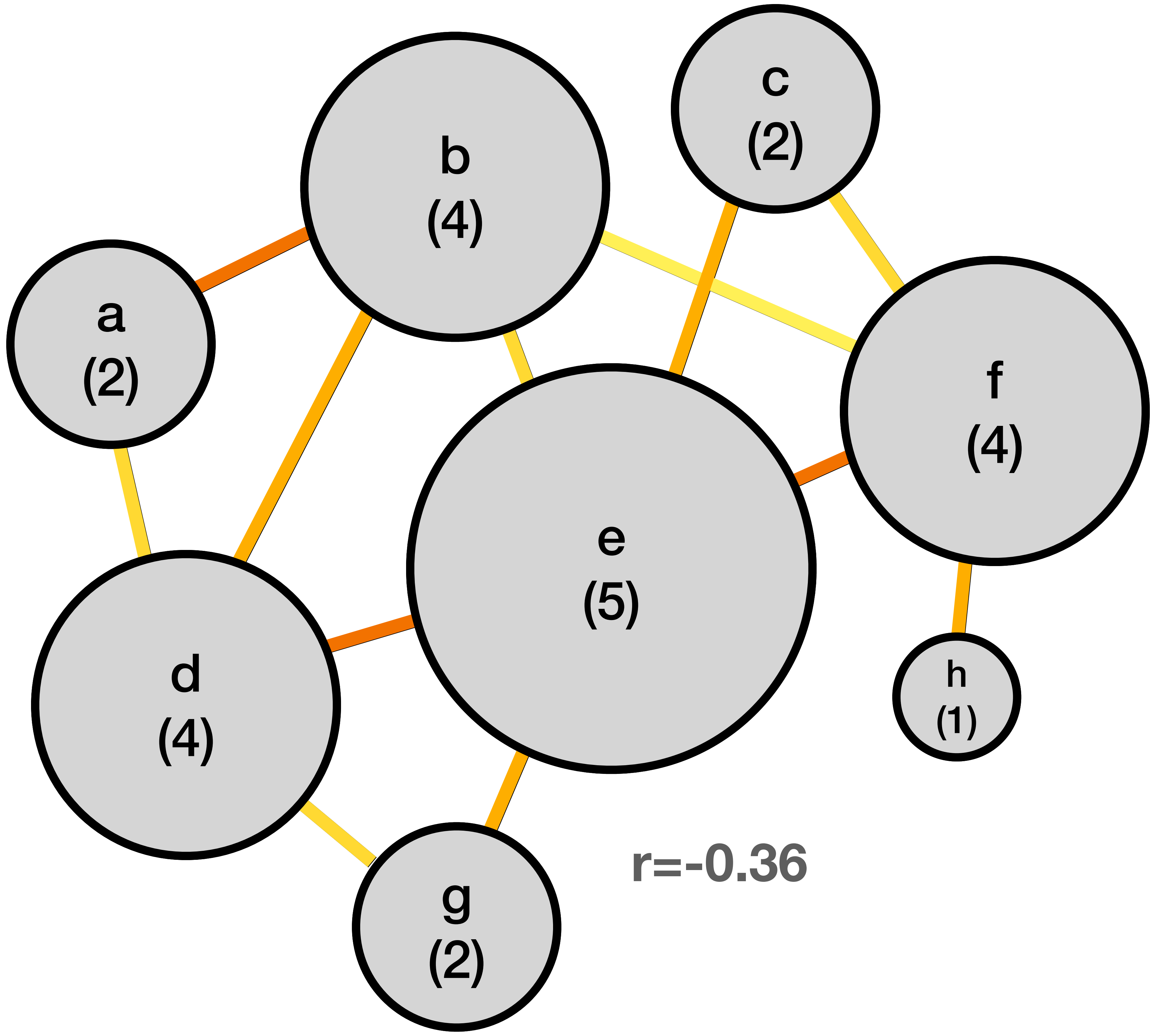
Non-uniform degree distribution
e.g., skewed or Power law (high vs low degree)

Biases at a link (formation) level:

Preferential attachment (Barabasi Albert 1999)
(i.e., Power law degree distribution)

What is inequality (bias)…

…in (social) networks?



Biases at a node level:

Class imbalance (e.g., majority vs. minority)

Non-uniform degree distribution
e.g., skewed or Power law (high vs low degree)

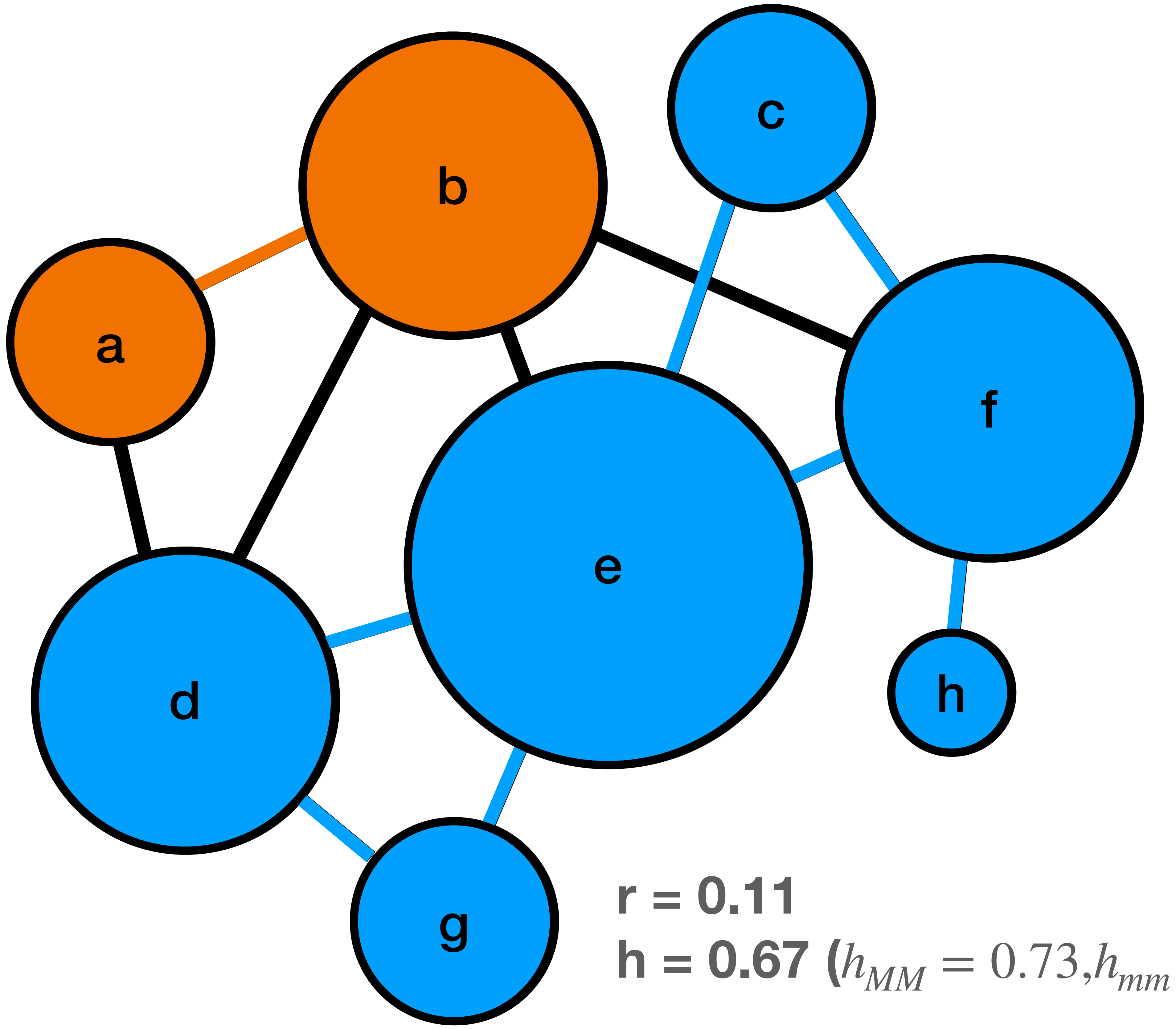
Biases at a link (formation) level:

Preferential attachment (Barabasi Albert 1999)
(i.e., Power law degree distribution)

+/- degree assortativity (Newman 2003)
Pearson corr. of degree between pairs of linked nodes $\in [-1, +1]$

What is inequality (bias)…

…in (social) networks?



r = 0.11
h = 0.67 ($h_{MM} = 0.73, h_{mm} = 0.60$)

Biases at a node level:

Class imbalance (e.g., majority vs. minority)

Non-uniform degree distribution
e.g., skewed or Power law (high vs low degree)

Biases at a link (formation) level:

Preferential attachment (Barabasi Albert 1999)
(i.e., Power law degree distribution)

+/- degree assortativity (Newman 2003)
Pearson corr. of degree between pairs of linked nodes $\in [-1, +1]$

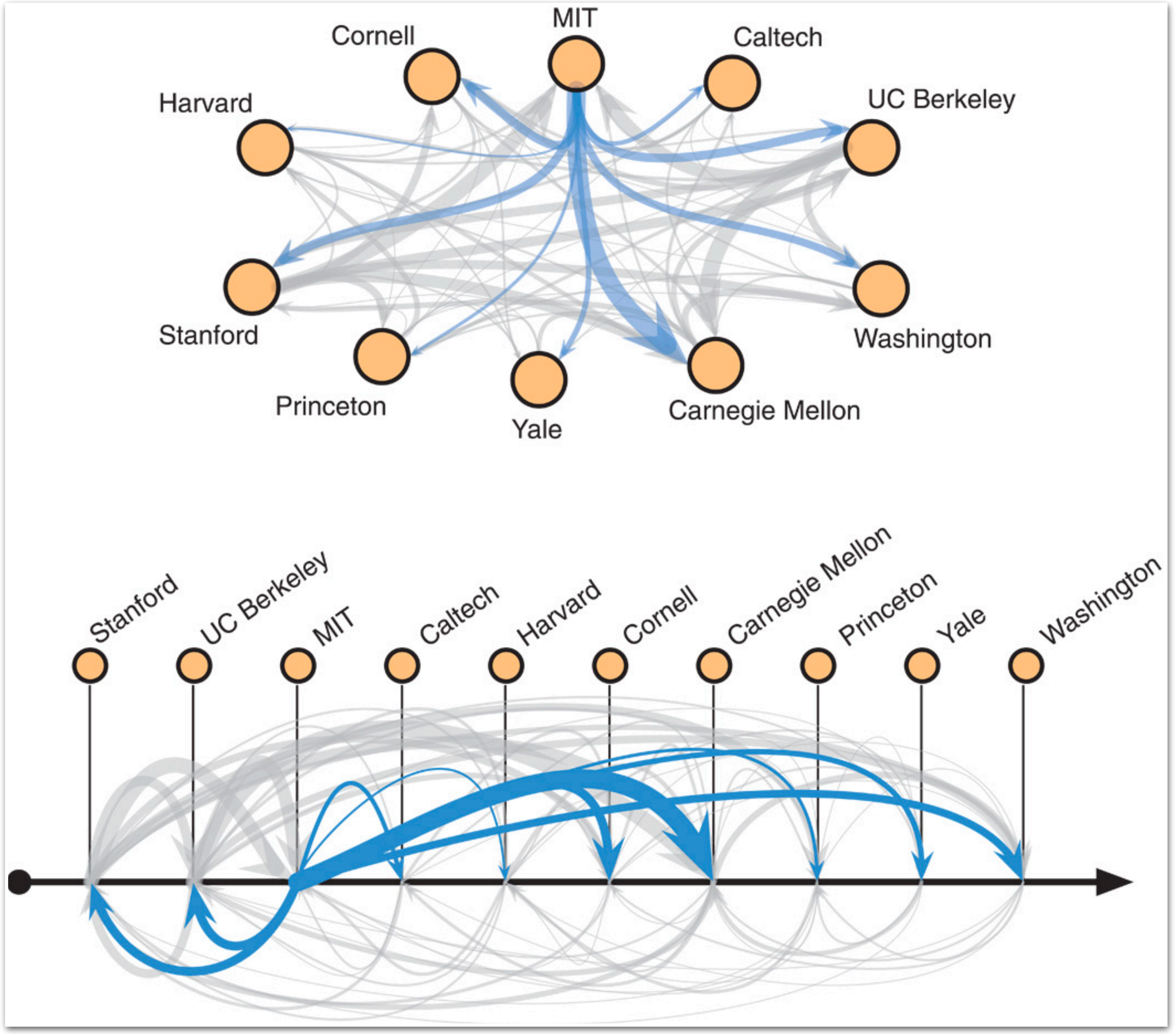
+/- attribute assortativity (Newman 2003)
High/low homophily with PA (Karimi et al. 2018)

Inequalities (biases) in networks

Examples

Inequalities (biases) in networks

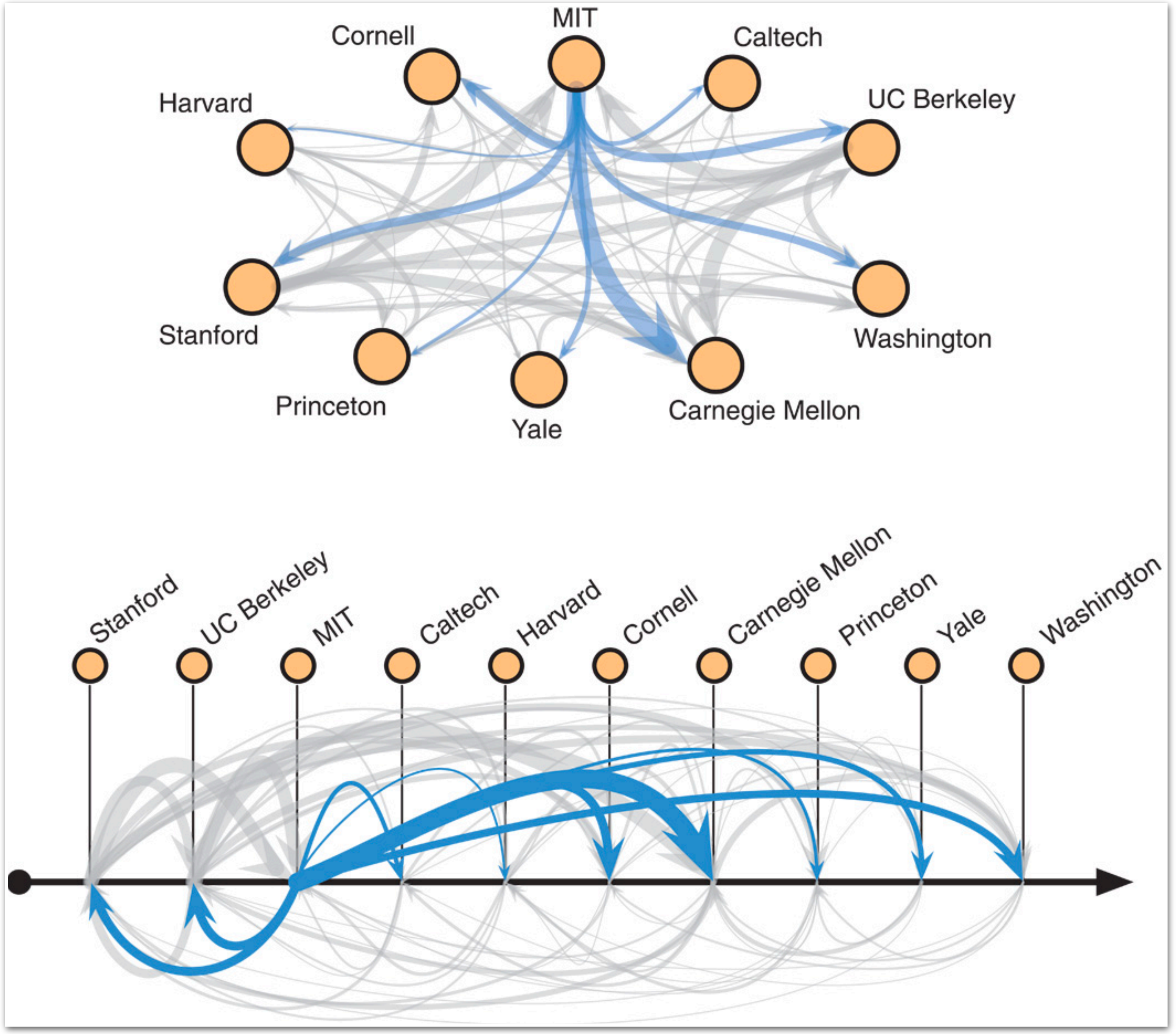
Examples



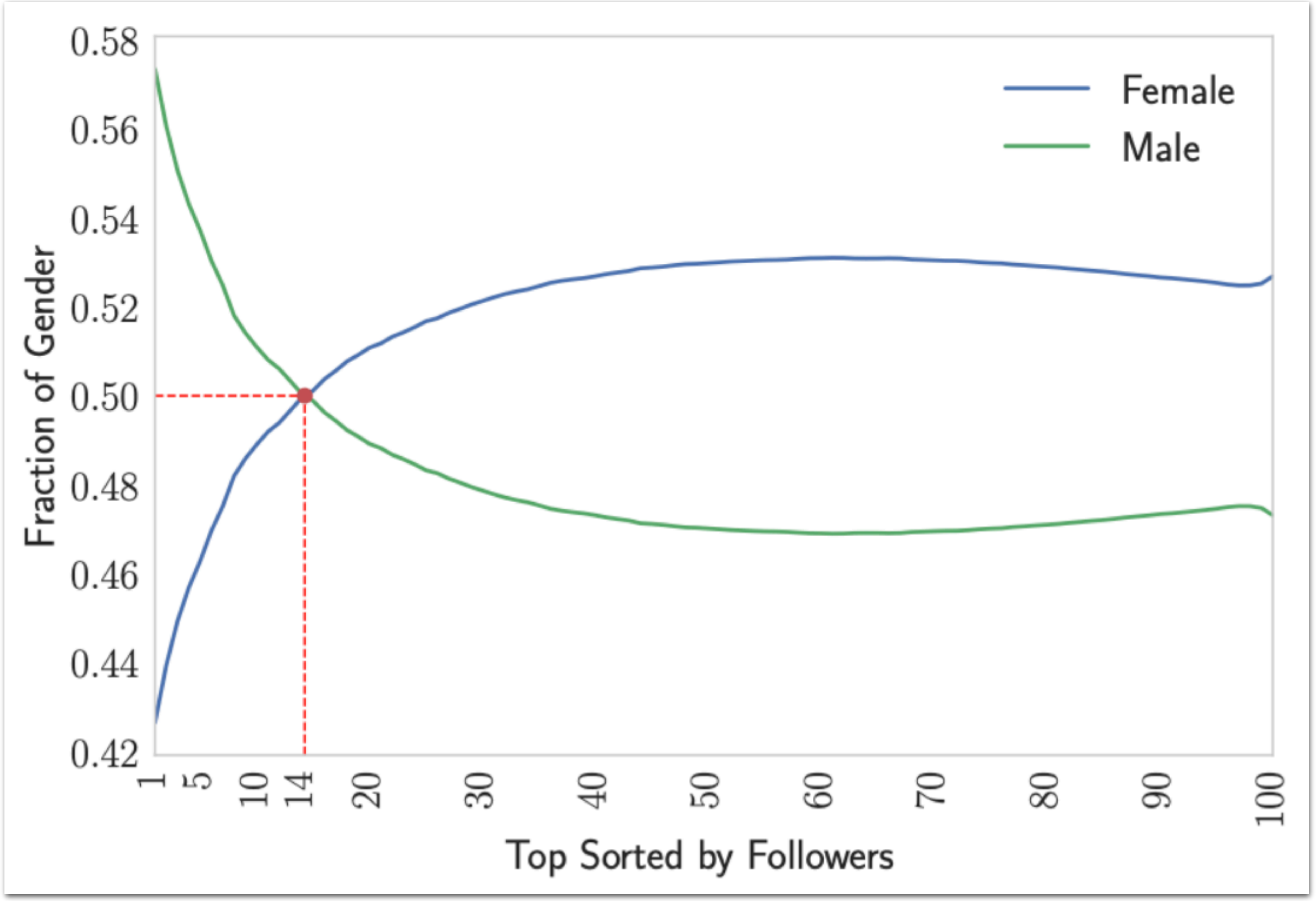
Systematic inequality and hierarchy in faculty hiring networks. Clauset et al. 2015.

Inequalities (biases) in networks

Examples



Systematic inequality and hierarchy in faculty hiring networks. Clauset et al. 2015.



White, Man, and Highly Followed: Gender and Race Inequalities in Twitter. Messias et al. 2017.

What is inequality (bias)…

…in (machine learning) algorithms?

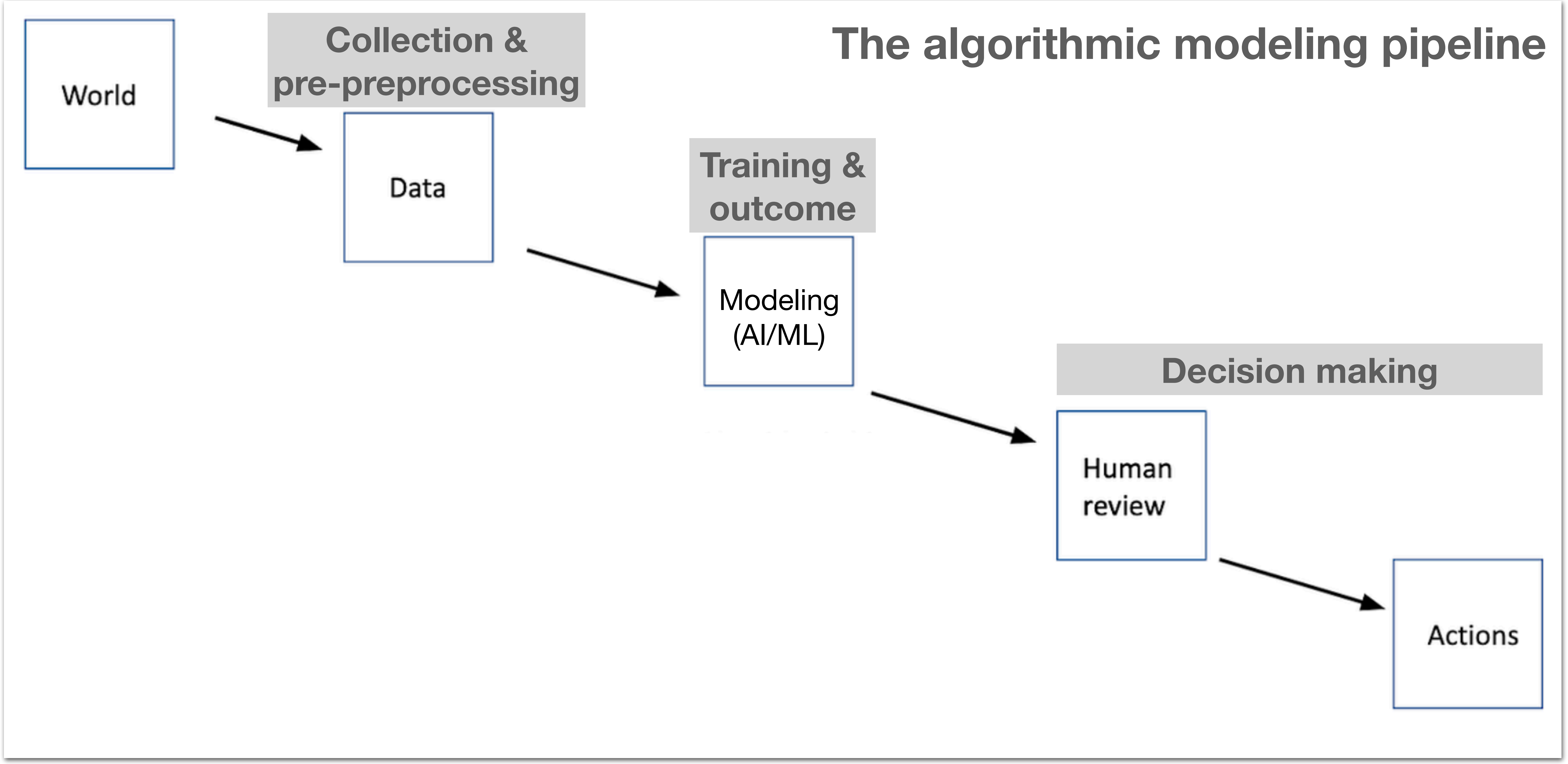


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

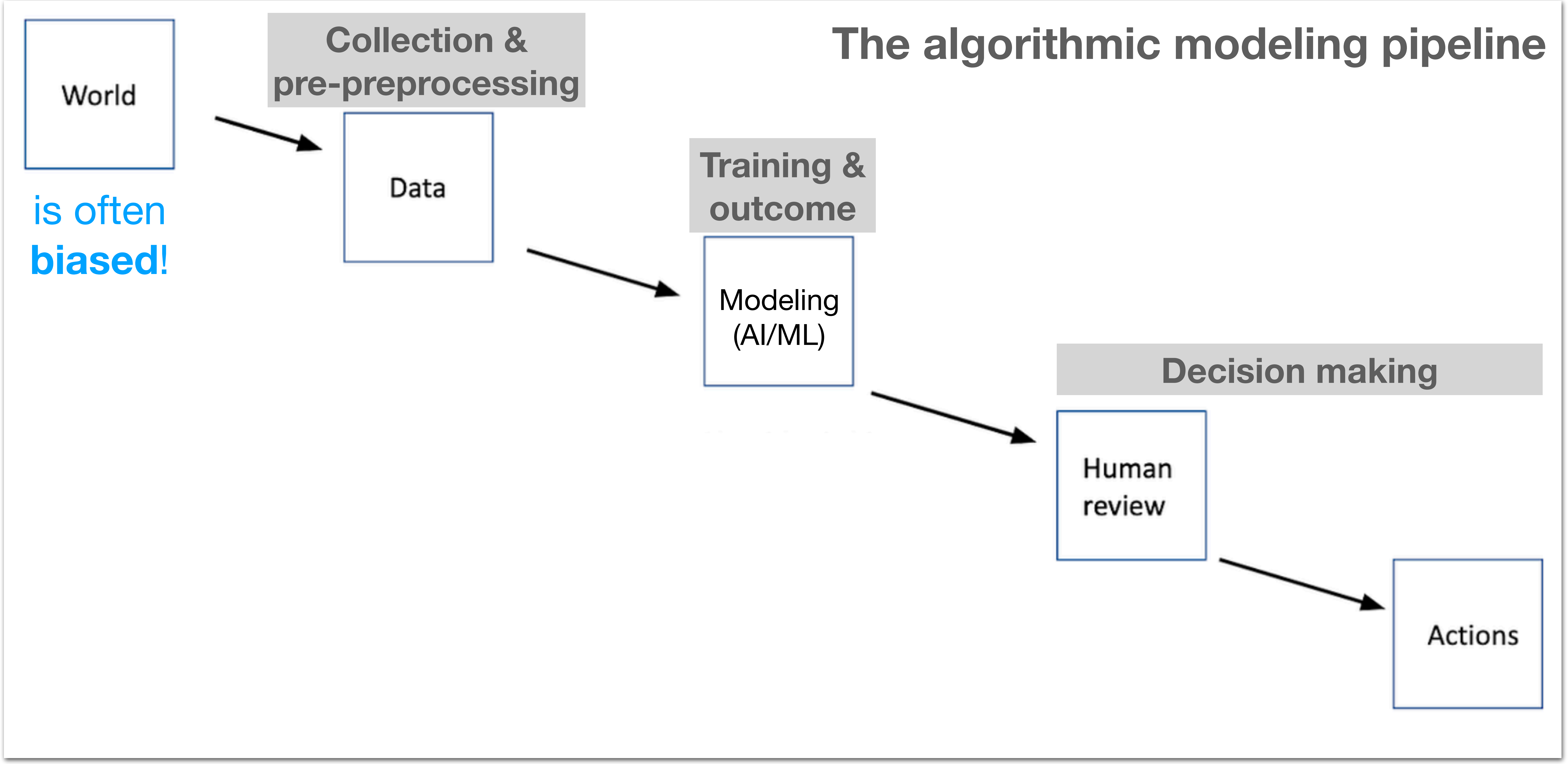


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

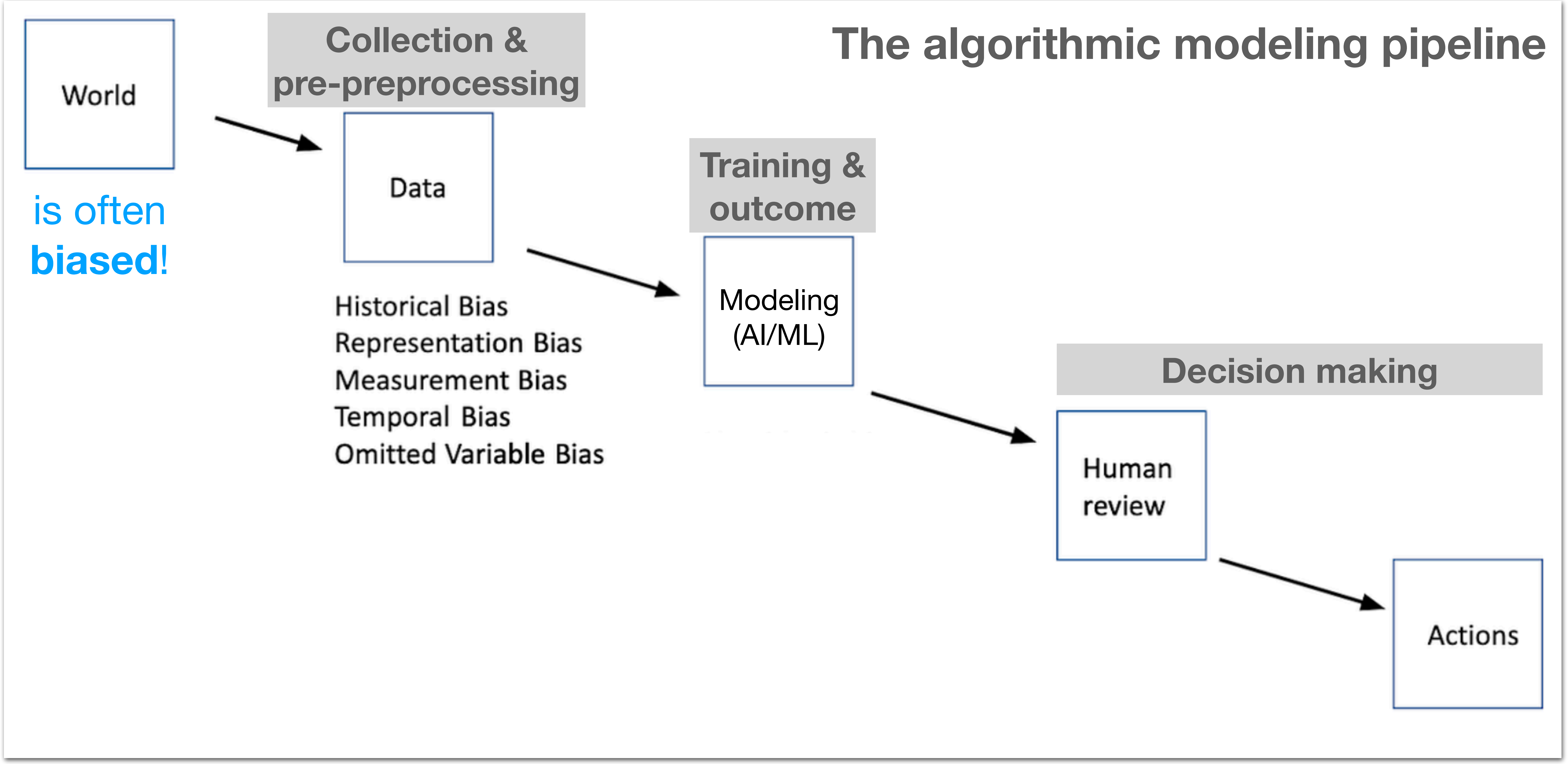


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

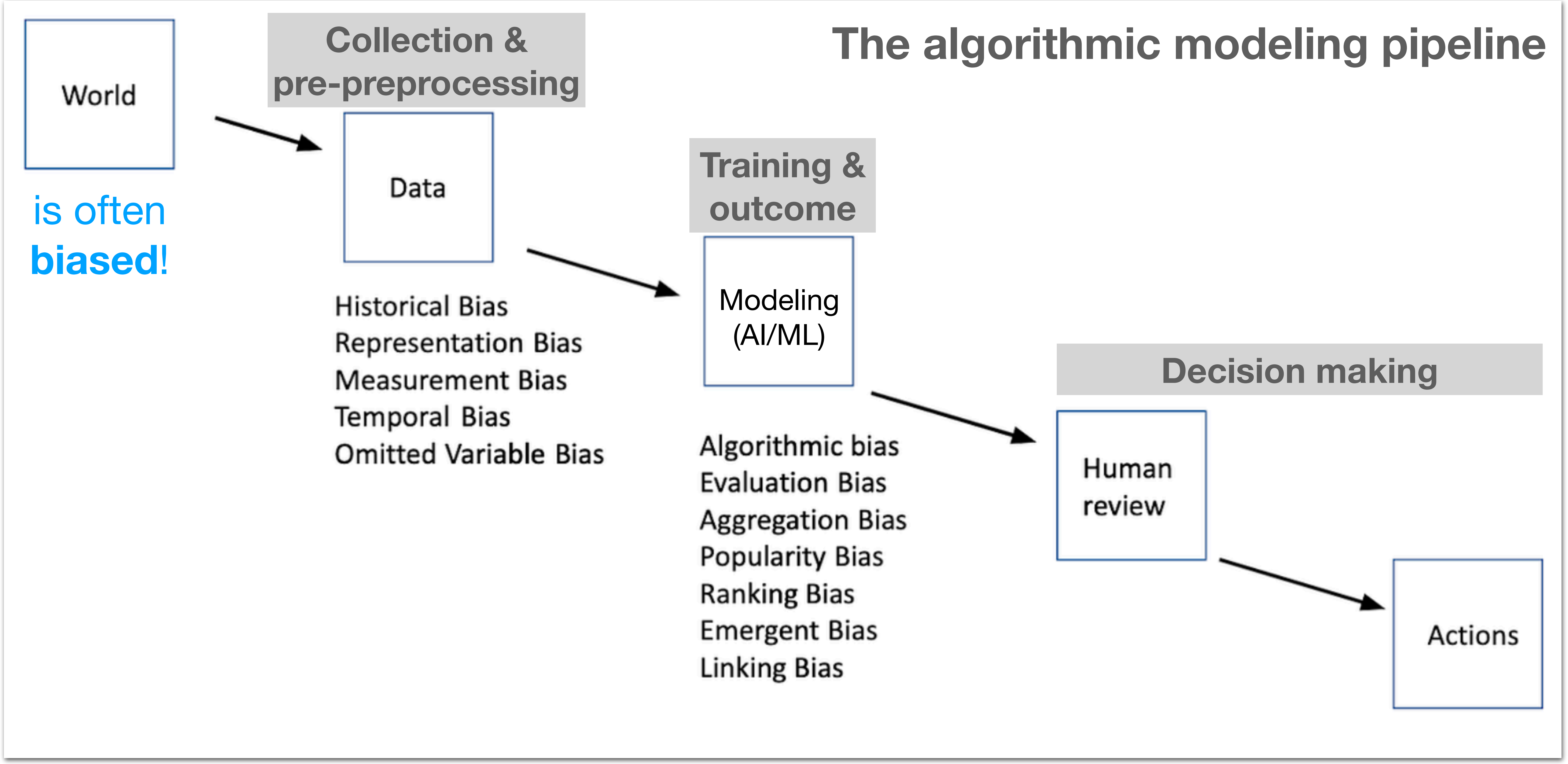


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

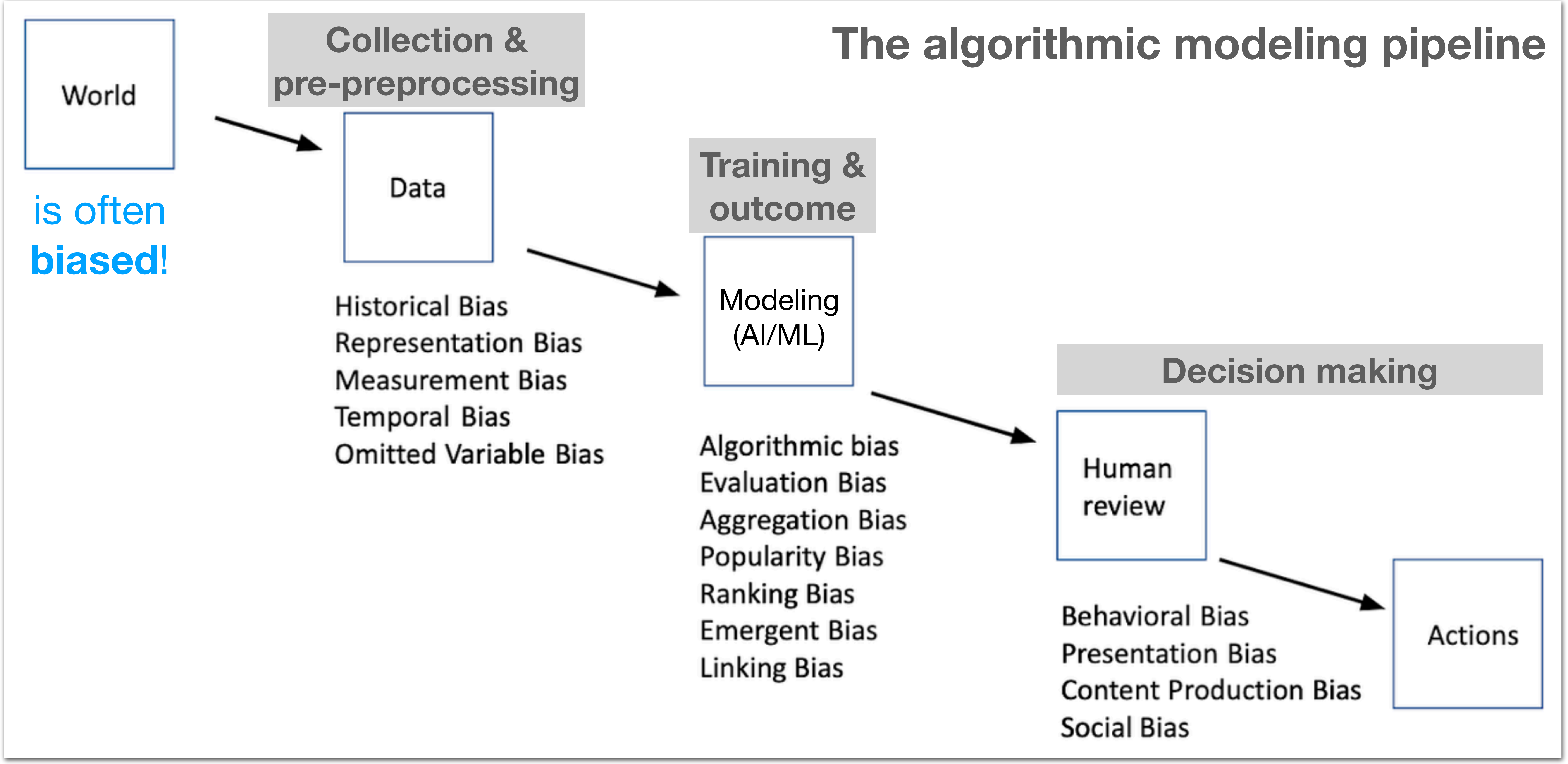


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

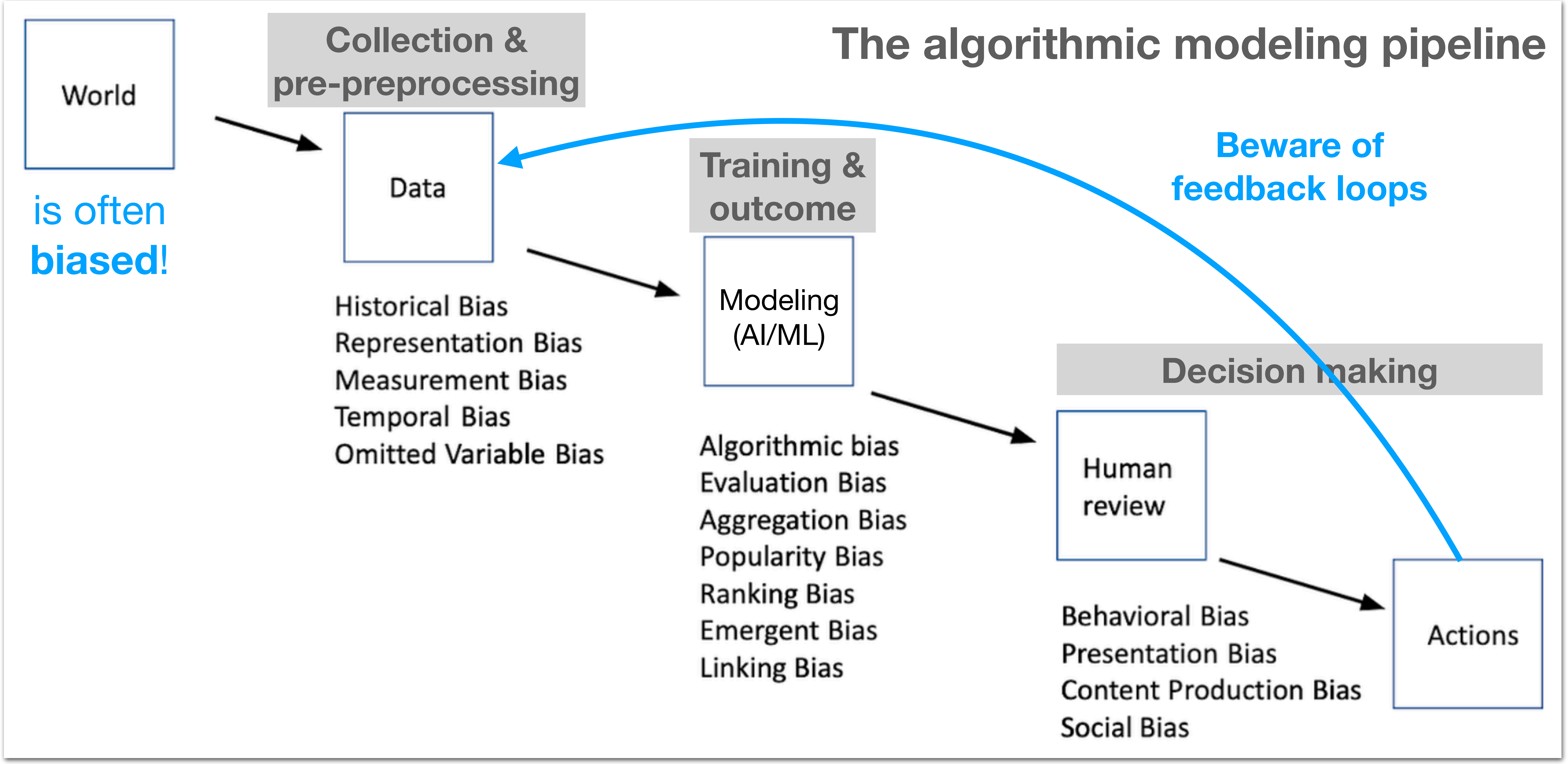


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

What is inequality (bias)…

…in (machine learning) algorithms?

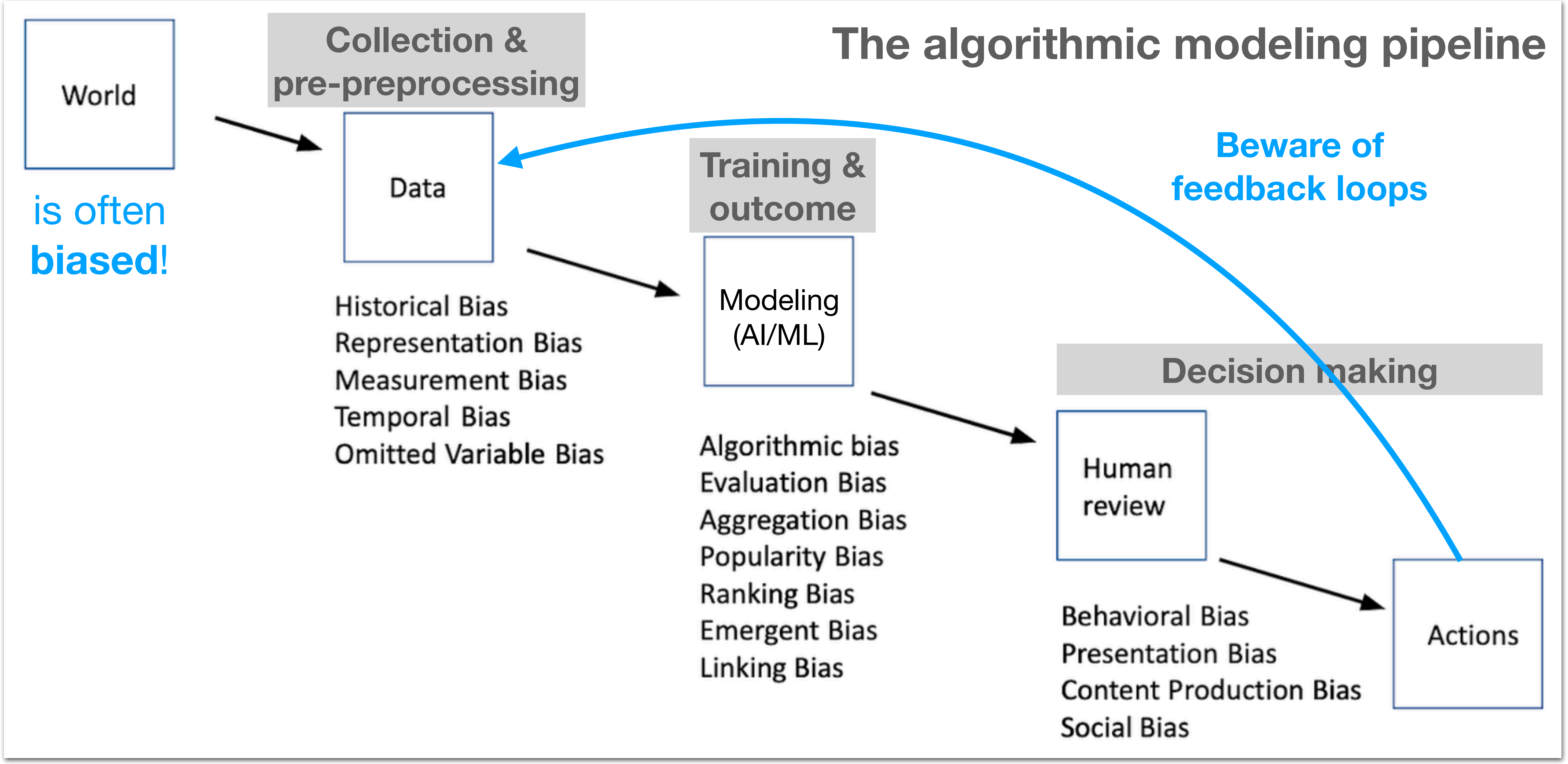
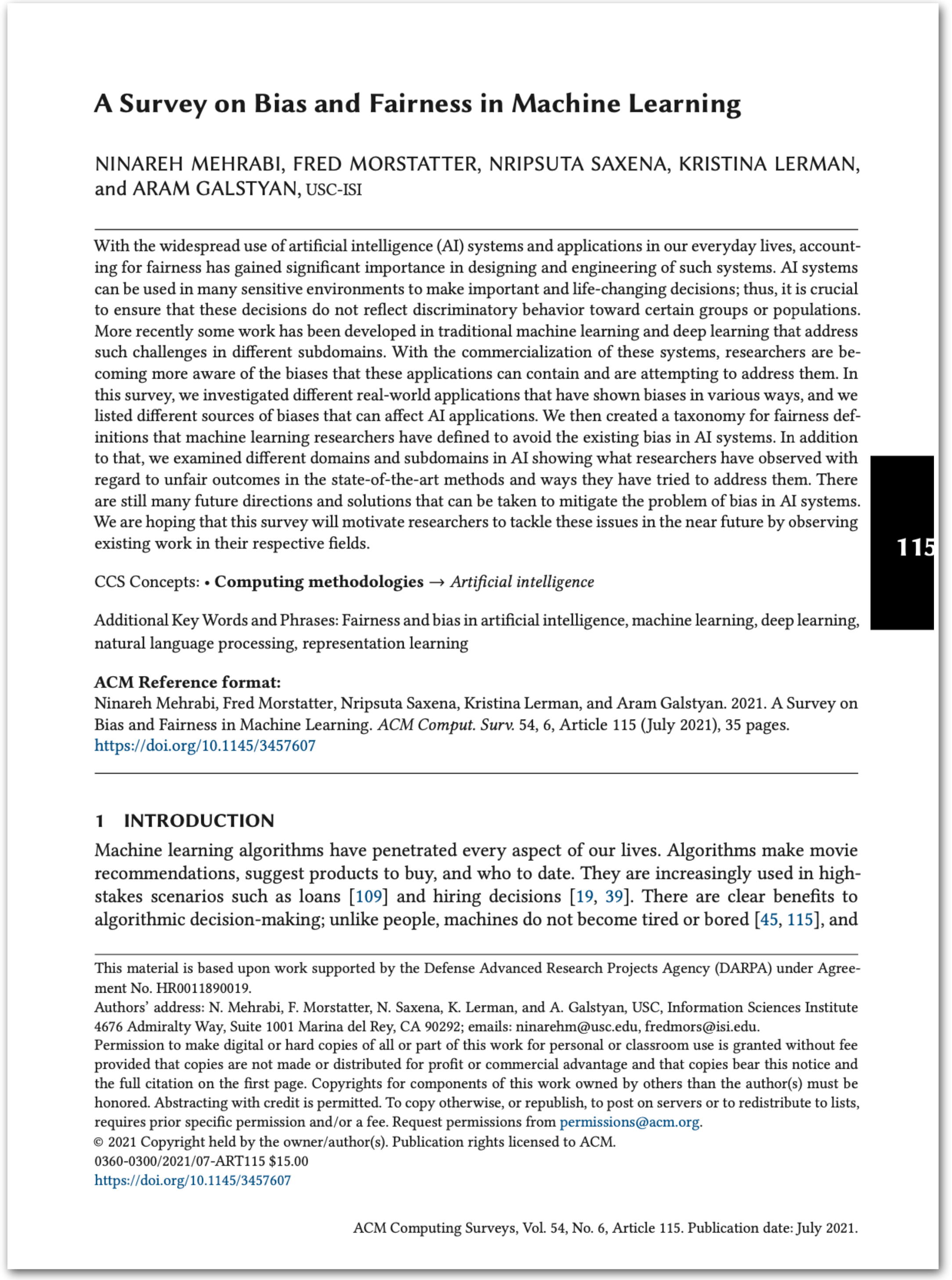


Image: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

Recommended reading:
Mehrabi et al. 2021



A Survey on Bias and Fairness in Machine Learning

NINAREH MEHRABI, FRED MORSTATTER, NRIPSUTA SAXENA, KRISTINA LERMAN, and ARAM GALSTYAN, USC-ISI

With the widespread use of artificial intelligence (AI) systems and applications in our everyday lives, accounting for fairness has gained significant importance in designing and engineering of such systems. AI systems can be used in many sensitive environments to make important and life-changing decisions; thus, it is crucial to ensure that these decisions do not reflect discriminatory behavior toward certain groups or populations. More recently some work has been developed in traditional machine learning and deep learning that address such challenges in different subdomains. With the commercialization of these systems, researchers are becoming more aware of the biases that these applications can contain and are attempting to address them. In this survey, we investigated different real-world applications that have shown biases in various ways, and we listed different sources of biases that can affect AI applications. We then created a taxonomy for fairness definitions that machine learning researchers have defined to avoid the existing bias in AI systems. In addition to that, we examined different domains and subdomains in AI showing what researchers have observed with regard to unfair outcomes in the state-of-the-art methods and ways they have tried to address them. There are still many future directions and solutions that can be taken to mitigate the problem of bias in AI systems. We are hoping that this survey will motivate researchers to tackle these issues in the near future by observing existing work in their respective fields.

115

CCS Concepts: • **Computing methodologies** → *Artificial intelligence*

Additional Key Words and Phrases: Fairness and bias in artificial intelligence, machine learning, deep learning, natural language processing, representation learning

ACM Reference format:
Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (July 2021), 35 pages.
https://doi.org/10.1145/3457607

1 INTRODUCTION

Machine learning algorithms have penetrated every aspect of our lives. Algorithms make movie recommendations, suggest products to buy, and who to date. They are increasingly used in high-stakes scenarios such as loans [109] and hiring decisions [19, 39]. There are clear benefits to algorithmic decision-making; unlike people, machines do not become tired or bored [45, 115], and

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR0011890019.
Authors' address: N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, USC, Information Sciences Institute 4676 Admiralty Way, Suite 1001 Marina del Rey, CA 90292; emails: ninarehm@usc.edu, fredmors@isi.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
0360-0300/2021/07-ART115 $15.00
https://doi.org/10.1145/3457607

ACM Computing Surveys, Vol. 54, No. 6, Article 115. Publication date: July 2021.

Inequalities (biases) in ML algorithms

Examples

Inequalities (biases) in ML algorithms

Examples

Many Facial-Recognition Systems Are Biased, Says U.S. Study

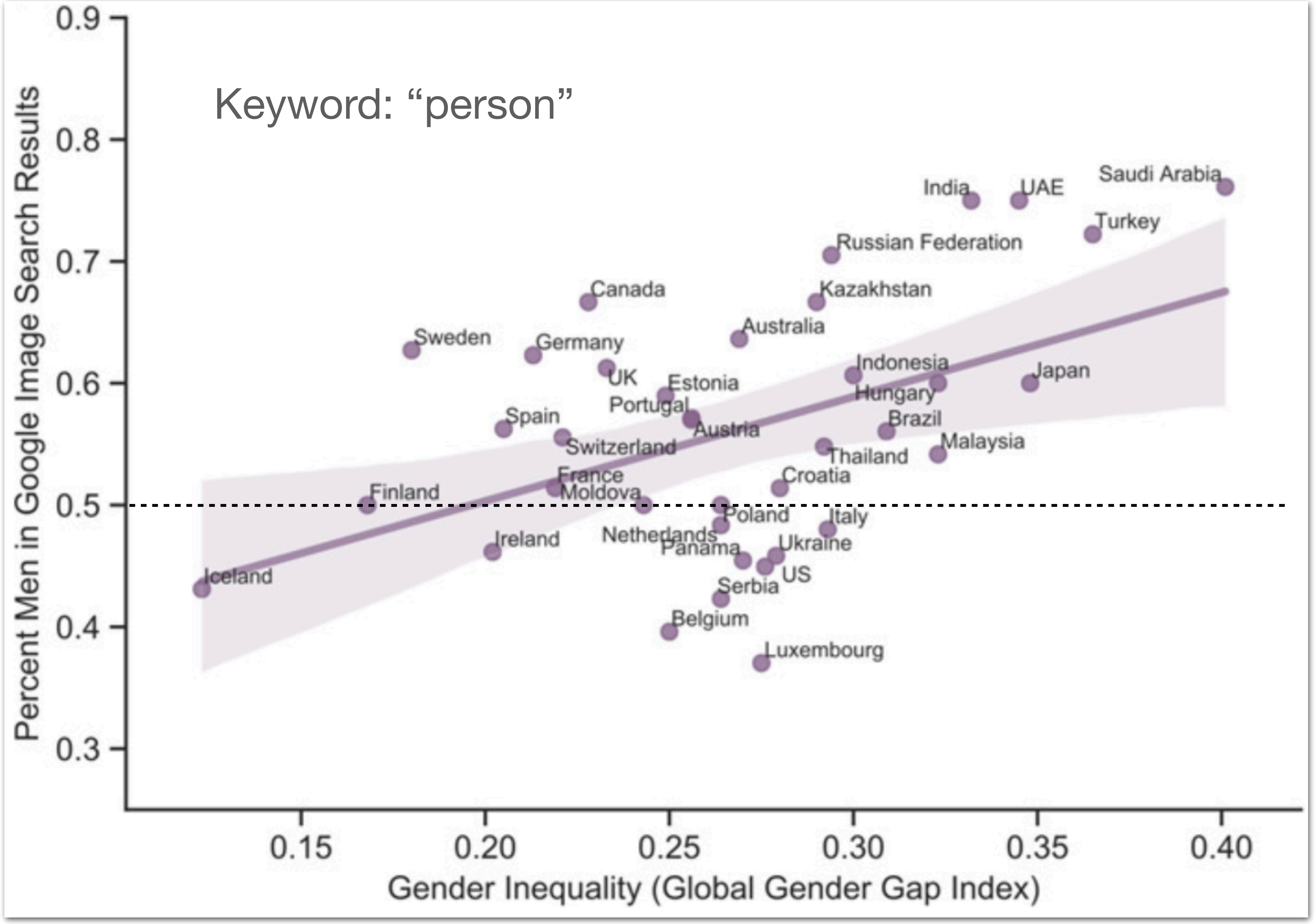
Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

Inequalities (biases) in ML algorithms

Examples

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



Propagation of societal gender inequality by internet search algorithms. Vlasceanu & Amodio 2022.

Inequalities (biases) in ML algorithms

Examples

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

All the Ways Hiring Algorithms Can Introduce Bias

by Miranda Bogen

Inequalities (biases) in ML algorithms

Examples

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

All the Ways Hiring Algorithms Can Introduce Bias

by Miranda Bogen

A.I. Is Sending People to Jail — and Getting It Wrong

Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past

Inequalities (biases) in ML algorithms

Examples

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

All the Ways Hiring Algorithms Can Introduce Bias

by Miranda Bogen

A.I. Is Sending People to Jail — and Getting It Wrong

Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past

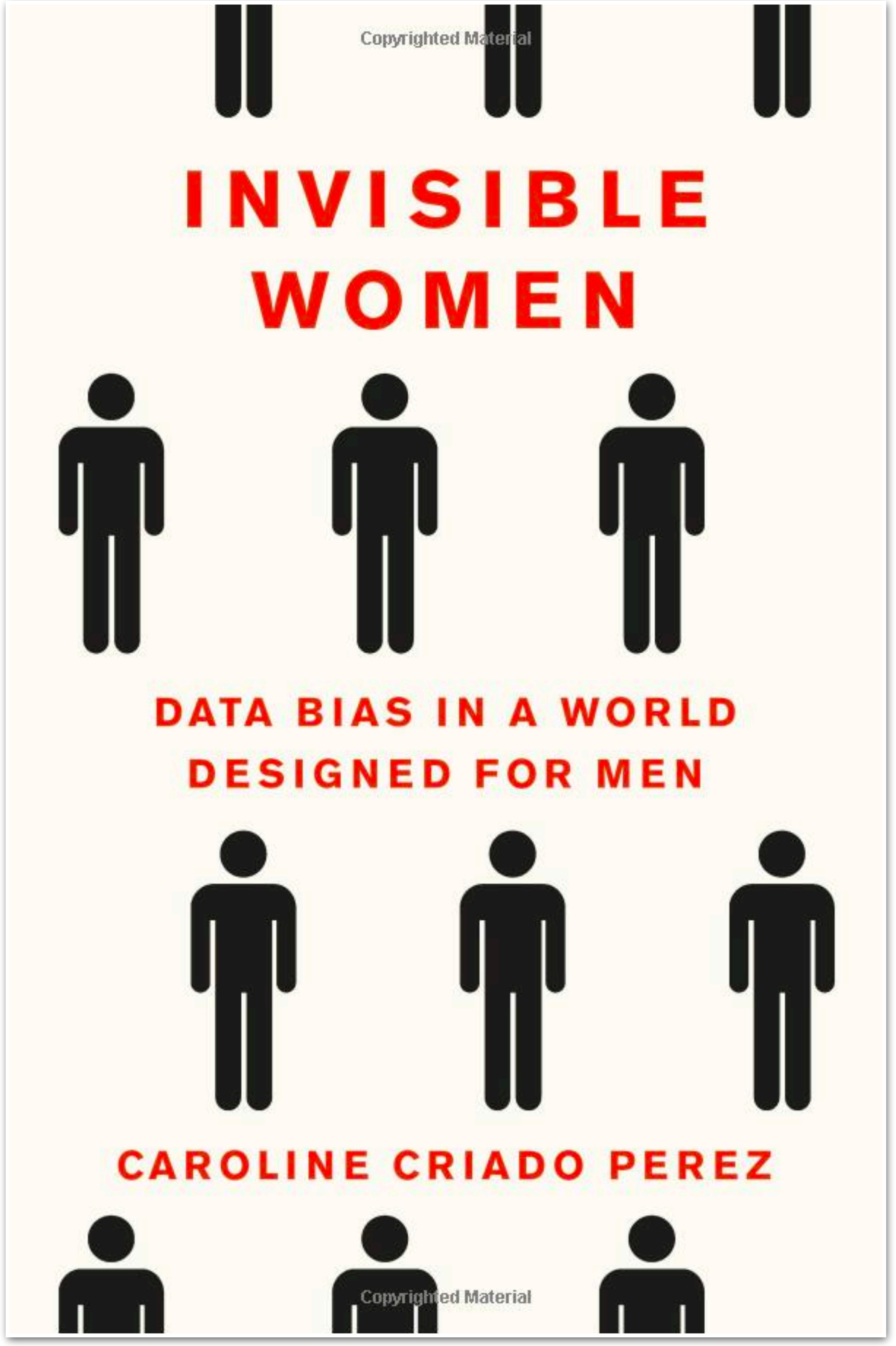
Comparative Study > Diabet Med. 2019 Oct;36(10):1234-1242. doi: 10.1111/dme.13979. Epub 2019 Jul 15.

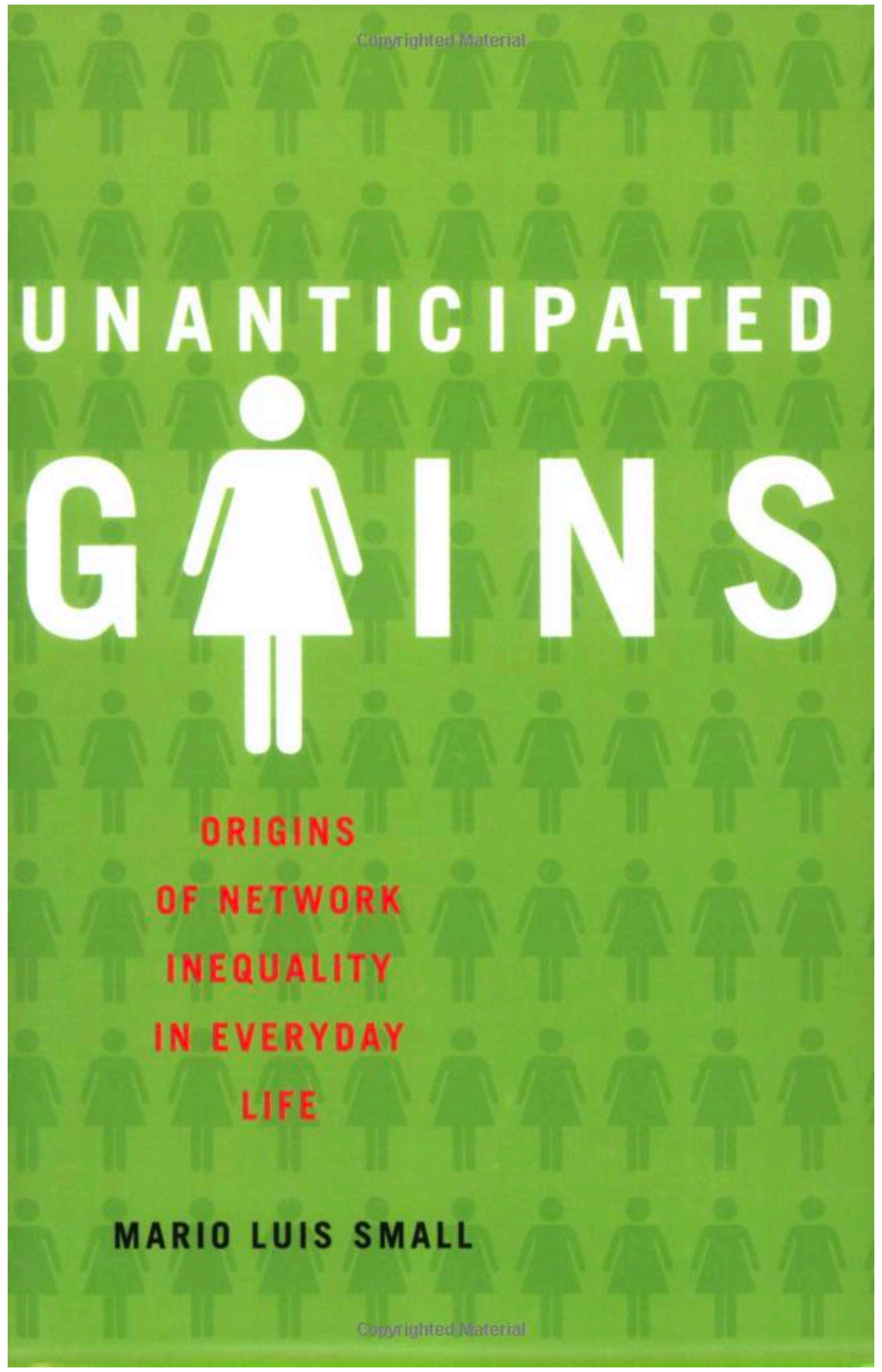
Racial differences in performance of HbA_{1c} for the classification of diabetes and prediabetes among US adults of non-Hispanic black and white race

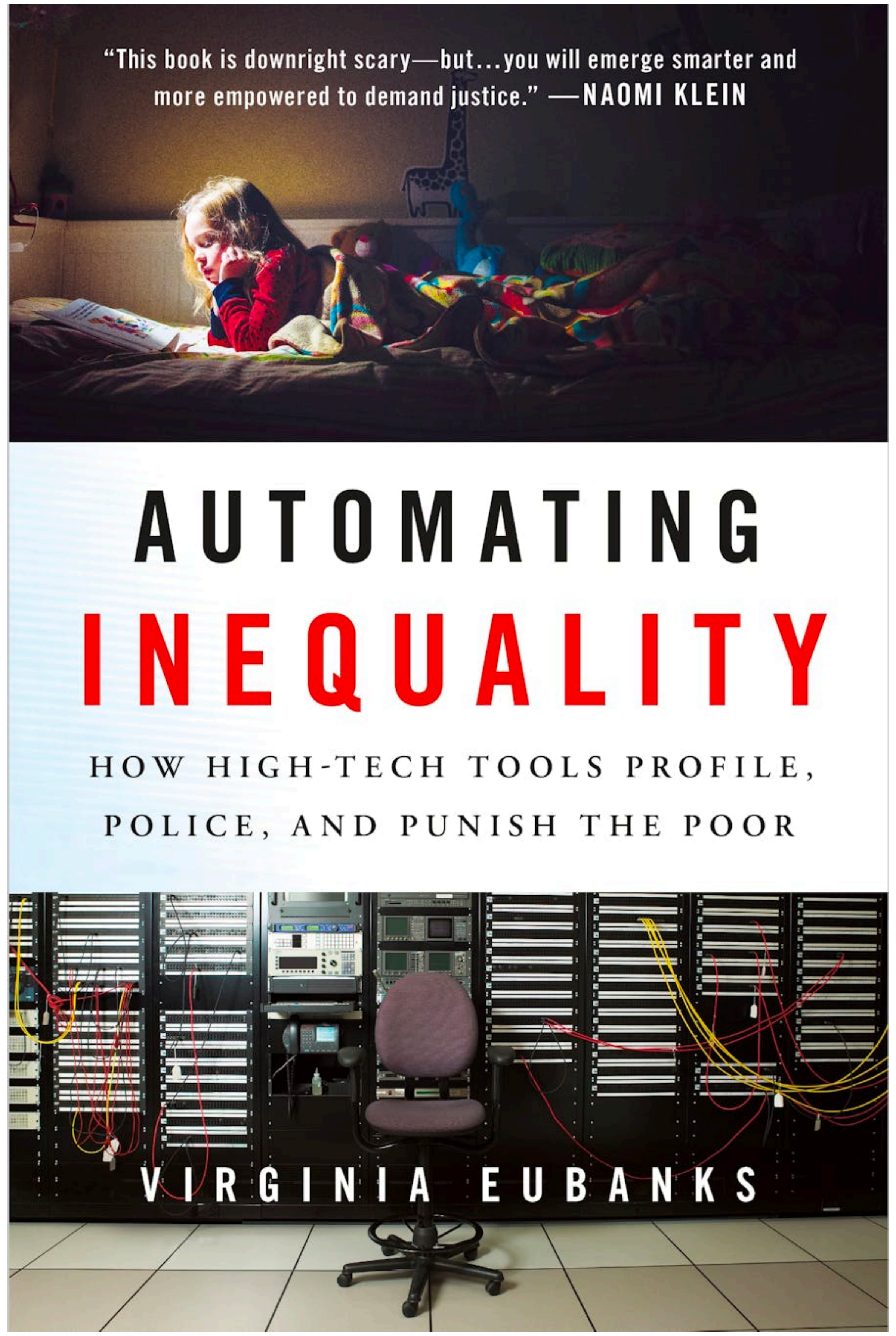
C N Ford [1], R W Leet [1 2], L M Kipling [1], M K Rhee [3 4], S L Jackson [5], P W F Wilson [3 4 6], L S Phillips [3 4], L R Staimez [1 2]

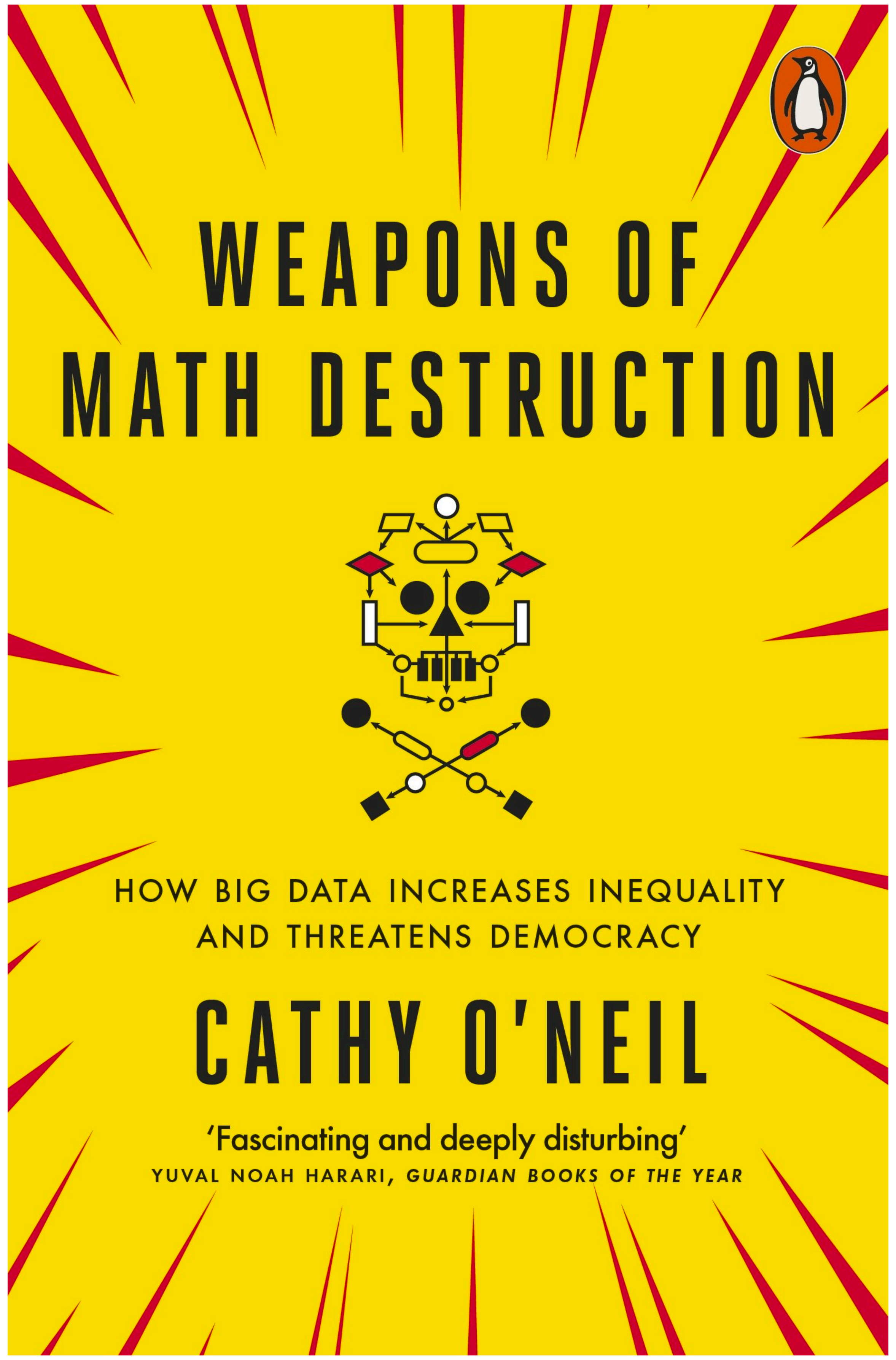
Data and algorithmic biases in the real world

Highly recommended books!

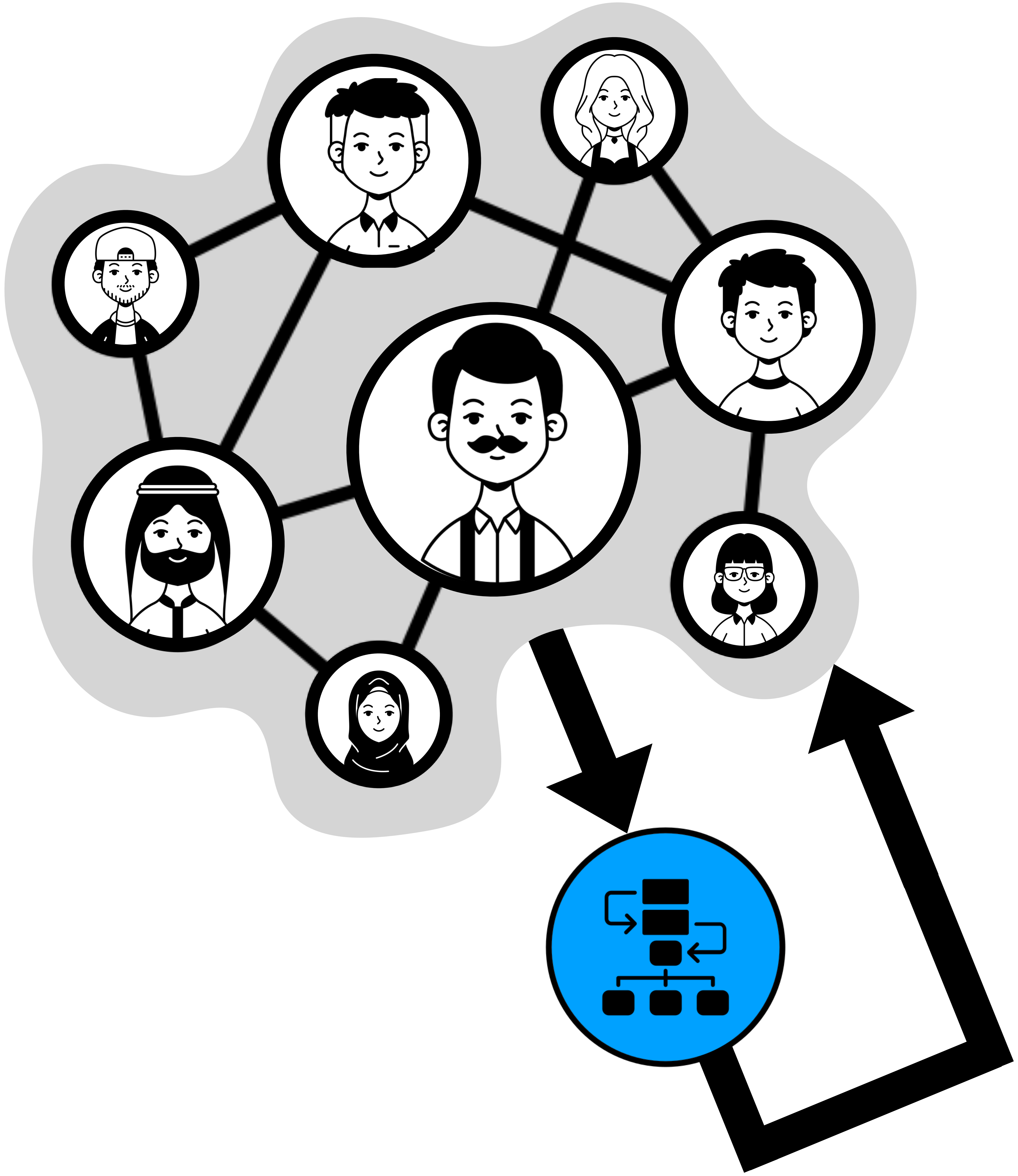








What about the **inequalities in network-based algorithms?**



Disclaimer:
A network-based algorithm is such that the data input is a network (graph).
It is NOT a deep neural network.

Network-based algorithms

Definition

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
  - **Influence maximization** (e.g., spreading information)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
  - **Influence maximization** (e.g., spreading information)
  - **Relational (node) classification** (e.g., your preferences = your friends' preferences)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
    - **Influence maximization** (e.g., spreading information)
    - **Relational (node) classification** (e.g., your preferences = your friends' preferences)
    - **Link prediction** (e.g., friends-of-friends)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
  - **Influence maximization** (e.g., spreading information)
  - **Relational (node) classification** (e.g., your preferences = your friends' preferences)
  - **Link prediction** (e.g., friends-of-friends)
  - **Node ranking** (e.g., based on structure such as centrality)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
    - **Influence maximization** (e.g., spreading information)
    - **Relational (node) classification** (e.g., your preferences = your friends' preferences)
    - **Link prediction** (e.g., friends-of-friends)
    - **Node ranking** (e.g., based on structure such as centrality)
    - **Node sampling** (e.g., based on structure such as random walk)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node information but also network structure (e.g., links, degree, etc).
  - **Influence maximization** (e.g., spreading information)
  - **Relational (node) classification** (e.g., your preferences = your friends' preferences)
  - **Link prediction** (e.g., friends-of-friends)
  - **Node ranking** (e.g., based on structure such as centrality)
  - **Node sampling** (e.g., based on structure such as random walk)
  - **Graph clustering** (e.g., based on structure such as edge-betweenness)

Network-based algorithms

Definition

- Any algorithm or process applied on networks that leverages not only node

Networks are biased, thus, these algorithms can be biased too!

- **Node ranking** (e.g., based on structure such as centrality)
- **Node sampling** (e.g., based on structure such as random walk)
- **Graph clustering** (e.g., based on structure such as edge-betweenness)

Today's class

Inequality in network-based algorithms

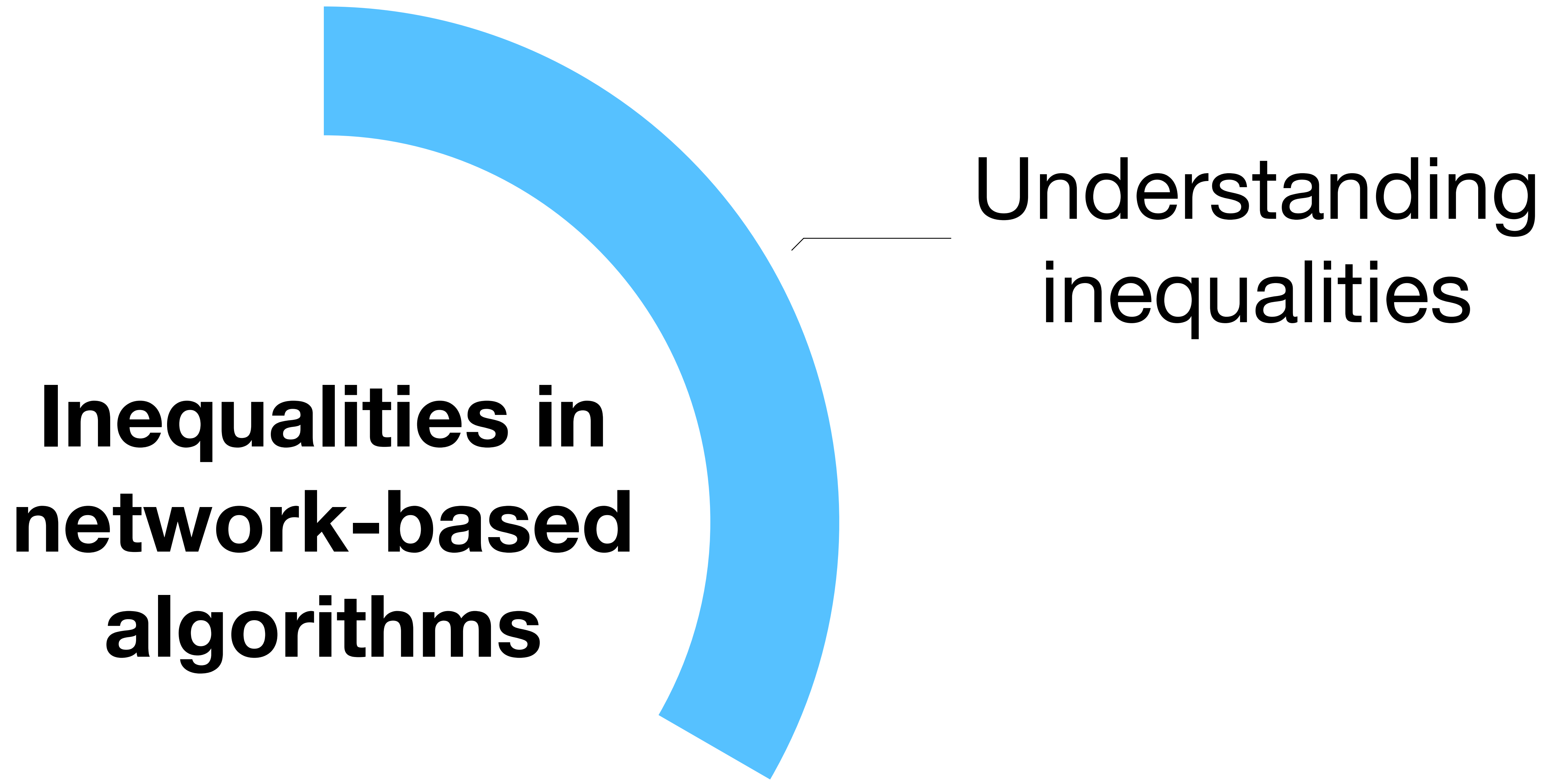
Today's class

Inequality in network-based algorithms

Inequalities in network-based algorithms

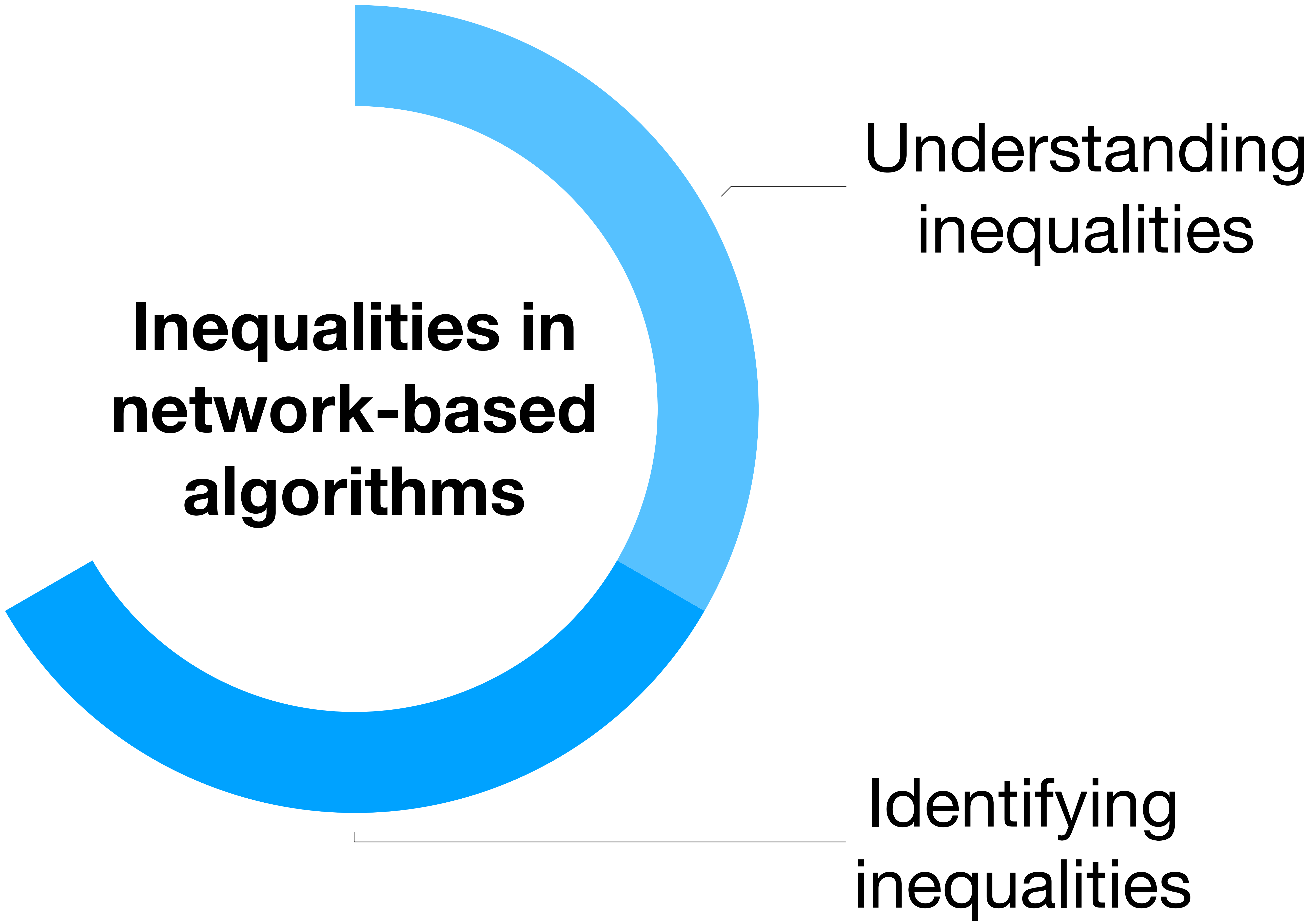
Today’s class

Inequality in network-based algorithms



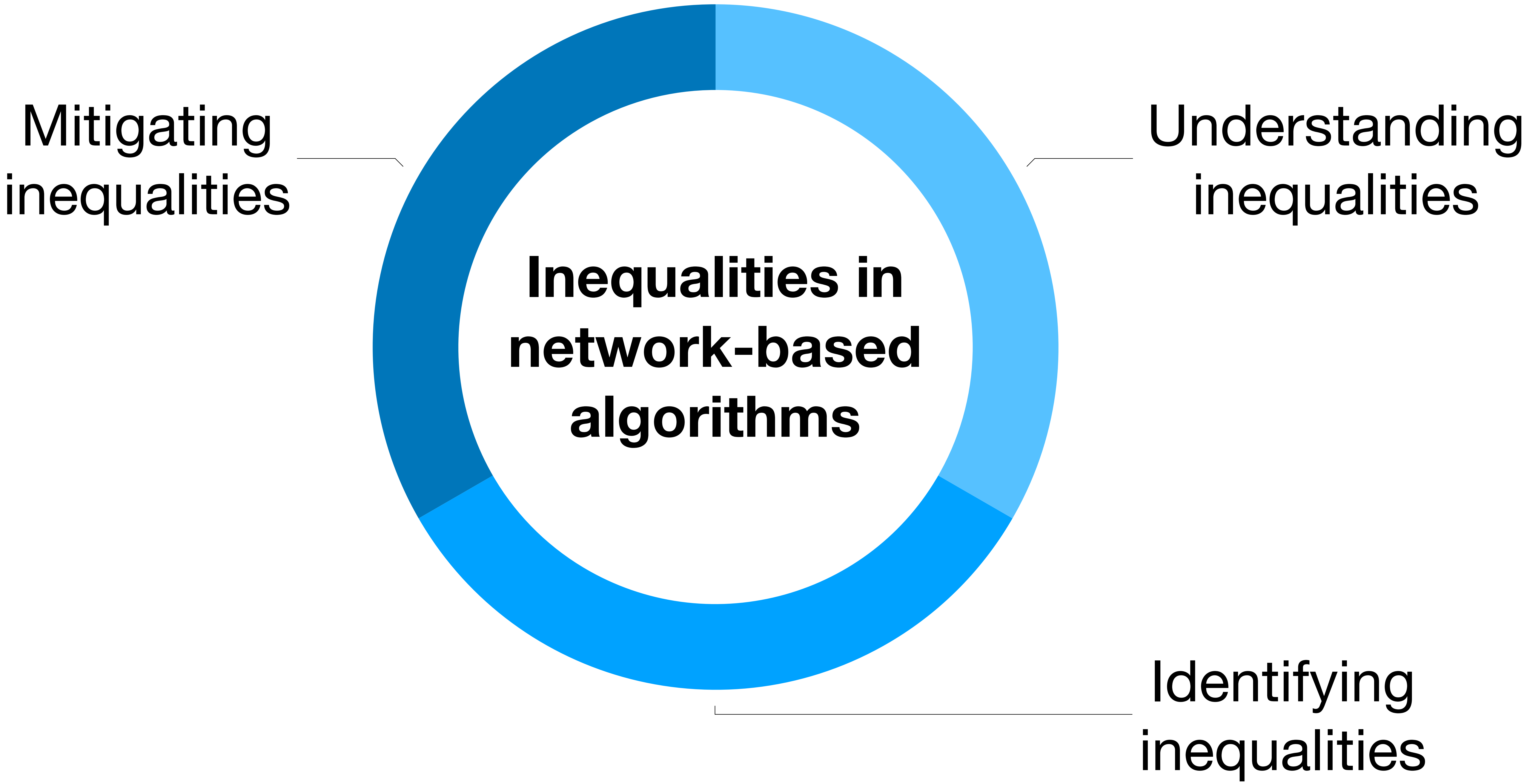
Today’s class

Inequality in network-based algorithms



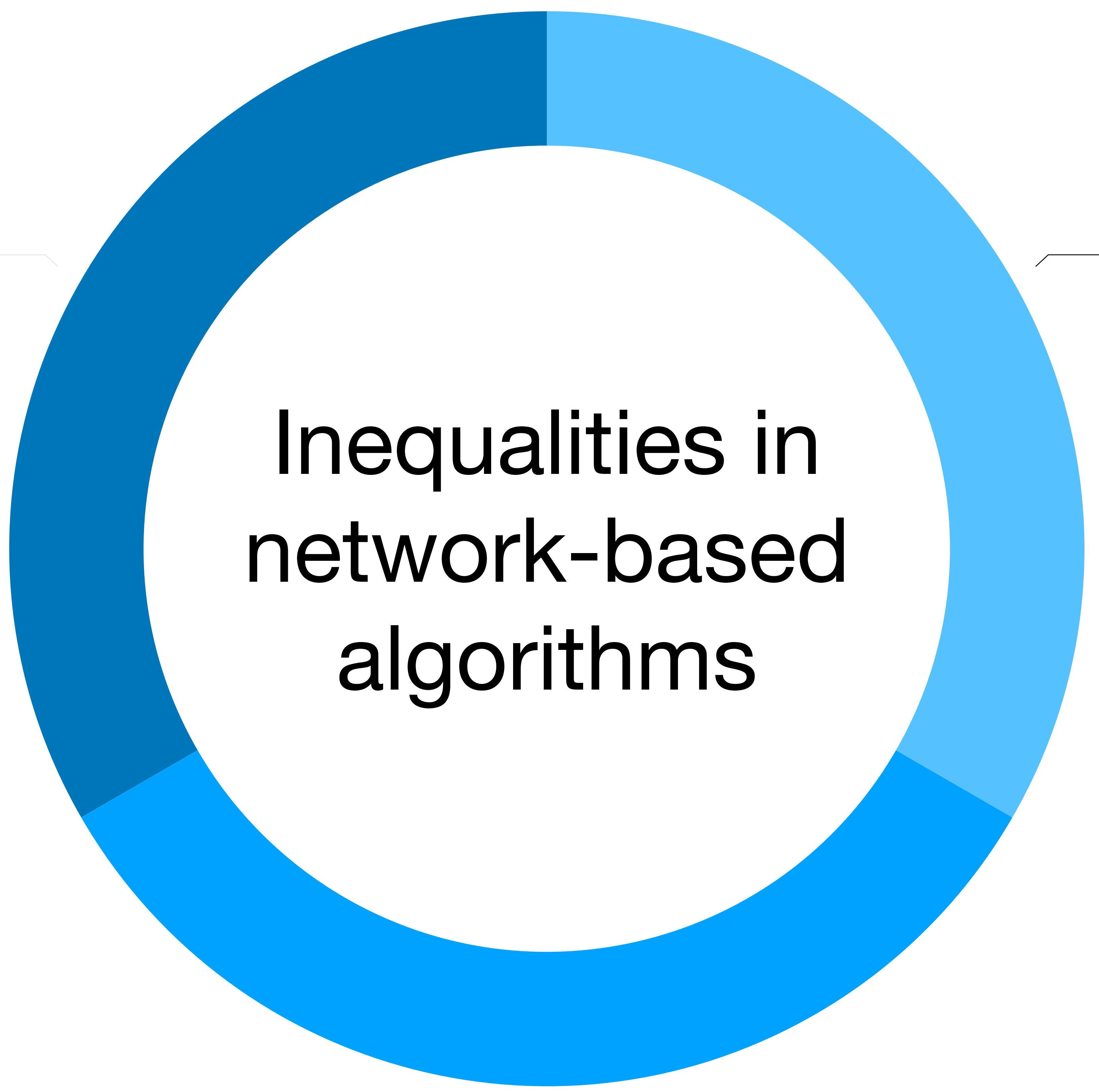
Today's class

Inequality in network-based algorithms



1.

Understanding inequalities



Understanding

Inequalities in network-based algorithms

Understanding

Inequalities in network-based algorithms

1. Understand the impact of these inequalities on society

Understanding

Inequalities in network-based algorithms

1. Understand the impact of these inequalities on society
2. Understand how these inequalities form

Understanding

Inequalities in network-based algorithms

1. Understand the impact of these inequalities on society
2. Understand how these inequalities form
    a. What are the **structure characteristics** and **mechanisms of link formation** in social networks that lead to such inequalities?

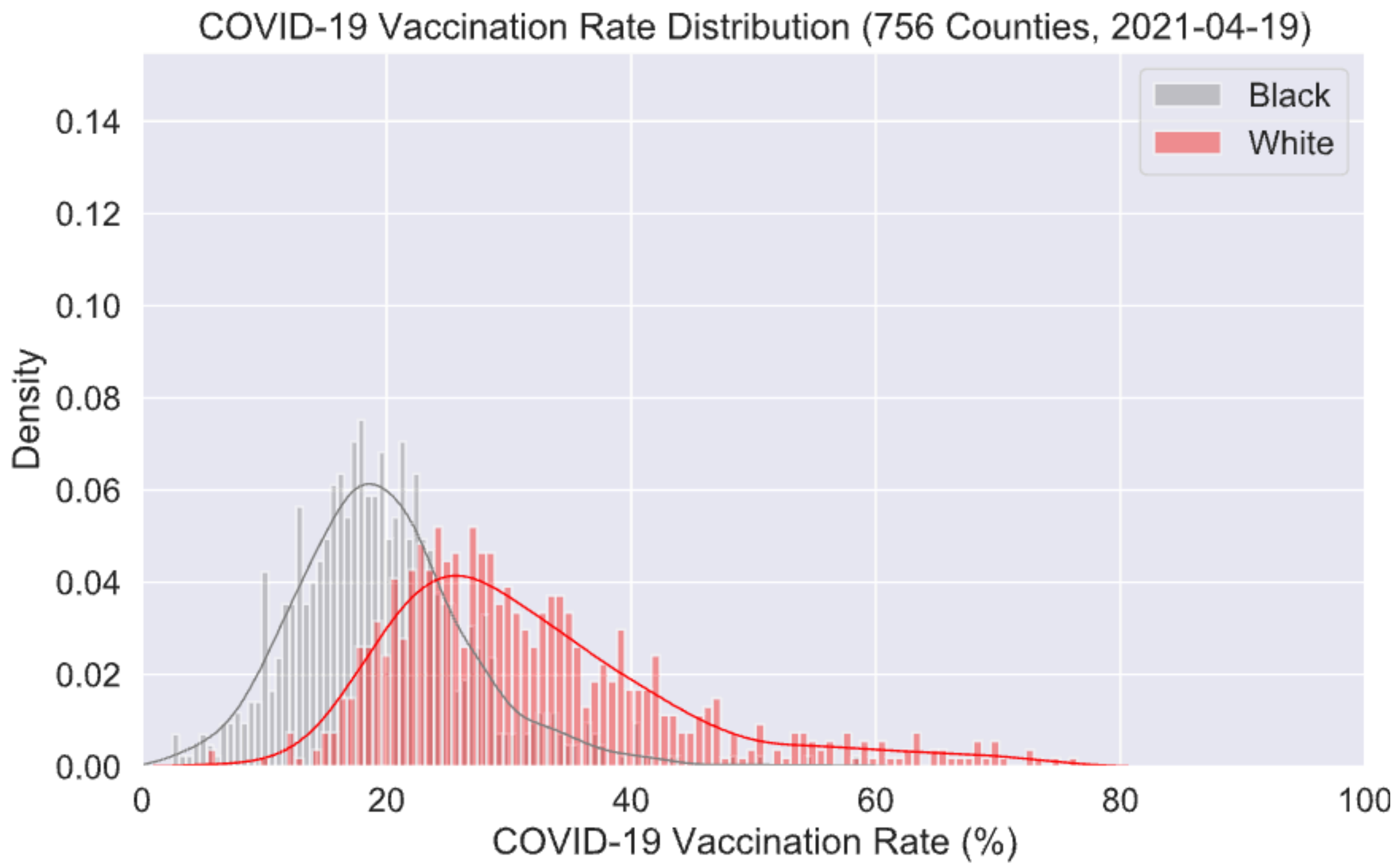
Understanding

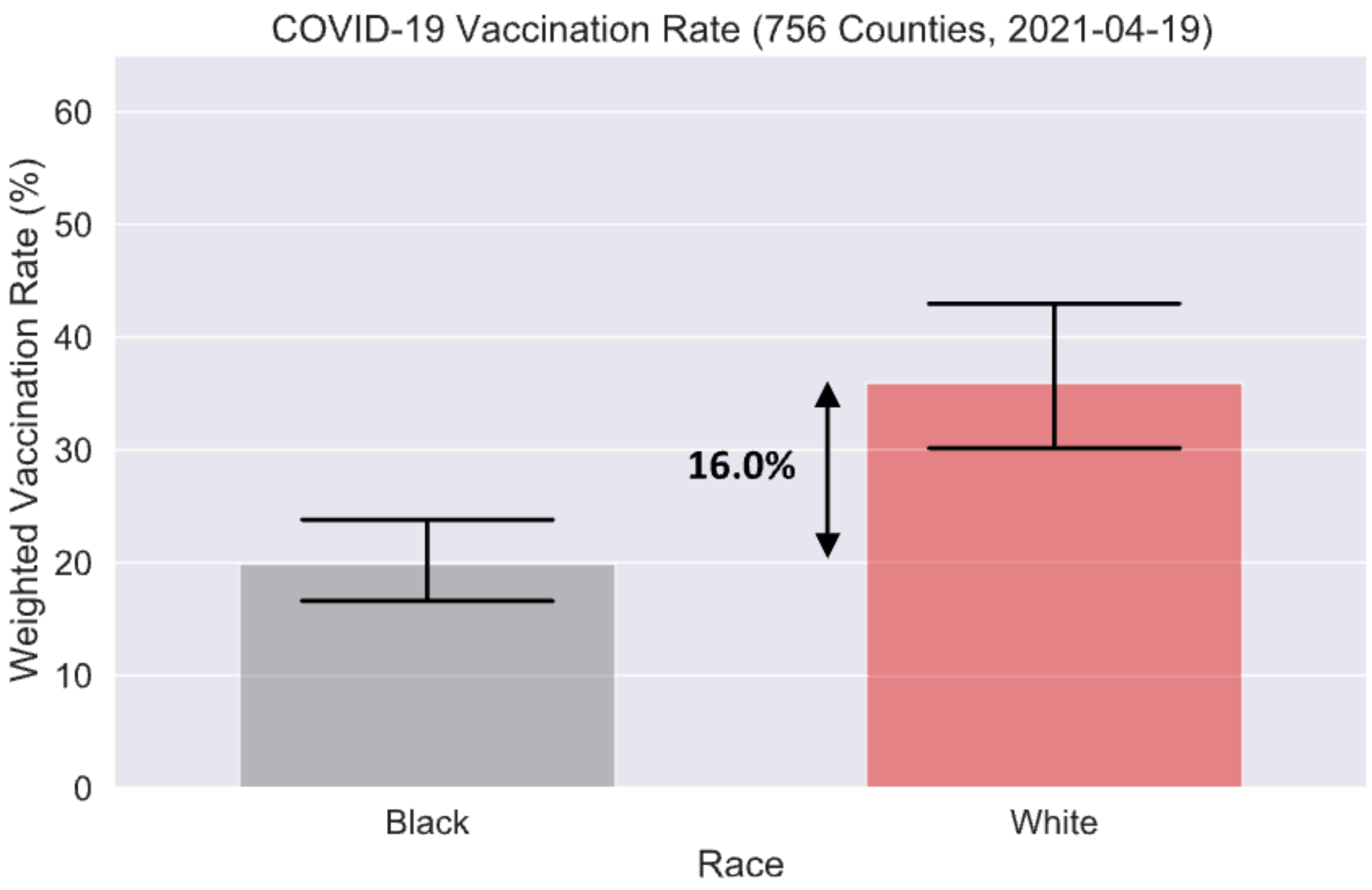
Inequalities in network-based algorithms

If we understand the impact of these inequalities on society, and the correlations with how people behave (connect with each other), we can propose better mitigation strategies to reduce such inequalities!

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?





Agarwal, Ritu, et al. "Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination." Proceedings of the National Academy of Sciences 118.33 (2021): e2107873118.

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?

Table 1. Regression estimates of relationship between social determinants and COVID-19

Variable category	Variable	**CVD = V_W - V_B**
Economic stability	Median income	−2.20* (0.99)
	Median income disparity	0.89† (0.44)
Education access and quality	High school graduation rate	1.22 (1.19)
	High school disparity	2.01*** (0.41)
Healthcare access and quality	Health facilities per capita	0.78 (0.76)
	COVID-19 cases per capita	−0.08 (0.75)
Neighborhood and built environment	Home IT rate	0.51 (0.77)
	Home IT disparity	0.20 (0.99)
	Urban	0.19 (1.23)
	Rate of vehicle ownership	2.07 (1.28)
Social and community context	Political ideology	−6.45** (1.73)
	Segregation index	1.43† (0.69)
	Racial bias	1.43† (0.73)
Constant		8.286*** (1.44)

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?

Table 1. Regression estimates of relationship between social determinants and COVID-19

Variable category	Variable	CVD = V_W - V_B
Economic stability	Median income	−2.20* (0.99)
	Median income disparity	0.89† (0.44)
Education access and quality	High school graduation rate	1.22 (1.19)
	High school disparity	2.01*** (0.41)
Healthcare access and quality	Health facilities per capita	0.78 (0.76)
	COVID-19 cases per capita	−0.08 (0.75)
Neighborhood and built environment	Home IT rate	0.51 (0.77)
	Home IT disparity	0.20 (0.99)
	Urban	0.19 (1.23)
	Rate of vehicle ownership	2.07 (1.28)
Social and community context	Political ideology	−6.45** (1.73)
	Segregation index	1.43† (0.69)
	Racial bias	1.43† (0.73)
Constant		8.286*** (1.44)

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?

1. More votes for the Republican candidate (Trump, 2020 elections) —> less White than Black people getting the vaccine

Table 1. Regression estimates of relationship between social determinants and COVID-19

Variable category	Variable	CVD = V_W - V_B
Economic stability	Median income	−2.20* (0.99)
	Median income disparity	0.89† (0.44)
Education access and quality	High school graduation rate	1.22 (1.19)
	High school disparity	2.01*** (0.41)
Healthcare access and quality	Health facilities per capita	0.78 (0.76)
	COVID-19 cases per capita	−0.08 (0.75)
Neighborhood and built environment	Home IT rate	0.51 (0.77)
	Home IT disparity	0.20 (0.99)
	Urban	0.19 (1.23)
	Rate of vehicle ownership	2.07 (1.28)
Social and community context	Political ideology	−6.45** (1.73)
	Segregation index	1.43† (0.69)
	Racial bias	1.43† (0.73)
Constant		8.286*** (1.44)

Agarwal, Rit... Proceedings ... of the National Academy of Sciences ...

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?

Table 1. Regression estimates of relationship between social determinants and COVID-19

Variable category	Variable	**CVD = V_W - V_B**
Economic stability	Median income	−2.20* (0.99)
	Median income disparity	0.89† (0.44)
Education access and quality	High school graduation rate	1.22 (1.19)
	High school disparity	2.01*** (0.41)
Health care access and quality	[illegible]	[illegible]
	COVID-19 cases per capita	−0.08 (0.75)
Neighborhood and built environment	Home IT rate	0.51 (0.77)
	Home IT disparity	0.20 (0.99)
	Urban	0.19 (1.23)
	Rate of vehicle ownership	2.07 (1.28)
Social and community context	Political ideology	−6.45** (1.73)
	Segregation index	1.43† (0.69)
	Racial bias	1.43† (0.73)
Constant		8.286*** (1.44)

1. More votes for the Republican candidate (Trump, 2020 elections) —> less White than Black people getting the vaccine

2. More segregated the area (higher homophily) —> more White than Black people getting the vaccine

Agarwal, Rit[illegible]
Proceedings [illegible]

Why do racial disparities in vaccination

persist despite the increased availability of vaccines?

Table 1. Regression estimates of relationship between social determinants and COVID-19

Variable category	Variable	CVD = V_W - V_B
Economic stability	Median income	−2.20* (0.99)
	Median income disparity	0.89† (0.44)
Education access and quality	High school graduation rate	[illegible]
	High school disparity	2.01*** (0.41)
[illegible]	[illegible]	[illegible]
	COVID-19 cases per capita	−0.08 (0.75)
Neighborhood and built environment	Home IT rate	0.51 (0.77)
	Home IT disparity	0.20 (0.99)
	Urban	0.19 (1.23)
	Rate of vehicle ownership	2.07 (1.28)
Social and community context	Political ideology	−6.45** (1.73)
	Segregation index	1.43† (0.69)
	Racial bias	1.43† (0.73)
Constant		8.286*** (1.44)

1. More votes for the Republican candidate (Trump, 2020 elections) —> less White than Black people getting the vaccine

2. More segregated the area (higher homophily) —> more White than Black people getting the vaccine

3. More bias against Black people —> more White than Black people getting the vaccine

Agarwal, Rit... Proceedings of the National Academy of Sciences 118.33 (2021): e2107873118.

Why do racial disparities in vaccination persist despite the increased availability of vaccines?

If we don't control for these disparities in network-based algorithms, we can exacerbate discrimination against already marginalized groups.

Agarwal, Ritu, et al. "Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination." Proceedings of the National Academy of Sciences 118.33 (2021): e2107873118.

Optimization vs. Fairness

In influence maximization

Rahmattalabi, Aida, et al. "Fair influence maximization: A welfare optimization approach." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 13. 2021.

Optimization vs. Fairness

In influence maximization

- Public health interventions, such as COVID-19 or suicide/HIV prevention, and community preparedness against natural disasters, leverage social network information to maximize outreach.

Rahmattalabi, Aida, et al. "Fair influence maximization: A welfare optimization approach." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 13. 2021.

Optimization vs. Fairness

In influence maximization

- Public health interventions, such as COVID-19 or suicide/HIV prevention, and community preparedness against natural disasters, leverage social network information to maximize outreach.

- **Algorithmic influence maximization** techniques have been proposed to aid with the choice of “peer leaders” or “influencers” in such interventions.

Rahmattalabi, Aida, et al. "Fair influence maximization: A welfare optimization approach." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 13. 2021.

Optimization vs. Fairness

In influence maximization

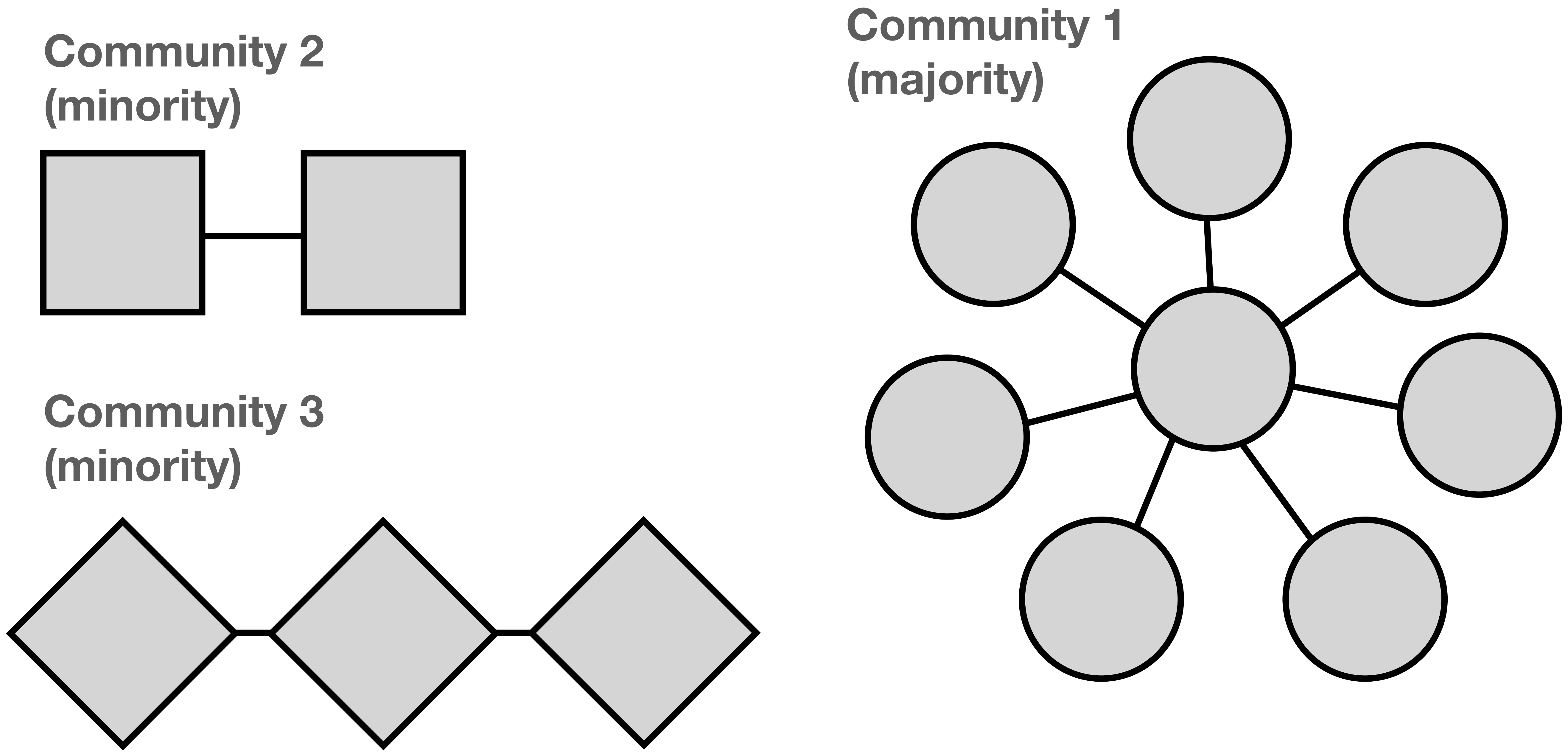
- Public health interventions, such as COVID-19 or suicide/HIV prevention, and community preparedness against natural disasters, leverage social network information to maximize outreach.
- **Algorithmic influence maximization** techniques have been proposed to aid with the choice of “peer leaders” or “influencers” in such interventions.
- Yet, traditional algorithms for influence maximization have not been designed with fair / diversity interventions in mind. As a result, they may **disproportionately exclude minority communities from the benefits of the intervention**.

Rahmattalabi, Aida, et al. "Fair influence maximization: A welfare optimization approach." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 13. 2021.

Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)

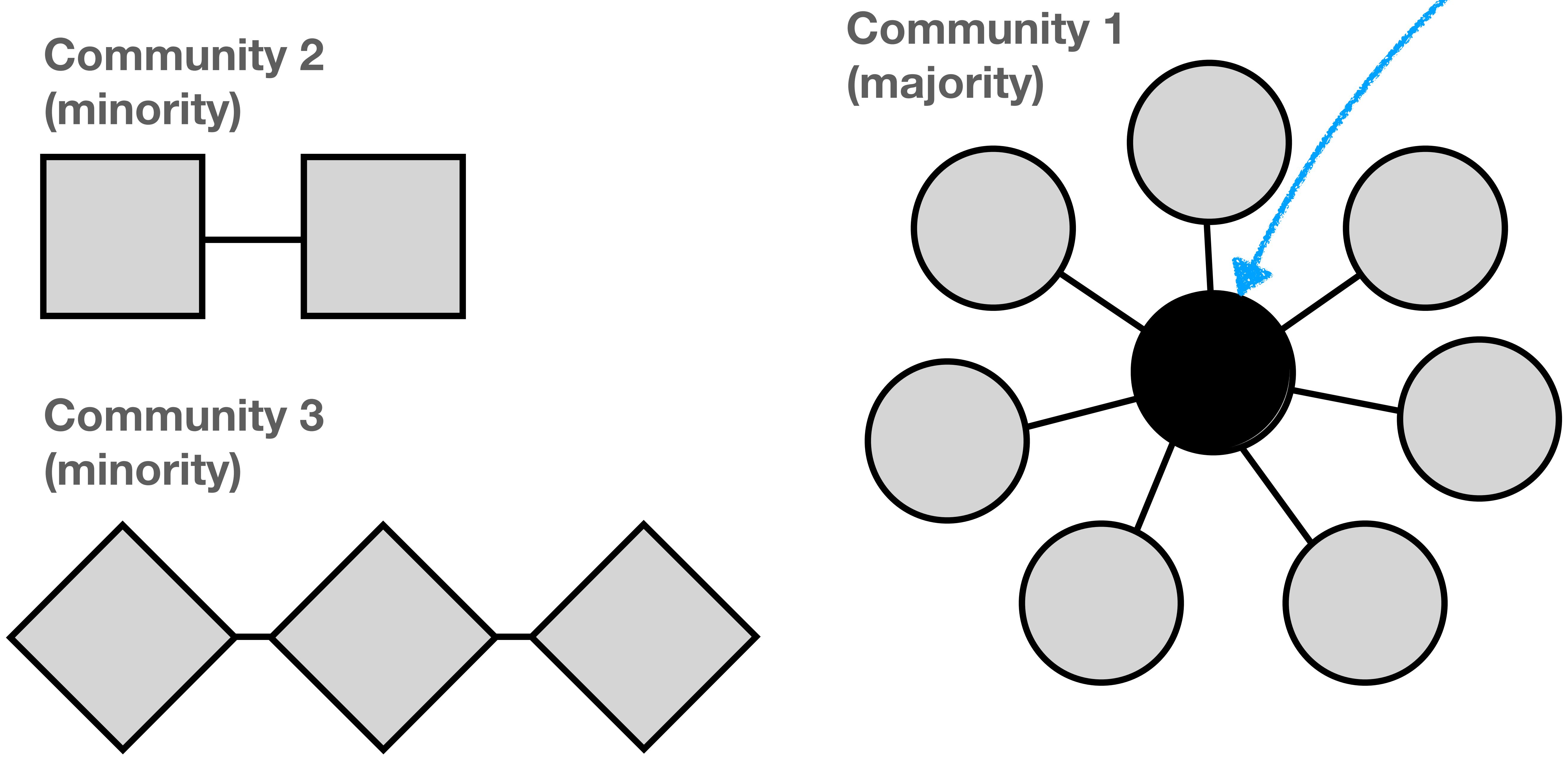


Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)

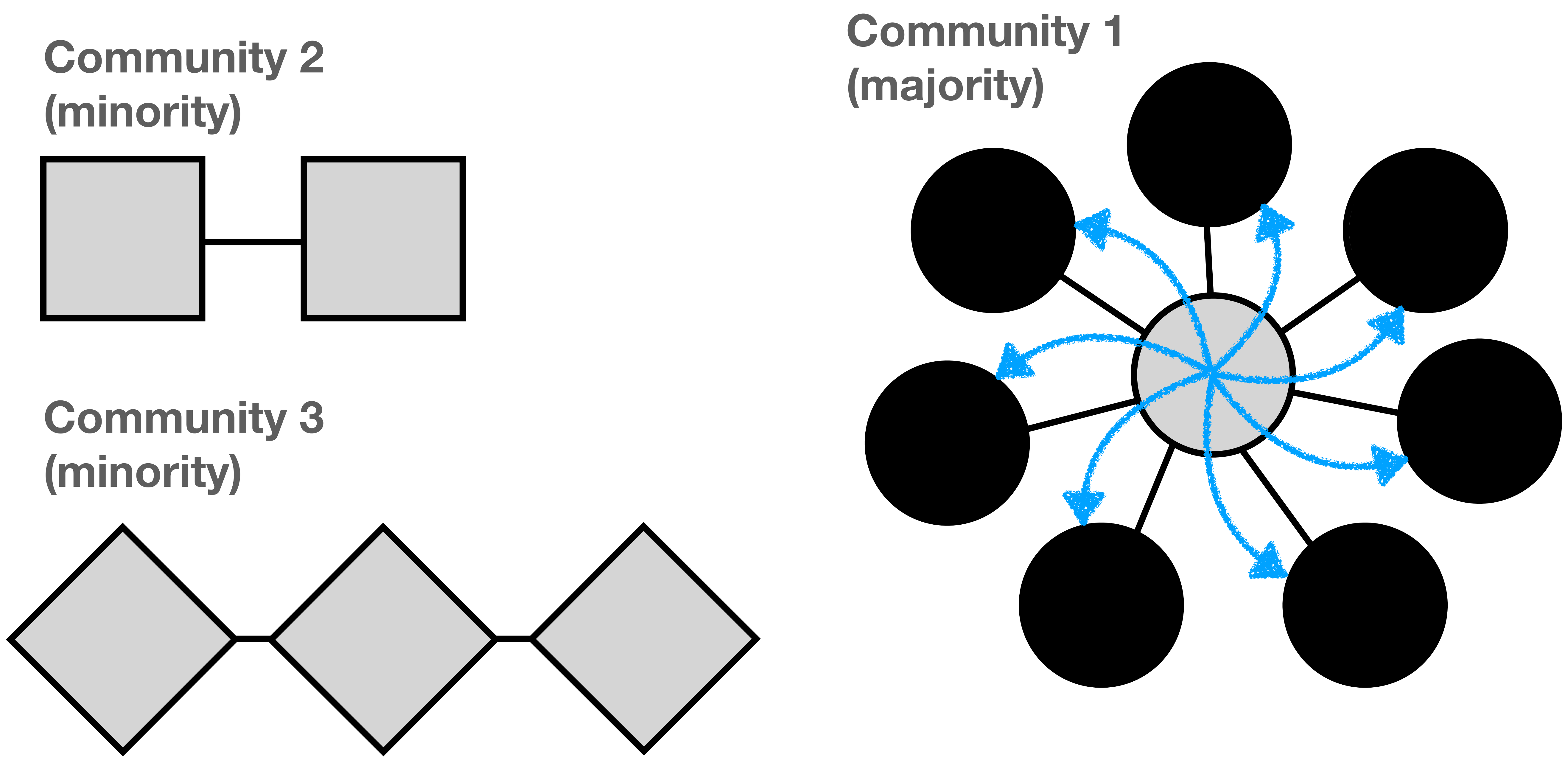
STEP 1: Select seed nodes
Seed, spreader, influential node



Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)



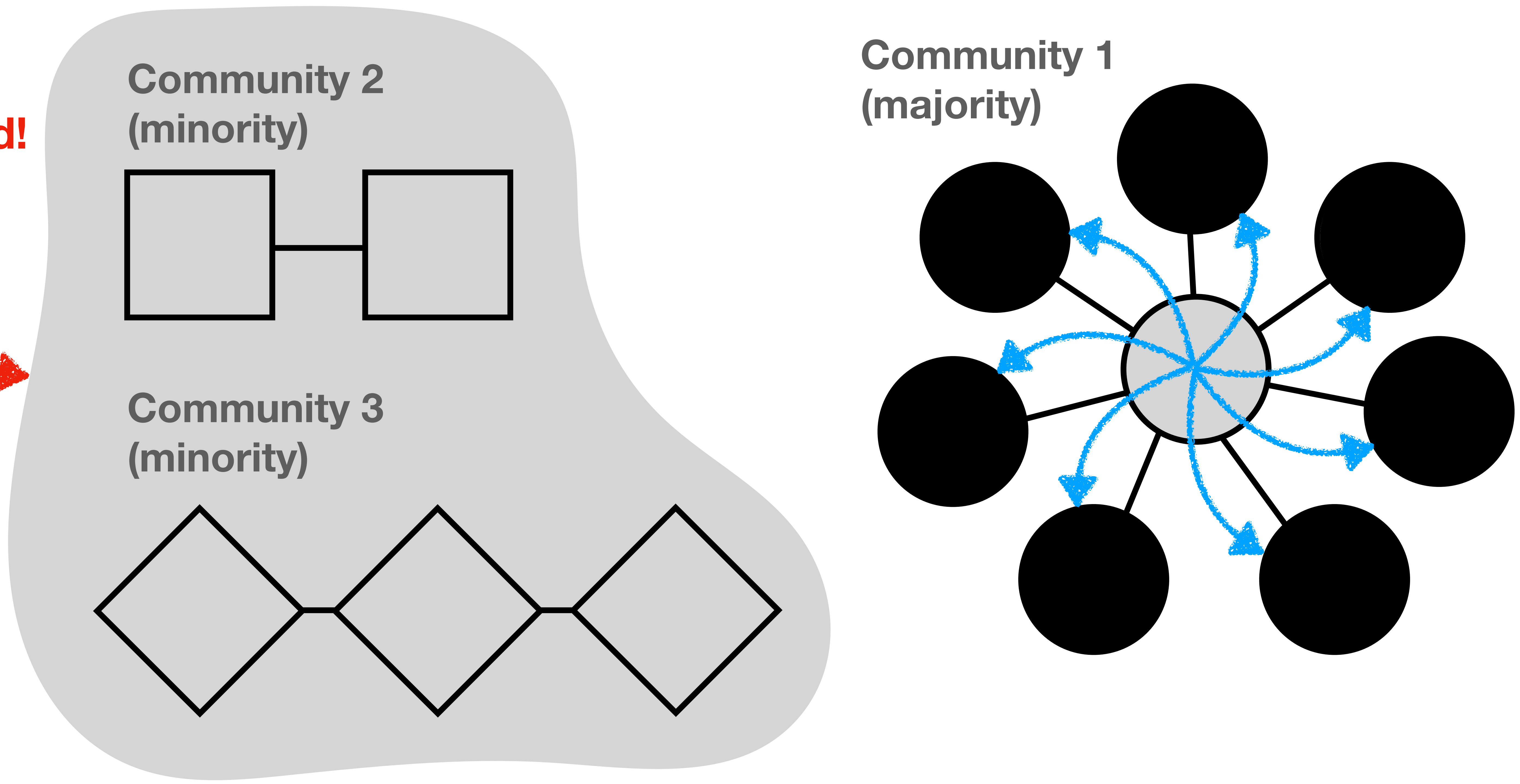
STEP 2: Spread influence
- Not getting infected
- Sharing a resource / info

Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)

These communities did not benefit from the “influential” spread!

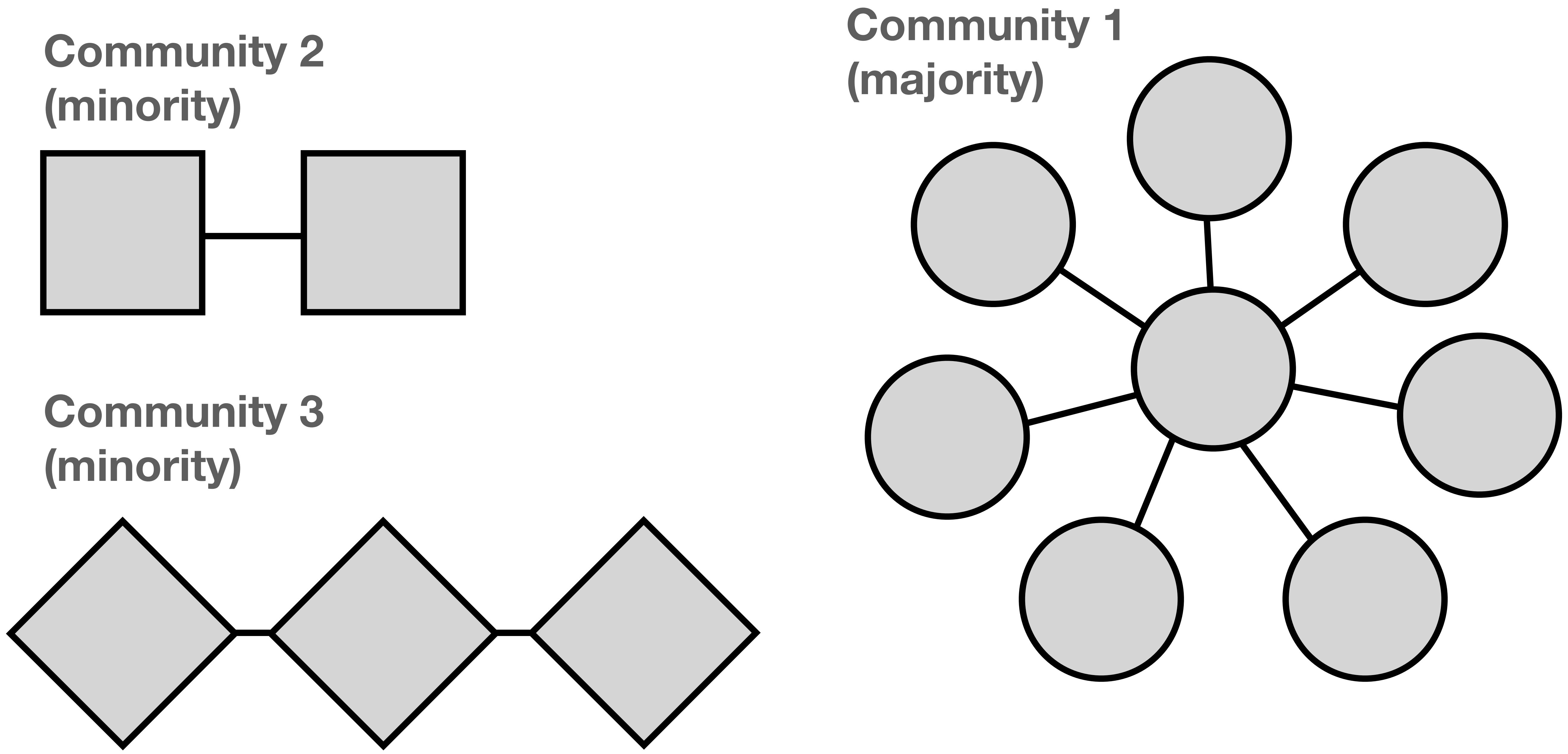


STEP 2: Spread influence
- Not getting infected
- Sharing a resource / info

Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)

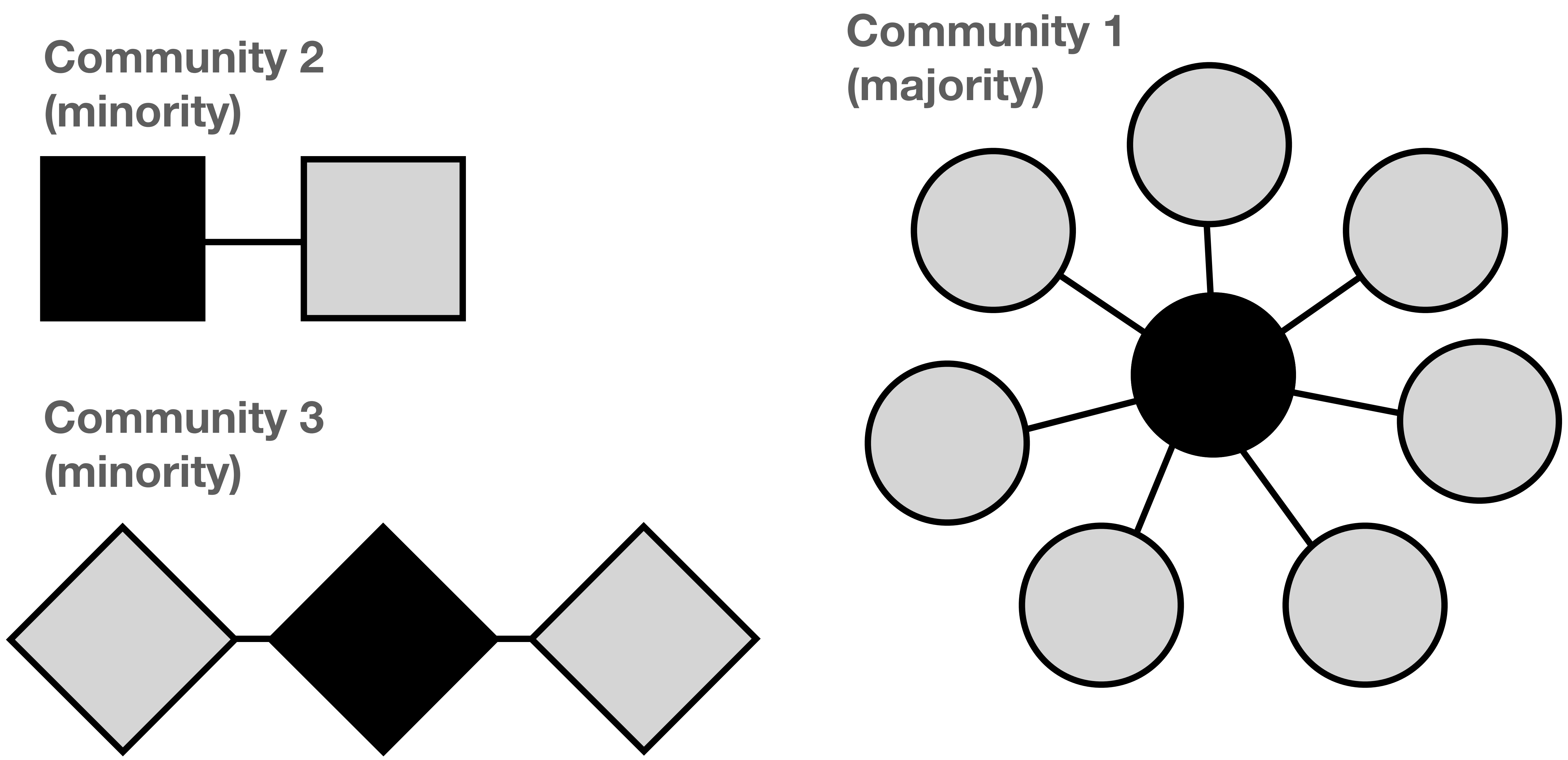


Optimization vs. Fairness

In influence maximization (example)

Mitigation strategy #1
Group fairness (more diverse outreach)

High segregation
(Like in the example before)



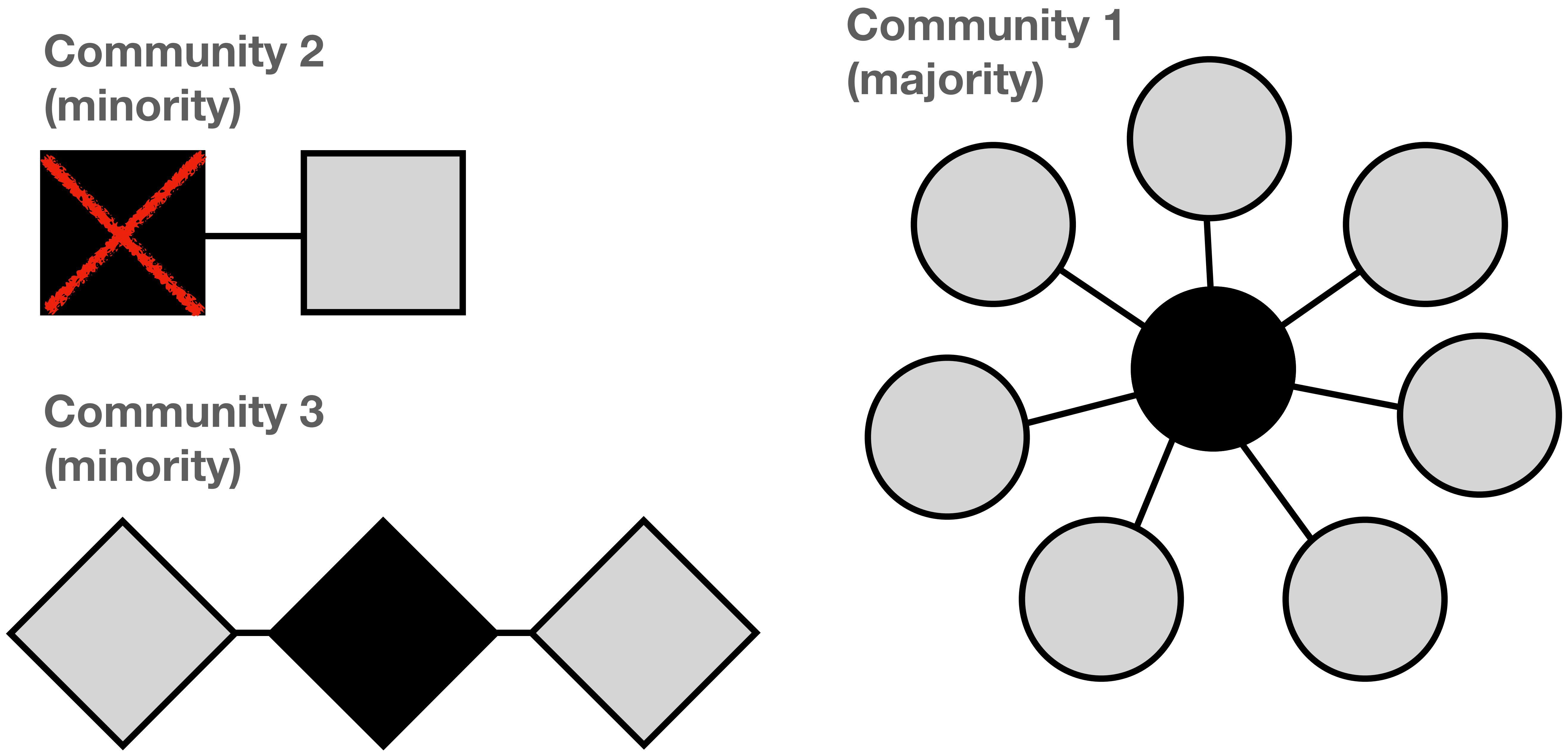
Optimization vs. Fairness

In influence maximization (example)

Mitigation strategy #1
Group fairness (more diverse outreach)

Trade-off between
- **budget** (number of seeds to be selected)
- And **utility** (outreach)

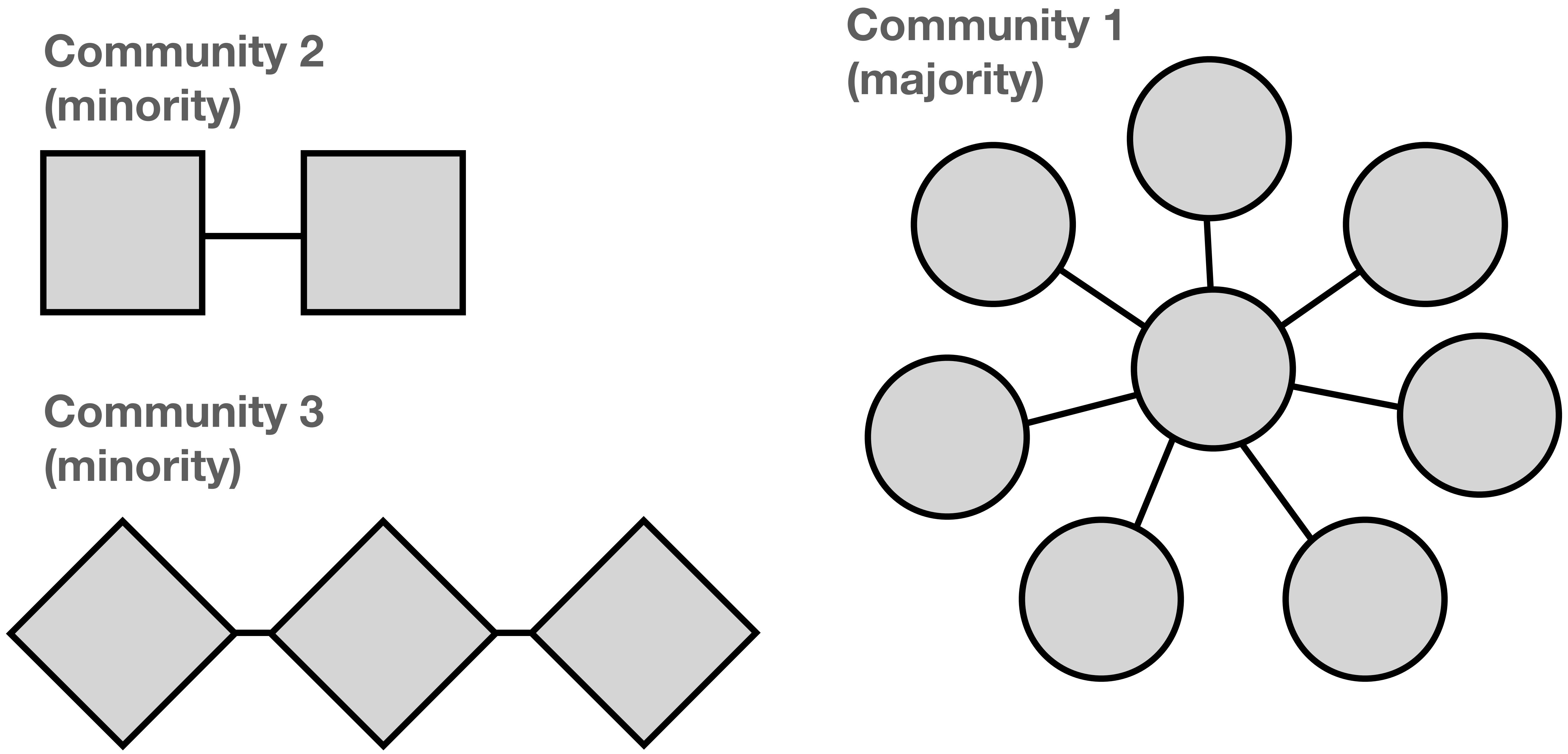
High segregation
(Like in the example before)



Optimization vs. Fairness

In influence maximization (example)

High segregation
(Like in the example before)

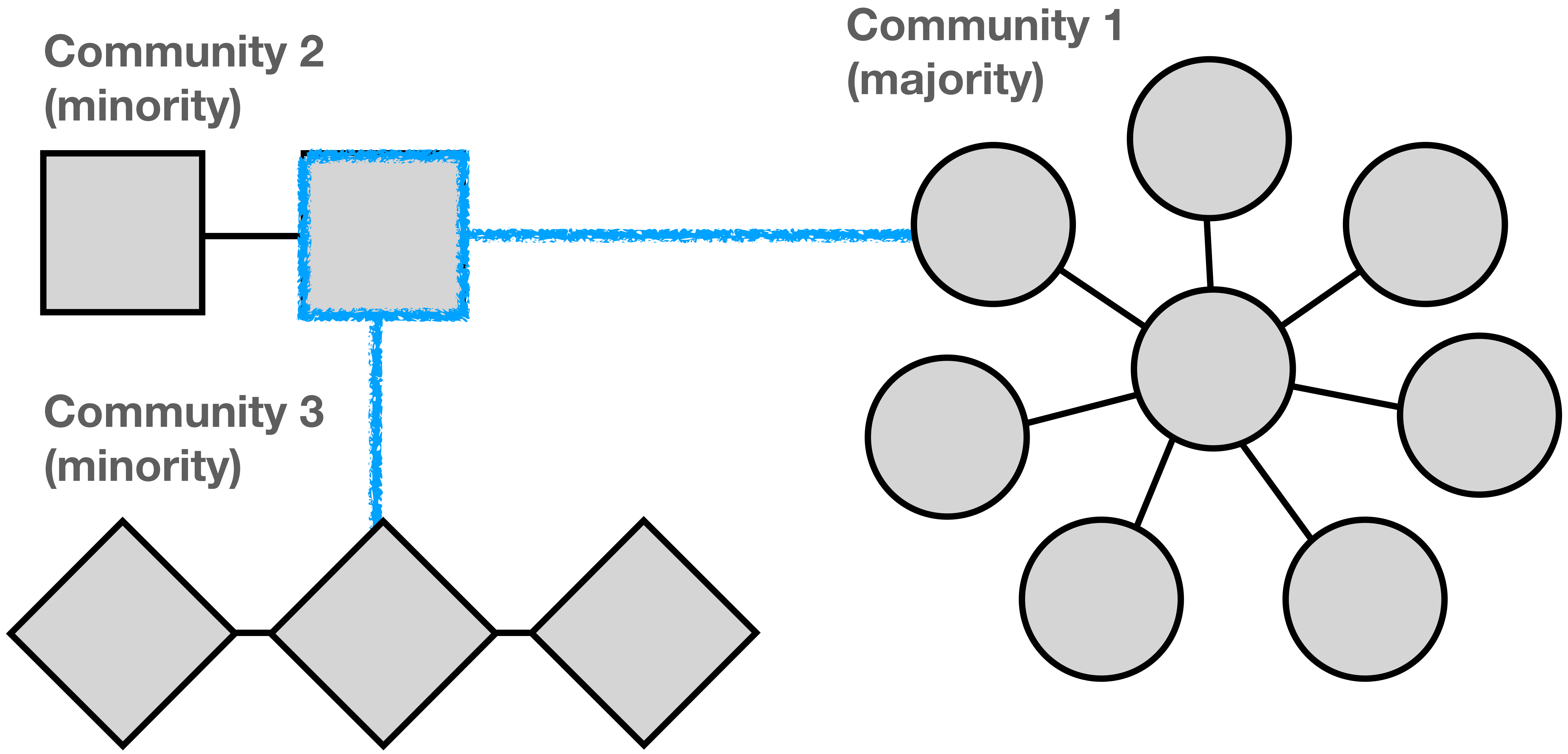


Optimization vs. Fairness

In influence maximization (example)

Mitigation strategy #2
Reducing segregation
by changing the structure of the network

High segregation
(Like in the example before)

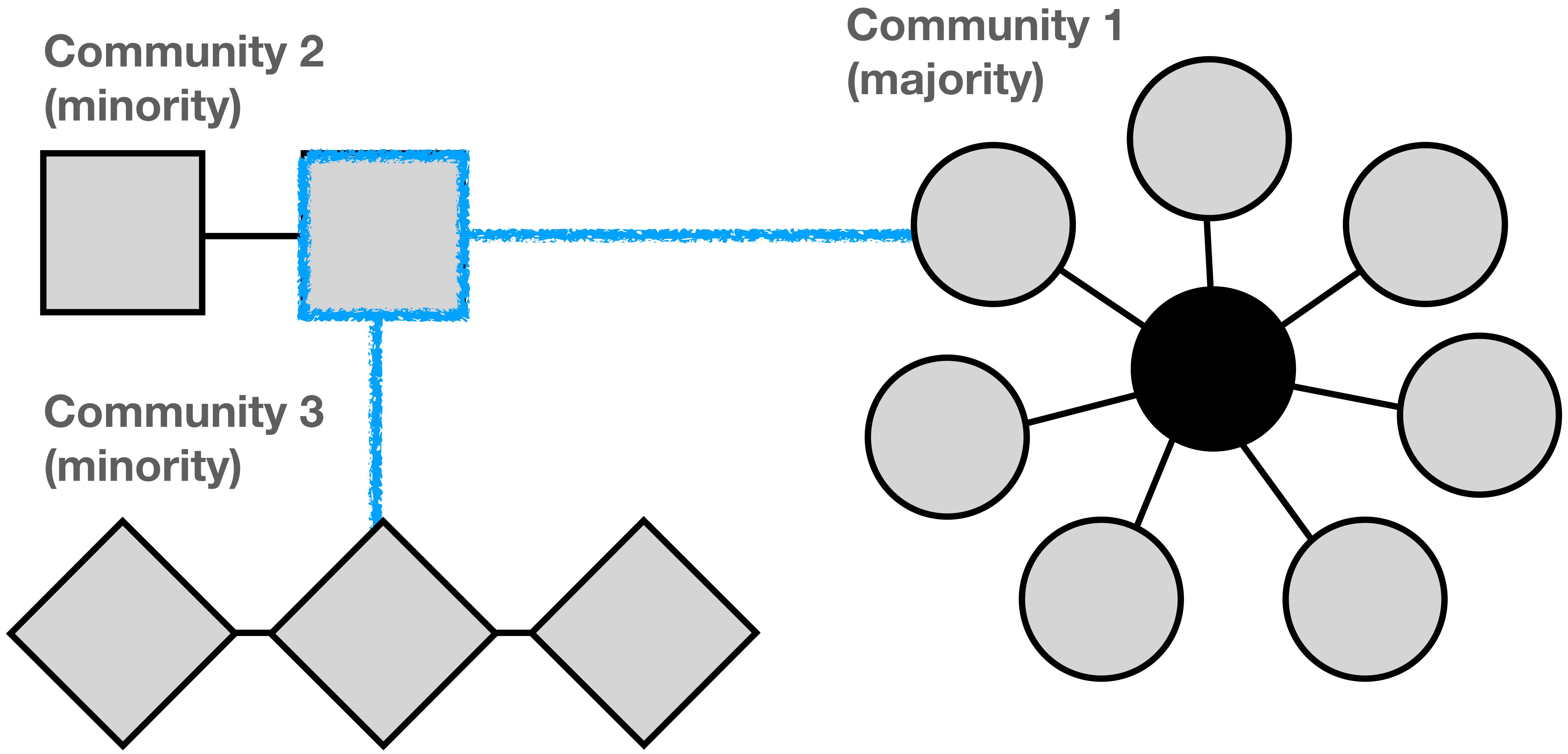


Optimization vs. Fairness

In influence maximization (example)

Mitigation strategy #2
Reducing segregation
by changing the structure of the network

High segregation
(Like in the example before)

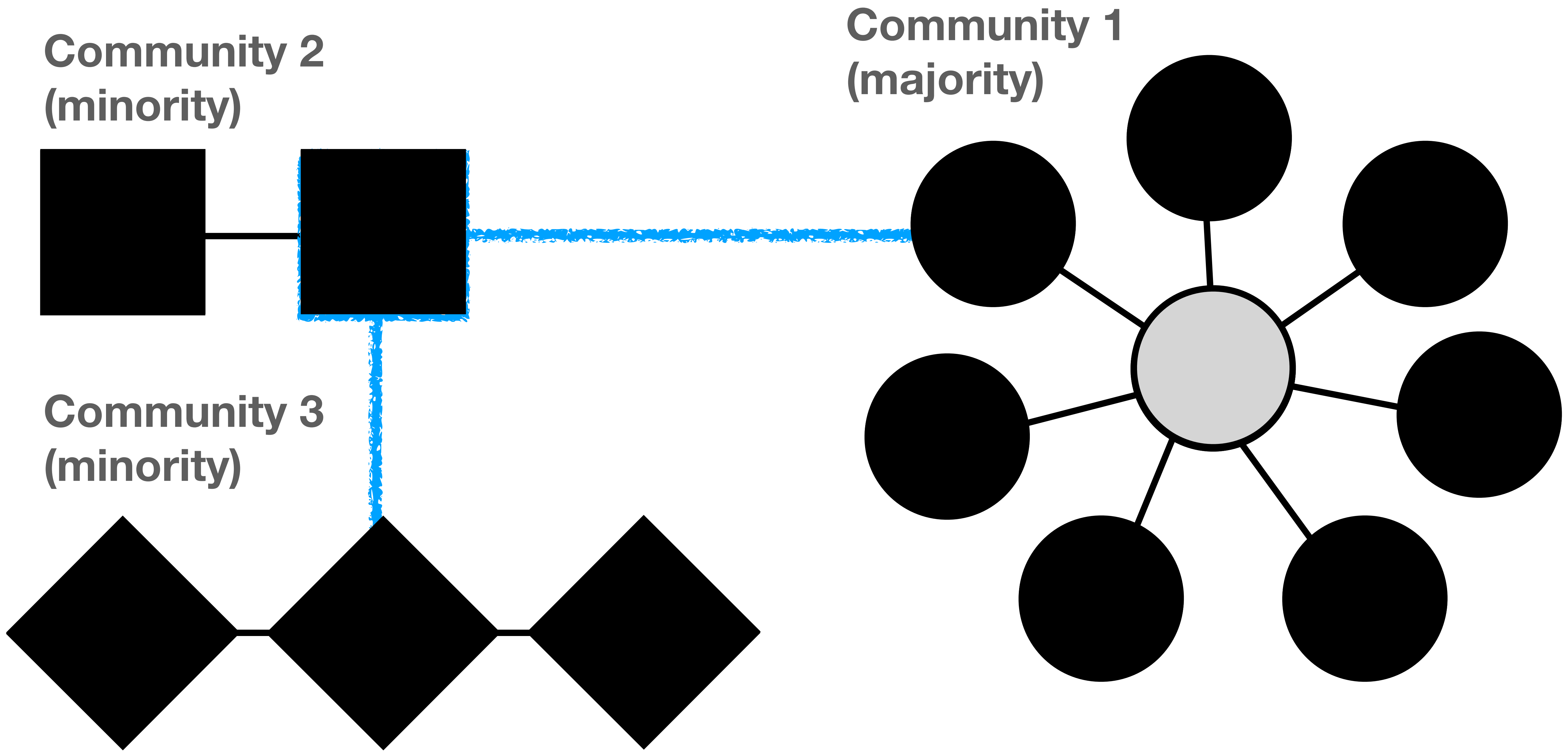


Optimization vs. Fairness

In influence maximization (example)

Mitigation strategy #2
Reducing segregation
by changing the structure of the network

High segregation
(Like in the example before)



Tools to understand…

…how inequalities in network-based algorithms emerge

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
  - What if the network is …

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
  - What if the network is …
    - Big/small, dense/sparse, homophilic wrt. gender…

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
  - What if the network is …
    - Big/small, dense/sparse, homophilic wrt. gender…
  - What if people have these characteristics …

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
    - What if the network is …
        - Big/small, dense/sparse, homophilic wrt. gender…
    - What if people have these characteristics …
        - 49% are female, 50% are men, 1% are non-binary

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
    - What if the network is …
        - Big/small, dense/sparse, homophilic wrt. gender…
    - What if people have these characteristics …
        - 49% are female, 50% are men, 1% are non-binary
        - Or 19% are female, 80% are men, 1% are non-binary

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
  - What if the network is …
    - Big/small, dense/sparse, homophilic wrt. gender…
  - What if people have these characteristics …
    - 49% are female, 50% are men, 1% are non-binary
    - Or 19% are female, 80% are men, 1% are non-binary
- Where do we get all these different networks?

Tools to understand…

…how inequalities in network-based algorithms emerge

- Via “What if scenarios”
  - What if the network is …
    - Big/small, dense/sparse, homophilic wrt. gender…
  - What if people have these characteristics …
    - 49% are female, 50% are men, 1% are non-binary
    - Or 19% are female, 80% are men, 1% are non-binary
- Where do we get all these different networks?
  - Using **synthetic data** (random network generators)

Synthetic networks

…and where to find them

Synthetic networks

…and where to find them

- What are synthetic networks?

Synthetic networks

…and where to find them

- What are synthetic networks?
    - Networks that are generated using a graph model

Synthetic networks

…and where to find them

- What are synthetic networks?
    - Networks that are generated using a graph model
    - The generation of nodes and links often mimics some properties of real-world networks, e.g.,

Synthetic networks

…and where to find them

- What are synthetic networks?
  - Networks that are generated using a graph model
  - The generation of nodes and links often mimics some properties of real-world networks, e.g.,
    - Nodes with attributes and class (im)balance

Synthetic networks

…and where to find them

- What are synthetic networks?
  - Networks that are generated using a graph model
  - The generation of nodes and links often mimics some properties of real-world networks, e.g.,
    - Nodes with attributes and class (im)balance
    - Biased links towards: popular nodes, similar nodes, friends-of-friends, etc.

Synthetic networks

…and where to find them

- What are synthetic networks?
  - Networks that are generated using a graph model
  - The generation of nodes and links often mimics some properties of real-world networks, e.g.,
    - Nodes with attributes and class (im)balance
    - Biased links towards: popular nodes, similar nodes, friends-of-friends, etc.
    - (We can also model completely random non-attributed graphs)

Model Class	Examples	Characteristics
Static Models	Erdos–Rényi Watts-Strogatz	• N fixed • pk exponentially bounded • Static, time independent topologies
Generative Models	Configuration Model Hidden Parameter Model	• Arbitrary pre-defined p_k • Static, time independent topologies
Evolving Network Models	Barabási–Albert Model Bianconi-Barabási Model Initial Attractiveness Model Internal Links Model Node Deletion Model Accelerated Growth Model Aging Model	• p_k is determined by the processes that contribute to the network's evolution. • Time-varying network topologies

Table 6.1

Classes of Models in Network Science

The table summarizes the three main modeling frameworks used in network science, together with their distinguishing features.

http://networksciencebook.com

Model Class	Examples	Characteristics
Static Models	Erdos–Rényi Watts-Strogatz	• N fixed • pk exponentially bounded • Static, time independent topologies
Generative Models	Configuration Model Hidden Parameter Model	• Arbitrary pre-defined p_k • Static, time independent topologies
Evolving Network Models	Barabási–Albert Model Bianconi-Barabási Model Initial Attractiveness Model Internal Links Model Node Deletion Model Accelerated Growth Model Aging Model	• p_k is determined by the processes that contribute to the network's evolution. • Time-varying network topologies

Today, we will cover this graph model and some variations

Table 6.1

Classes of Models in Network Science

The table summarizes the three main modeling frameworks used in network science, together with their distinguishing features.

http://networksciencebook.com

Mechanisms of edge formation

Found in real-world social networks

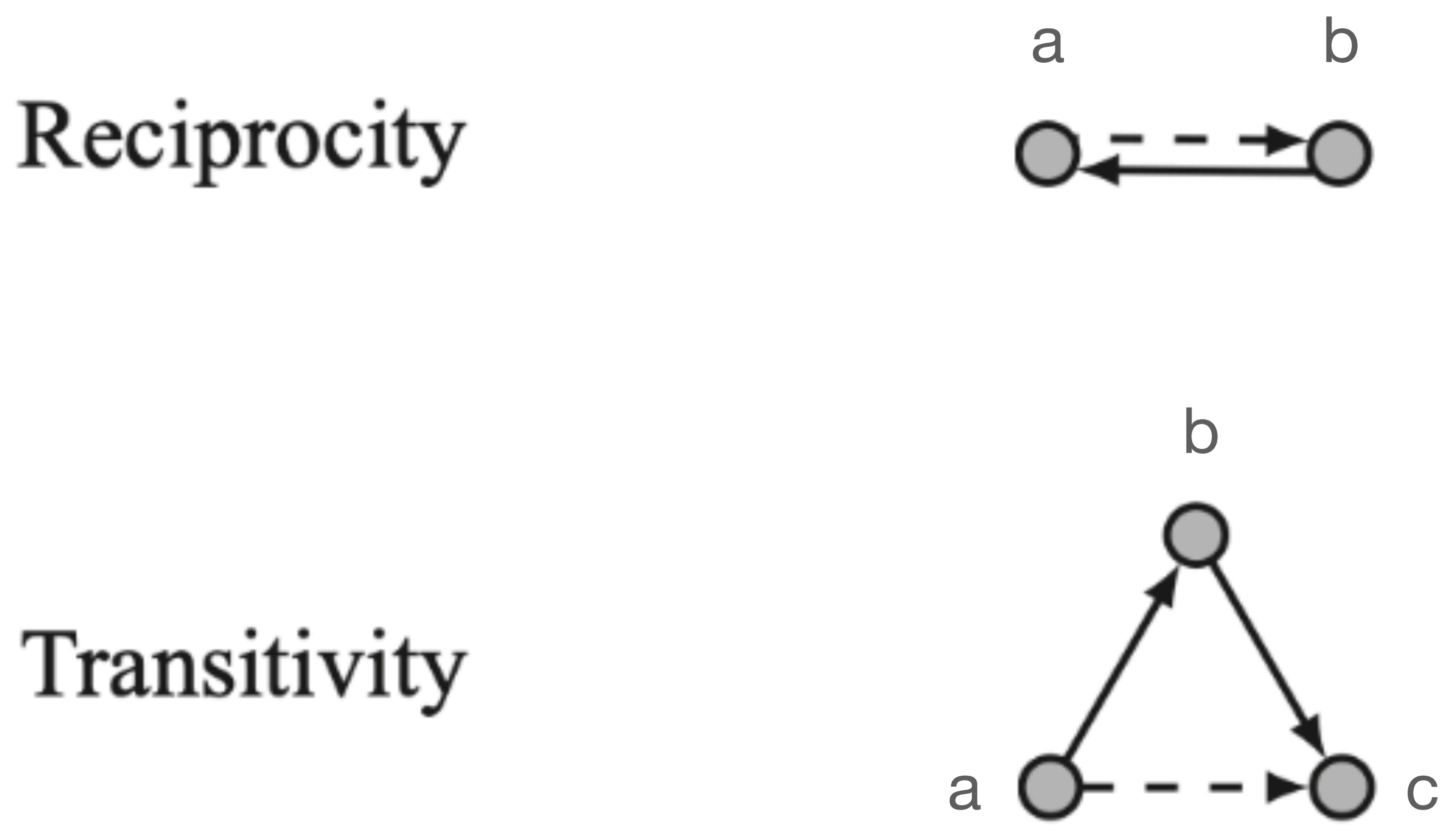
Mechanisms of edge formation

Found in real-world social networks

Reciprocity

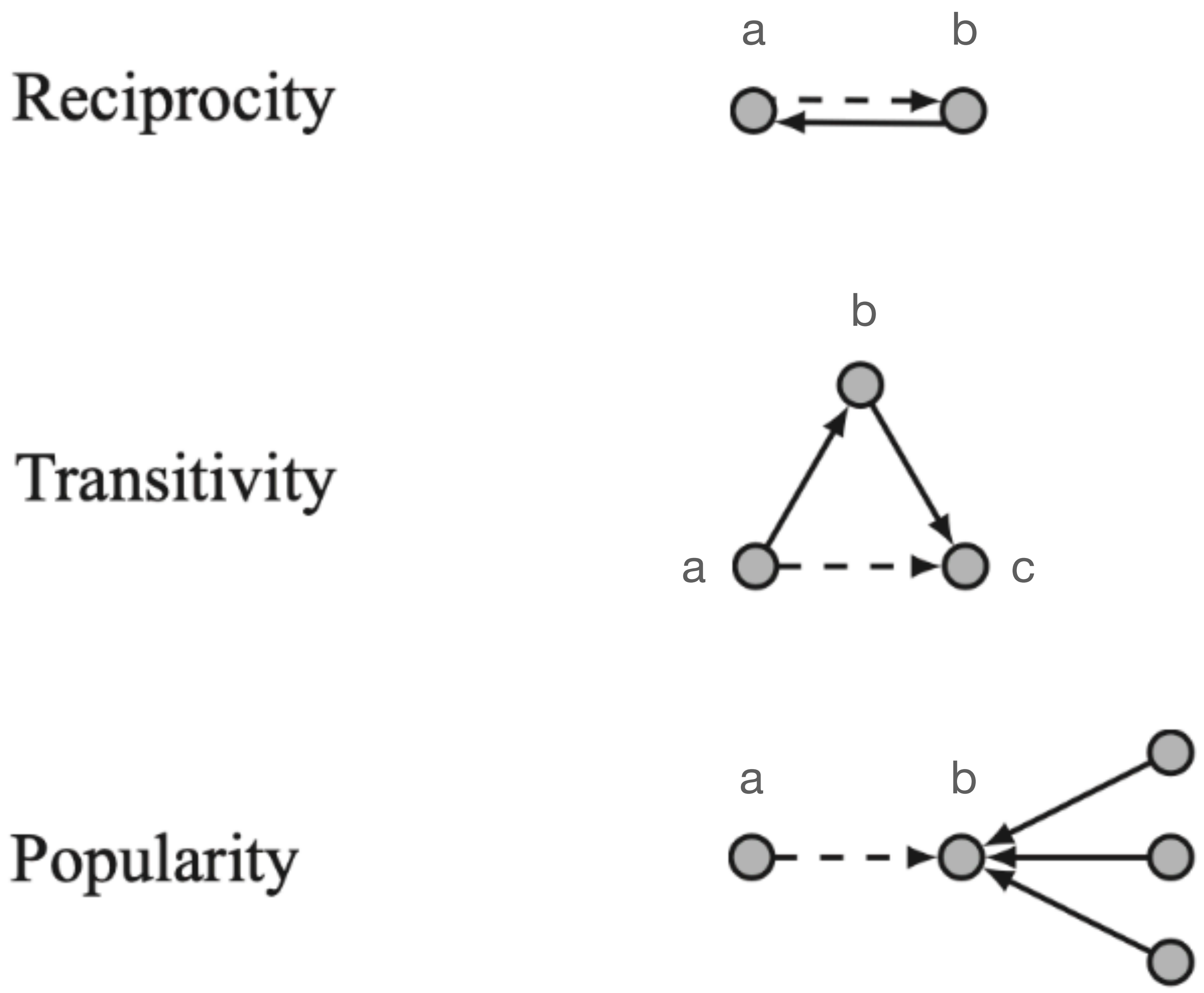
Mechanisms of edge formation

Found in real-world social networks



Mechanisms of edge formation

Found in real-world social networks



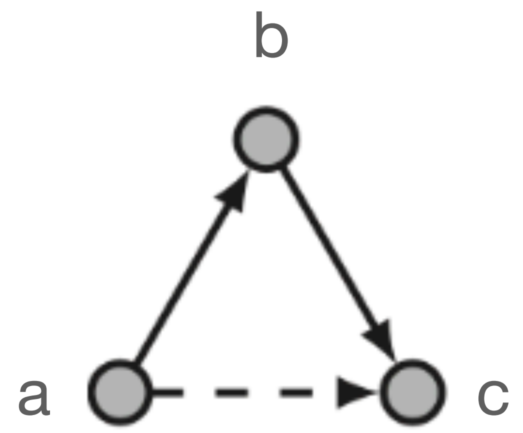
Mechanisms of edge formation

Found in real-world social networks

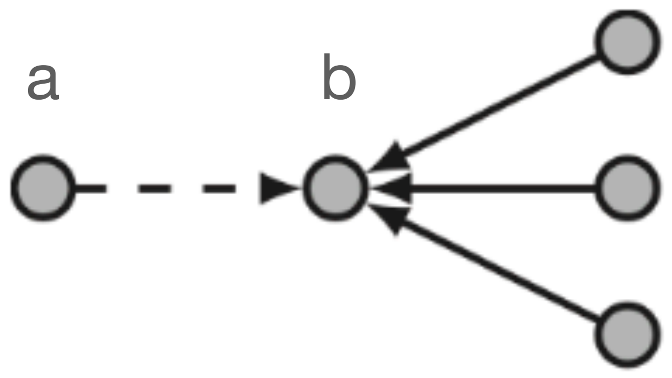
Reciprocity

Activity

Transitivity



Popularity

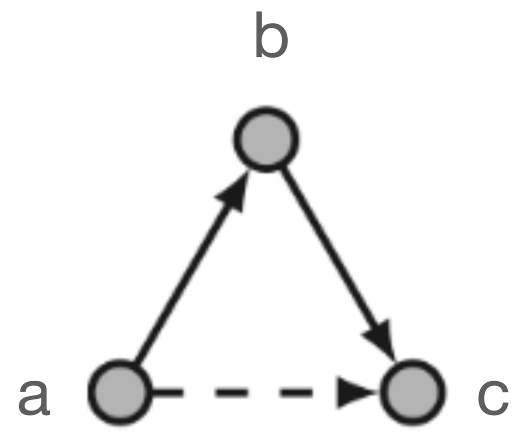


Mechanisms of edge formation

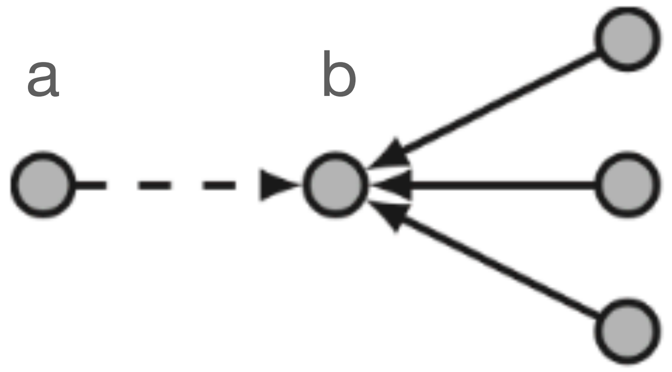
Found in real-world social networks

Reciprocity

Transitivity



Popularity



Activity

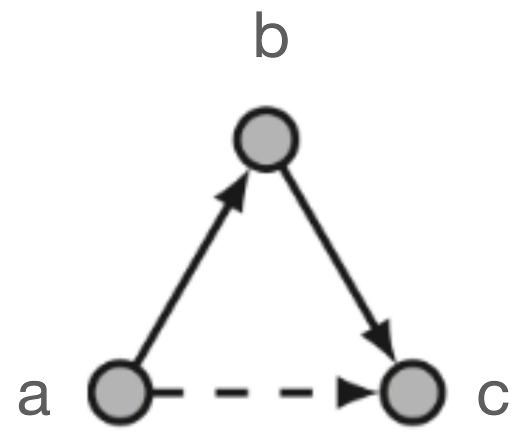
Attraction

Mechanisms of edge formation

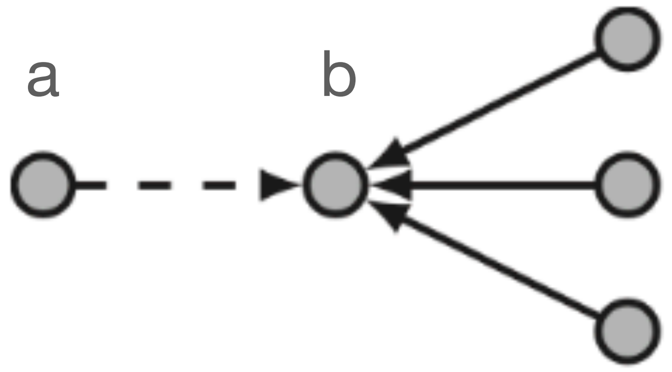
Found in real-world social networks

Reciprocity

Transitivity



Popularity



Activity

Attraction

Homophily

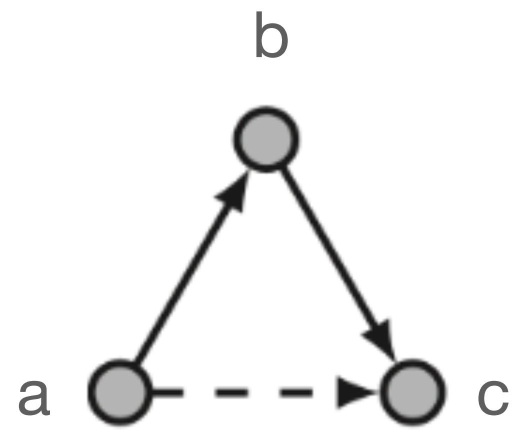
Mechanisms of edge formation

Found in real-world social networks

Reciprocity

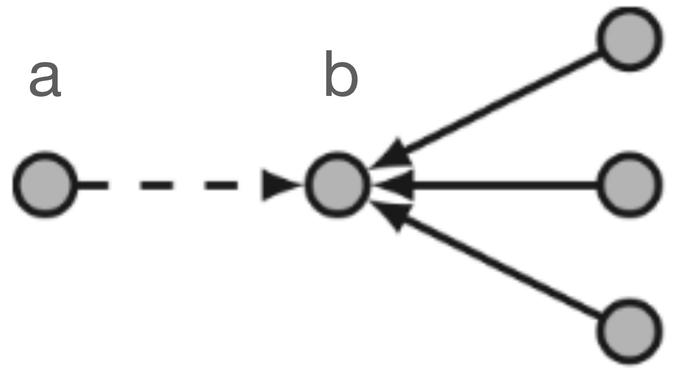
Activity

Transitivity



Attraction

Popularity

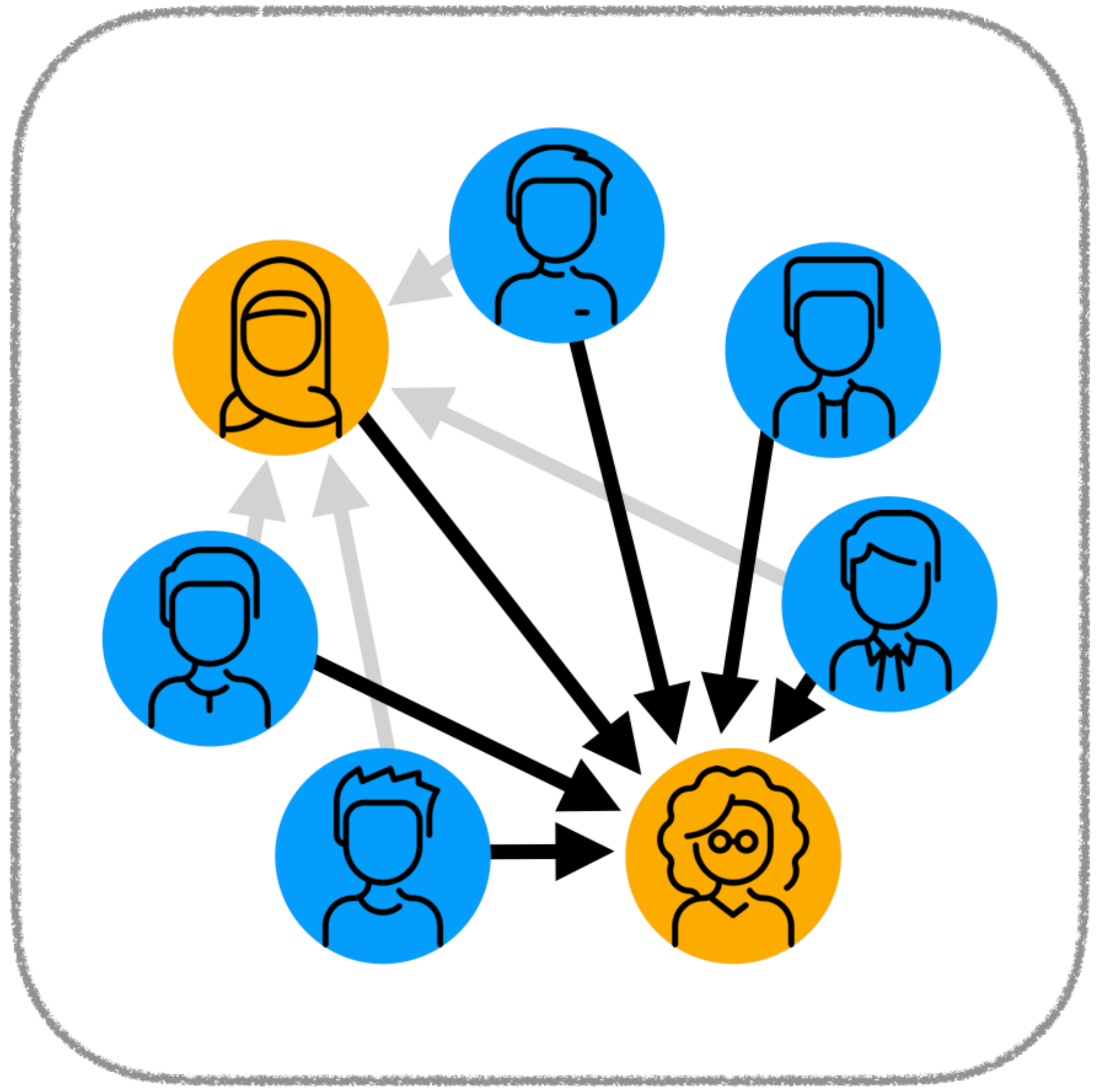


Homophily

We'll cover this today!

Popularity

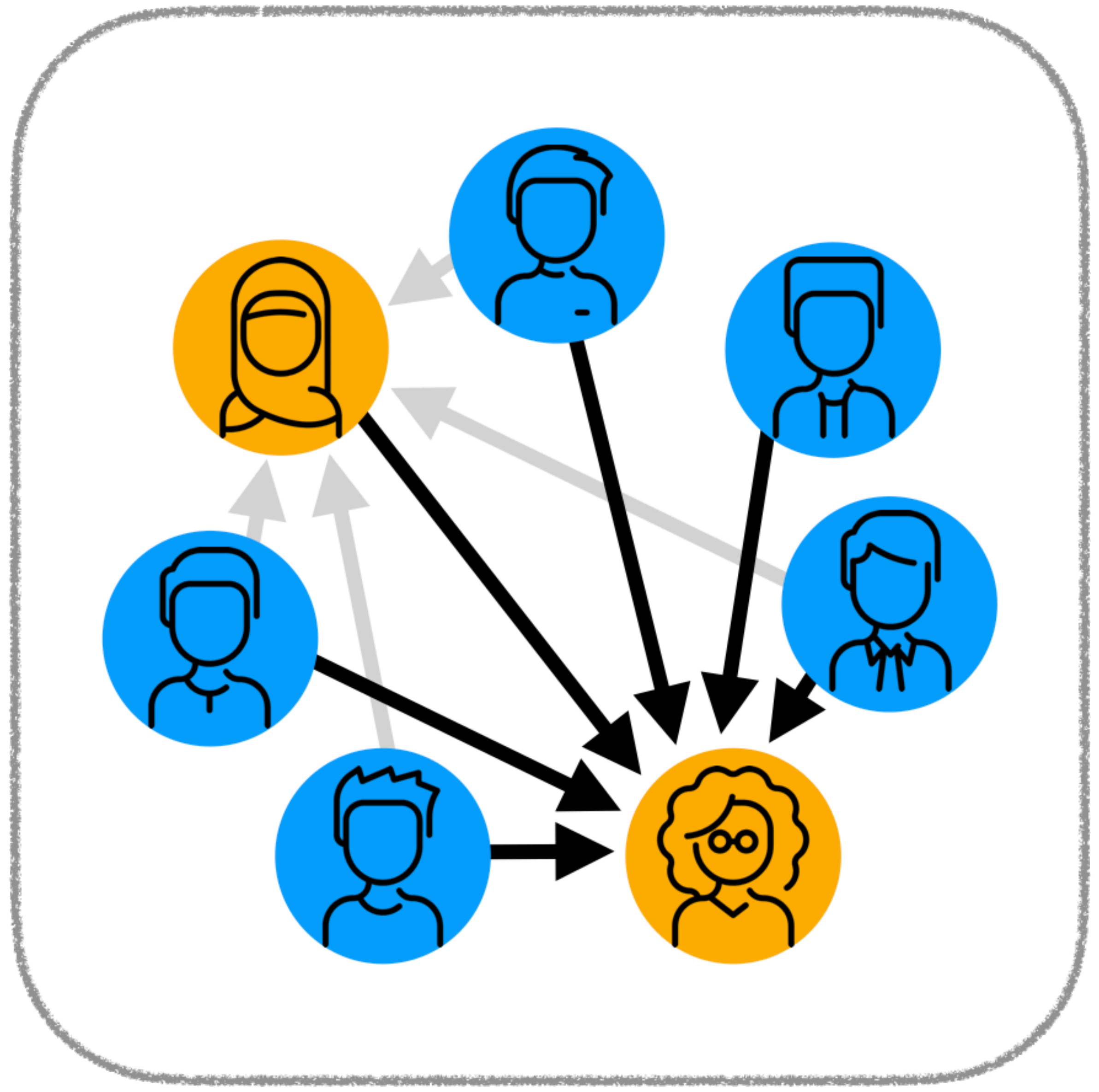
a.k.a. preferential attachment



Popularity

Popularity

a.k.a. preferential attachment

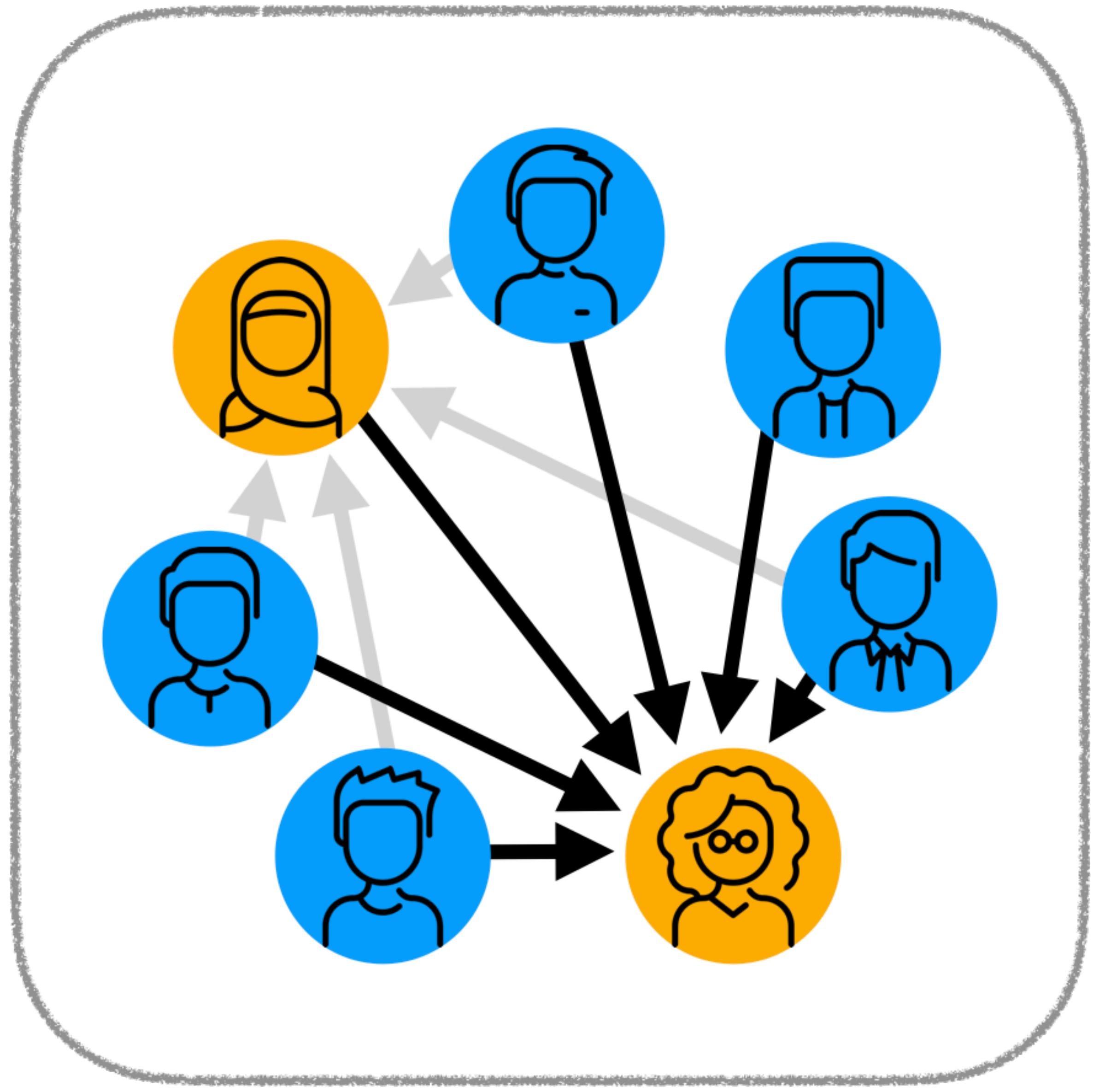


Popularity

- The tendency of actors **to connect to those who receive links from many others.** The structural position is defined by the **in-degree** of the receiver (target node) of an explained tie.

Popularity

a.k.a. preferential attachment

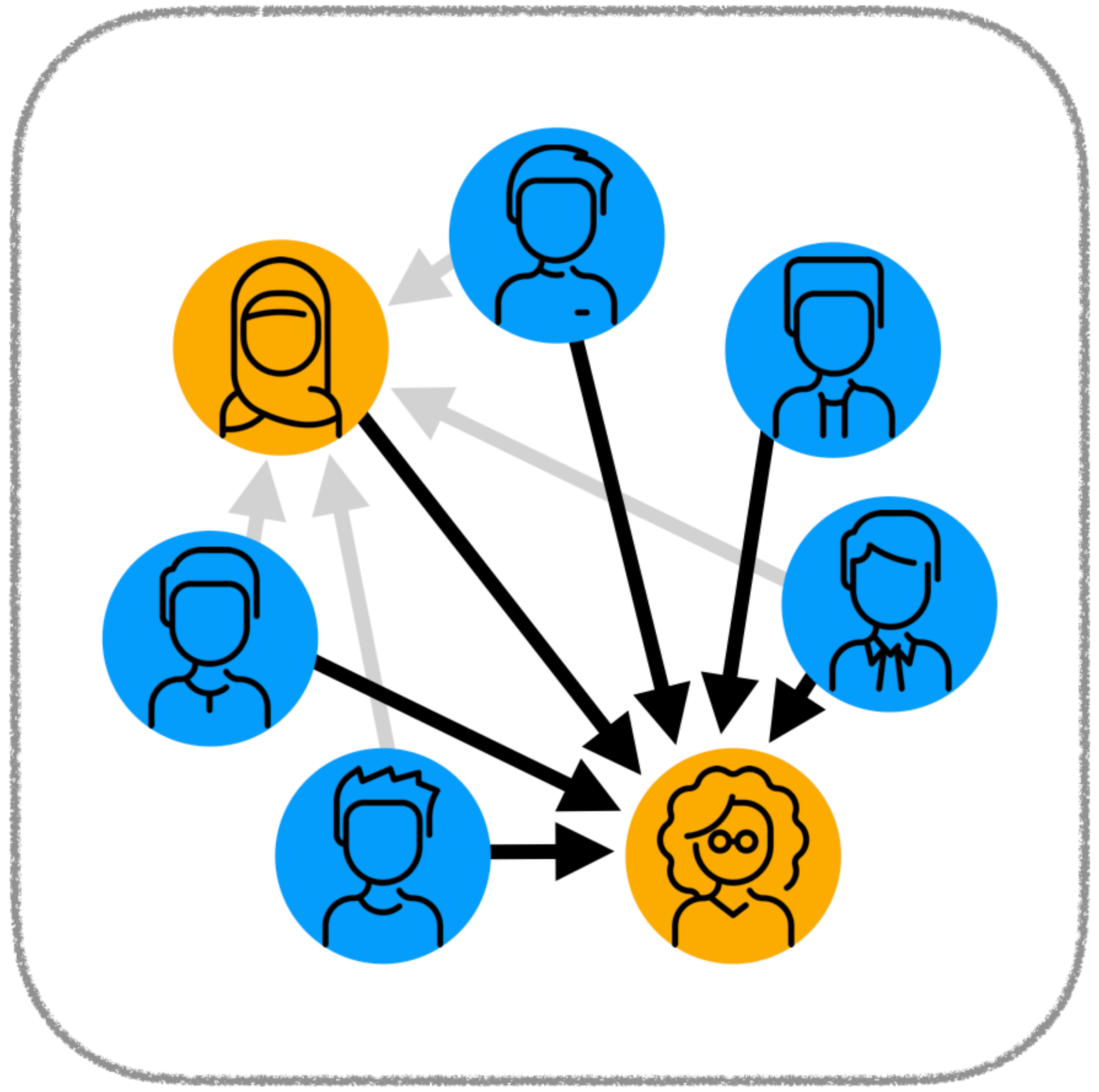


Popularity

- The tendency of actors **to connect to those who receive links from many others**. The structural position is defined by the **in-degree** of the receiver (target node) of an explained tie.
  - **Matthew effect or rich-gets-richer effect** mechanism in science discussed by Merton (1968), Price (1976) and others. Increasing recognition of an actor's scientific work (e.g. number of ties in a citation network).

Popularity

a.k.a. preferential attachment

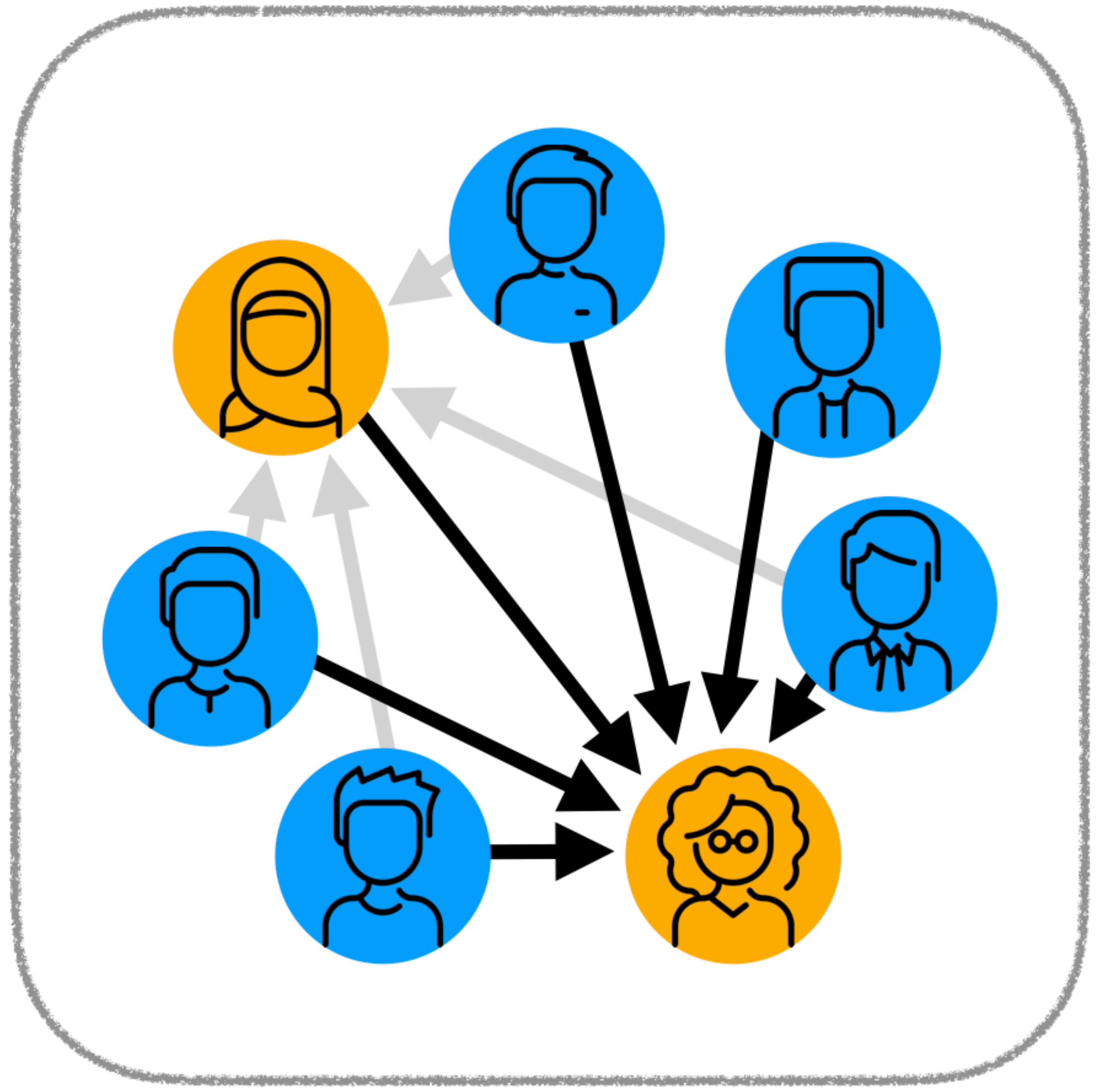


Popularity

- The tendency of actors **to connect to those who receive links from many others**. The structural position is defined by the **in-degree** of the receiver (target node) of an explained tie.
  - **Matthew effect or rich-gets-richer effect** mechanism in science discussed by Merton (1968), Price (1976) and others. Increasing recognition of an actor's scientific work (e.g. number of ties in a citation network).
  - **Preferential attachment** in the work of Barabási and Albert (1999) operationalised this mechanism in networks.

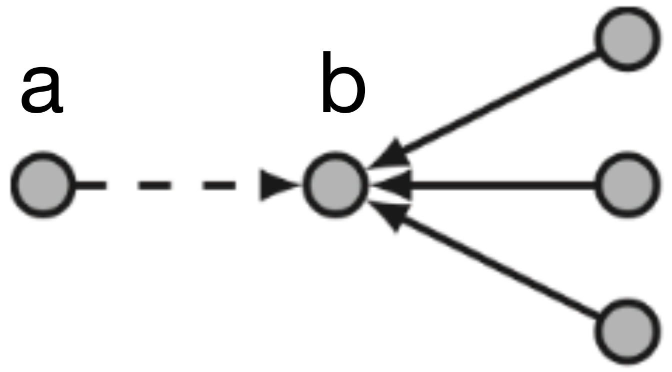
Popularity

a.k.a. preferential attachment



Popularity

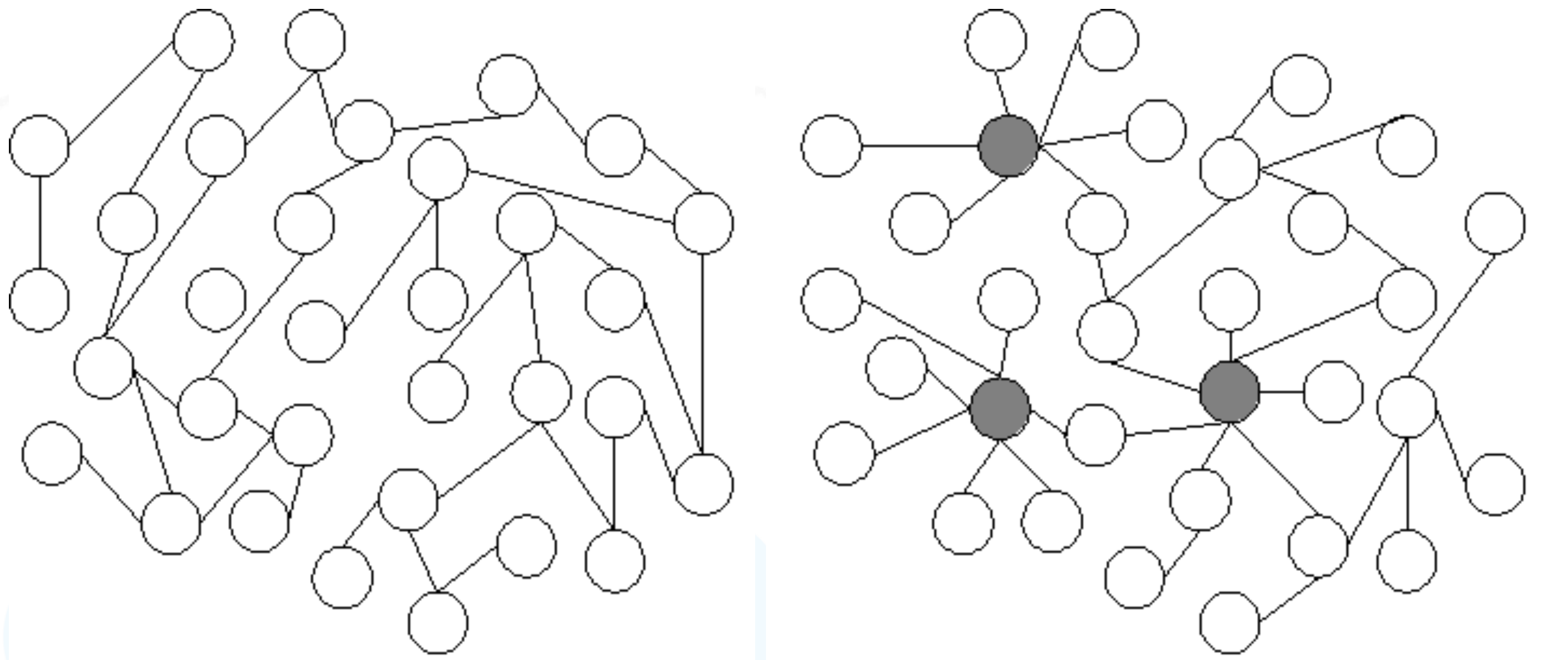
- The tendency of actors **to connect to those who receive links from many others.** The structural position is defined by the **in-degree** of the receiver (target node) of an explained tie.
  - **Matthew effect or rich-gets-richer effect** mechanism in science discussed by Merton (1968), Price (1976) and others. Increasing recognition of an actor's scientific work (e.g. number of ties in a citation network).
  - **Preferential attachment** in the work of Barabási and Albert (1999) operationalised this mechanism in networks.



$$P(a \to b) = P(b|a) = p_{ab} = \frac{k_b}{\sum_c k_c}$$

Popularity

a.k.a. preferential attachment



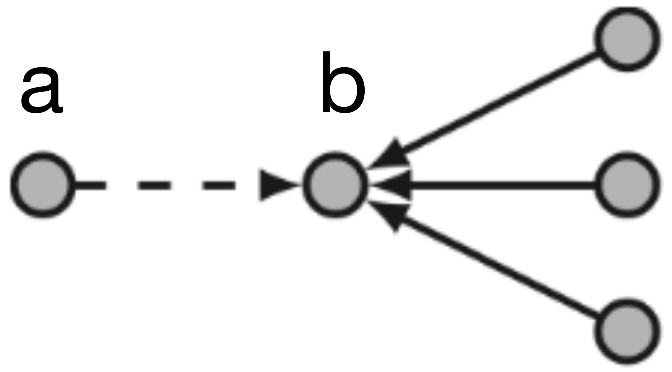
(a) Random network (b) Scale-free network

of actors **to connect to those who receive** …**ny others**. The structural position is defined by … of the receiver (target node) of an explained tie.

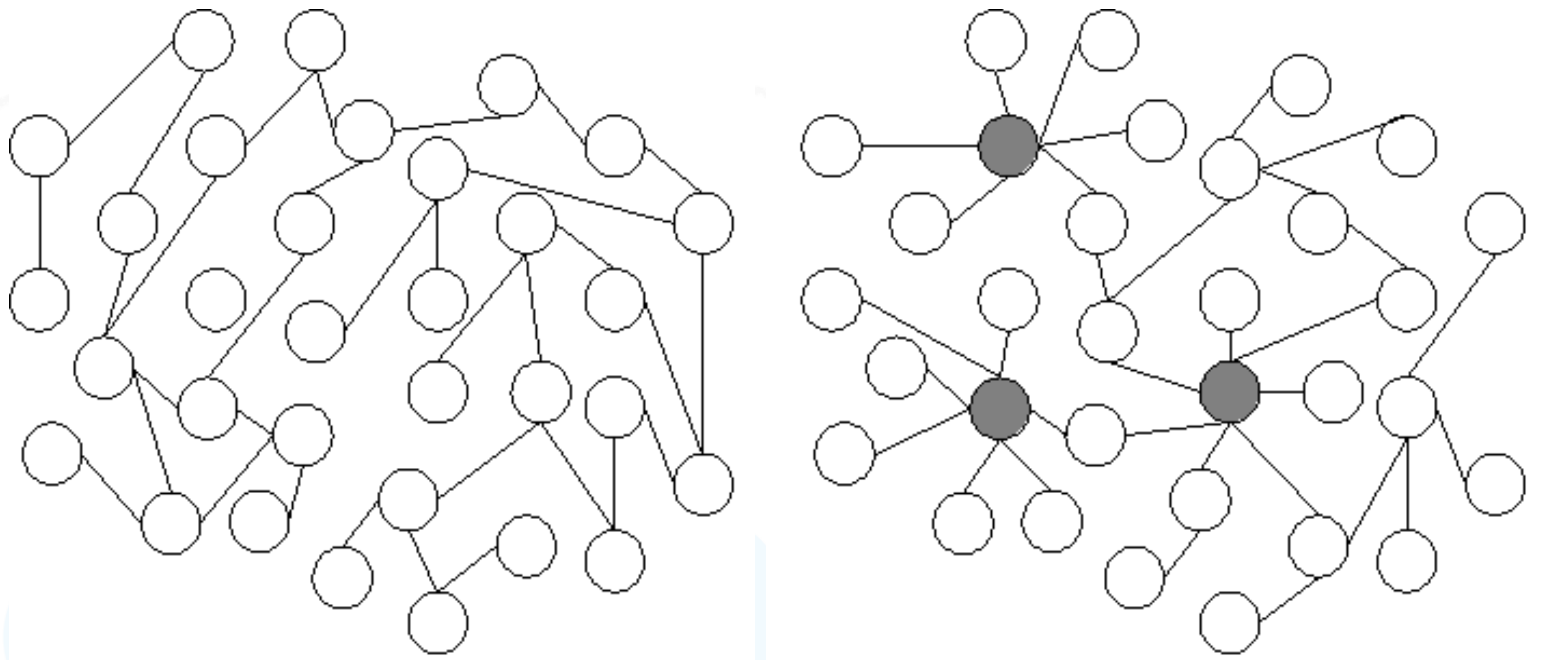
…**ffect or rich-gets-richer effect** mechanism in …cussed by Merton (1968), Price (1976) and …easing recognition of an actor's scientific work …r of ties in a citation network).

Popularity

- **Preferential attachment** in the work of Barabási and Albert (1999) operationalised this mechanism in networks.

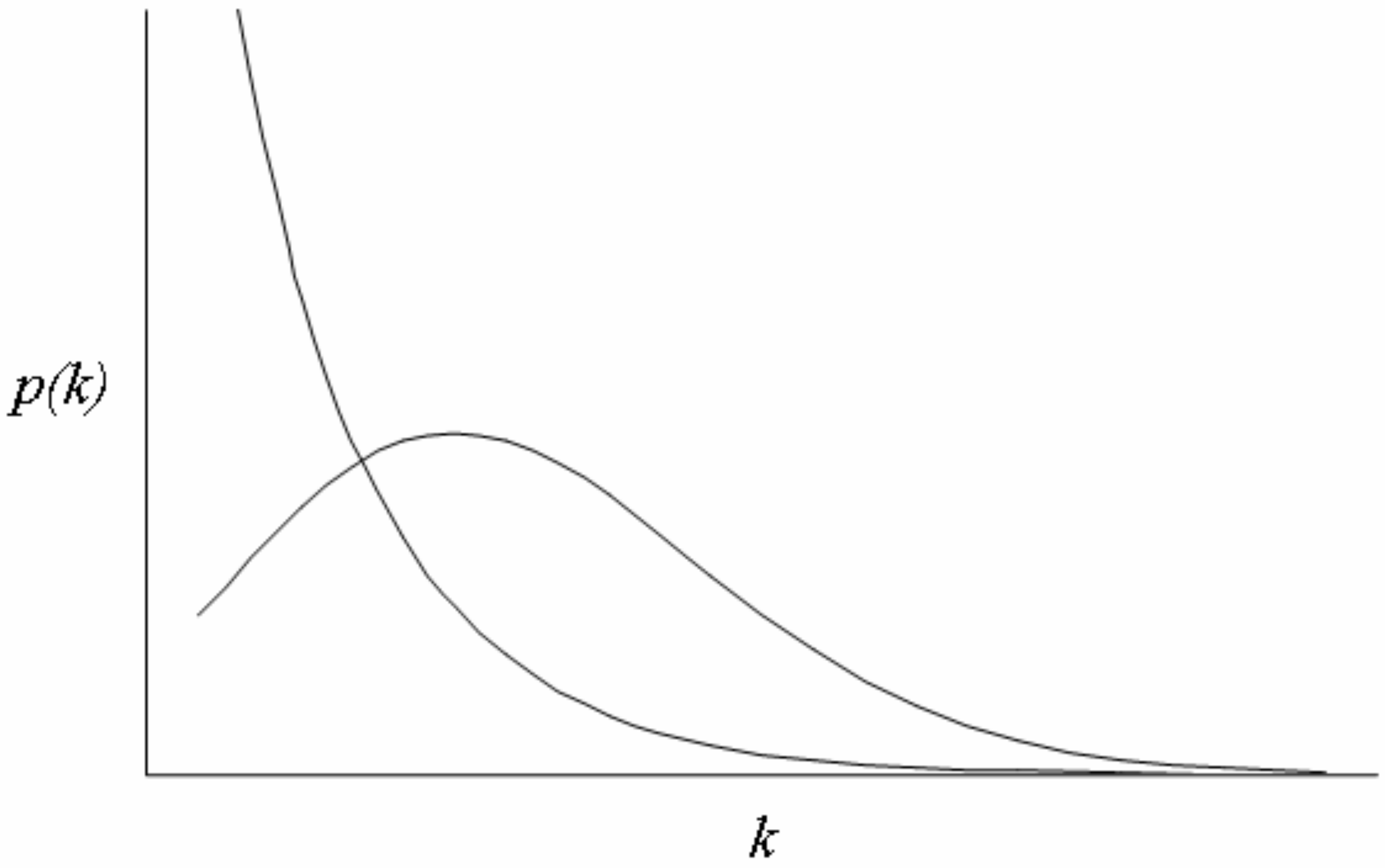


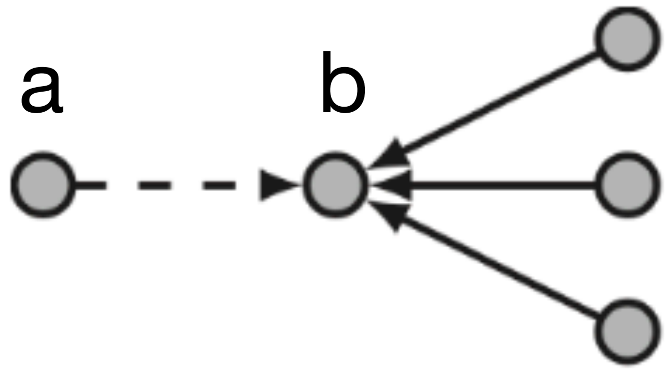
$$P(a \to b) = P(b|a) = p_{ab} = \frac{k_b}{\sum_c k_c}$$



(a) Random network (b) Scale-free network

Degree distribution plot

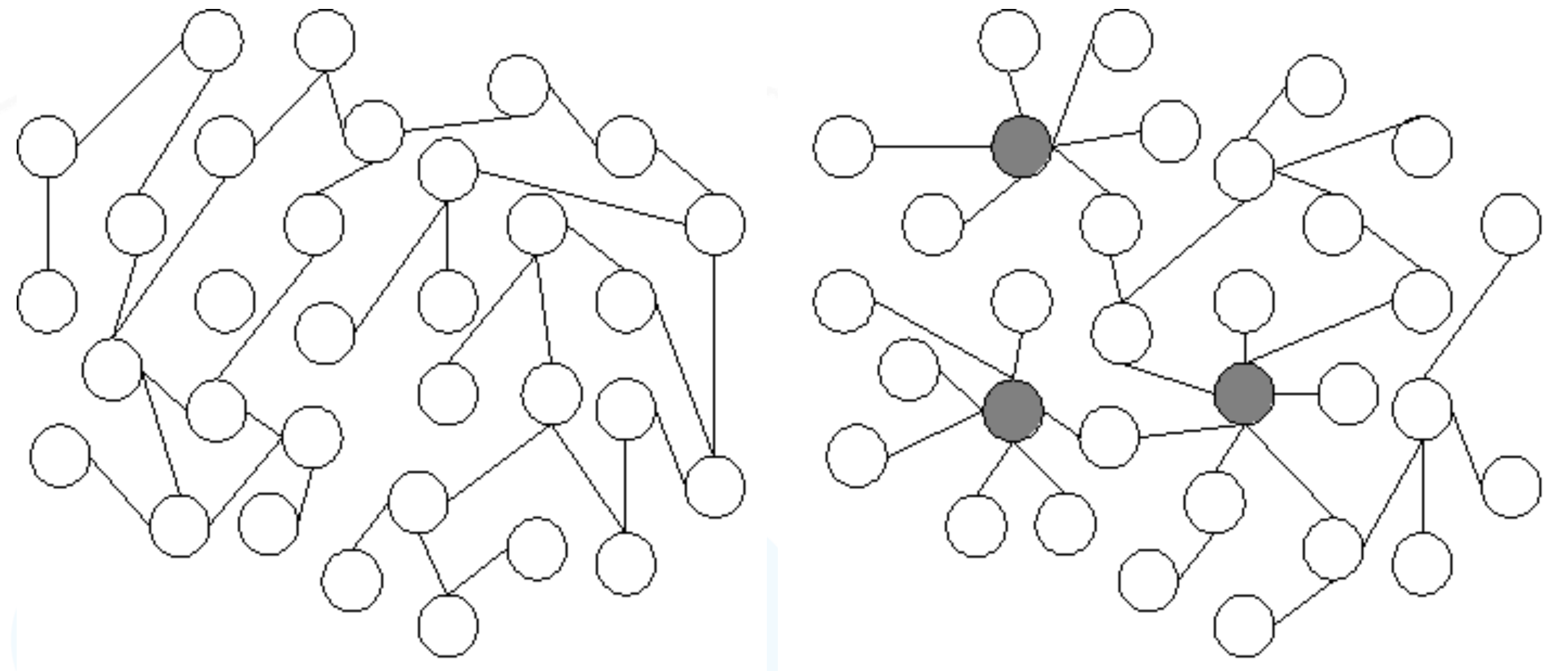




$$P(a \to b) = P(b|a) = p_{ab} = \frac{k_b}{\sum_c k_c}$$

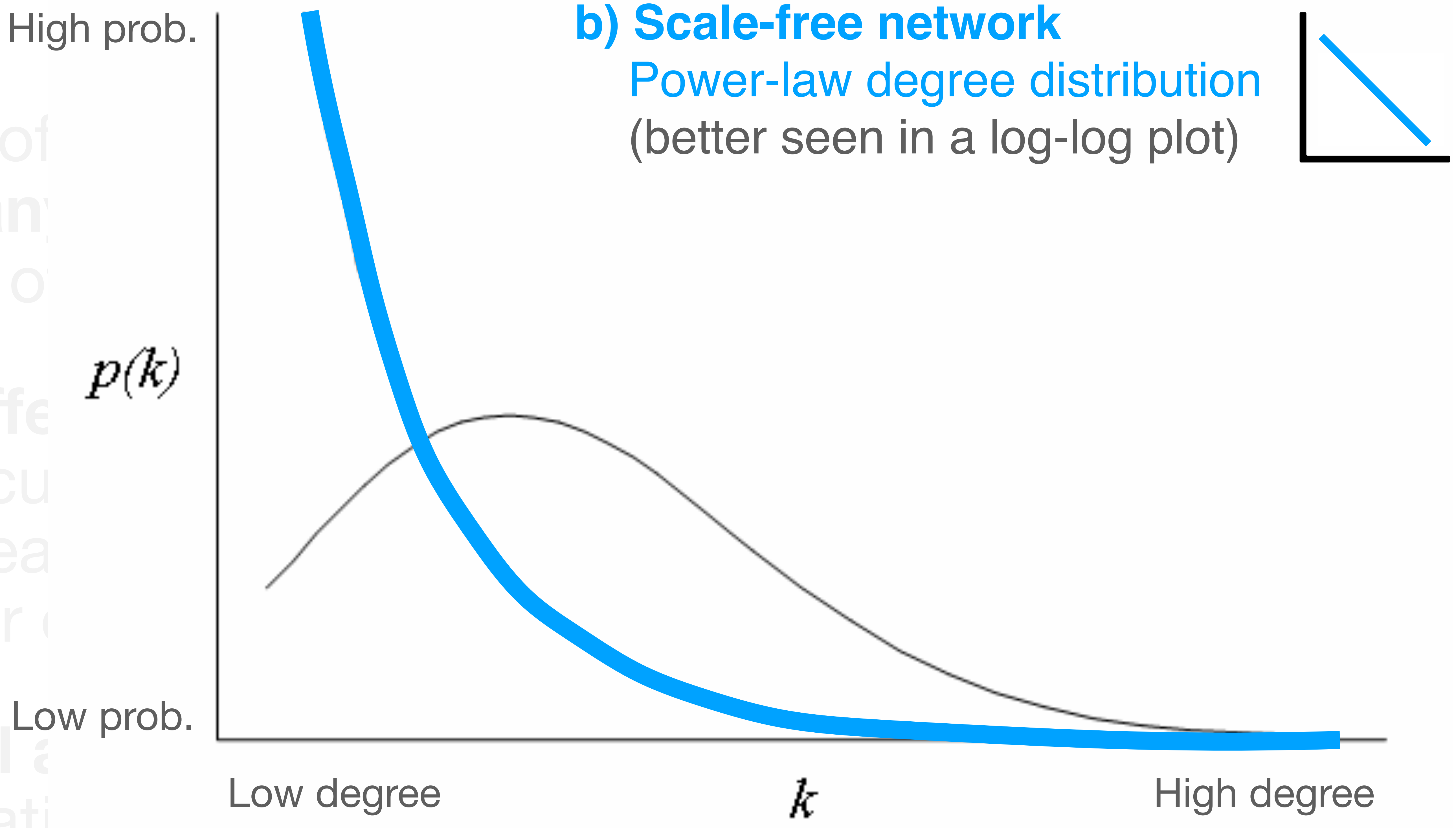
Popularity

a.k.a. preferential attachment

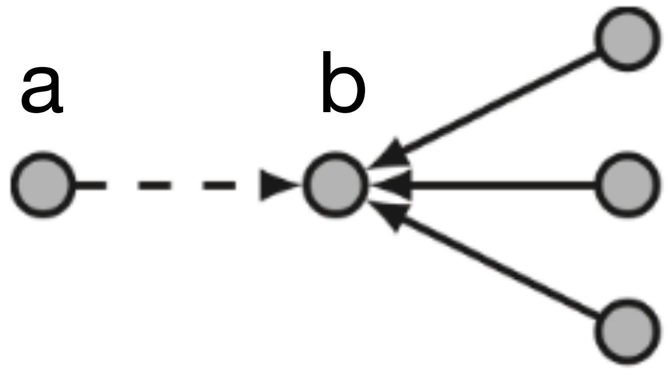


(a) Random network (b) Scale-free network

Degree distribution plot



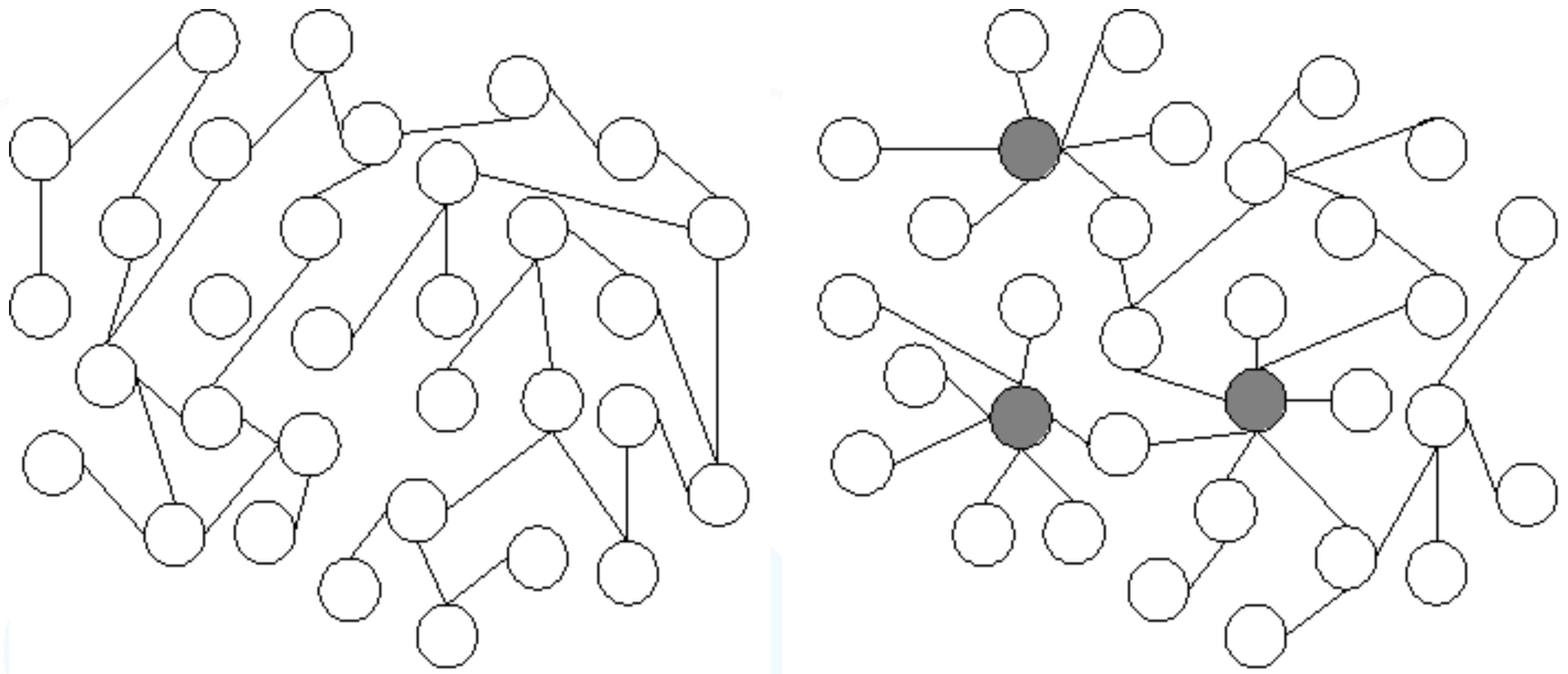
b) Scale-free network
Power-law degree distribution
(better seen in a log-log plot)



$$P(a \rightarrow b) = P(b|a) = p_{ab} = \frac{k_b}{\sum_c k_c}$$

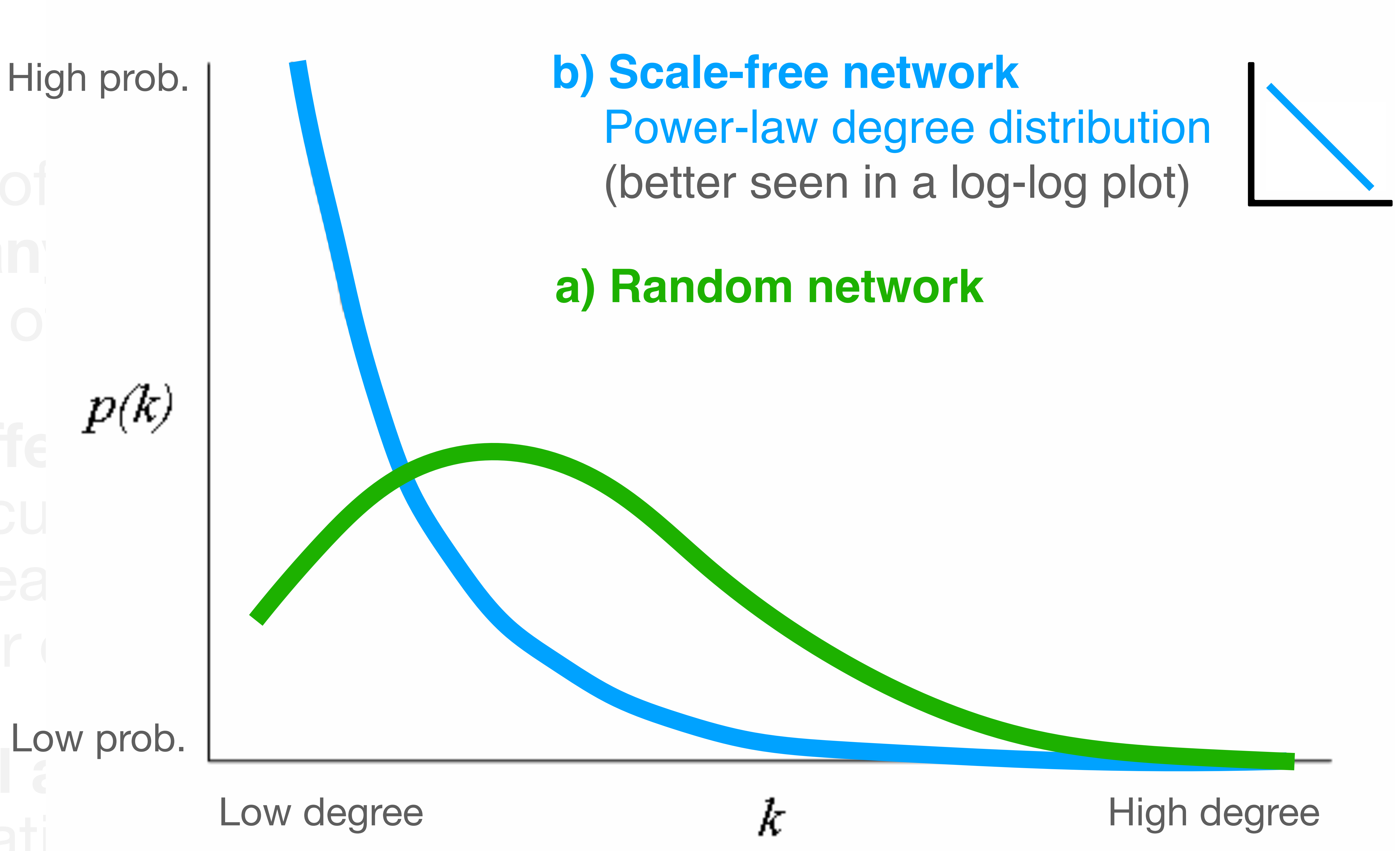
Popularity

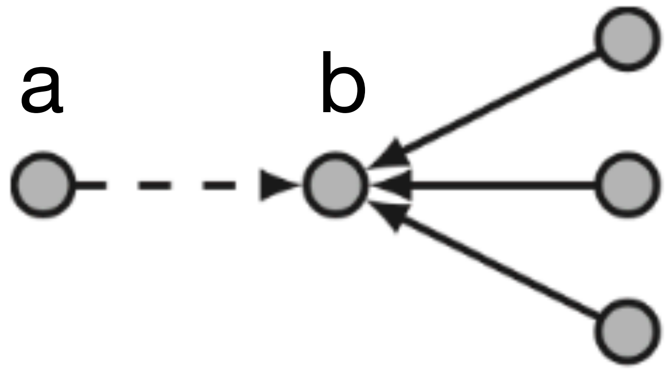
a.k.a. preferential attachment



(a) Random network (b) Scale-free network

Degree distribution plot





$$P(a \rightarrow b) = P(b|a) = p_{ab} = \frac{k_b}{\sum_c k_c}$$

Modeling “preferential attachment”

in node-attributed undirected networks

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.
“ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.
“ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.
“ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

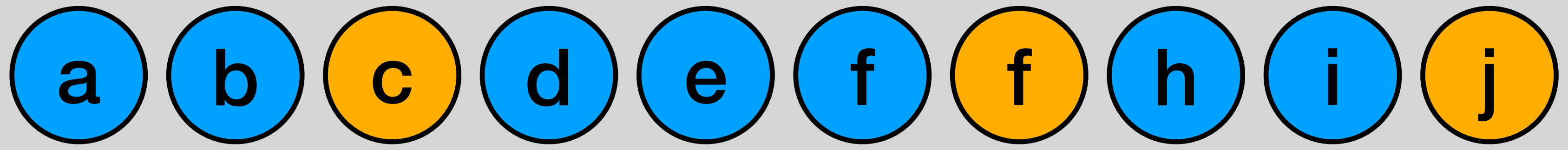
k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

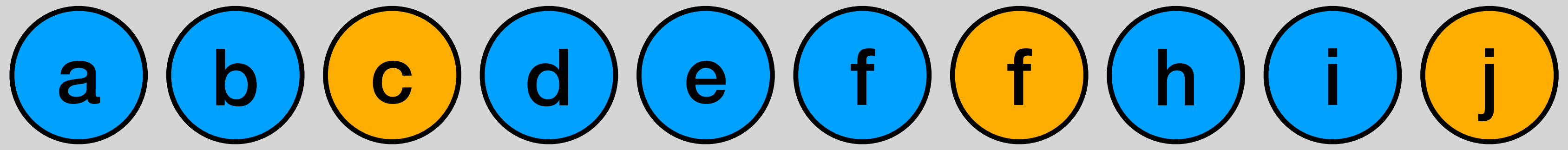
k: int

minimum degree of nodes (minimum=1)

f_m: float

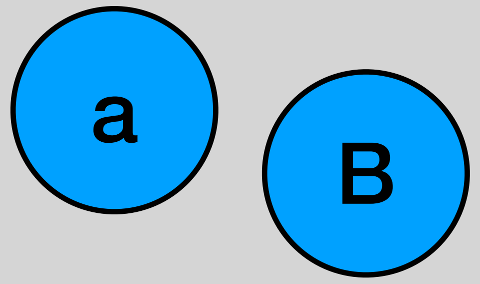
fraction of minorities (minimum=1/n, maximum=(n-1)/n)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$



STEP 2: Add k nodes

$k = 2$



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.
“ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

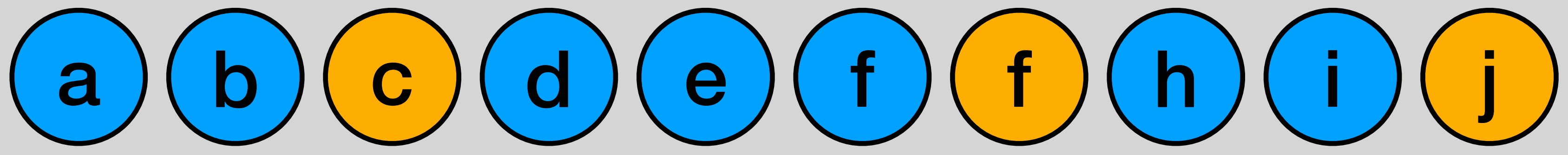
k: int

minimum degree of nodes (minimum=1)

f_m: float

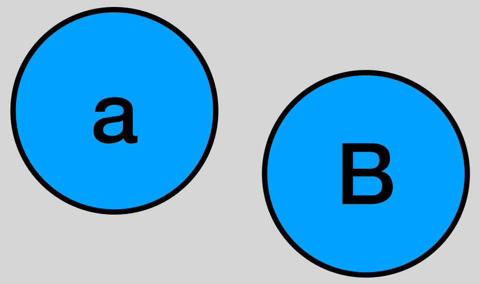
fraction of minorities (minimum=1/n, maximum=(n-1)/n)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

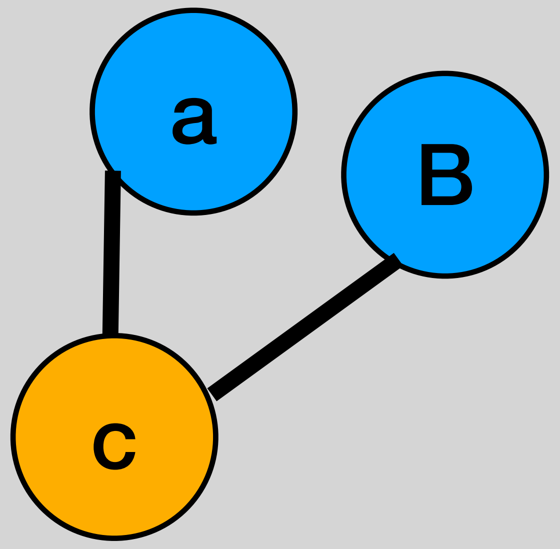


STEP 2: Add k nodes

$k = 2$



STEP 3: Add 1 new node and connect to k existing nodes.



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Modeling “preferential attachment”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

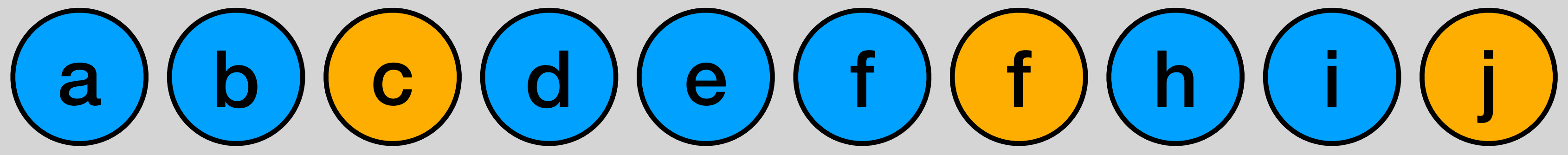
k: int

minimum degree of nodes (minimum=1)

f_m: float

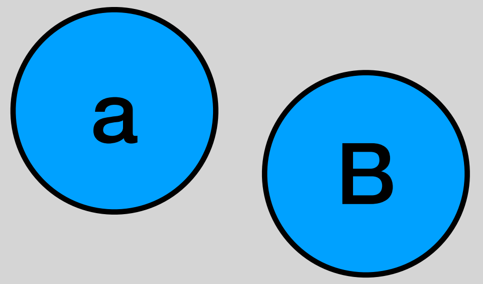
fraction of minorities (minimum=1/n, maximum=(n-1)/n)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

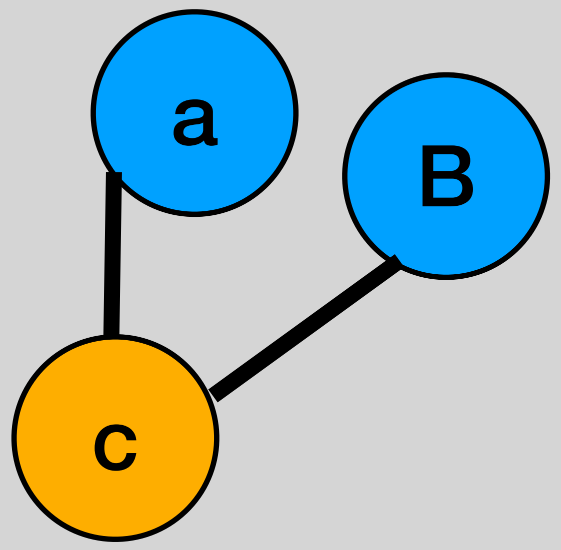


STEP 2: Add k nodes

$k = 2$



STEP 3: Add 1 new node and connect to k existing nodes.



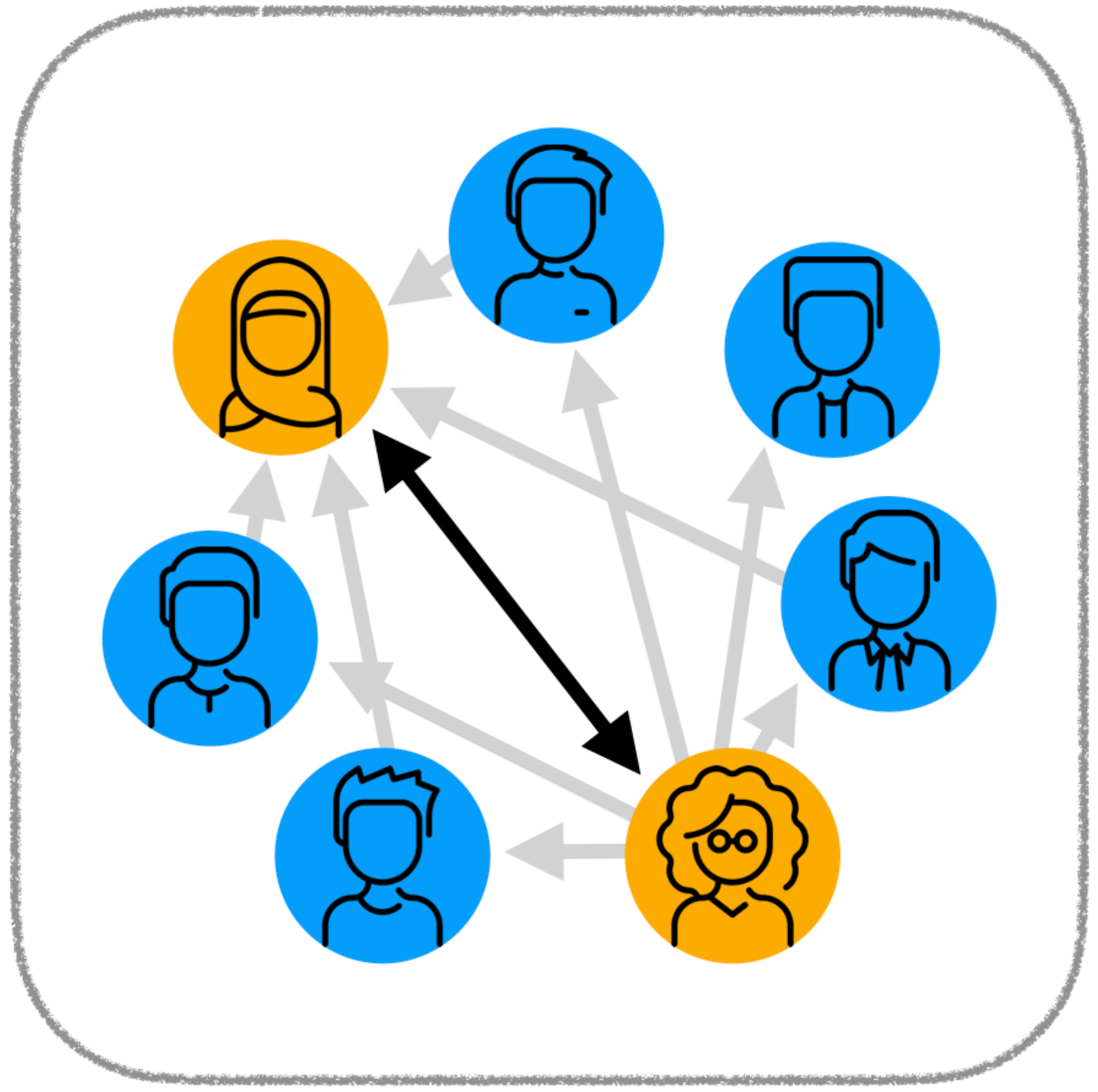
STEP 4: Repeat step 3 using $P(i \rightarrow j) \propto k_j$

Stop once n nodes have been added.

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Homophily

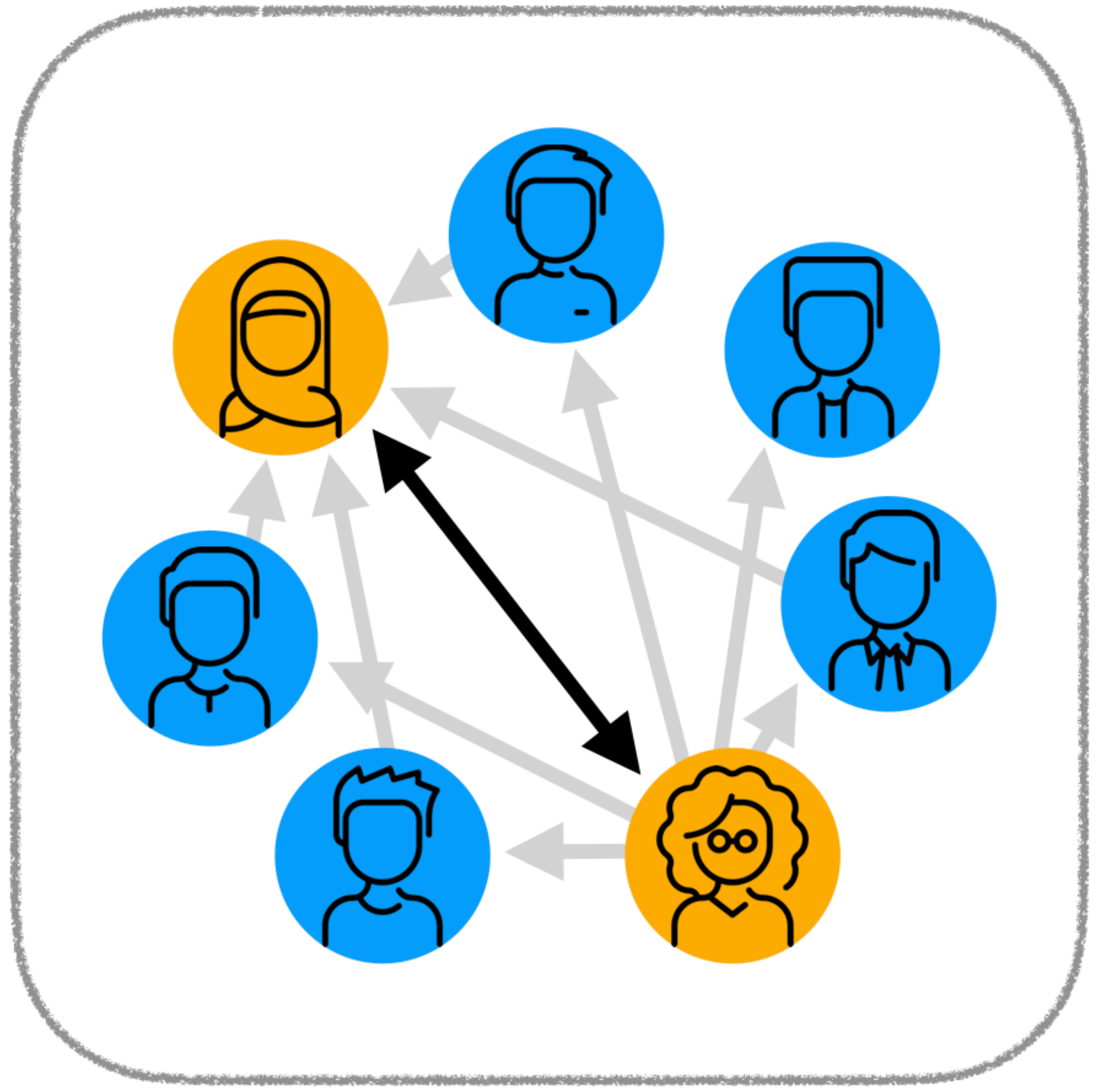
a.k.a. similarity



Homophily

Homophily

a.k.a. similarity

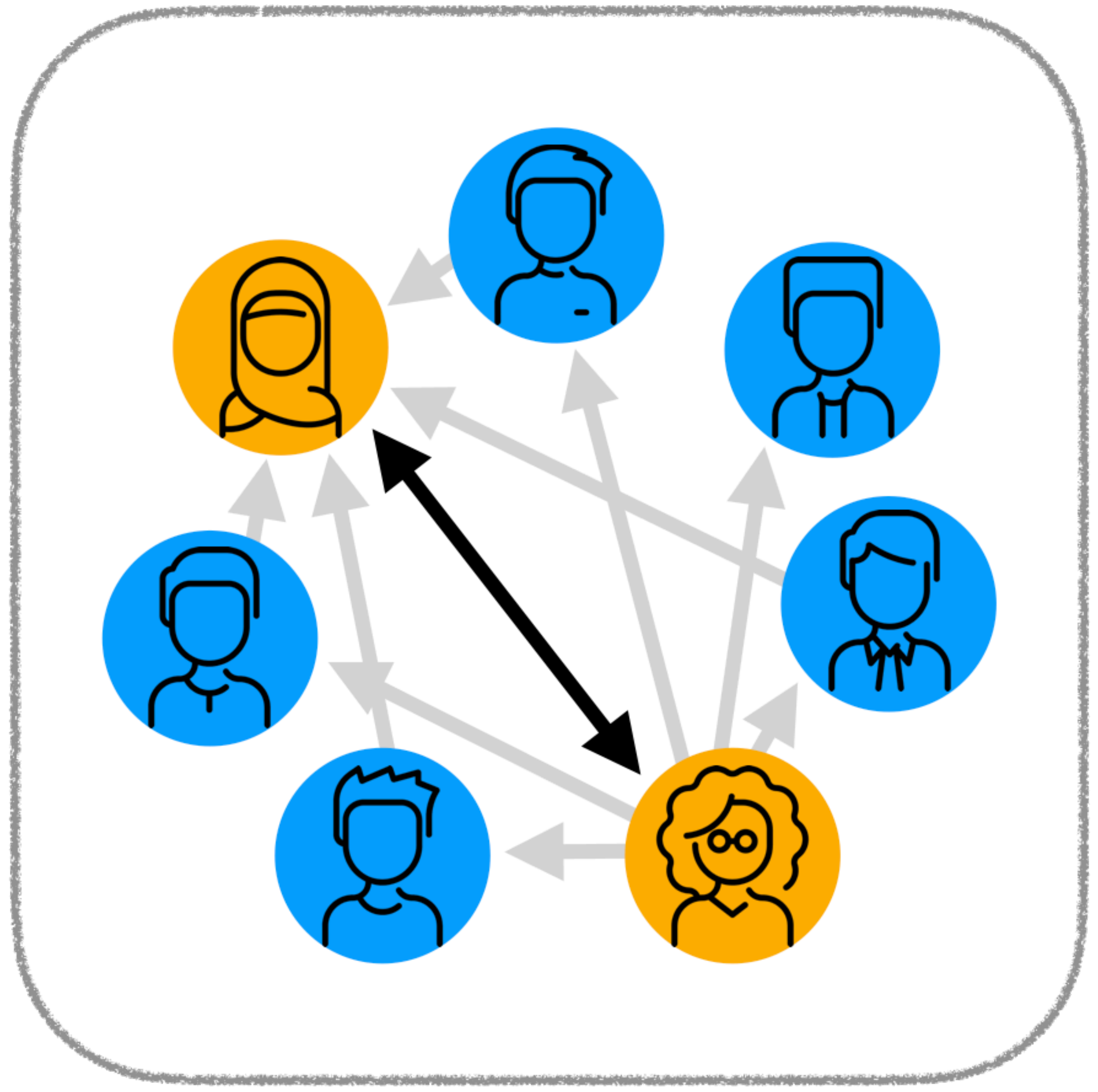


Homophily

- The tendency that **similar actors are more likely to connect** (Lazarsfeld, et al. 1954; McPherson et al. 2001). Its structural position is defined by the **attribute similarity** of the sender and receiver of the explained tie.

Homophily

a.k.a. similarity

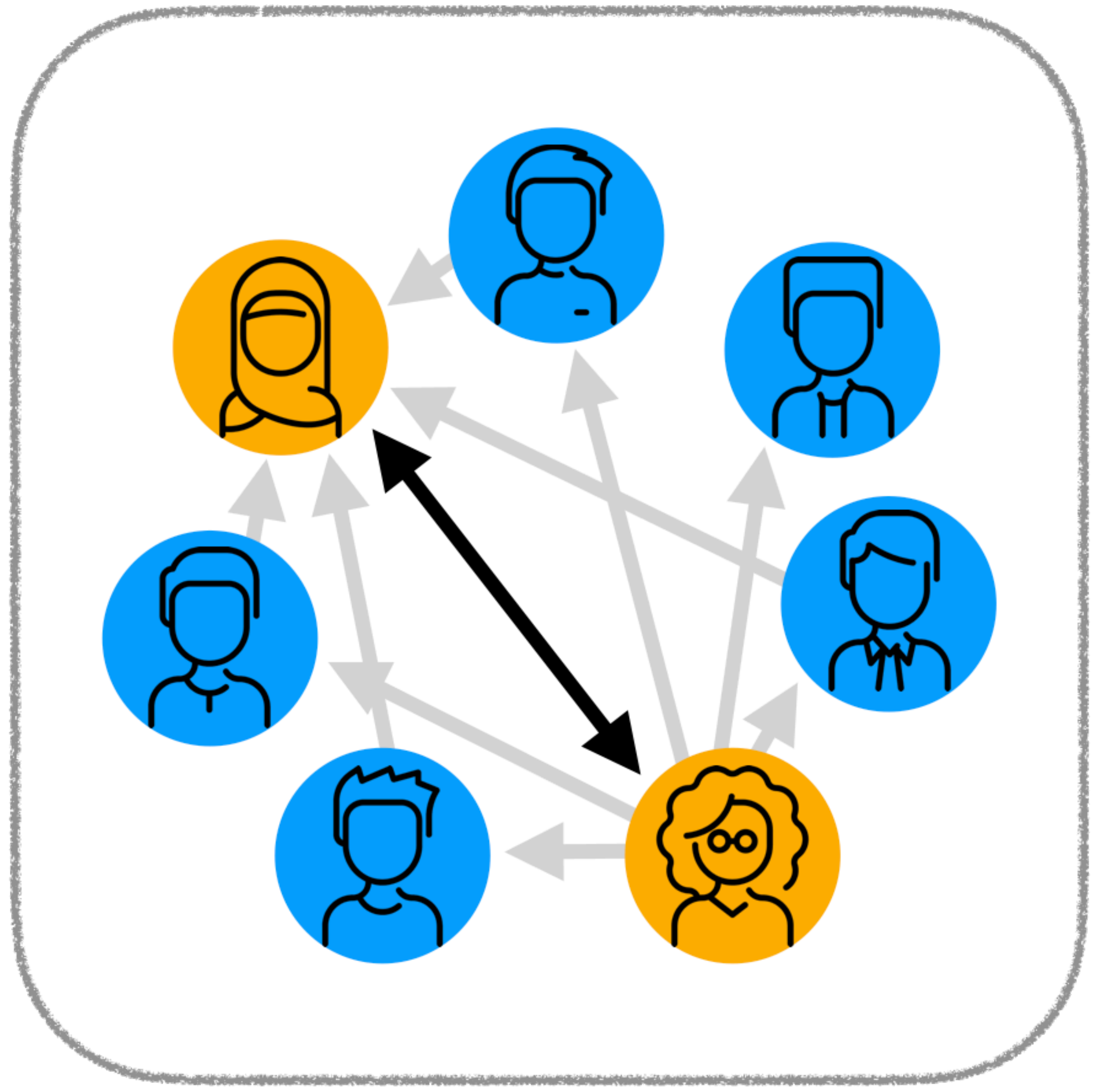


Homophily

- The tendency that **similar actors are more likely to connect** (Lazarsfeld, et al. 1954; McPherson et al. 2001). Its structural position is defined by the **attribute similarity** of the sender and receiver of the explained tie.
- Causal explanations:

Homophily

a.k.a. similarity

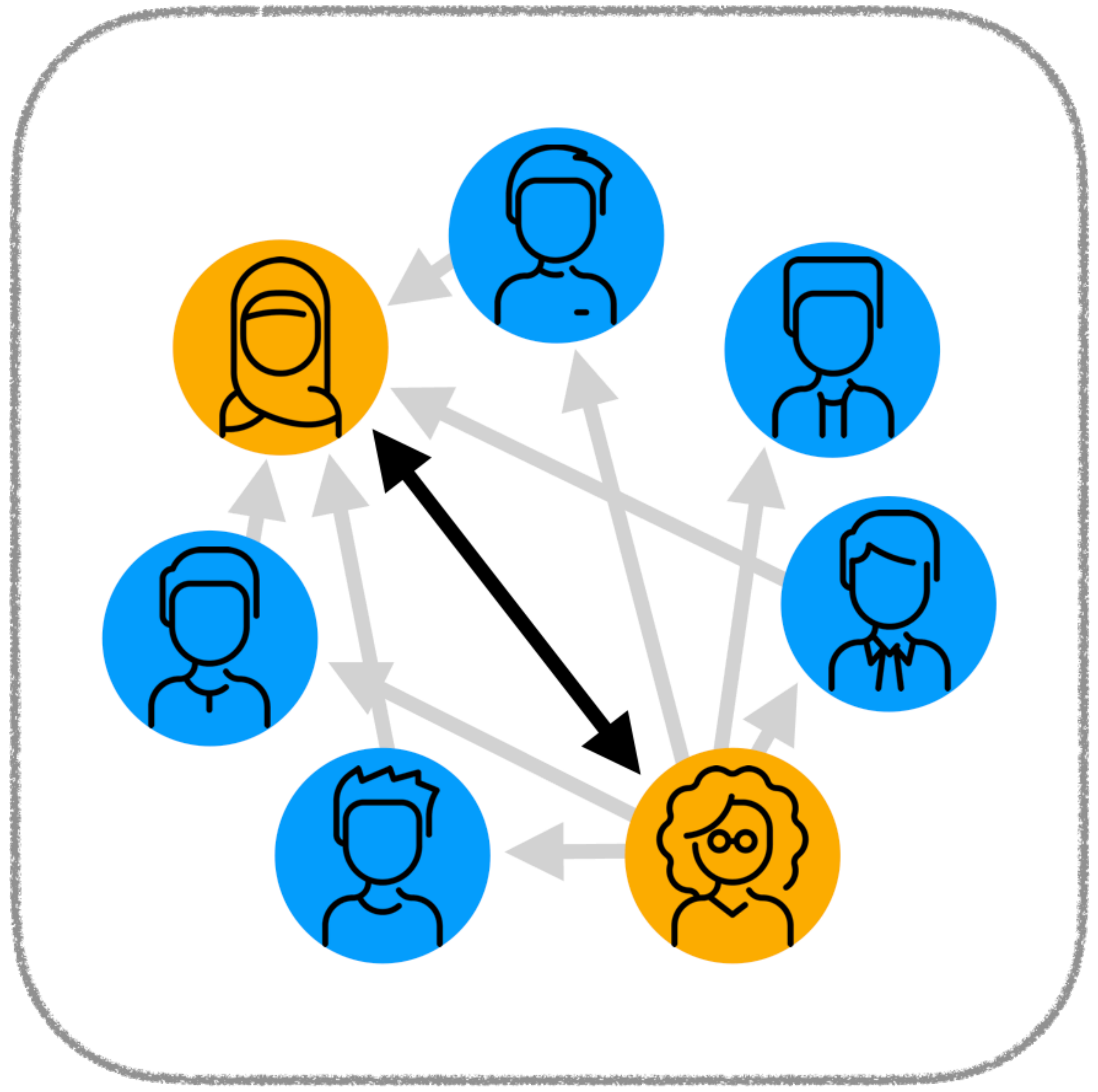


Homophily

- The tendency that **similar actors are more likely to connect** (Lazarsfeld, et al. 1954; McPherson et al. 2001). Its structural position is defined by the **attribute similarity** of the sender and receiver of the explained tie.
- Causal explanations:
  - These may be cognitive processes about similarity attraction (Huston and Levinger 1978)

Homophily

a.k.a. similarity

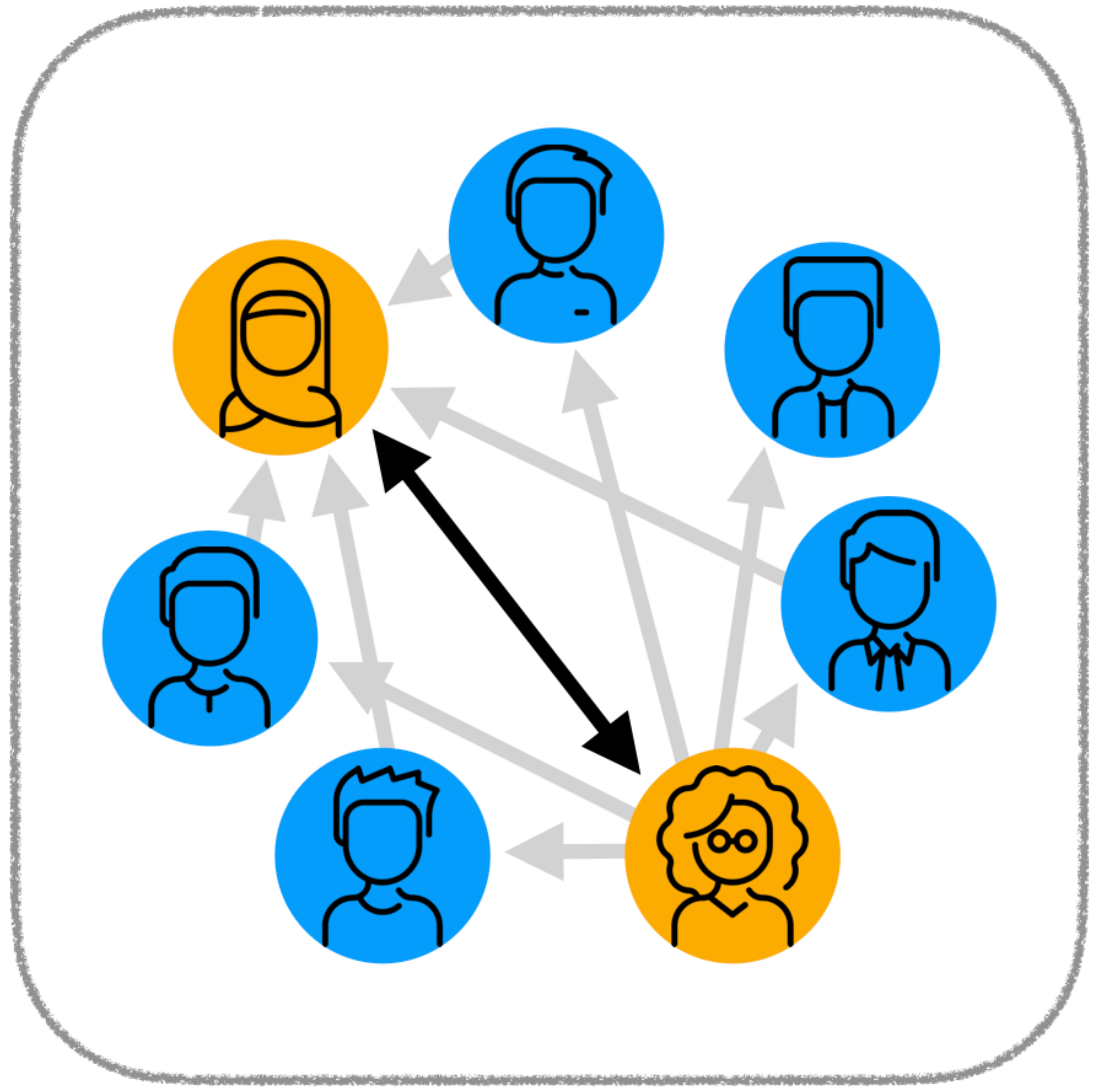


Homophily

- The tendency that **similar actors are more likely to connect** (Lazarsfeld, et al. 1954; McPherson et al. 2001). Its structural position is defined by the **attribute similarity** of the sender and receiver of the explained tie.
- Causal explanations:
    - These may be cognitive processes about similarity attraction (Huston and Levinger 1978)
    - Structural processes that are affected by existing baseline segregation and social distances in the social setting under study.

Homophily

a.k.a. similarity



Homophily

- The tendency that **similar actors are more likely to connect** (Lazarsfeld, et al. 1954; McPherson et al. 2001). Its structural position is defined by the **attribute similarity** of the sender and receiver of the explained tie.
- Causal explanations:
  - These may be cognitive processes about similarity attraction (Huston and Levinger 1978)
  - Structural processes that are affected by existing baseline segregation and social distances in the social setting under study.

$$P(a \to b) = P(b\,|\,a) = p_{ab} = similarity(a, b) = h_{a,b}$$

Modeling “homophily”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.
“ Scientific reports 8.1 (2018): 11077.

Modeling “homophily”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

homophily (similarity) between minority nodes (minimum=0, maximum=1.)

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Modeling “homophily”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

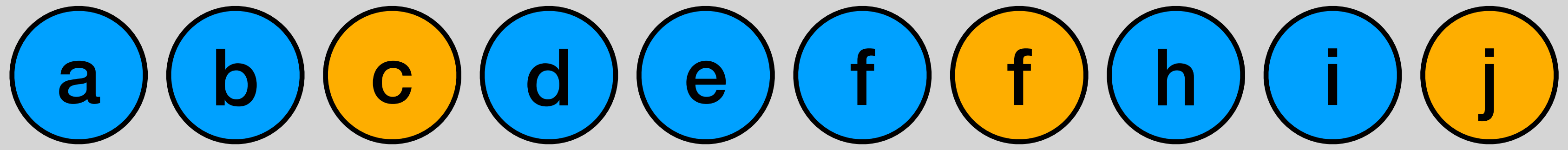
homophily (similarity) between minority nodes (minimum=0, maximum=1.)

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks.“ Scientific reports 8.1 (2018): 11077.

Modeling “homophily”

in node-attributed undirected networks

STEP 1: Initialize nodes $n = 10, f_m = 0.30$



netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

homophily (similarity) between minority nodes (minimum=0, maximum=1.)

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Modeling “homophily”

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

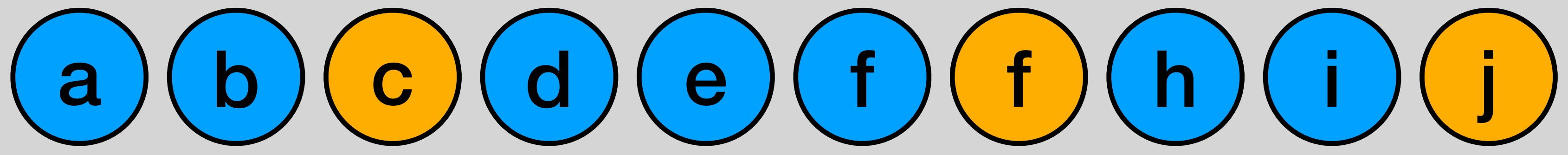
h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

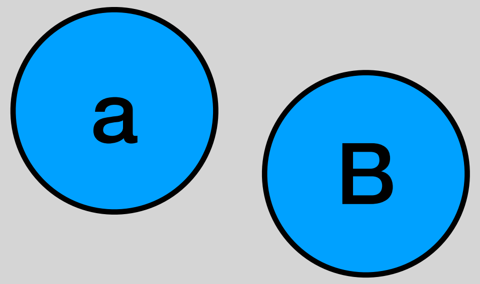
homophily (similarity) between minority nodes (minimum=0, maximum=1.)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$



STEP 2: Add k nodes

k=2



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. “ Scientific reports 8.1 (2018): 11077.

Modeling "homophily"

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

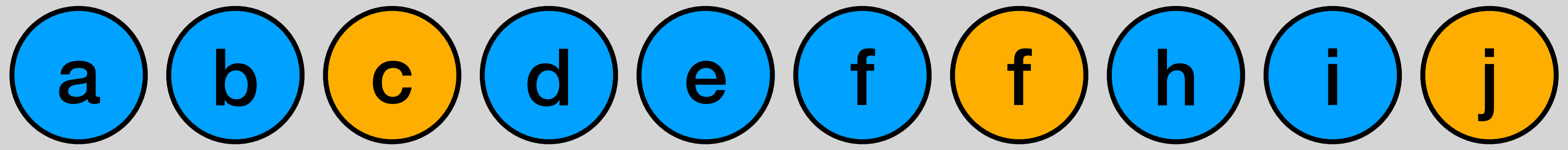
h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

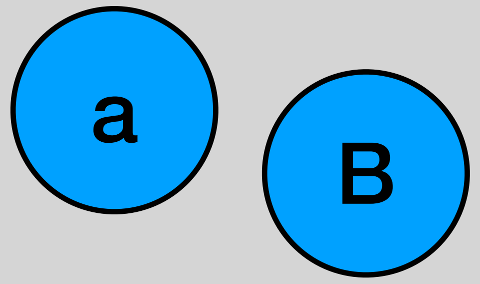
homophily (similarity) between minority nodes (minimum=0, maximum=1.)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

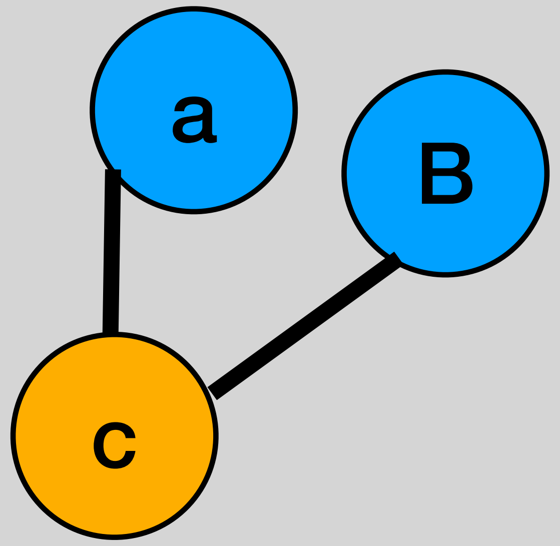


STEP 2: Add k nodes

k=2



STEP 3: Add 1 new node and connect to k existing nodes.



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. " Scientific reports 8.1 (2018): 11077.

Modeling "homophily"

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

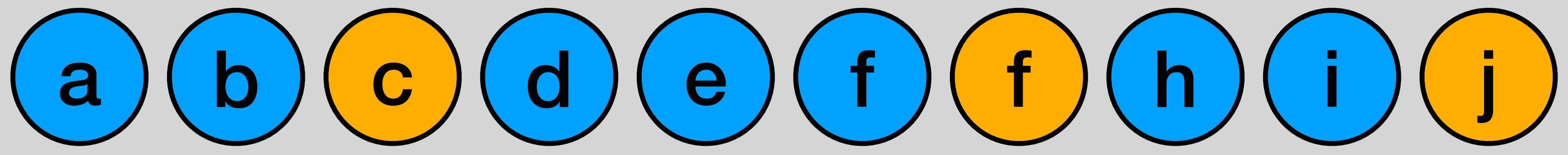
h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

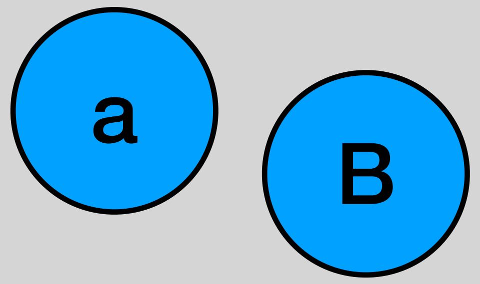
homophily (similarity) between minority nodes (minimum=0, maximum=1.)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

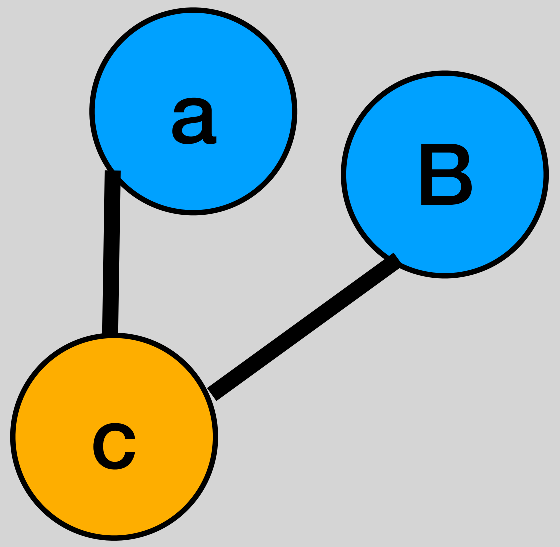


STEP 2: Add k nodes

k=2



STEP 3: Add 1 new node and connect to k existing nodes.



STEP 4: Repeat step 3 using $P(i \rightarrow j) \propto h_{ij}$

Stop once n nodes have been added.

Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. " Scientific reports 8.1 (2018): 11077.

Modeling "homophily & preferential attachment"

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int
  number of nodes (minimum=2)

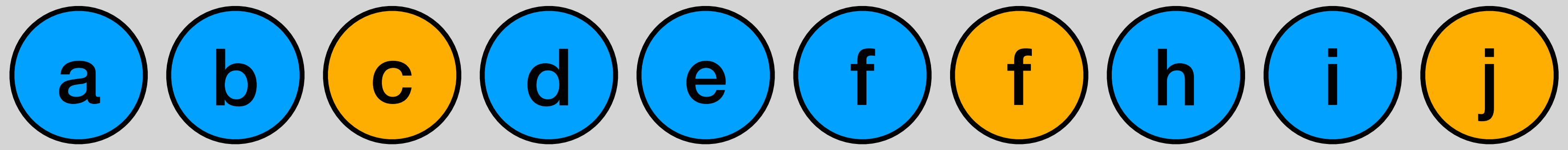
k: int
  minimum degree of nodes (minimum=1)

f_m: float
  fraction of minorities (minimum=1/n, maximum=(n-1)/n)

h_MM: float
  homophily (similarity) between majority nodes (minimum=0, maximum=1.)

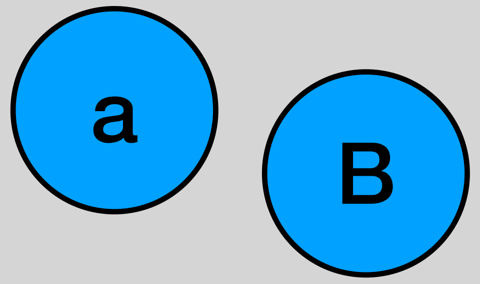
h_mm: float
  homophily (similarity) between minority nodes (minimum=0, maximum=1.)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

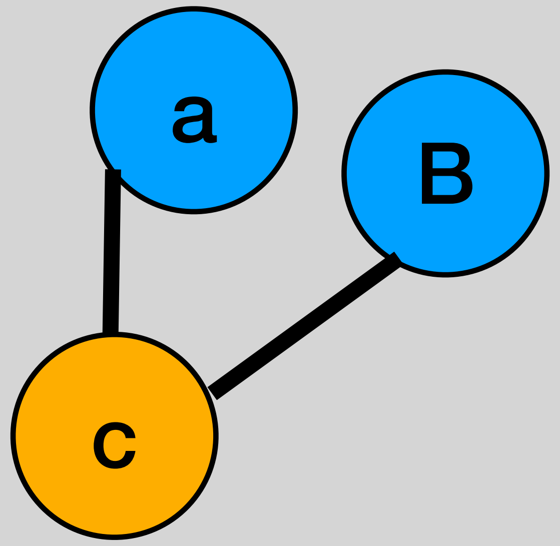


STEP 2: Add k nodes

k=2



STEP 3: Add 1 new node and connect to k existing nodes.



Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. " Scientific reports 8.1 (2018): 11077.

Modeling "homophily & preferential attachment"

in node-attributed undirected networks

netin 1.0.7

```
pip install netin
```

Parameters

n: int

number of nodes (minimum=2)

k: int

minimum degree of nodes (minimum=1)

f_m: float

fraction of minorities (minimum=1/n, maximum=(n-1)/n)

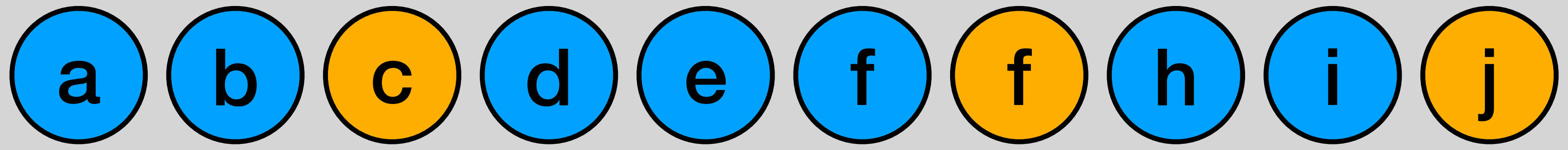
h_MM: float

homophily (similarity) between majority nodes (minimum=0, maximum=1.)

h_mm: float

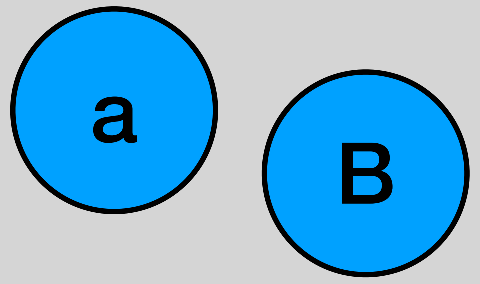
homophily (similarity) between minority nodes (minimum=0, maximum=1.)

STEP 1: Initialize nodes $n = 10, f_m = 0.30$

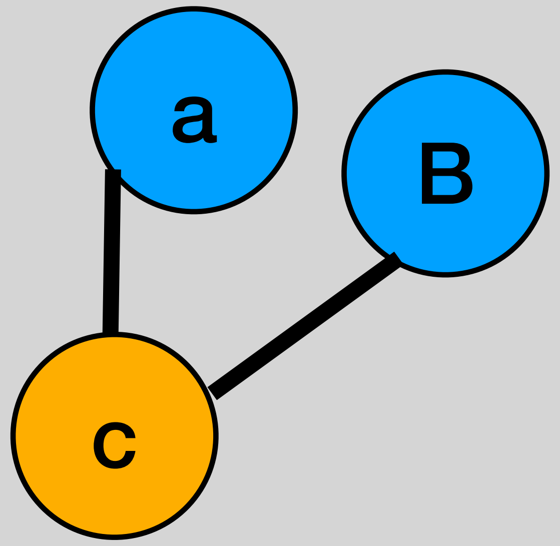


STEP 2: Add k nodes

k=2



STEP 3: Add 1 new node and connect to k existing nodes.



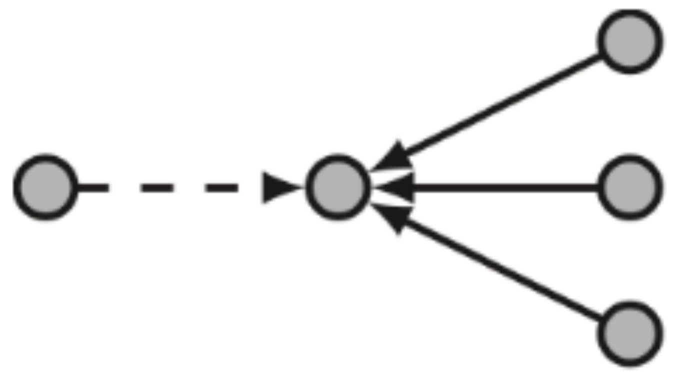
STEP 4: Repeat step 3 using $P(i \rightarrow j) \propto k_j h_{ij}$

Stop once n nodes have been added.

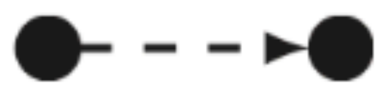
Karimi, Fariba, et al. "Homophily influences ranking of minorities in social networks. " Scientific reports 8.1 (2018): 11077.

How can these **mechanisms of edge formation** help us understand social phenomena?

Popularity



Homophily



Understanding influence maximization

Using synthetic networks

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Using synthetic networks

Goal: Understand the correlation between network structure and information access (spreading) equality
Who gets more access to information? majority or minority? Both equally?

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models."
Available at SSRN 3880680 (2021).

Understanding influence maximization

Using synthetic networks

Goal: Understand the correlation between network structure and information access (spreading) equality
Who gets more access to information? majority or minority? Both equally?

Tasks:

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Using synthetic networks

Goal: Understand the correlation between network structure and information access (spreading) equality
Who gets more access to information? majority or minority? Both equally?

Tasks:

1. Generate synthetic networks using multiple models
    1. Random (baseline)
    2. PA: Preferential attachment
    3. H: Homophily
    4. PA+H: Preferential attachment + Homophily

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Using synthetic networks

Goal: Understand the correlation between network structure and information access (spreading) equality
Who gets more access to information? majority or minority? Both equally?

Tasks:

1. Generate synthetic networks using multiple models
    1. Random (baseline)
    2. PA: Preferential attachment
    3. H: Homophily
    4. PA+H: Preferential attachment + Homophily
2. Compare their degree equality

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Using synthetic networks

Goal: Understand the correlation between network structure and information access (spreading) equality
Who gets more access to information? majority or minority? Both equally?

Tasks:

1. Generate synthetic networks using multiple models
    1. Random (baseline)
    2. PA: Preferential attachment
    3. H: Homophily
    4. PA+H: Preferential attachment + Homophily
2. Compare their degree equality
3. Run a contagion simulation and report the "spreading equality" for each network
    1. Simple contagion: a spreading process induced by a single exposure to a contagious entity.
    2. Complex contagion: demands multiple exposures for transmission.

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

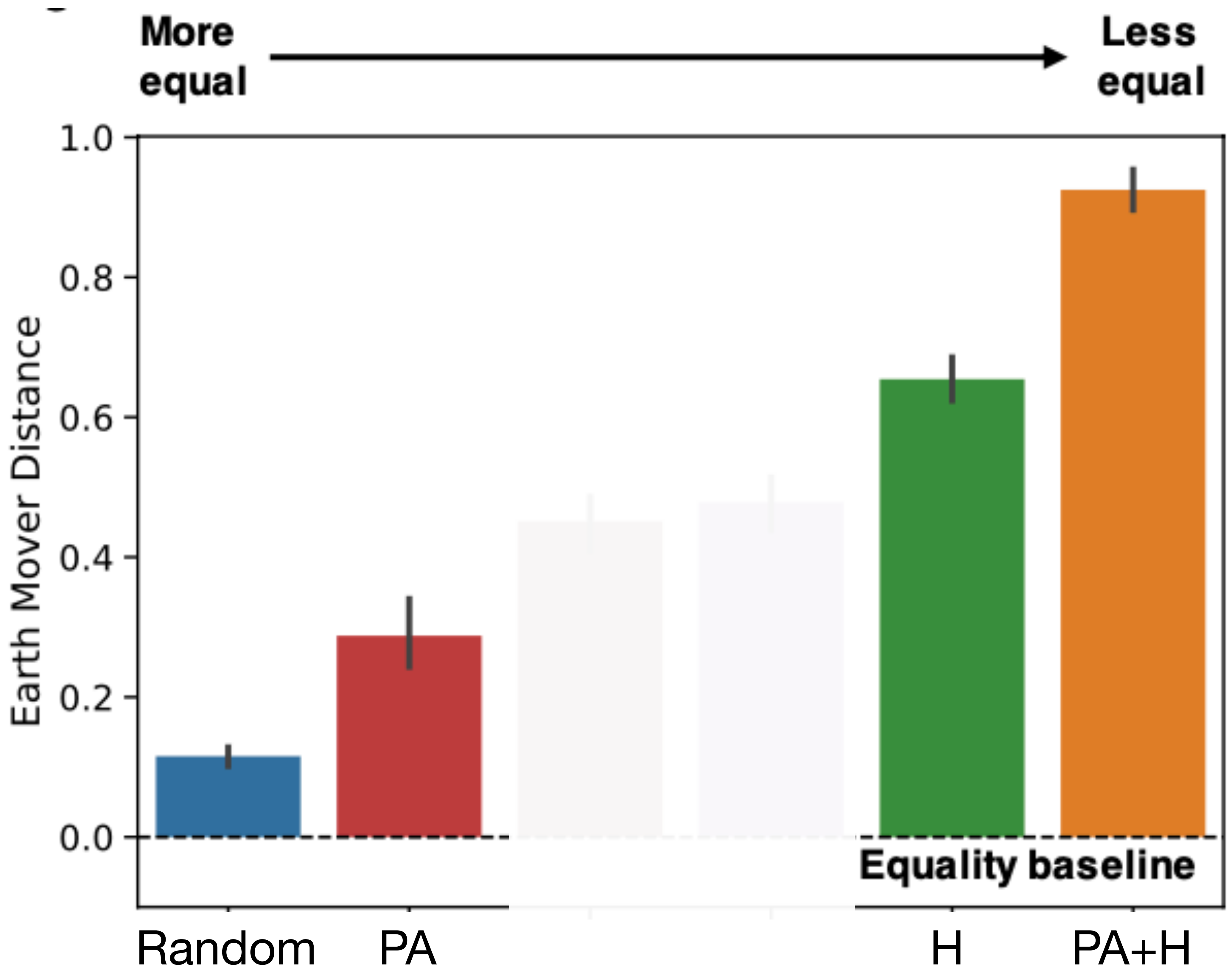
Understanding influence maximization

Task 2. Degree equality (social capital)

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

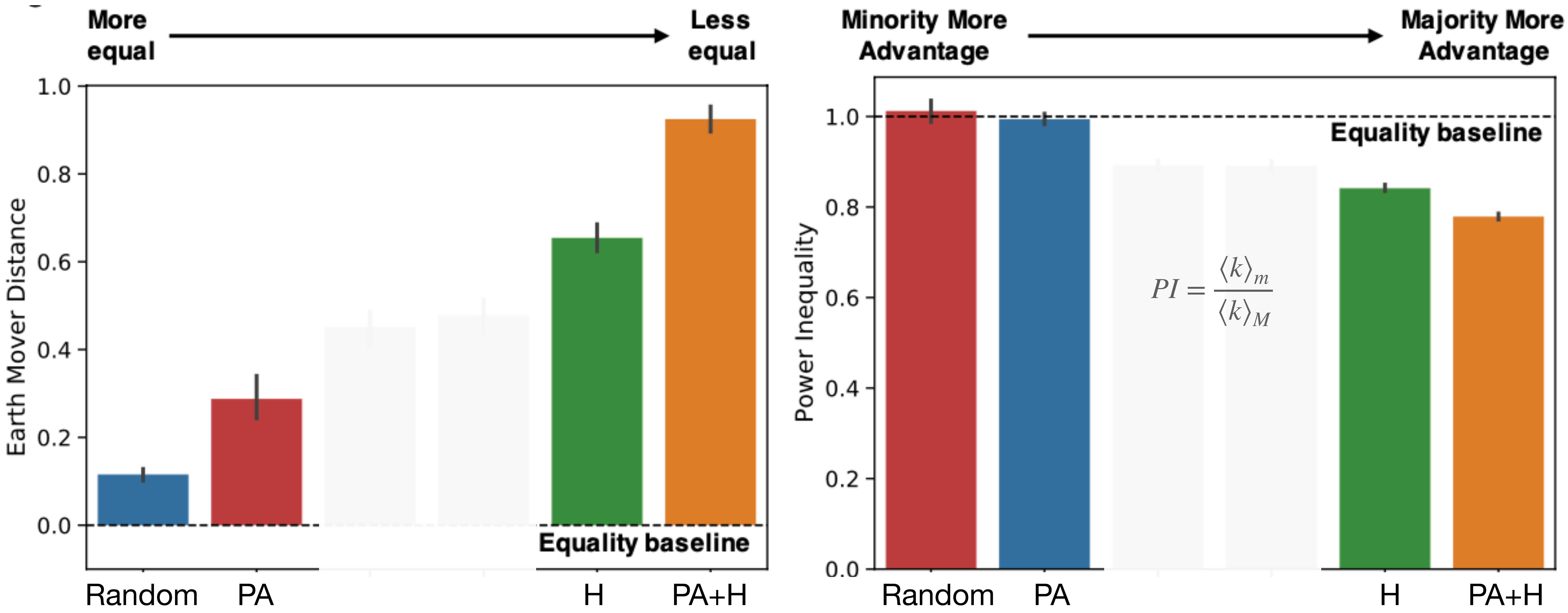
Task 2. Degree equality (social capital)



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

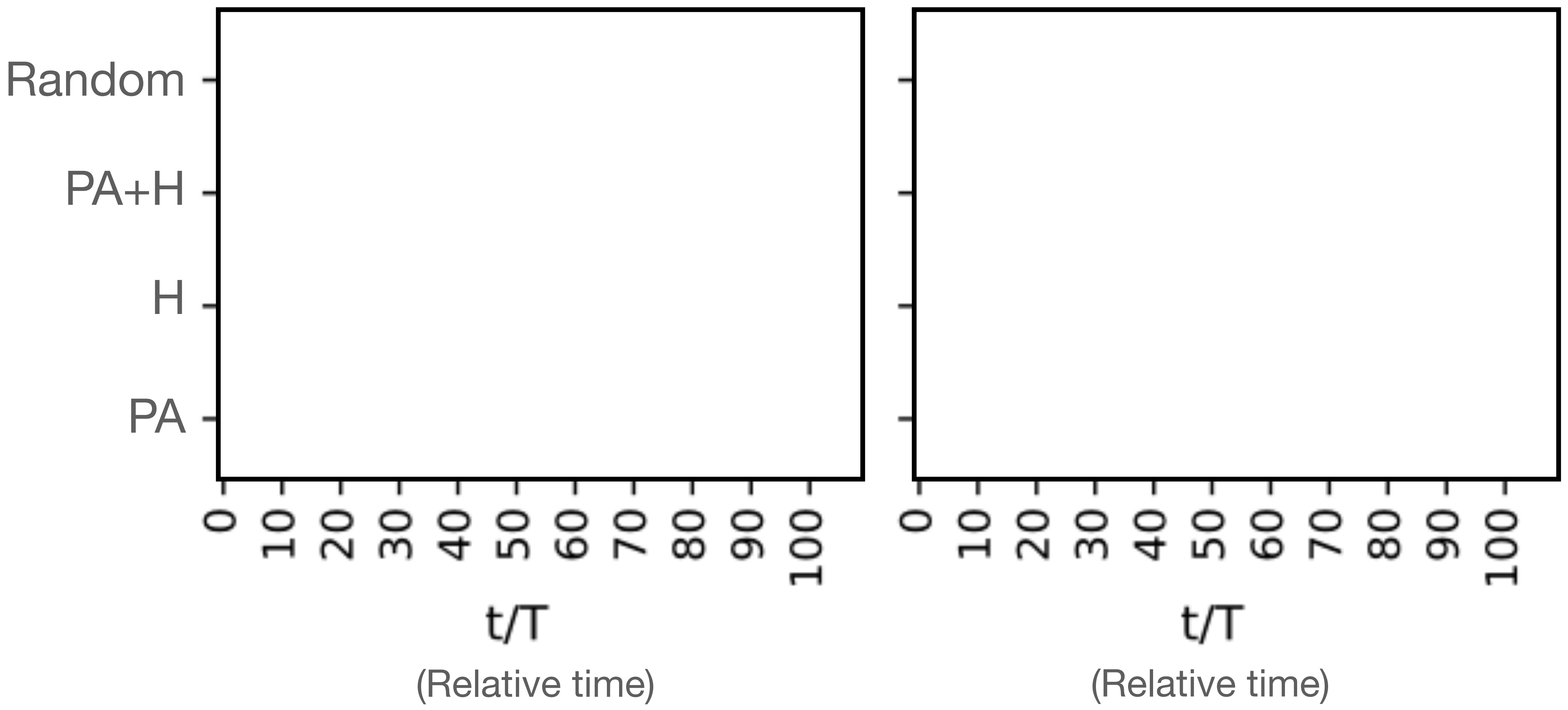
Task 2. Degree equality (social capital)



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Task 3. Information access equality (simple contagion)

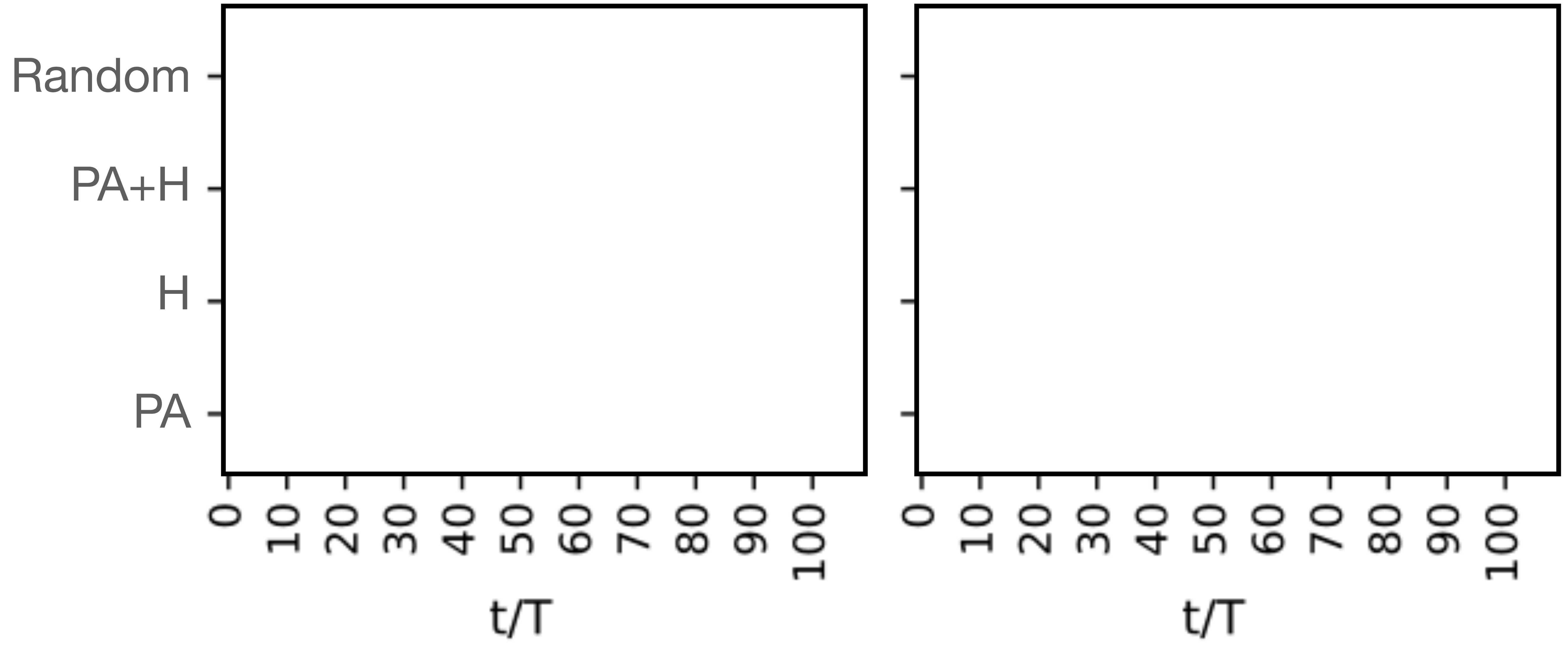


Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models."
Available at SSRN 3880680 (2021).

Understanding influence maximization

Task 3. Information access equality (simple contagion)

Low seeding
<30% min in seed sample

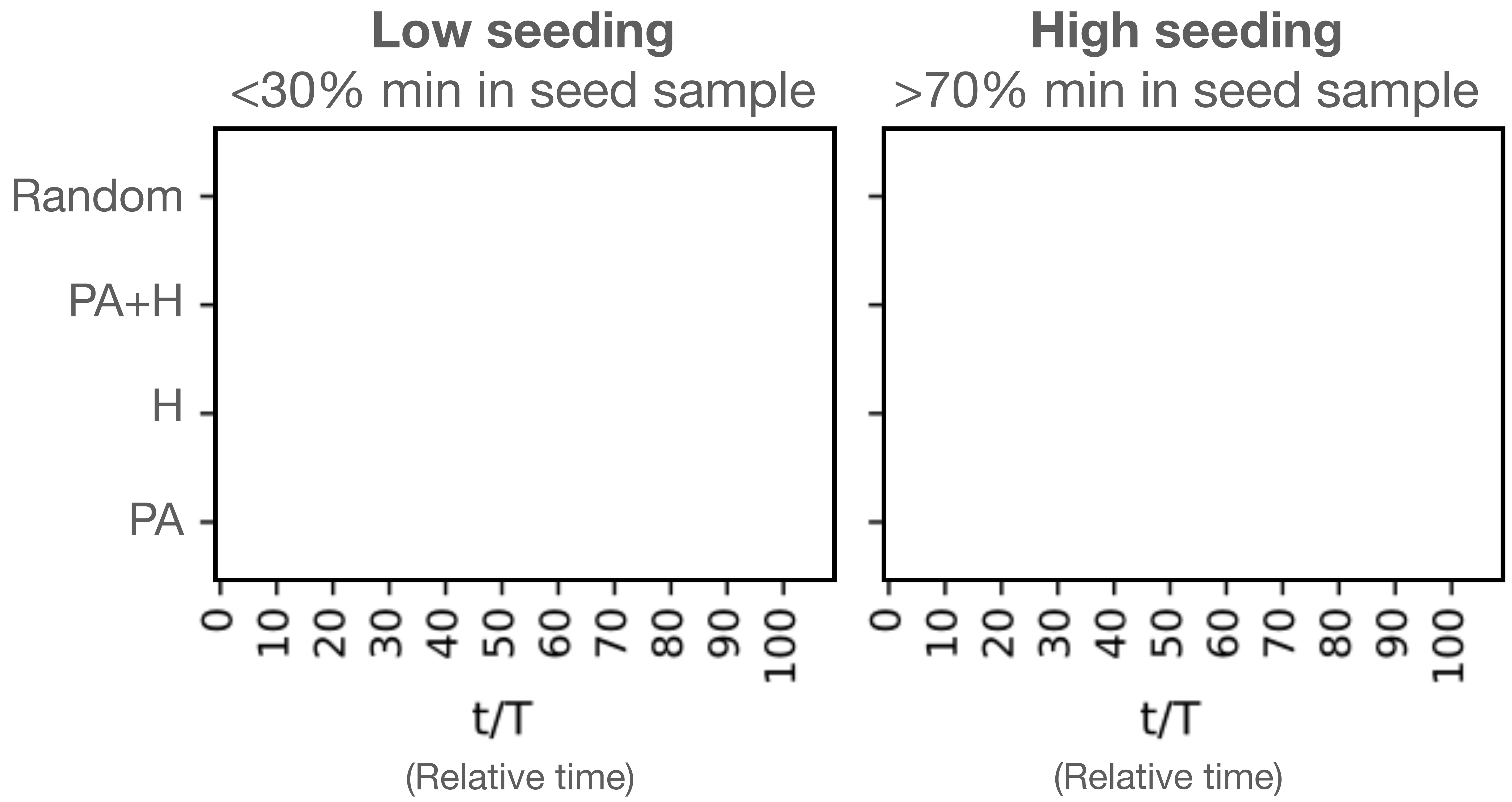


(Relative time) (Relative time)

Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

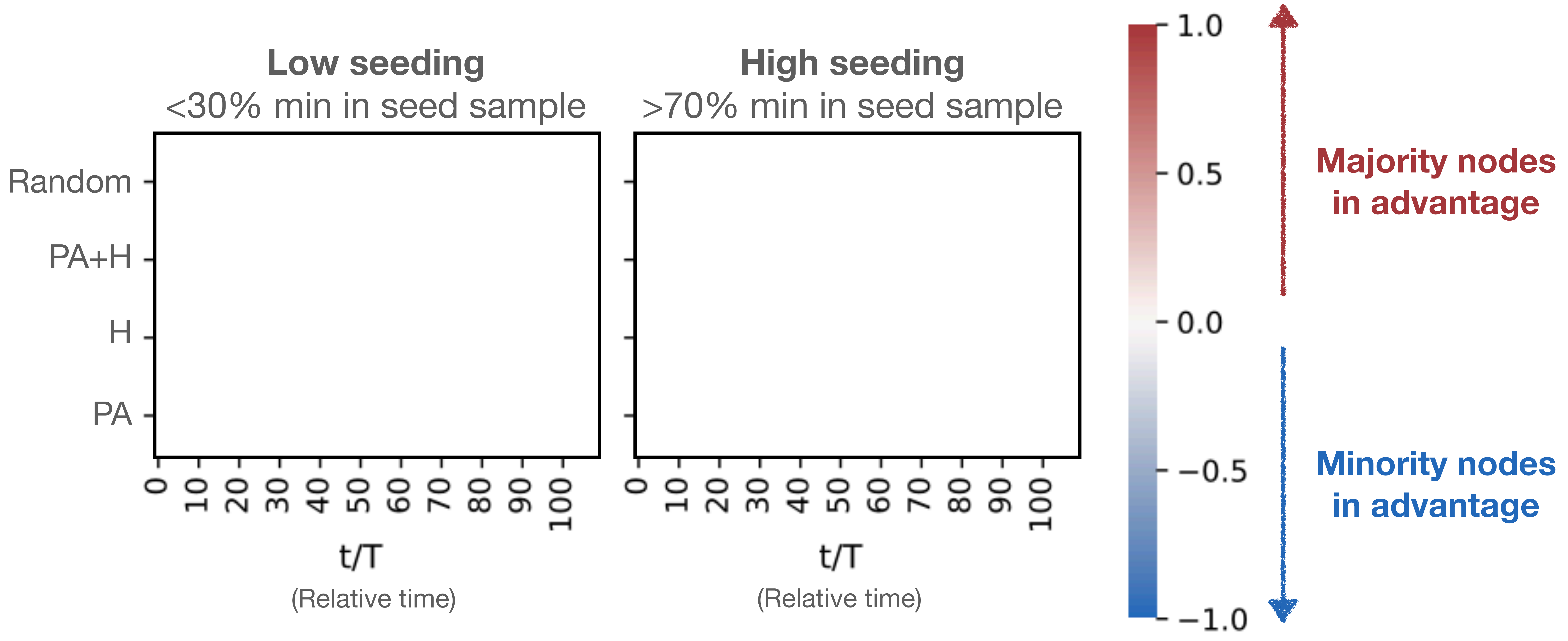
Task 3. Information access equality (simple contagion)



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

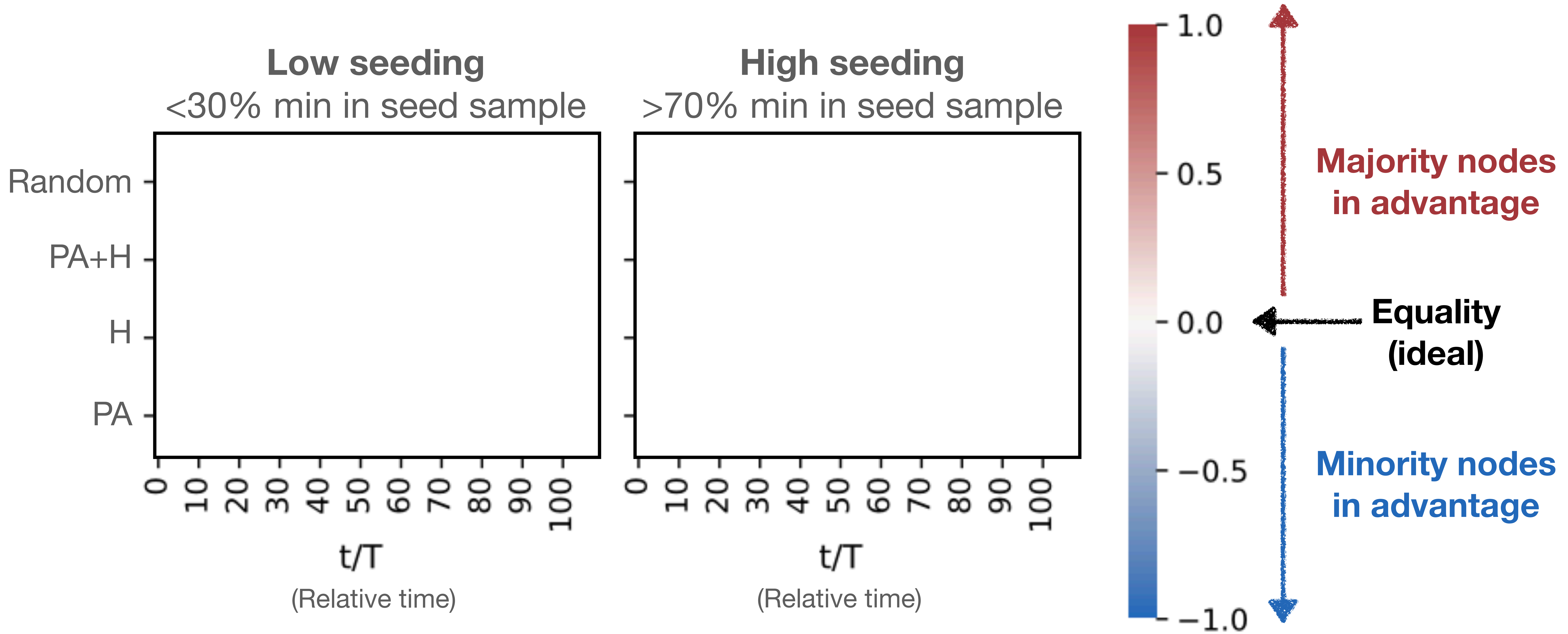
Task 3. Information access equality (simple contagion)



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

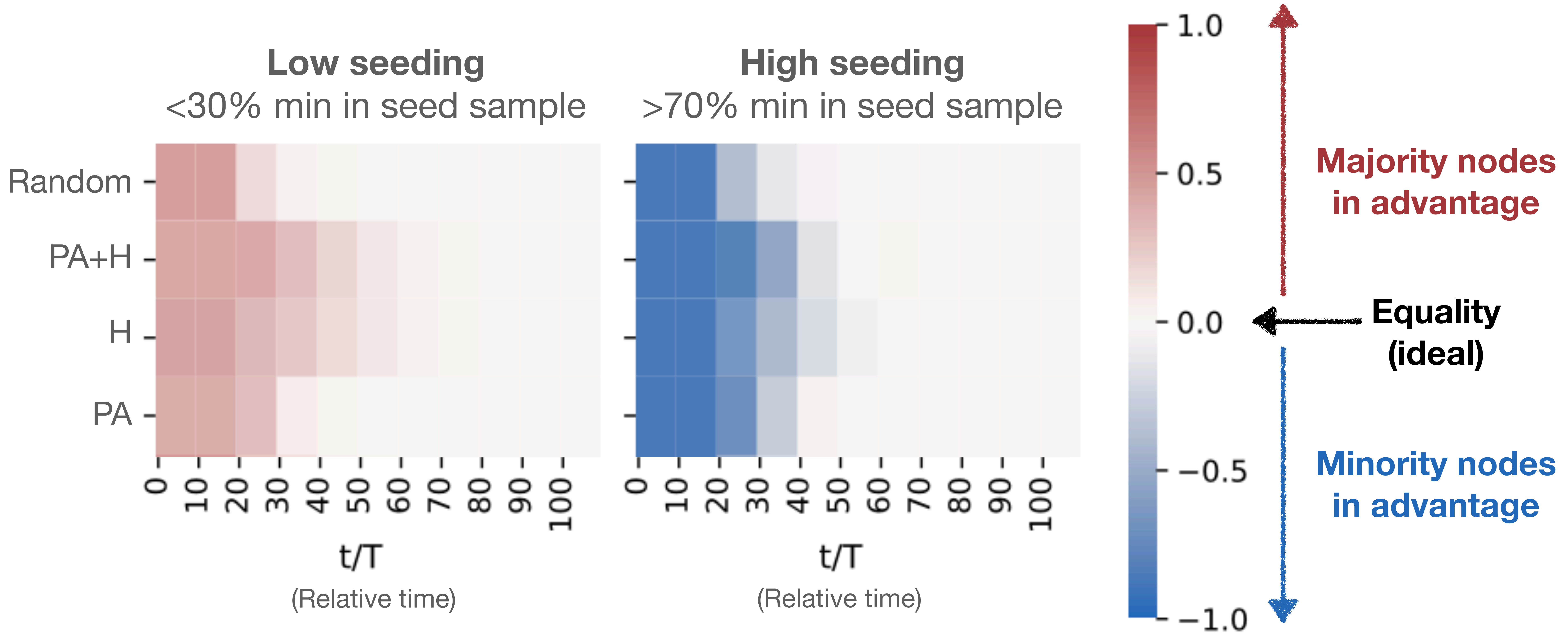
Task 3. Information access equality (simple contagion)



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Task 3. Information access equality (simple contagion)

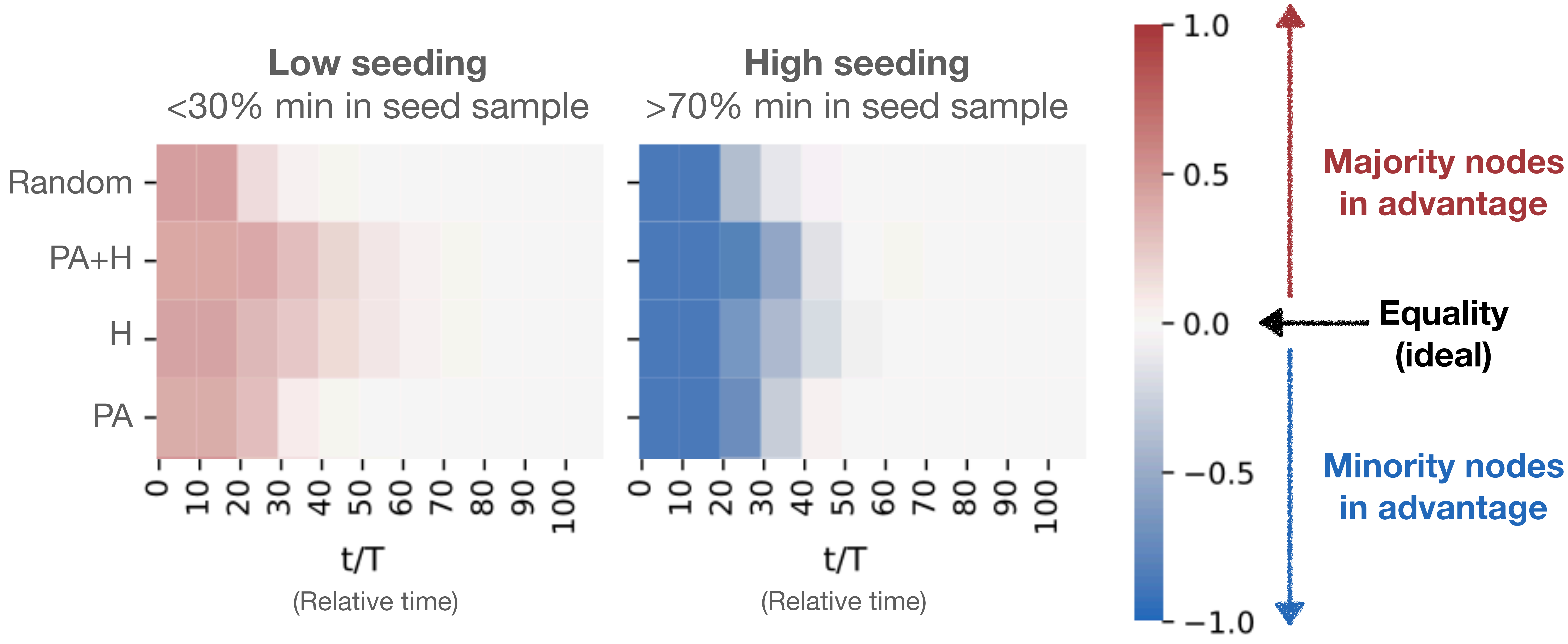


Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Task 3. Information access equality (simple contagion)

Low seeding initially benefits the nodes in the majority group



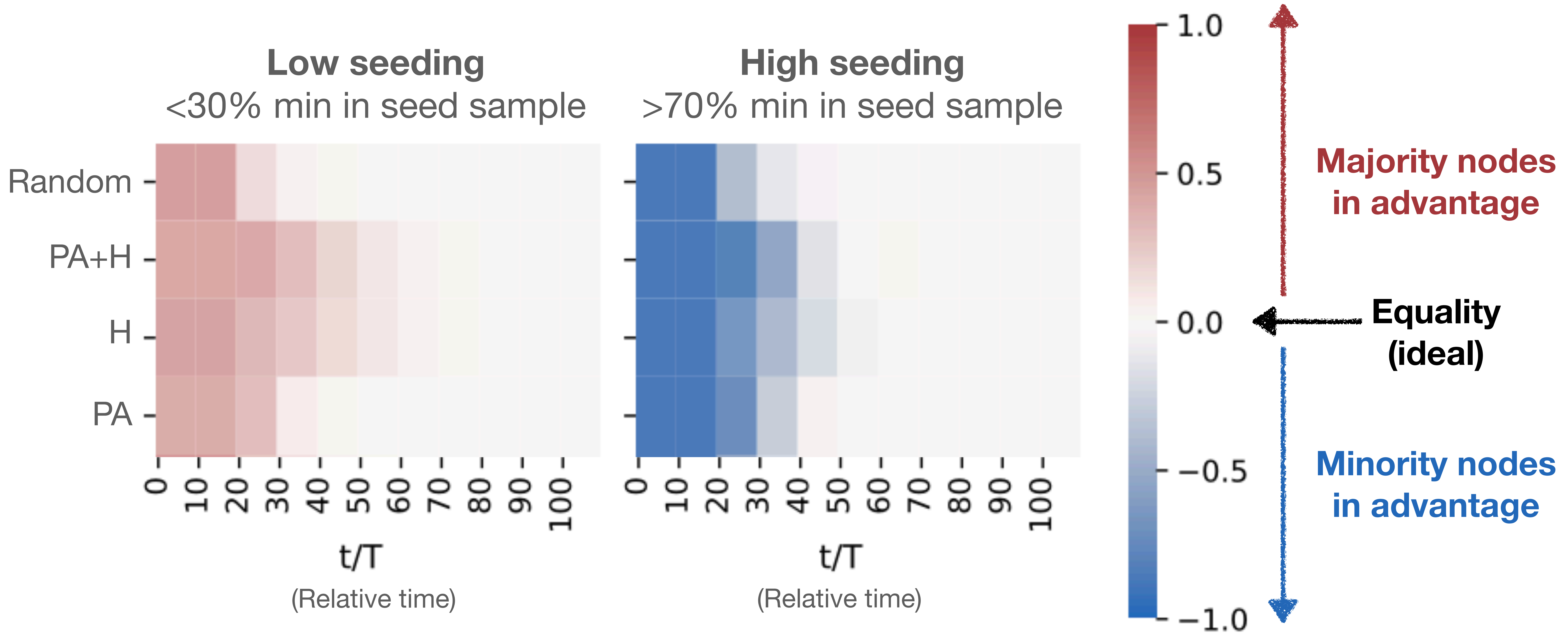
Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Understanding influence maximization

Task 3. Information access equality (simple contagion)

Low seeding initially benefits the nodes in the majority group

High seeding initially benefits the nodes in the minority group



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

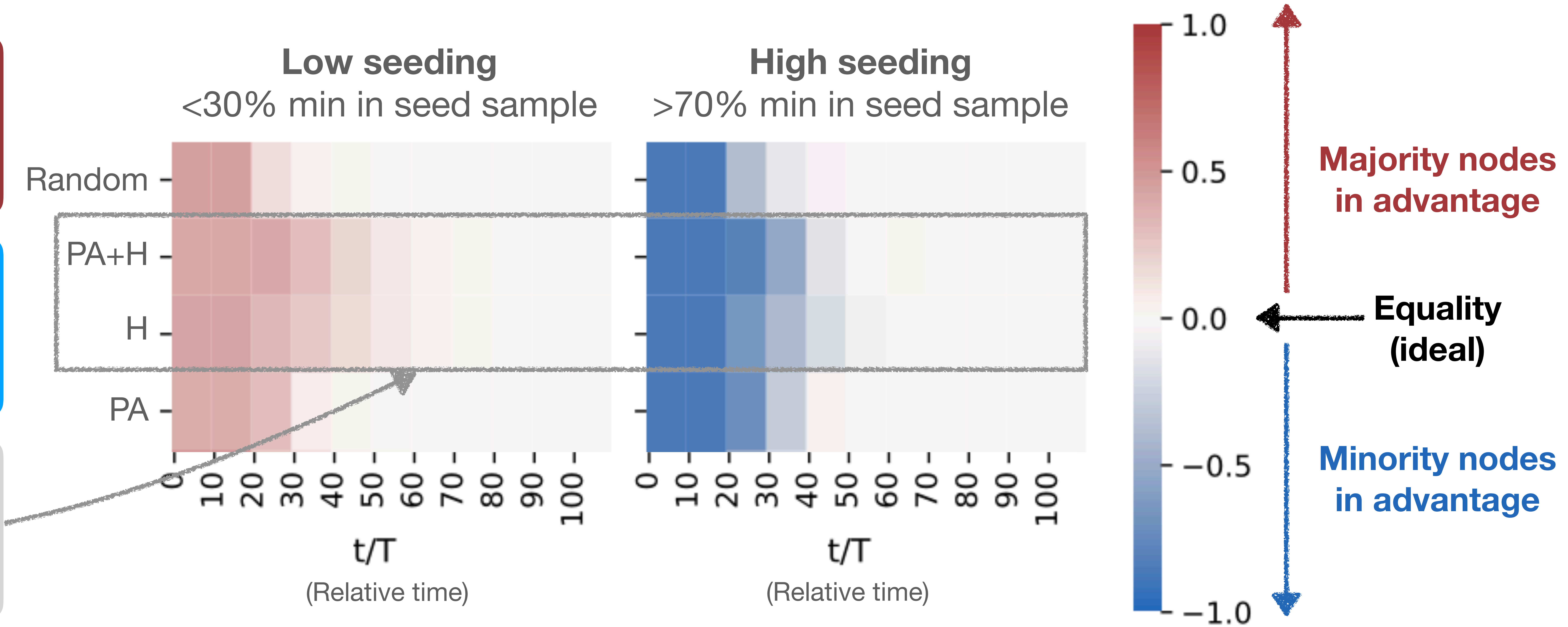
Understanding influence maximization

Task 3. Information access equality (simple contagion)

Low seeding initially benefits the nodes in the majority group

High seeding initially benefits the nodes in the minority group

Networks with homophily take longer to reach "spreading equality"



Wang, Xindi, Onur Varol, and Tina Eliassi-Rad. "Information Access Equality on Network Generative Models." Available at SSRN 3880680 (2021).

Other simulations to understand

Inequalities in network-based algorithms (“algorithmic auditing”)

Other simulations to understand

Inequalities in network-based algorithms (“algorithmic auditing”)

- Fairness in Social Influence Maximization (diverse out-reach and seeds)
  Stoica et al 2019

Other simulations to understand

Inequalities in network-based algorithms ("algorithmic auditing")

- Fairness in Social Influence Maximization (diverse out-reach and seeds)
  Stoica et al 2019

- Glass ceiling (under-representation of minorities) in top-k ranks
  Karimi et al. 2018, Stoica et al. 2018, Fabbri et al. 2020, Espin-Noboa et al. 2022

Other simulations to understand

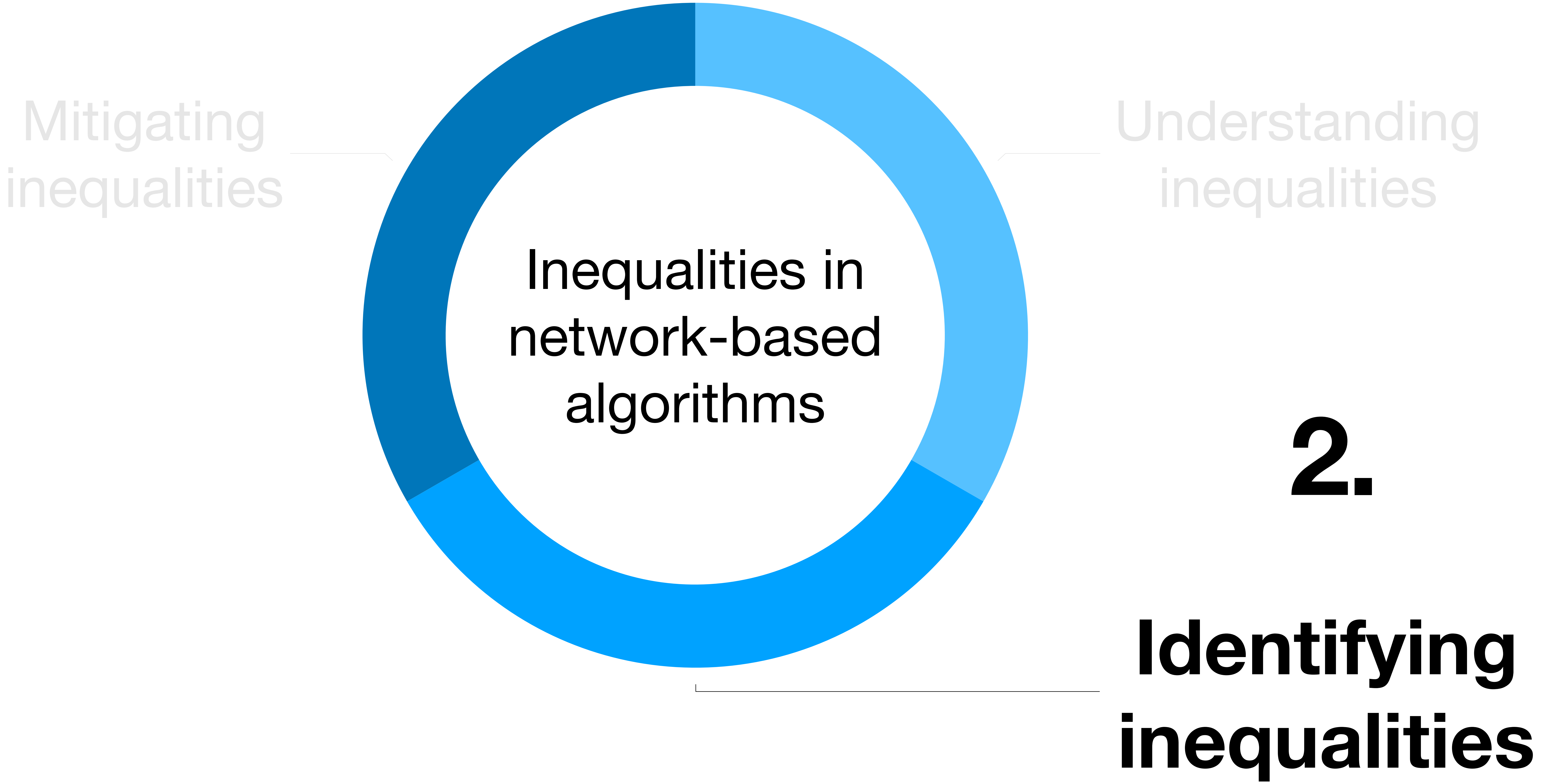
Inequalities in network-based algorithms (“algorithmic auditing”)

- Fairness in Social Influence Maximization (diverse out-reach and seeds)
  Stoica et al 2019
- Glass ceiling (under-representation of minorities) in top-k ranks
  Karimi et al. 2018, Stoica et al. 2018, Fabbri et al. 2020, Espin-Noboa et al. 2022
- How network structure affects node sampling and relational (node) classification in social networks
  Espin-Noboa et al. 2018, Espin-Noboa et al. 2021

Other simulations to understand

Inequalities in network-based algorithms (“algorithmic auditing”)

- Fairness in Social Influence Maximization (diverse out-reach and seeds)
  Stoica et al 2019
- Glass ceiling (under-representation of minorities) in top-k ranks
  Karimi et al. 2018, Stoica et al. 2018, Fabbri et al. 2020, Espin-Noboa et al. 2022
- How network structure affects node sampling and relational (node) classification in social networks
  Espin-Noboa et al. 2018, Espin-Noboa et al. 2021
- How network structure affects node link prediction algorithms (and vice versa)
  Espin-Noboa et al. 2022, Ferrara et al. 2022


Mitigating inequalities
Understanding inequalities
Inequalities in network-based algorithms
2.
Identifying inequalities

Evaluation benchmarks in “Relational classification”

INPUT
uncolored graph

SAMPLING
subgraph with labeled nodes

CLASSIFICATION
node-label inference

OUTPUT
colored graph

How does network **structure** influence **performance** & **bias**?

Espín-Noboa et al. MatNet@TheWebConf’2018, APN’2021

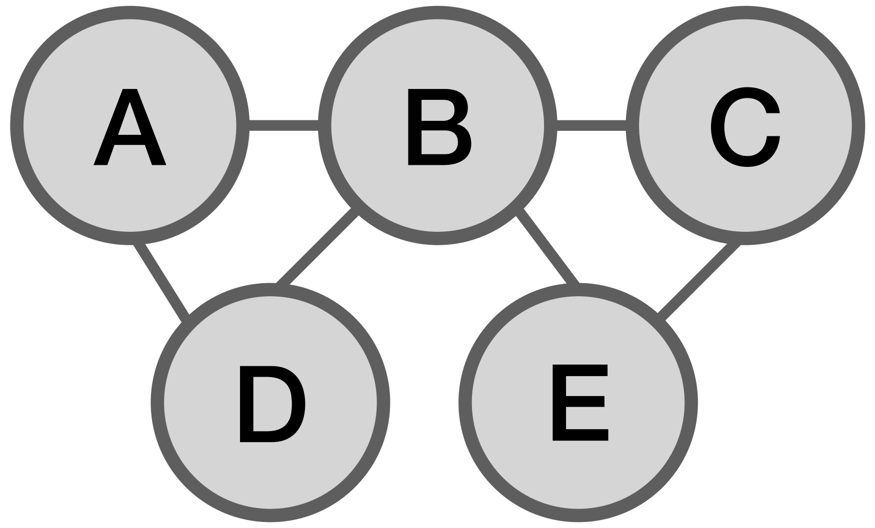
Relational Classification

Machine learning on networked data

Relational Classification

Machine learning on networked data

INPUT
uncolored
graph

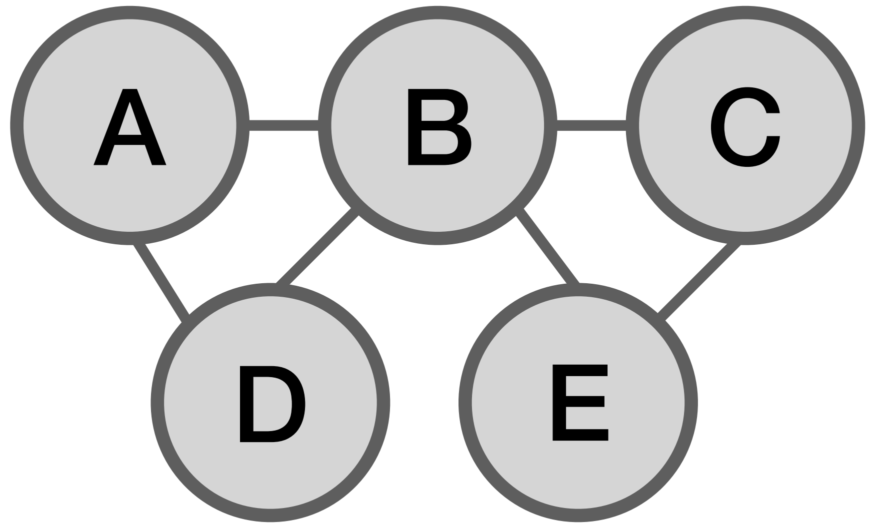


Relational Classification

Machine learning on networked data

INPUT
uncolored
graph

TASK
infer the colors of a node based on the color of its neighbors



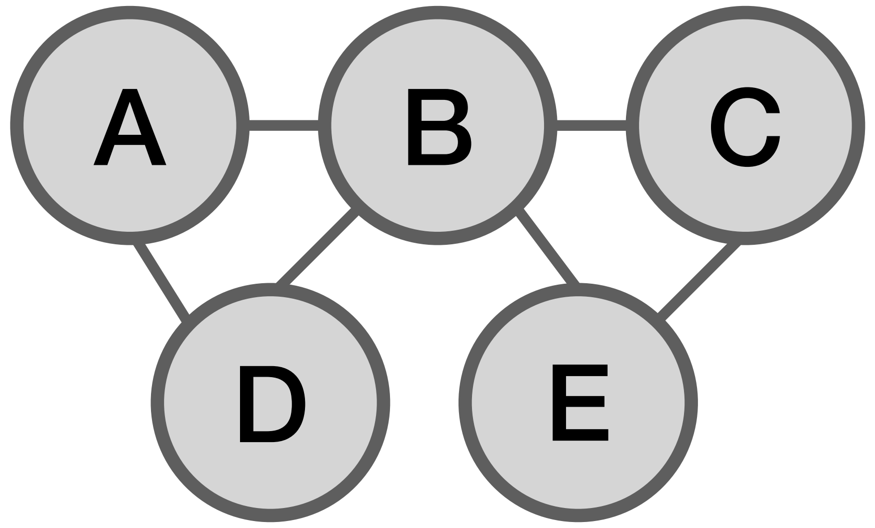
Relational Classification

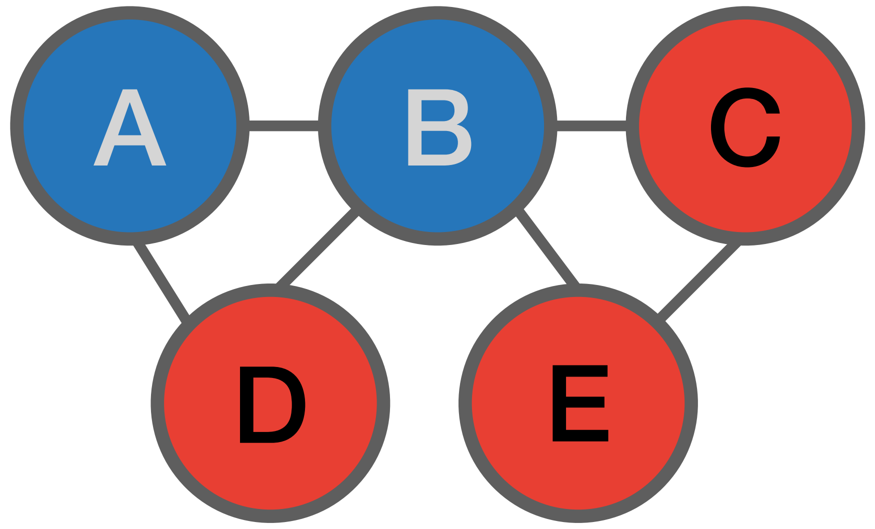
Machine learning on networked data

INPUT
uncolored graph

TASK
infer the colors of a node based on the color of its neighbors

OUTPUT
colored graph





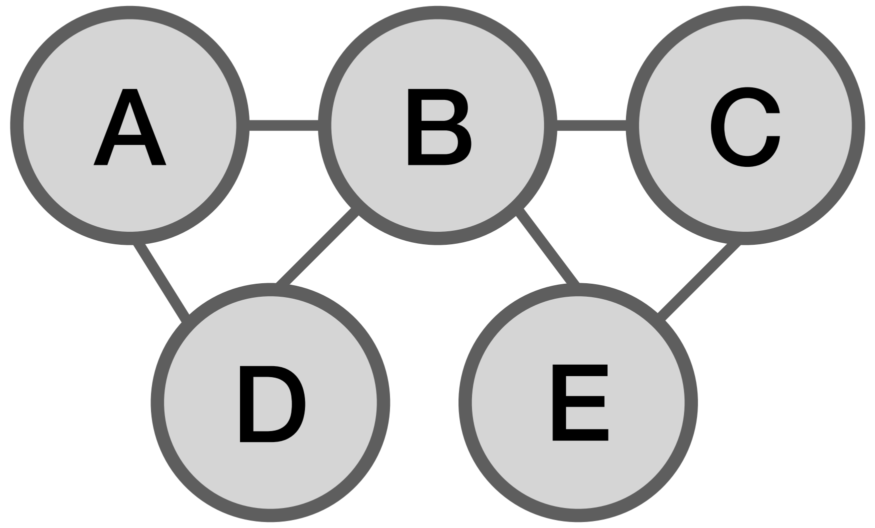
Relational Classification

Machine learning on networked data

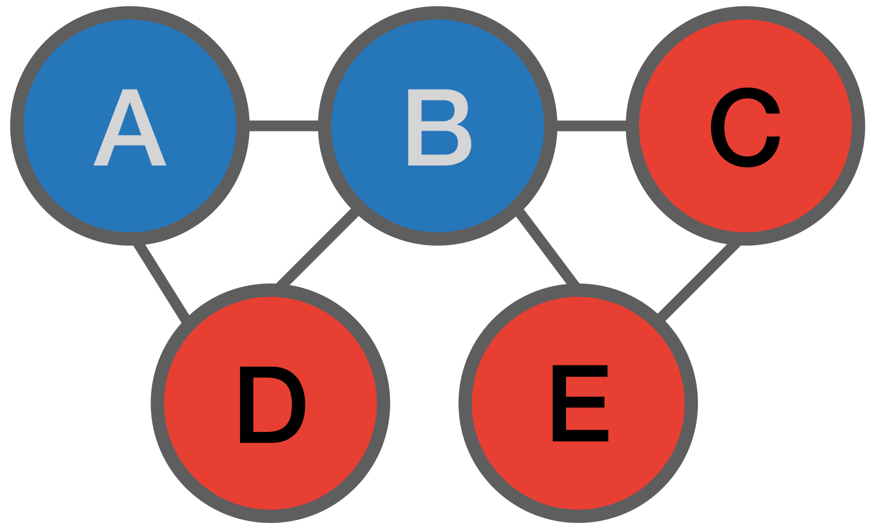
INPUT
uncolored graph

TASK
infer the colors of a node based on the color of its neighbors

OUTPUT
colored graph



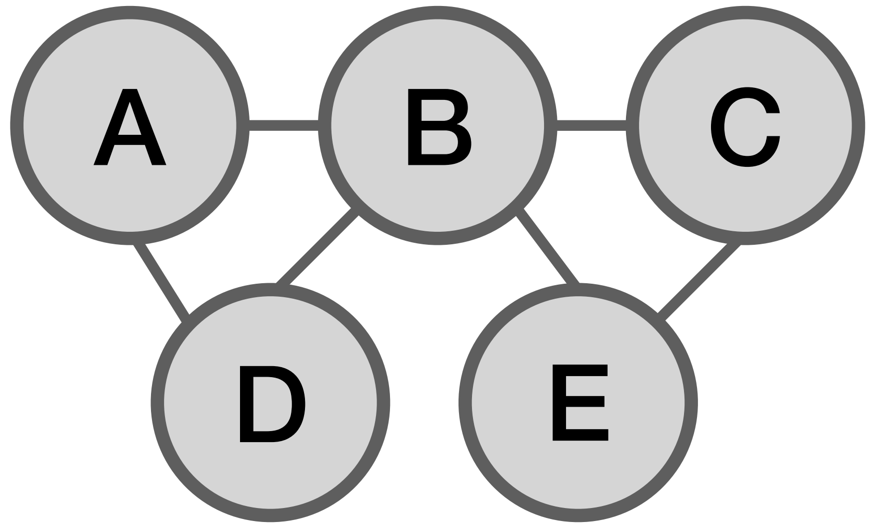
Are these results correct?
Or, are they biased against blue nodes?

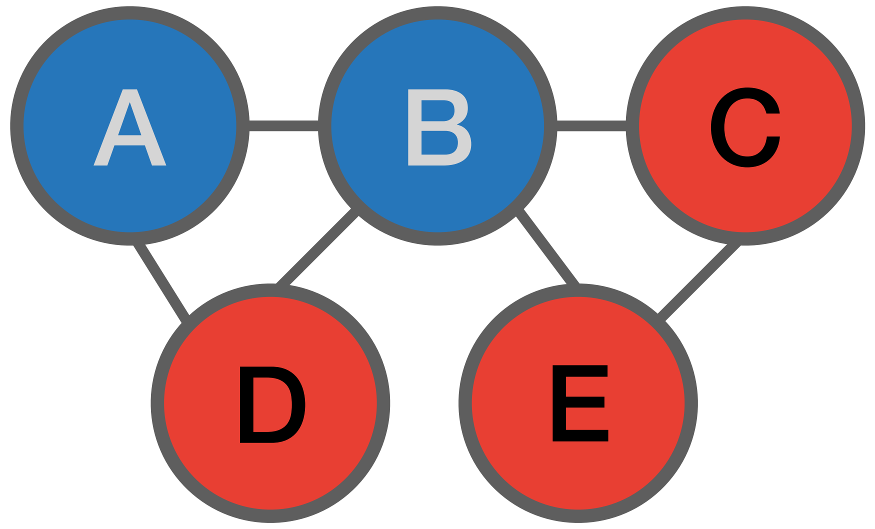


Relational Classification

Machine learning on networked data

INPUT uncolored graph → **SAMPLING** subgraph → **MODELING** parameter estimation → **INFERENCE** posteriors → **OUTPUT** colored graph





Relational Classification

Machine learning on networked data

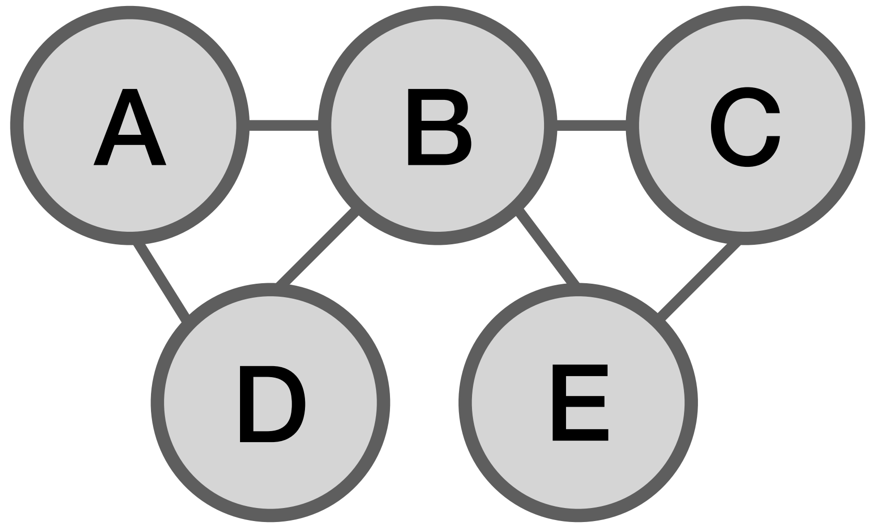
INPUT
uncolored
graph

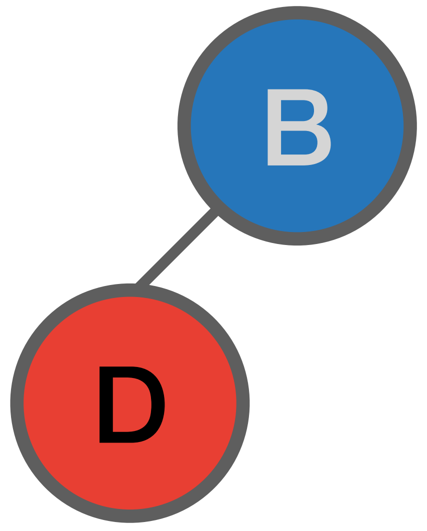
SAMPLING
subgraph

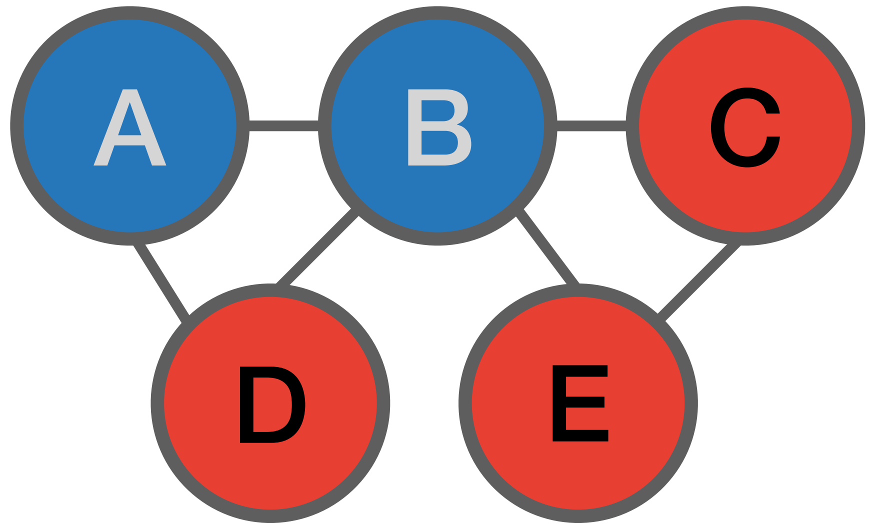
MODELING
parameter
estimation

INFERENCE
posteriors

OUTPUT
colored
graph



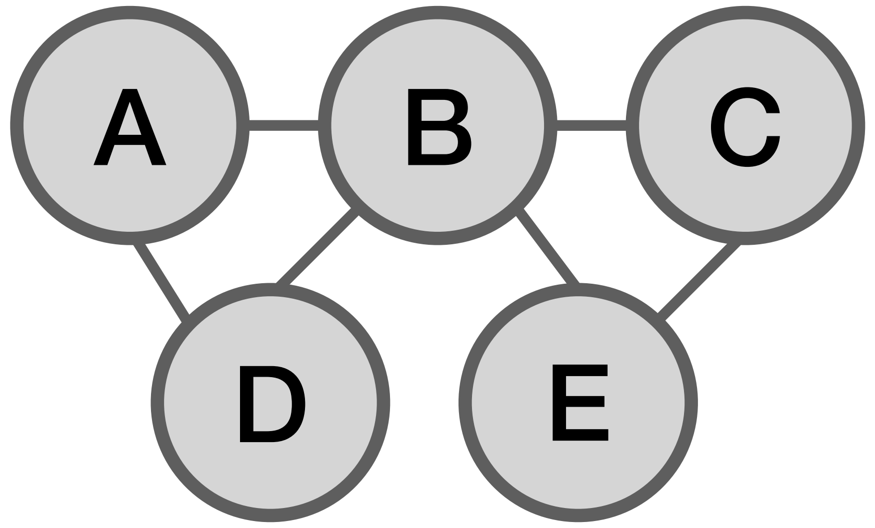


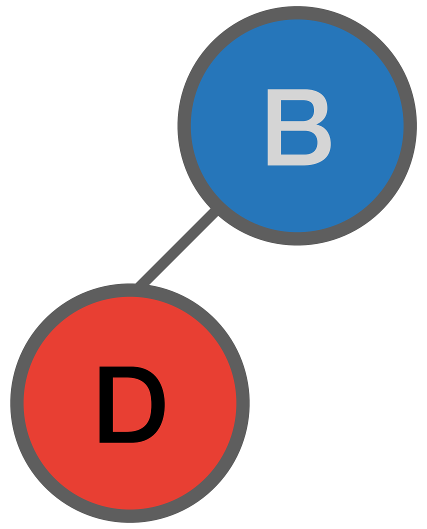


Relational Classification

Machine learning on networked data

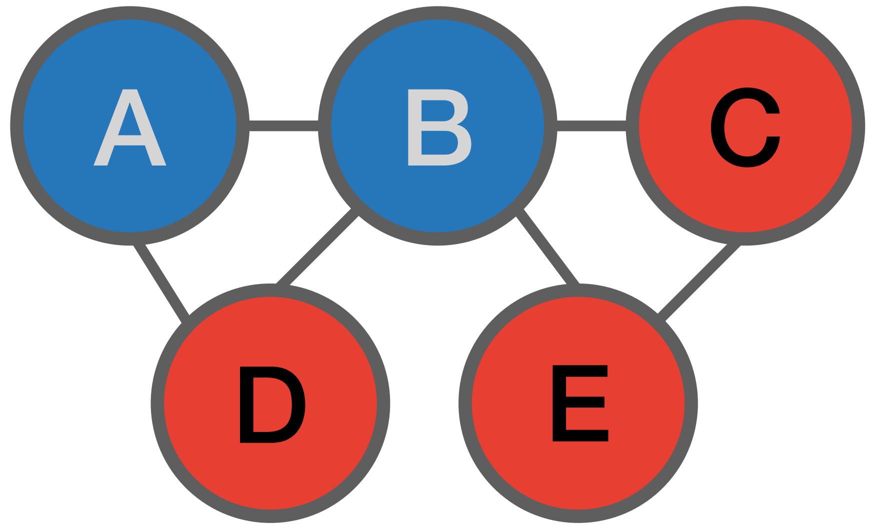
INPUT uncolored graph → **SAMPLING** subgraph → **MODELING** parameter estimation → **INFERENCE** posteriors → **OUTPUT** colored graph





$$P(v_i = \text{red}) = 0.5$$

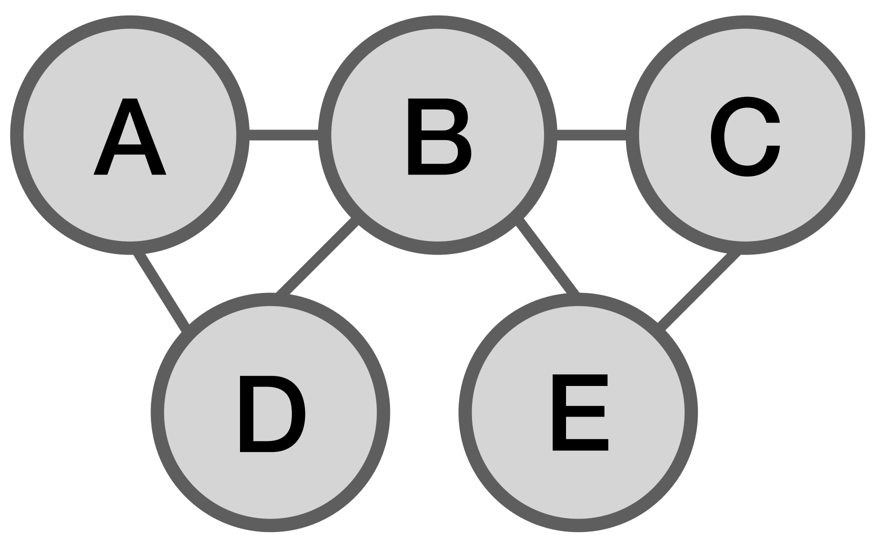
$$P(v_i = \text{red} \mid v_j = \text{blue}) = 1.0$$

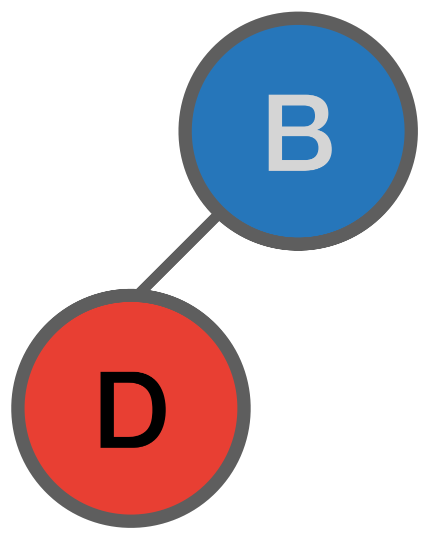


Relational Classification

Machine learning on networked data

INPUT uncolored graph → **SAMPLING** subgraph → **MODELING** parameter estimation → **INFERENCE** posteriors → **OUTPUT** colored graph

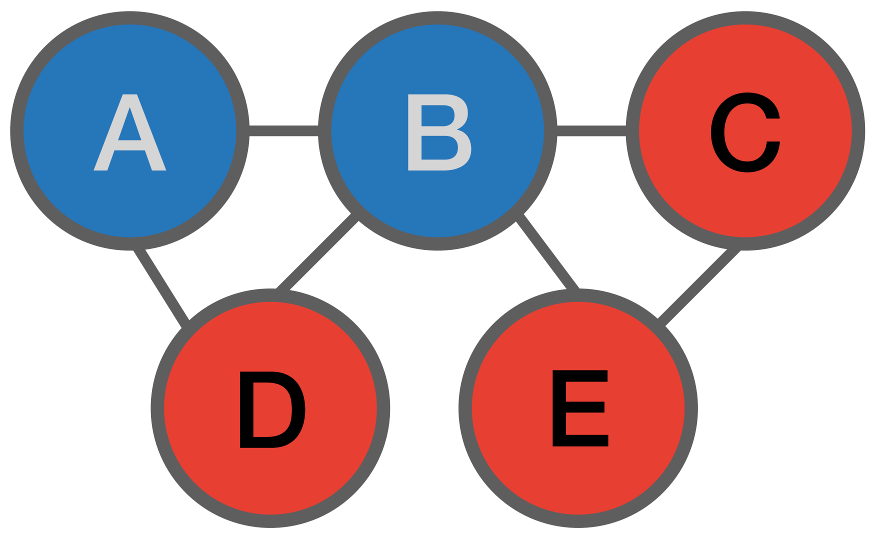




$$P(v_i = \text{red}) = 0.5$$

$$P(v_i = \text{red} \,|\, v_j = \text{blue}) = 1.0$$

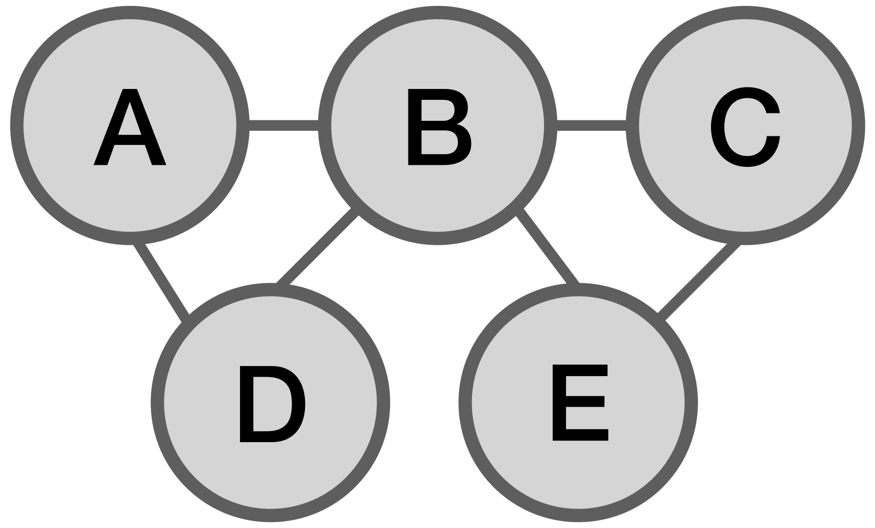
$$P(A_b \,|\, B_b, D_r)$$

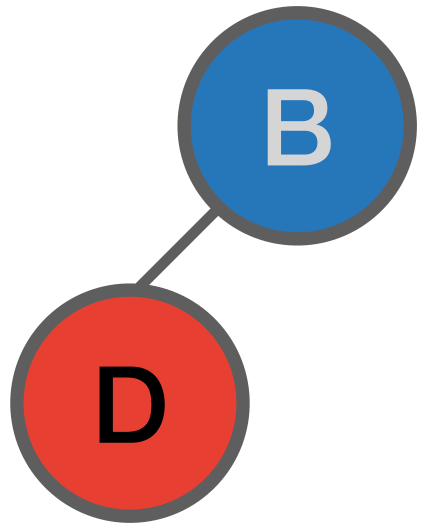


Relational Classification

Machine learning on networked data

?	?	?	?	
INPUT uncolored graph	**SAMPLING** subgraph	**MODELING** parameter estimation	**INFERENCE** posteriors	**OUTPUT** colored graph

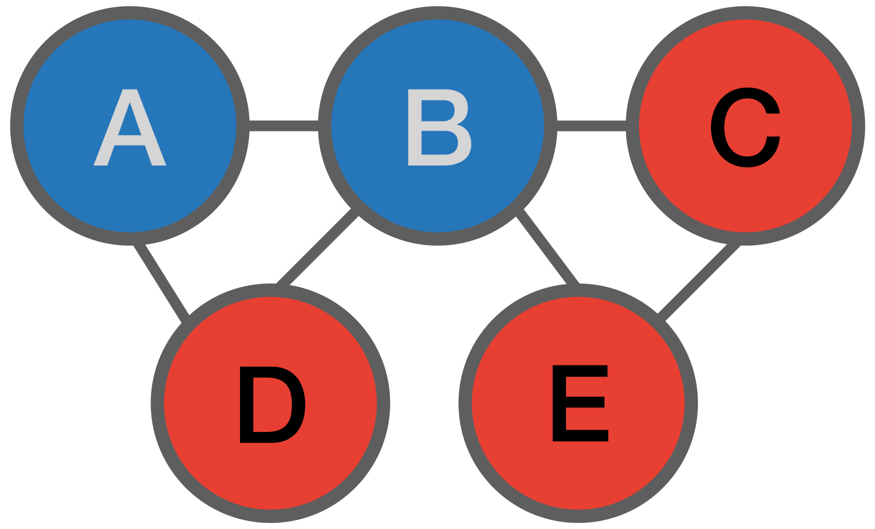




$$P(v_i = \text{red}) = 0.5$$

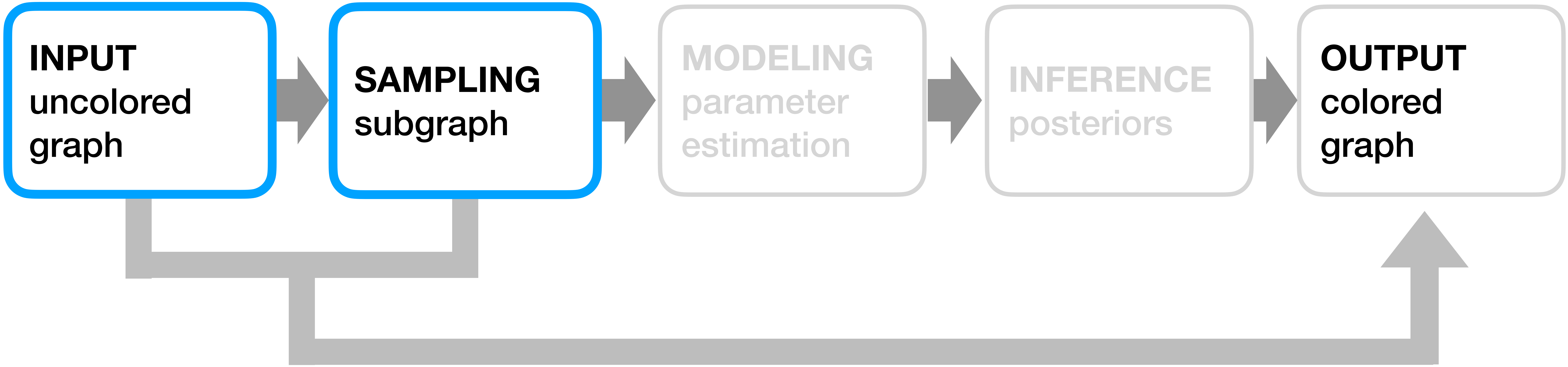
$$P(v_i = \text{red} \mid v_j = \text{blue}) = 1.0$$

$$P(A_b \mid B_b, D_r)$$



Relational Classification

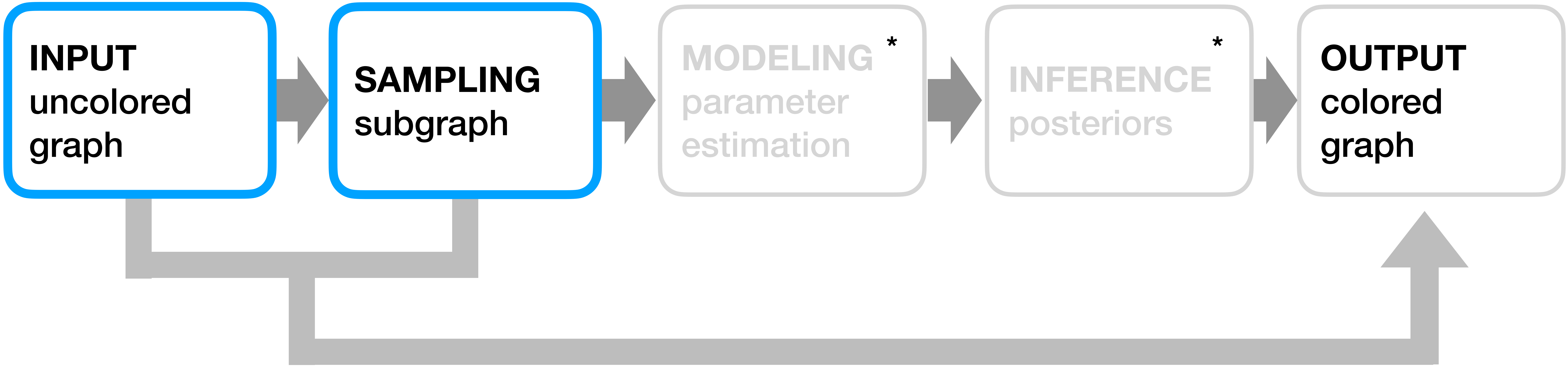
We quantify biases coming from data!



We explain the output performance and bias
with the characteristics of the input and training sample!

Relational Classification

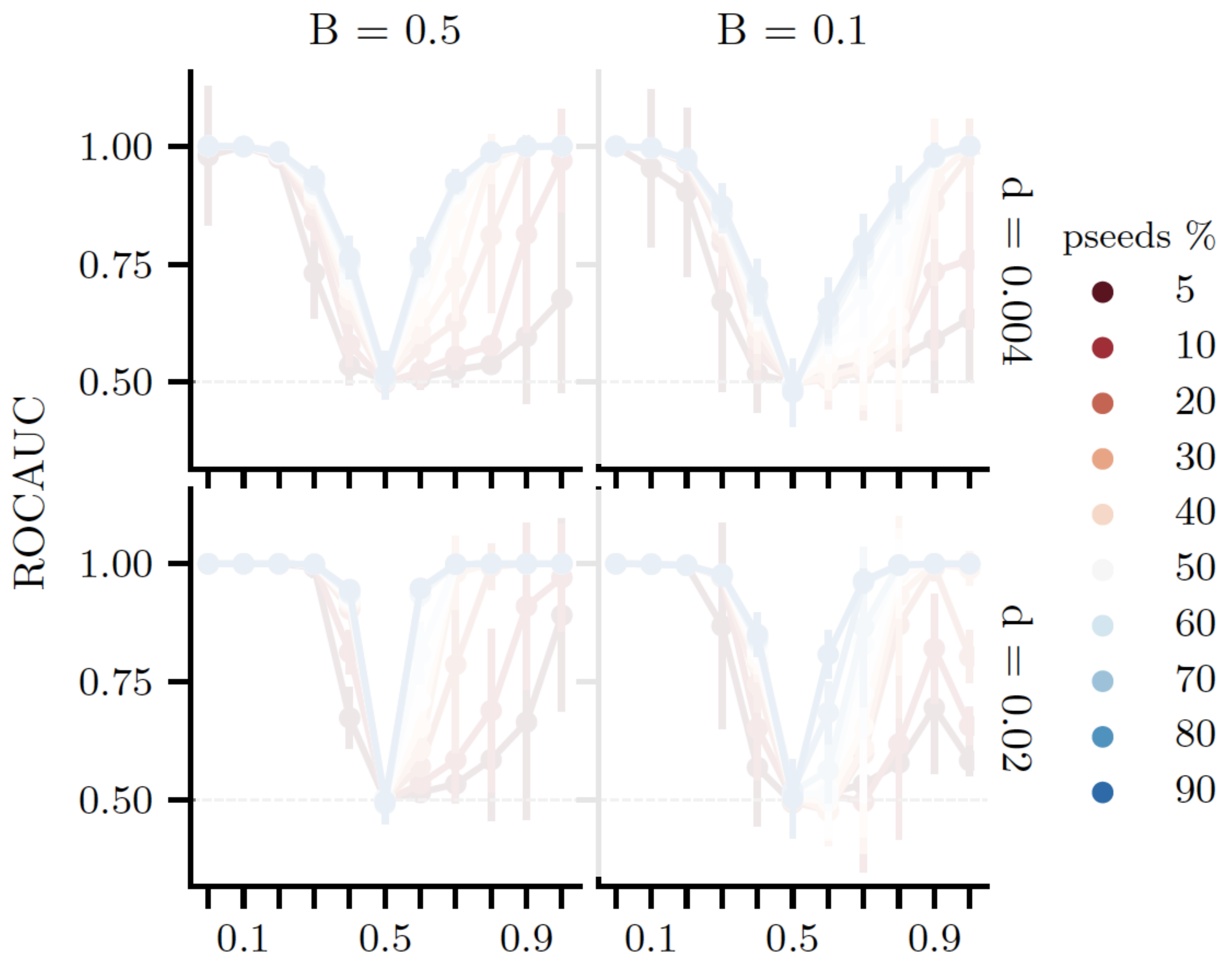
We quantify biases coming from data!



We explain the output performance and bias
with the characteristics of the input and training sample!

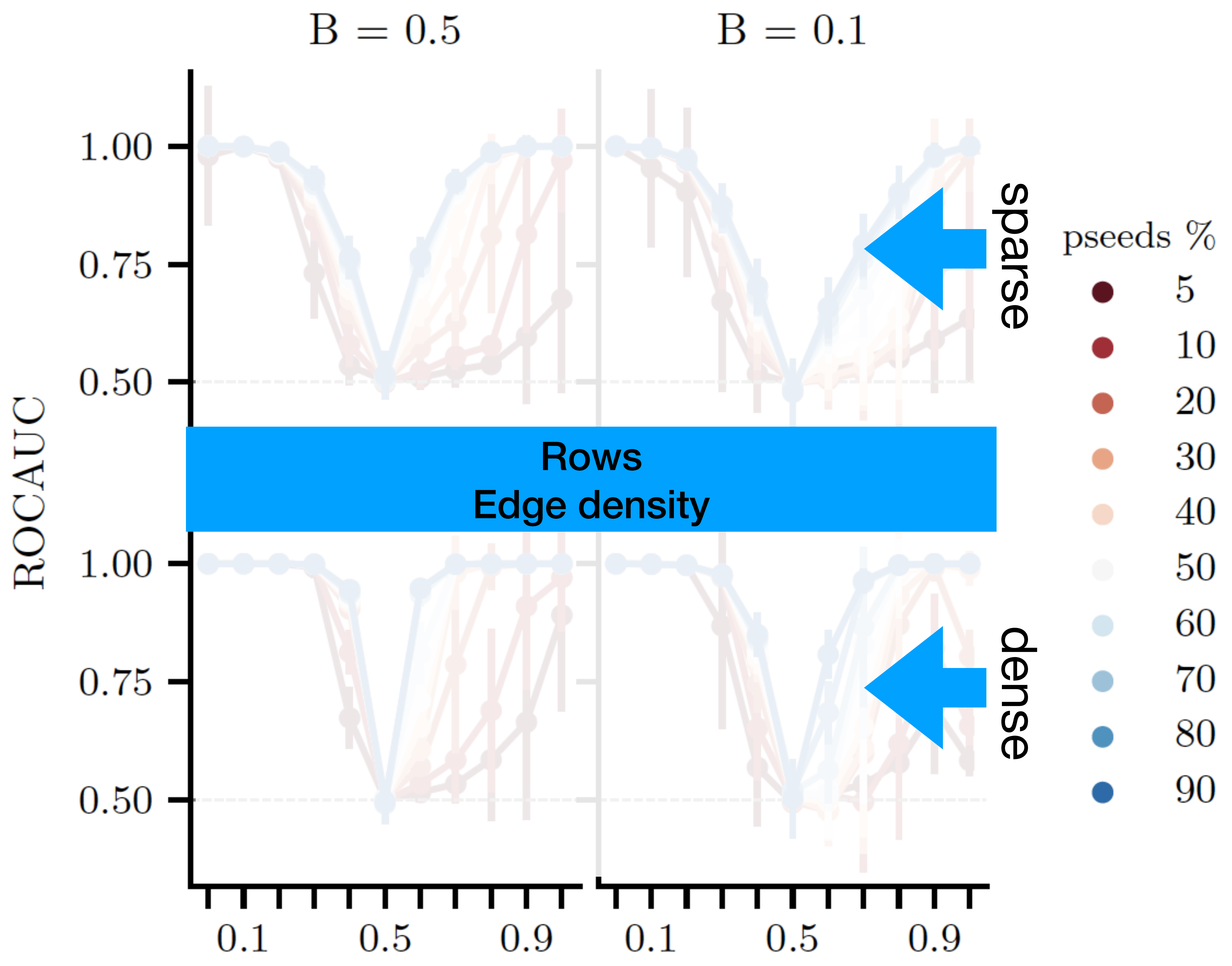
* Modeling + Inference in [Macskassy and Provost 2007]
(network-only Bayes classifier + relaxation labeling)

Network structure vs. classification performance



Espín-Noboa et al. 2021

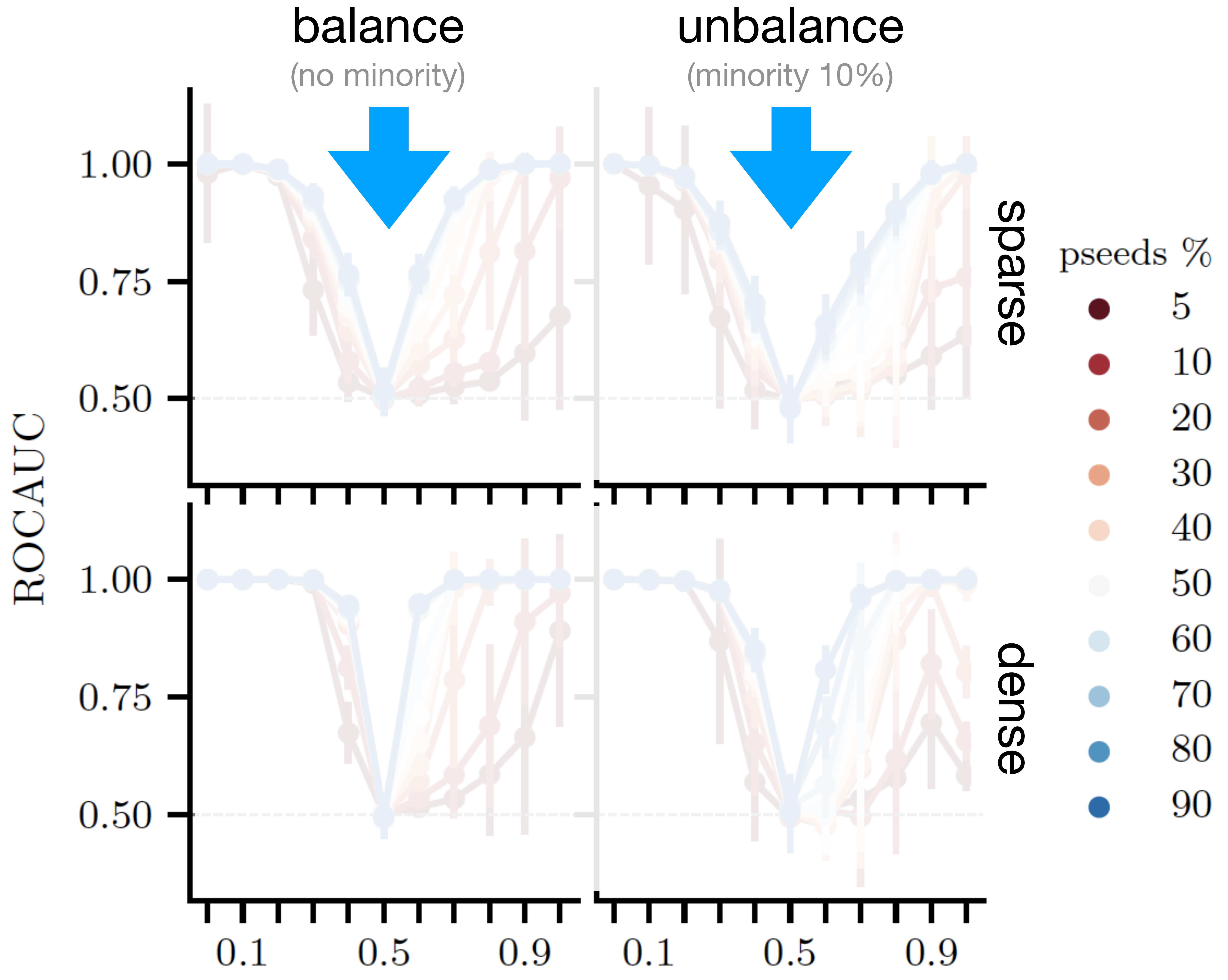
Network structure vs. classification performance



Espín-Noboa et al. 2021

Network structure vs. cla[...]ance

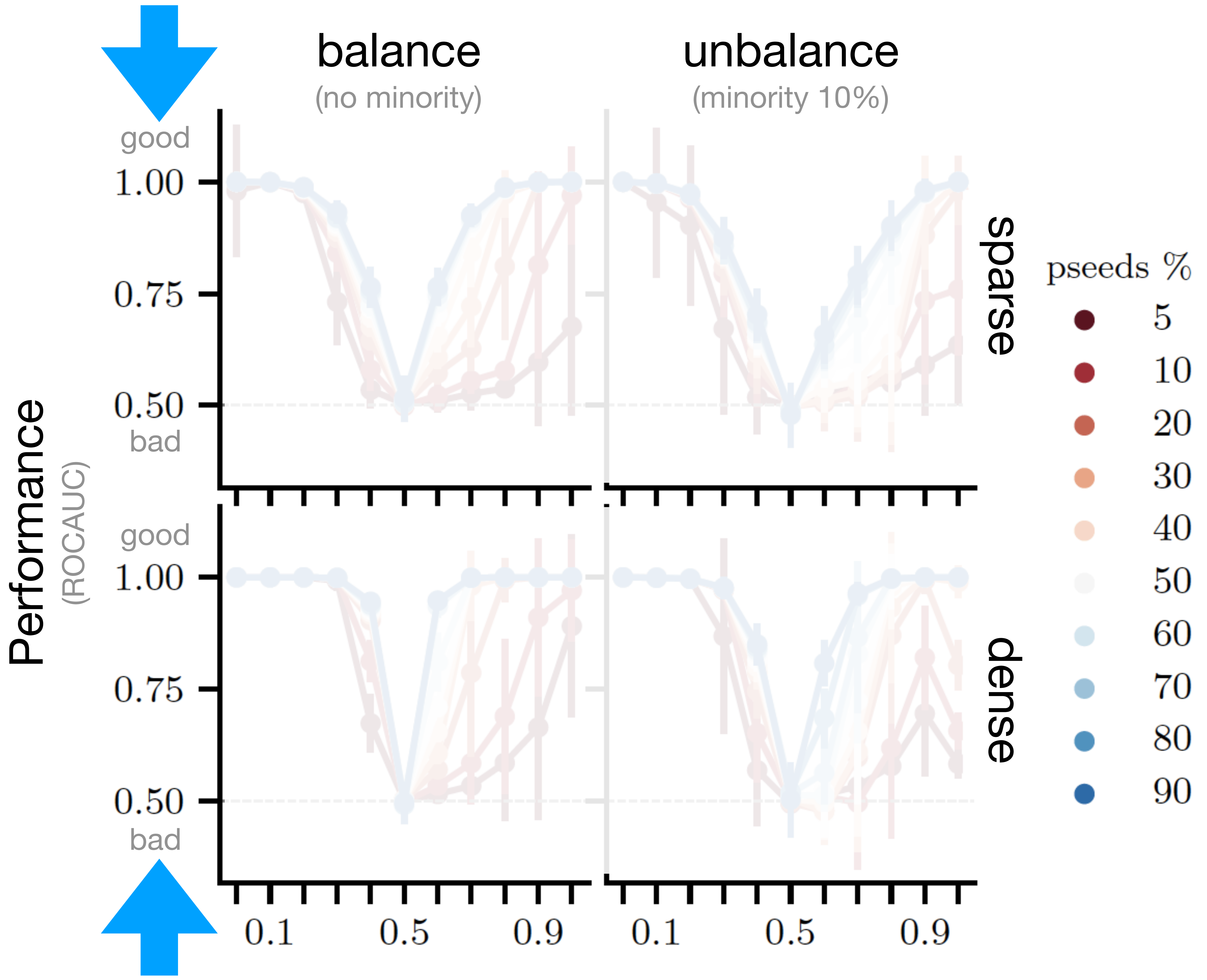
Columns
Class balance or fraction of minority group



Espín-Noboa et al. 2021

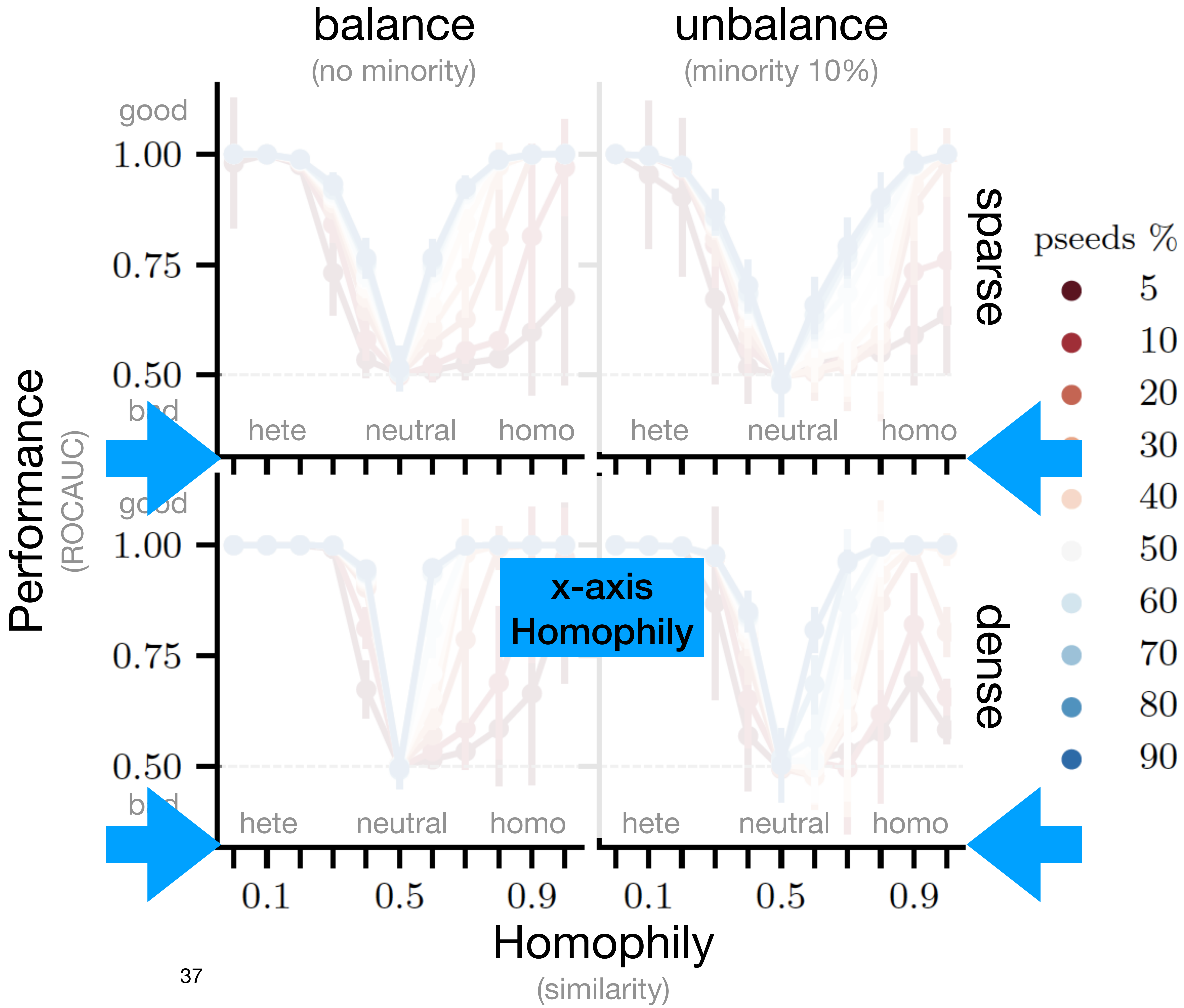
Network structure vs. classification performance

y-axis
ROCAUC



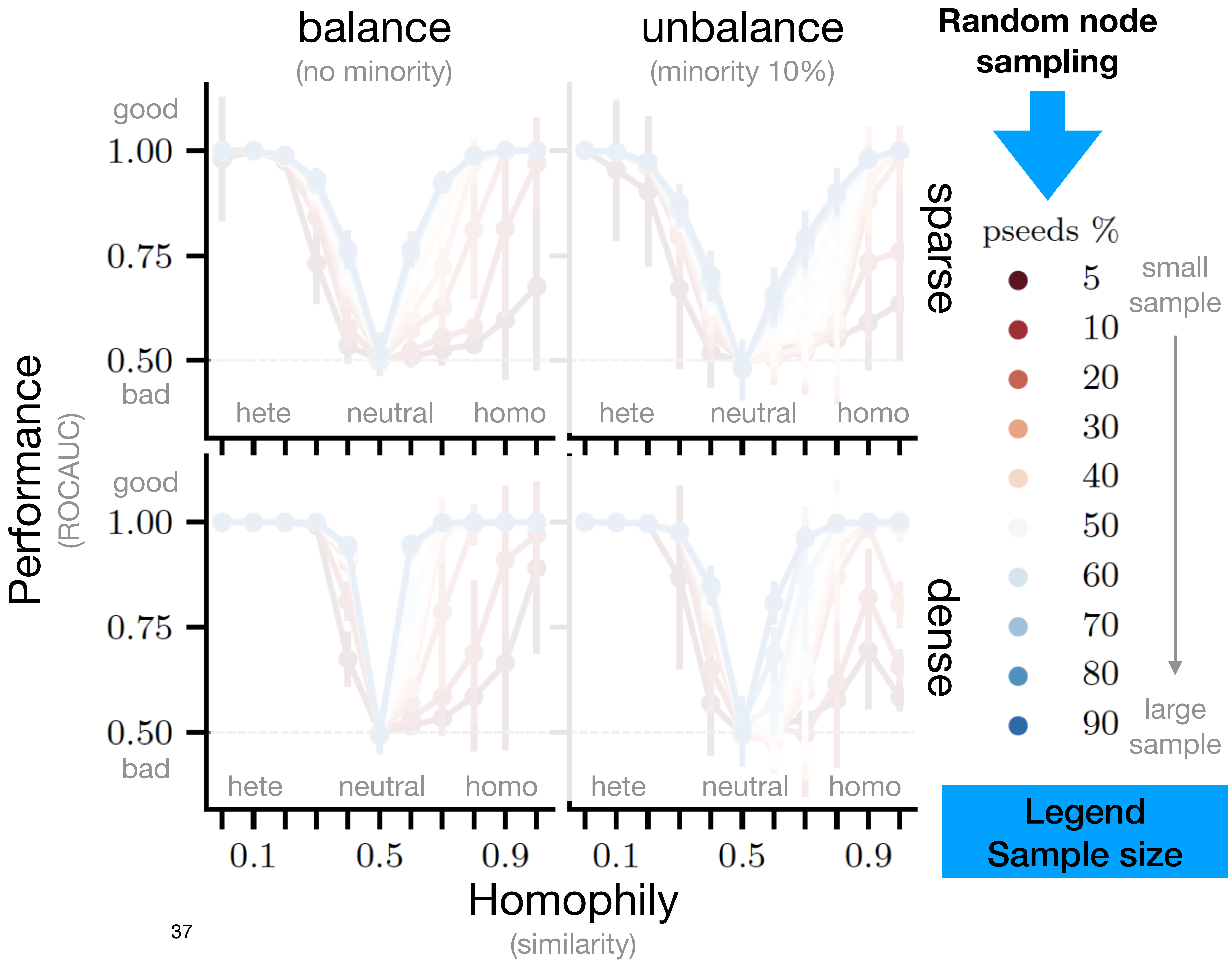
Espín-Noboa et al. 2021

Network structure vs. classification performance



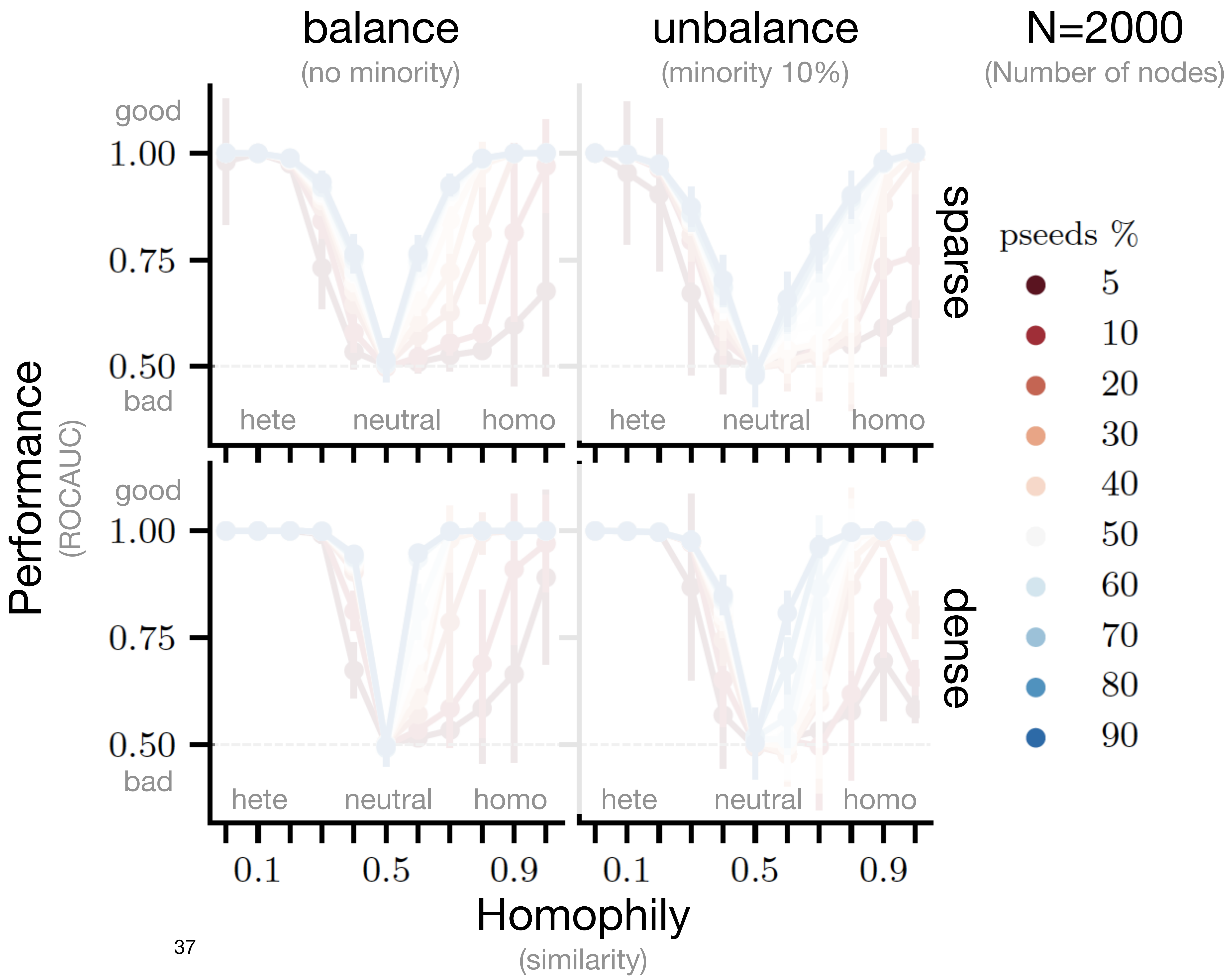
Espín-Noboa et al. 2021

Network structure vs. classification performance



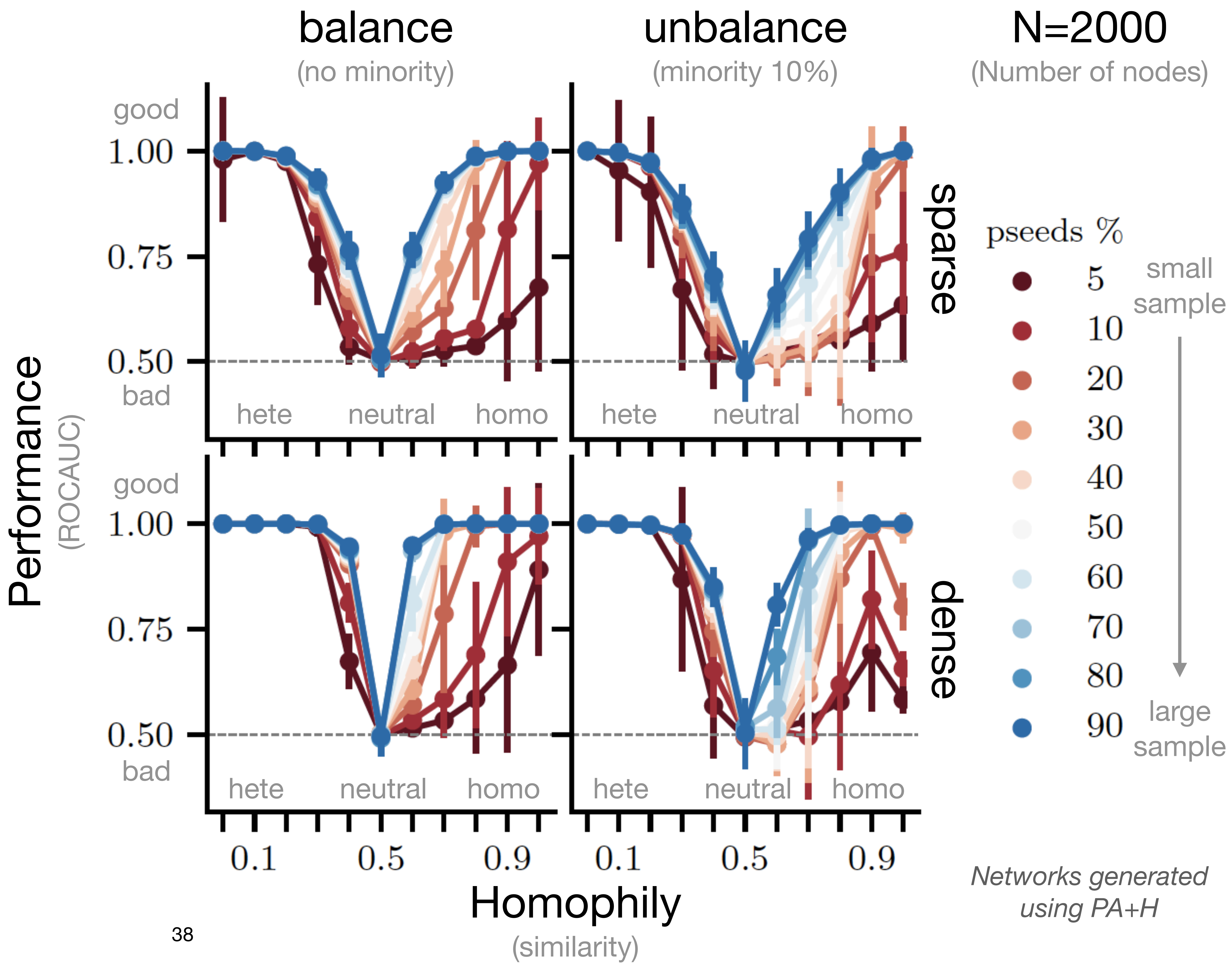
Espín-Noboa et al. 2021

Network structure vs. classification performance



Espín-Noboa et al. 2021

Network structure vs. classification performance

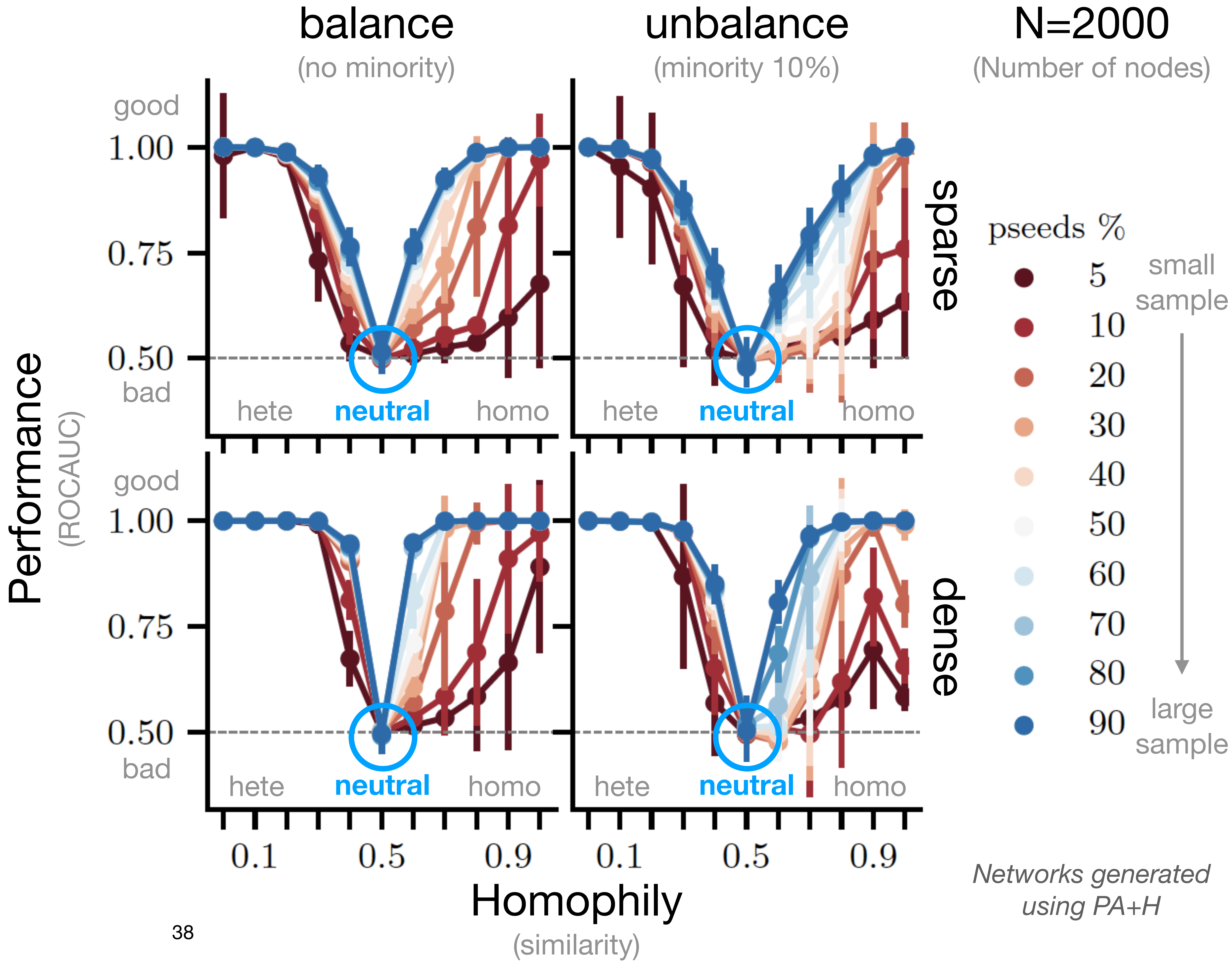


Networks generated using PA+H

Espín-Noboa et al. 2021

Network structure vs. classification performance

1. Neutral networks (H=0.5) cannot be classified better than a random classifier.



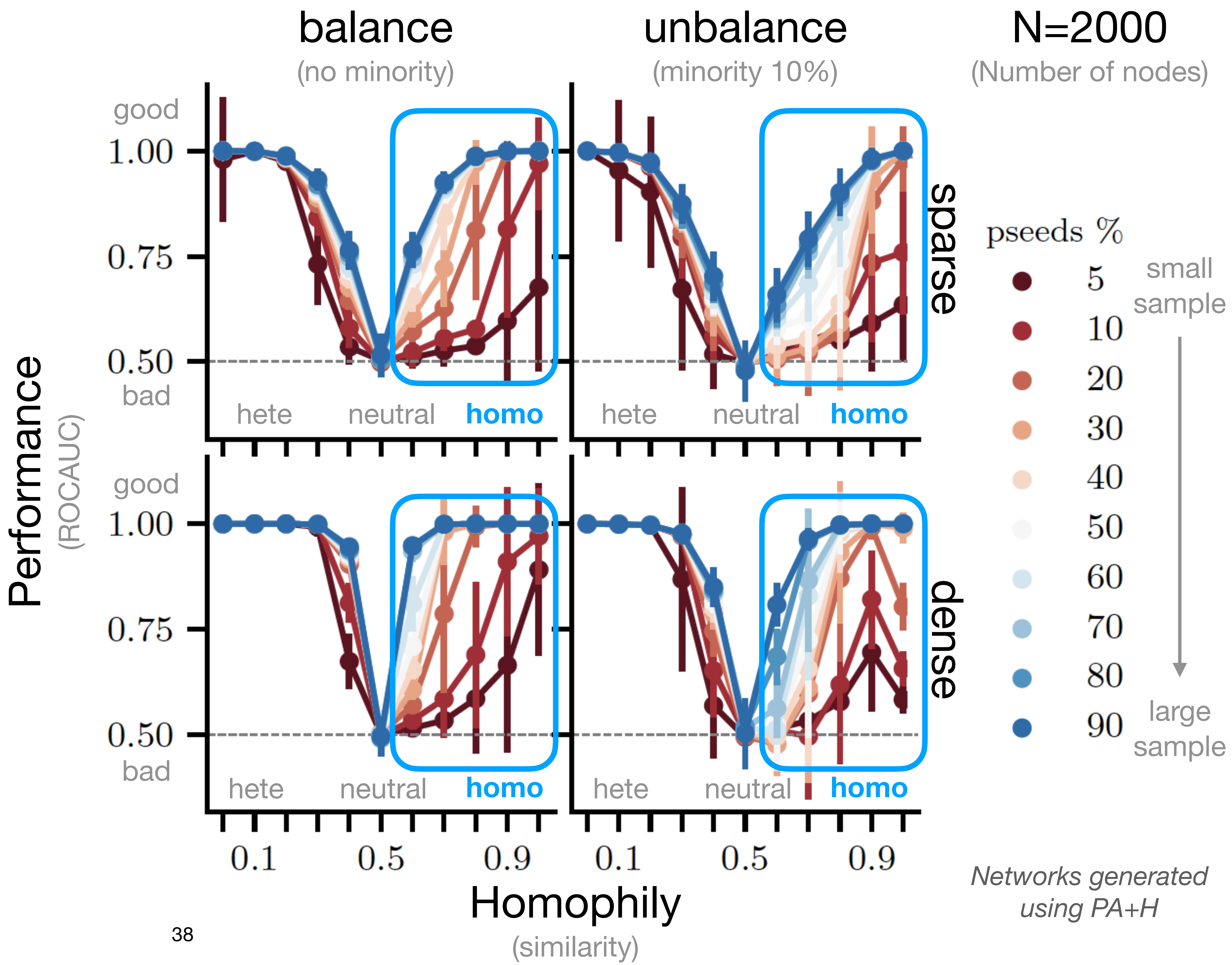
Networks generated using PA+H

Espín-Noboa et al. 2021

Network structure vs. classification performance

1. **Neutral networks** (H=0.5) cannot be classified better than a random classifier.

2. **Homophilic networks** (H>0.5) achieve lower performance than heterophilic networks when samples are small.



Networks generated using PA+H

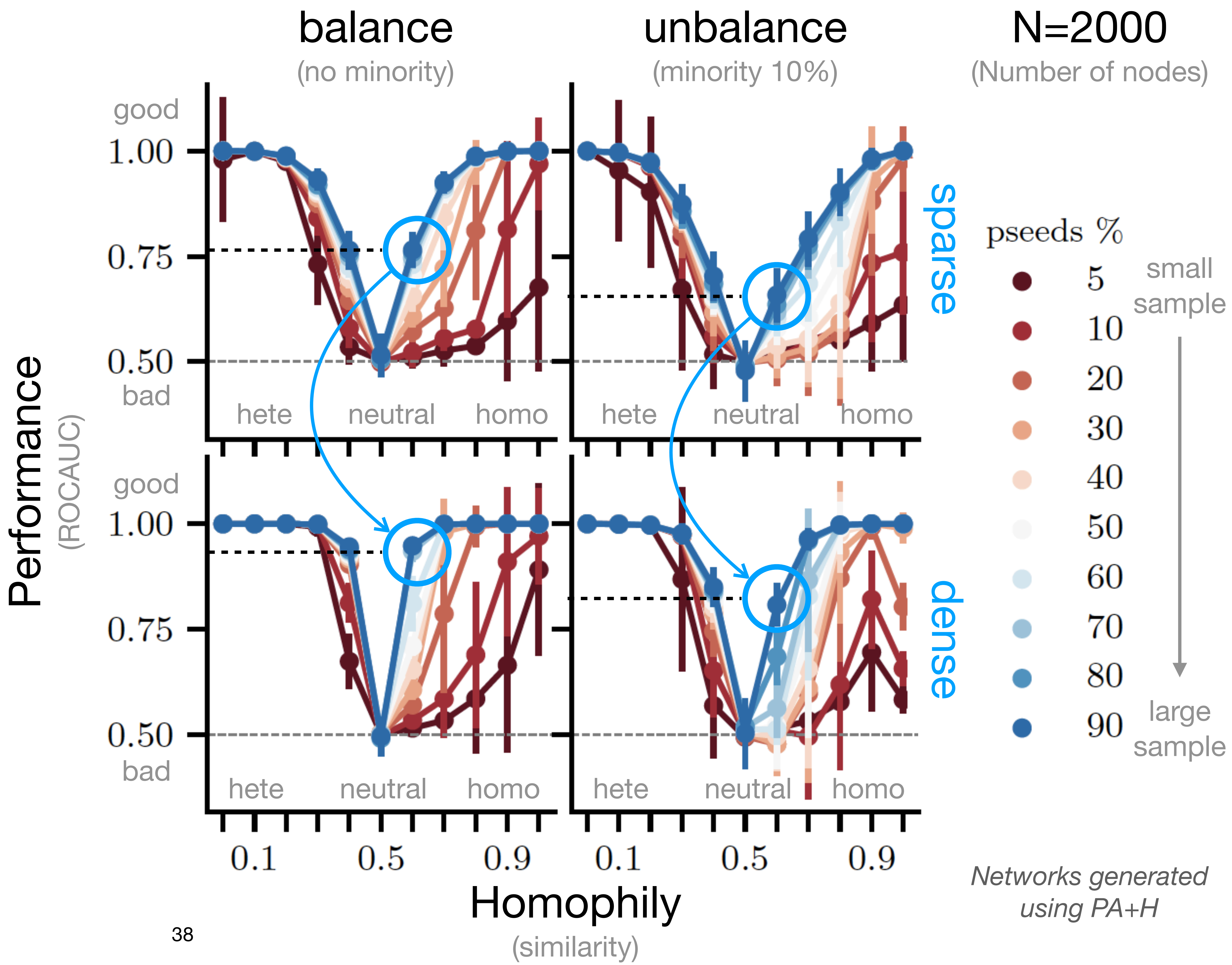
Espín-Noboa et al. 2021

Network structure vs. classification performance

1. **Neutral networks** (H=0.5) cannot be classified better than a random classifier.

2. **Homophilic networks** (H>0.5) achieve lower performance than heterophilic networks when samples are small.

3. **Denser networks** achieve higher performance compared to sparse networks.



Networks generated using PA+H

Espín-Noboa et al. 2021

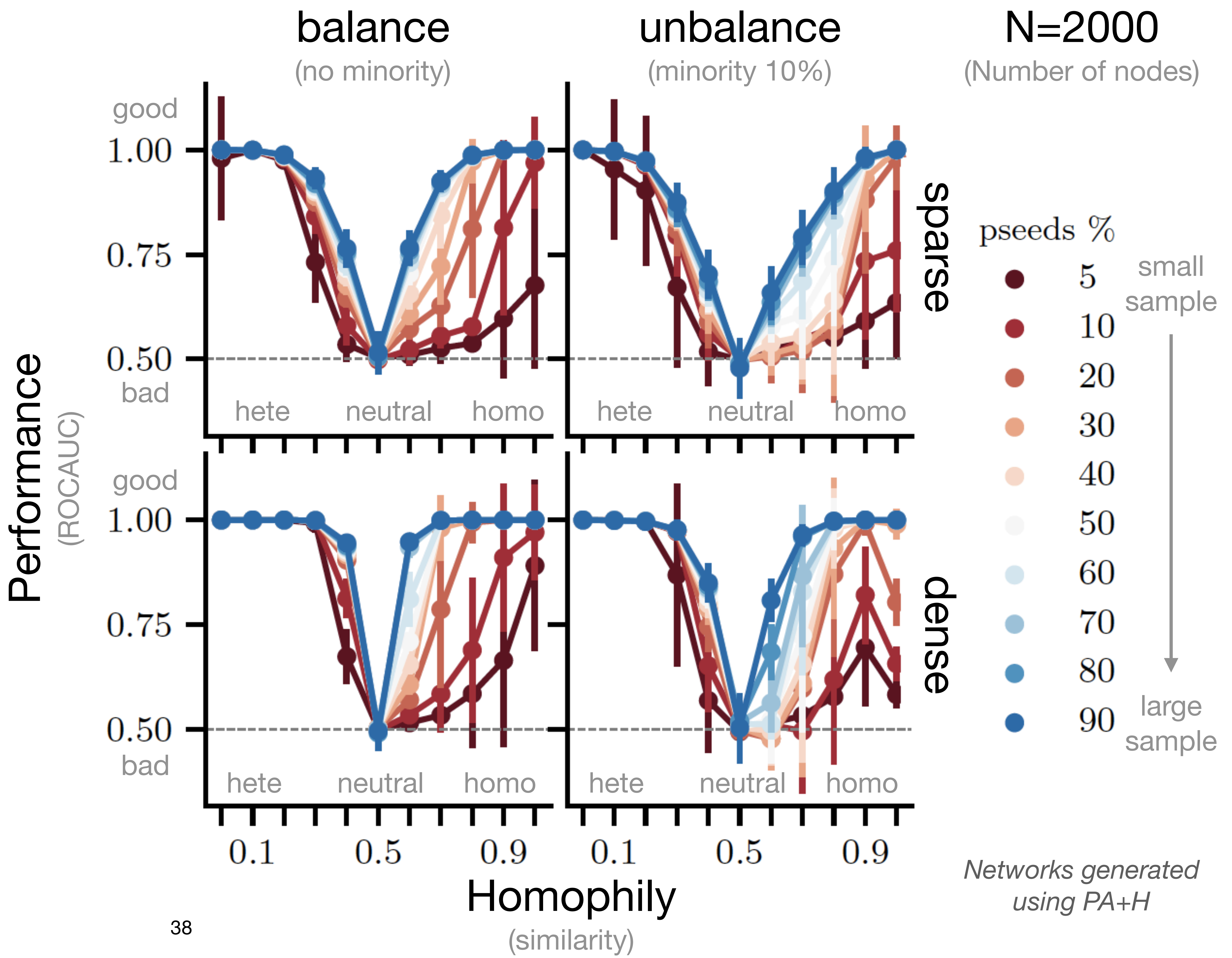
Network structure vs. classification performance

1. **Neutral networks** (H=0.5) cannot be classified better than a random classifier.

2. **Homophilic networks** (H>0.5) achieve lower performance than heterophilic networks when samples are small.

3. **Denser networks** achieve higher performance compared to sparse networks.

4. **Network size** mainly affects ROCAUC variance. Larger networks produce more stable results. (not shown here)



Networks generated using PA+H

Espín-Noboa et al. 2021

So what?

A real-world scenario

So what?

A real-world scenario

Real-world social network
(people with attributes)

Social Network
Mechanism of edge formation:
Homophily & Preferential Attachment

So what?

A real-world scenario

Real-world <u>social network</u>
(people with attributes)

What you get
(people with no attributes)

Social Network
Mechanism of edge formation:
Homophily & Preferential Attachment

So what?

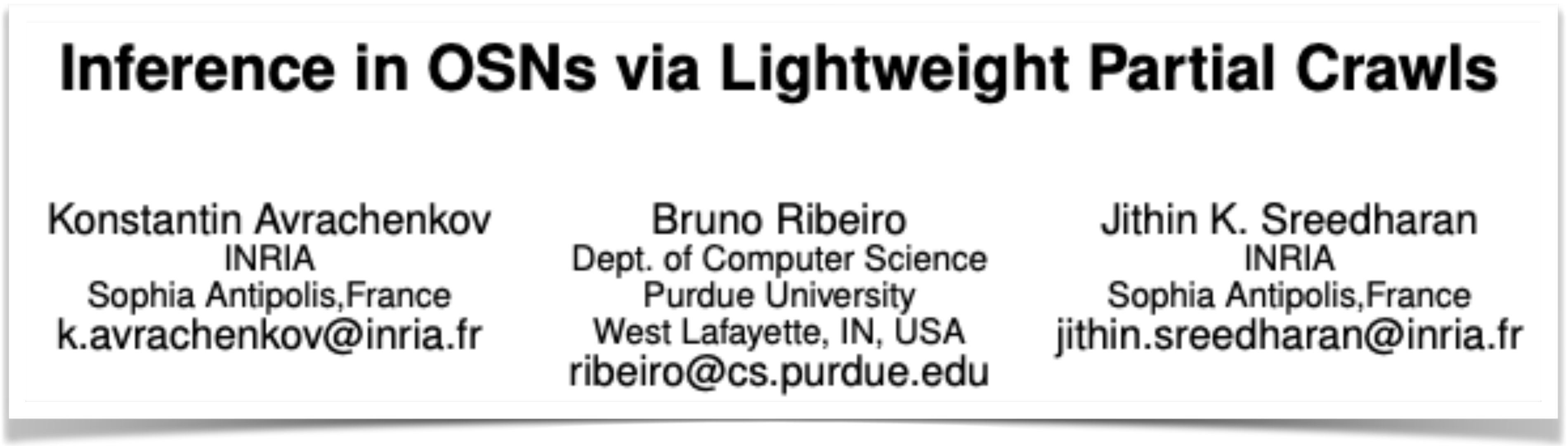
A real-world scenario

Real-world social network
(people with attributes)

What you get
(people with no attributes)

Social Network
Mechanism of edge formation:
Homophily & Preferential Attachment

1. Identify network structure



Inference in OSNs via Lightweight Partial Crawls

Konstantin Avrachenkov
INRIA
Sophia Antipolis,France
k.avrachenkov@inria.fr

Bruno Ribeiro
Dept. of Computer Science
Purdue University
West Lafayette, IN, USA
ribeiro@cs.purdue.edu

Jithin K. Sreedharan
INRIA
Sophia Antipolis,France
jithin.sreedharan@inria.fr

Class balance
B=0.3

Homophily
H=0.8

So what?

A real-world scenario

Real-world social network
(people with attributes)

What you get
(people with no attributes)

Social Network
Mechanism of edge formation:
Homophily & Preferential Attachment

1. Identify network structure

Inference in OSNs via Lightweight Partial Crawls

Konstantin Avrachenkov
INRIA
Sophia Antipolis,France
k.avrachenkov@inria.fr

Bruno Ribeiro
Dept. of Computer Science
Purdue University
West Lafayette, IN, USA
ribeiro@cs.purdue.edu

Jithin K. Sreedharan
INRIA
Sophia Antipolis,France
jithin.sreedharan@inria.fr

Class balance
B=0.3

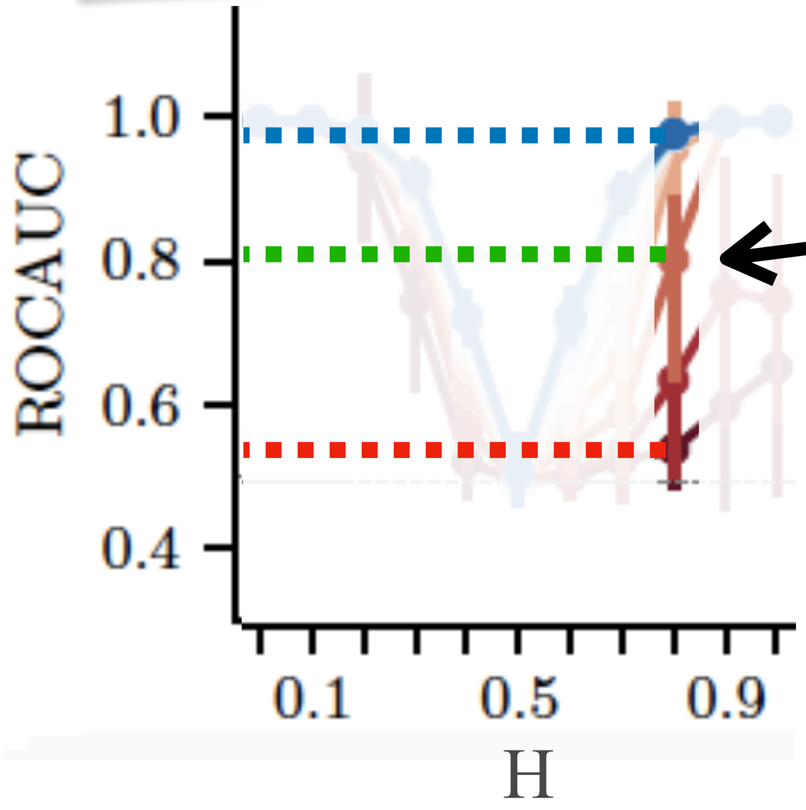
Homophily
H=0.8

2. Identify ROCAUC range for that network

RESEARCH

Explaining Classification Performance and Bias via Network Structure and Sampling Technique

Lisette Espin-Noboa, Fariba Karimi, Bruno Ribeiro, Kristina Lerman and Claudia Wagner



At least 20% of nodes in training sample to achieve ROCAUC > 0.8

Real-world (empirical) networks

Model fitting

Real-world (empirical) networks

Model fitting

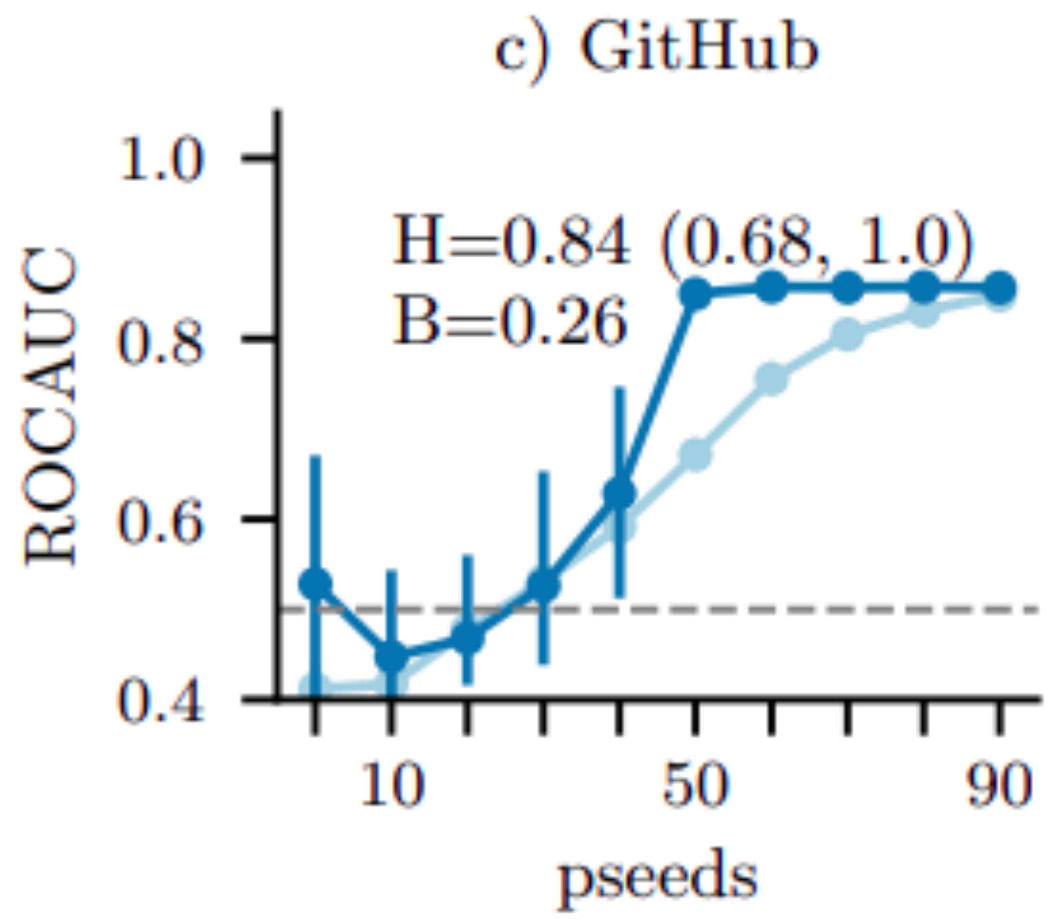
With symmetric homophily
$H_{mm} = H_{MM}$

Real-world (empirical) networks

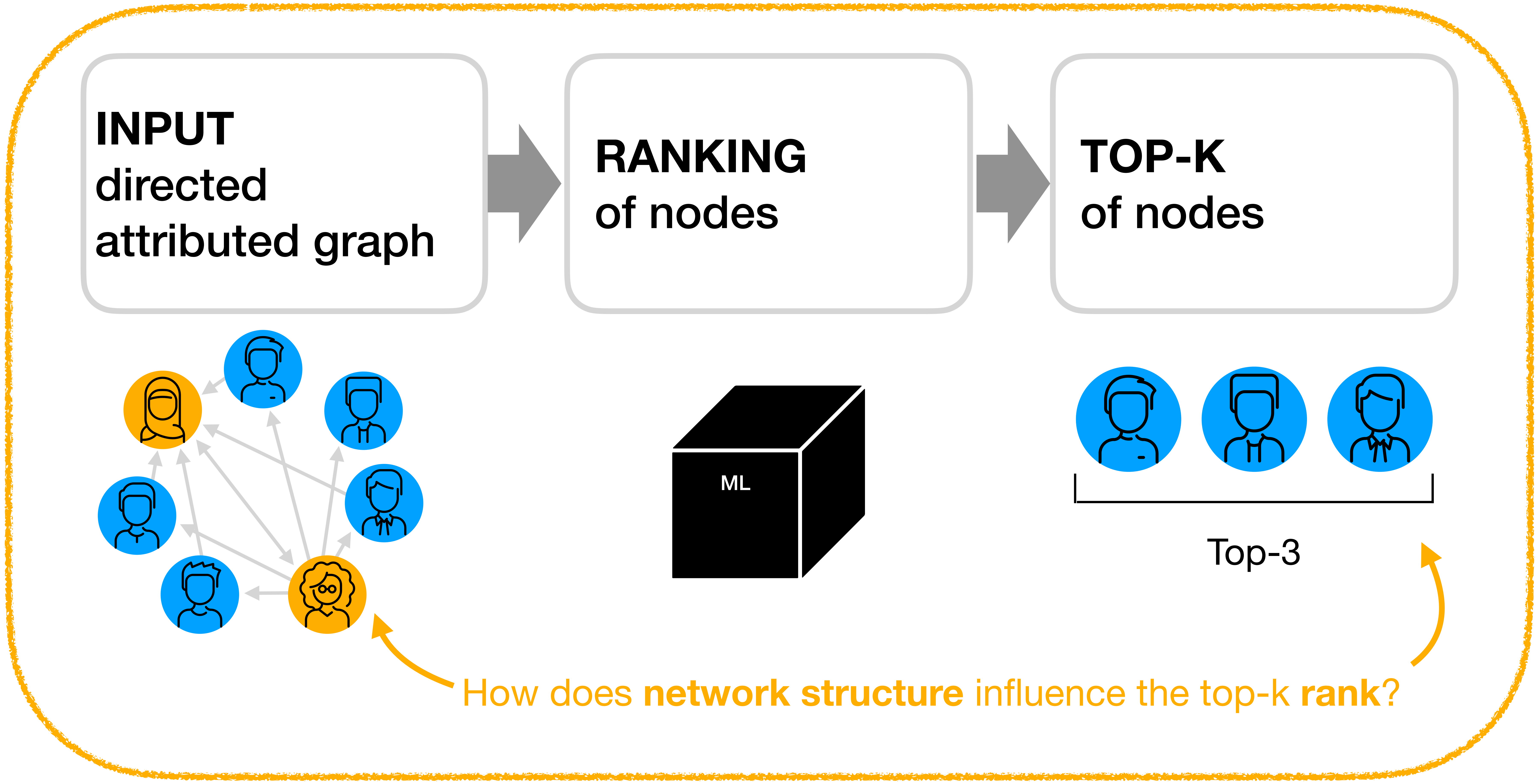
Model fitting

With symmetric homophily
$H_{mm} = H_{MM}$

With asymmetric homophily
$H_{mm} \neq H_{MM}$



Evaluation benchmarks in “ranking inequalities”



Espín-Noboa et al. SciRep’2022

Planets of disparity

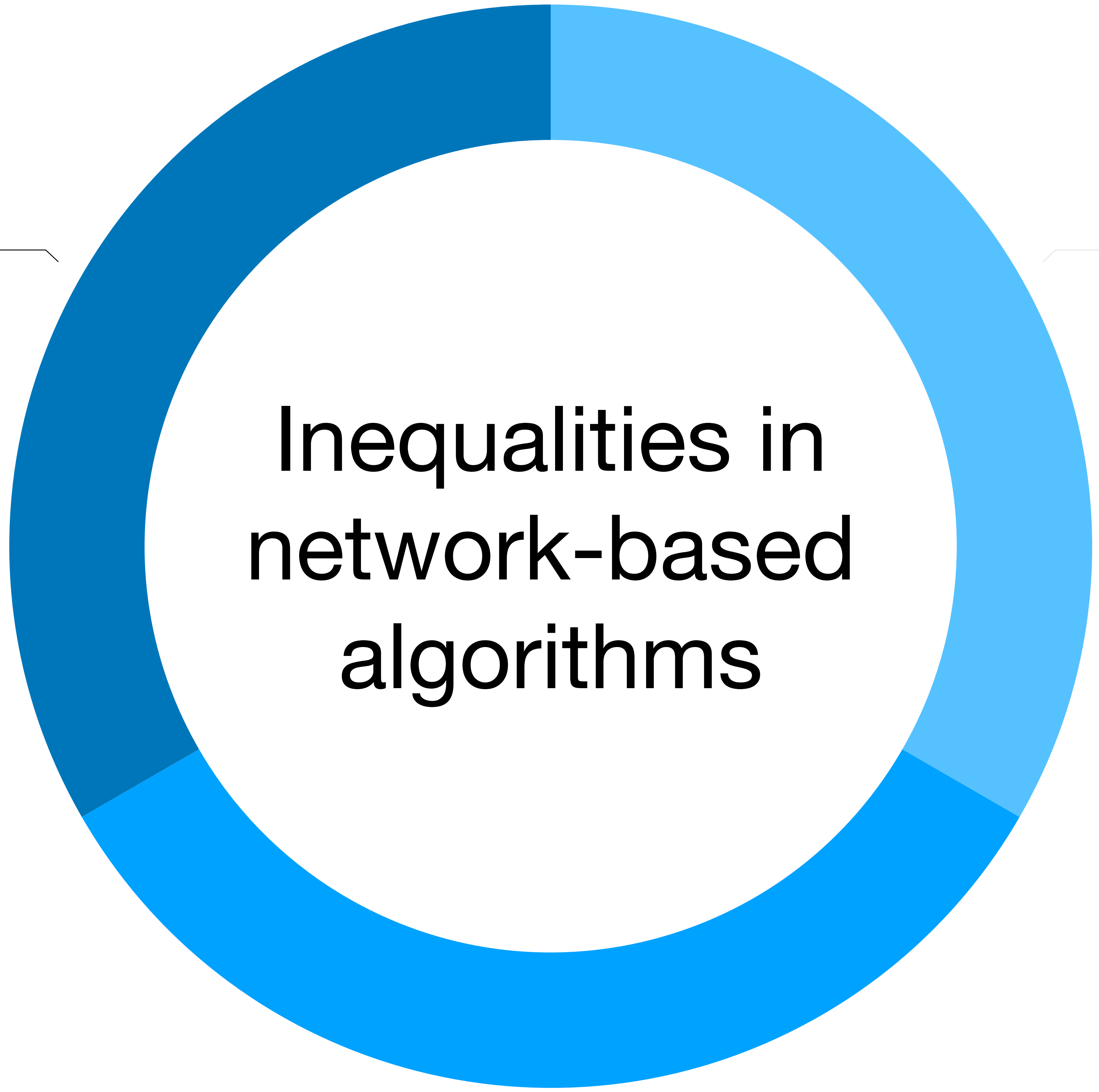
vis.csh.ac.at/planets-of-disparity

Homophily 0.5
Activity 1.5
x 10
% of Cats 10%
Edge density 0.01
x 90
Homophily 0.8
Activity 1.5
Inequality
Global Gini 0.43
Gini 0.3 in top 70 %
?
1
0.5
0
5
top-k%
100
Virtual planet #0010105081515
?
Inequality
moderate
Inequity
under-represented
Show Top 70%
from rankings
?
Rank
Network
Inequity
Mean error -0.07
8 % of cats in top 70 %
?
100
50
0
% of cats on planet
5
top-k%
100


Homophily 0.5
Activity 1.5
x 10
% of Cats 10%
Edge density 0.01
x 90
Homophily 0.8
Activity 1.5
Inequality Global Gini 0.43
Gini 0.3 in top 70 %
?
1
0.5
0
5
top-k%
100
Virtual planet #0010105081515
?
Inequality moderate
Inequity under-represented
Show Top 70% from rankings
?
Rank
Network
1 3 3 4 4 5 5 6 6 7 8 8
Inequity Mean error -0.07
8 % of cats in top 70 %
?
100
50
0
% of cats on planet
5
top-k%
100

3.

Mitigating inequalities on networks



Understanding inequalities on networks

Identifying inequalities on networks

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network
  - Control for data biases and adjust / calibrate the outputs of algorithms

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network
  - Control for data biases and adjust / calibrate the outputs of algorithms
    - Non-discrimination between groups

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network
  - Control for data biases and adjust / calibrate the outputs of algorithms
    - Non-discrimination between groups
- **Without algorithmic interventions**

We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network
  - Control for data biases and adjust / calibrate the outputs of algorithms
    - Non-discrimination between groups
- **Without algorithmic interventions**
  - Changing the structure of networks strategically

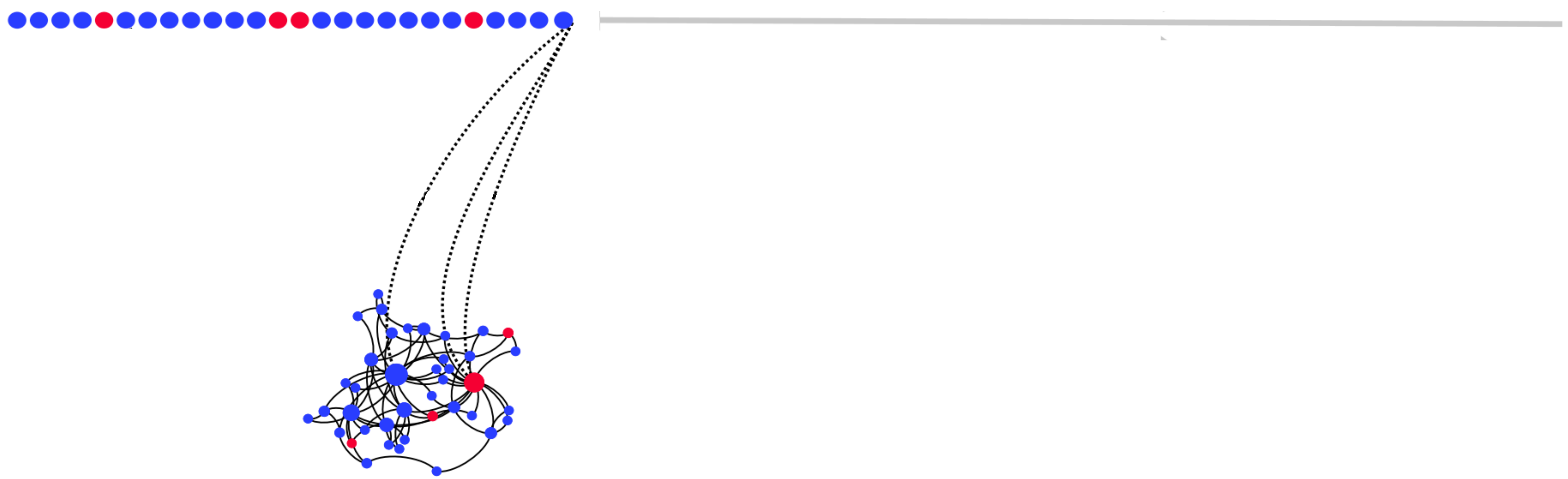
We should leverage the benchmarks

By identifying the conditions that lead to good outcomes

- **With algorithmic interventions**
  - Recommend new connections to shape the structure of the network
  - Control for data biases and adjust / calibrate the outputs of algorithms
    - Non-discrimination between groups
- **Without algorithmic interventions**
  - Changing the structure of networks strategically
    - It can be dangerous in the presence of bad actors!

PHASE 1

Network growth under PA+H
Homophily h1
fraction of minority m1



network at point of intervention
10% minority
heterophilic

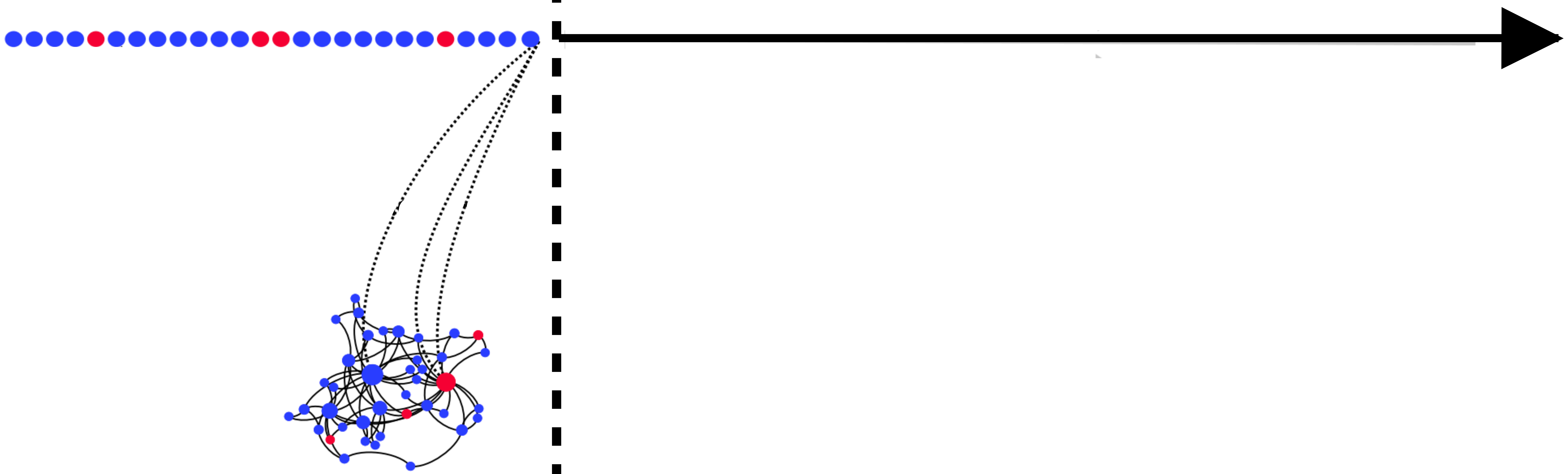
Neuhäuser, et al. 2023

PHASE 1

Network growth under PA+H
Homophily h1
fraction of minority m1

PHASE 2

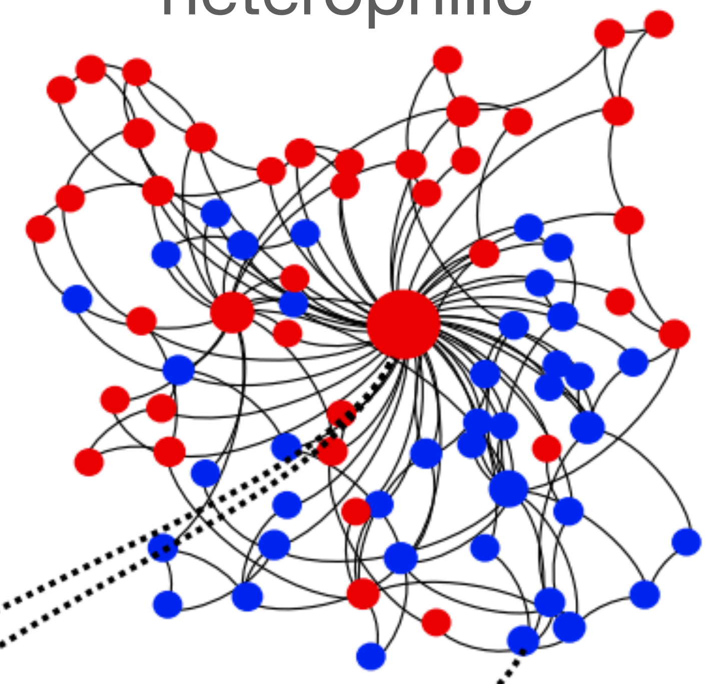
Intervention
Homophily h2
fraction of minority m2



network at point of intervention
10% minority
heterophilic

Neuhäuser, et al. 2023

final network
50% minority
heterophilic



PHASE 1

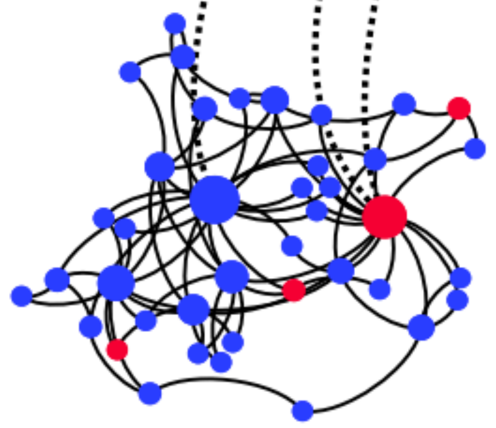
Network growth under PA+H
Homophily h1
fraction of minority m1

PHASE 2

Intervention
Homophily h2
fraction of minority m2

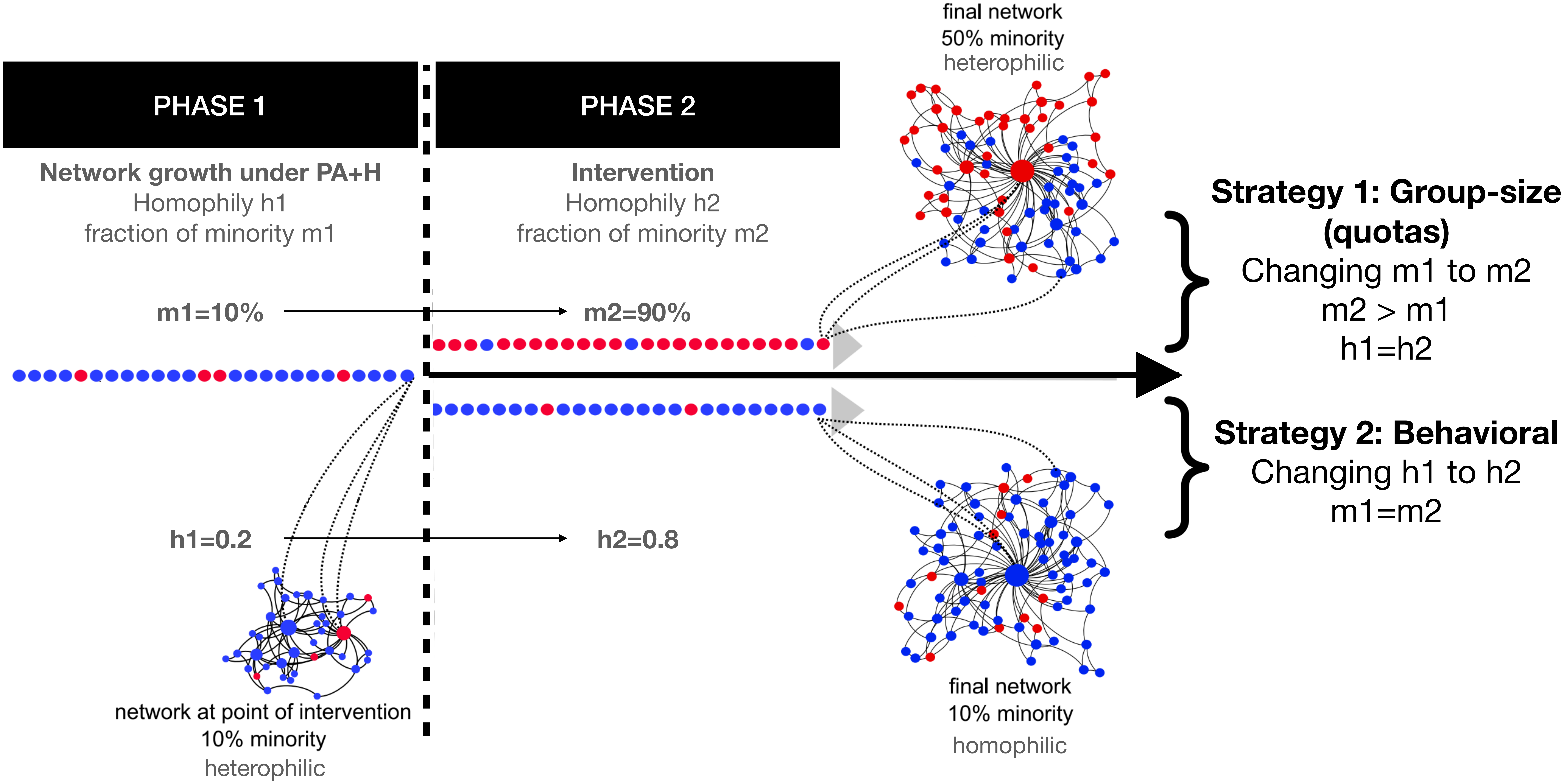
m1=10% → **m2=90%**

Strategy 1: Group-size (quotas)
Changing m1 to m2
m2 > m1
h1=h2



network at point of intervention
10% minority
heterophilic

Neuhäuser, et al. 2023



Neuhäuser, et al. 2023

The effects of changing homophily on minority representation in rankings

$h_1 = 0.1$ $h_1 = 0.5$ $h_1 = 0.9$

Heterophilic

m1 = m2 = 0.1

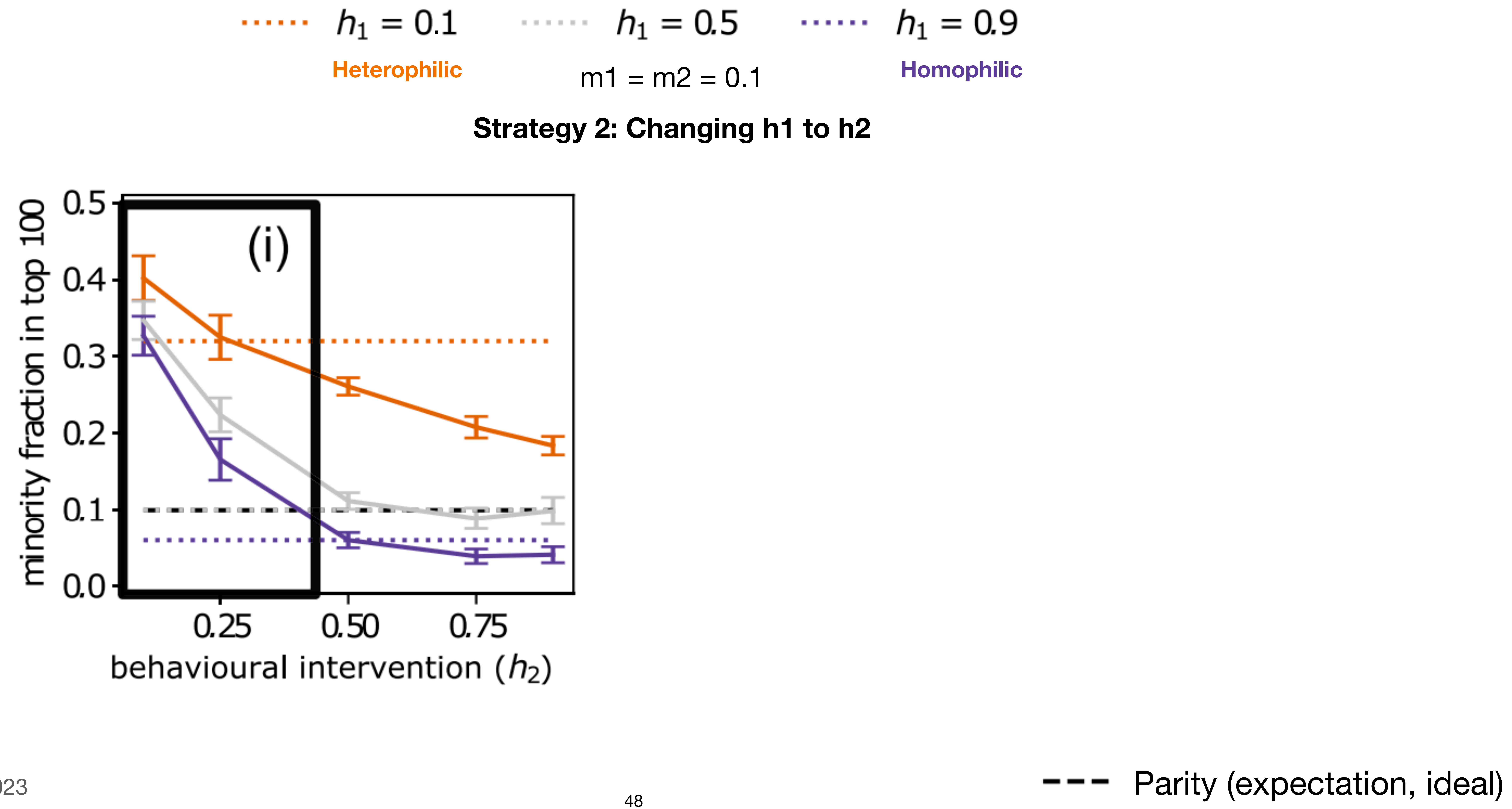
Homophilic

Strategy 2: Changing h1 to h2

Parity (expectation, ideal)

Neuhäuser, et al. 2023

The effects of changing homophily on minority representation in rankings



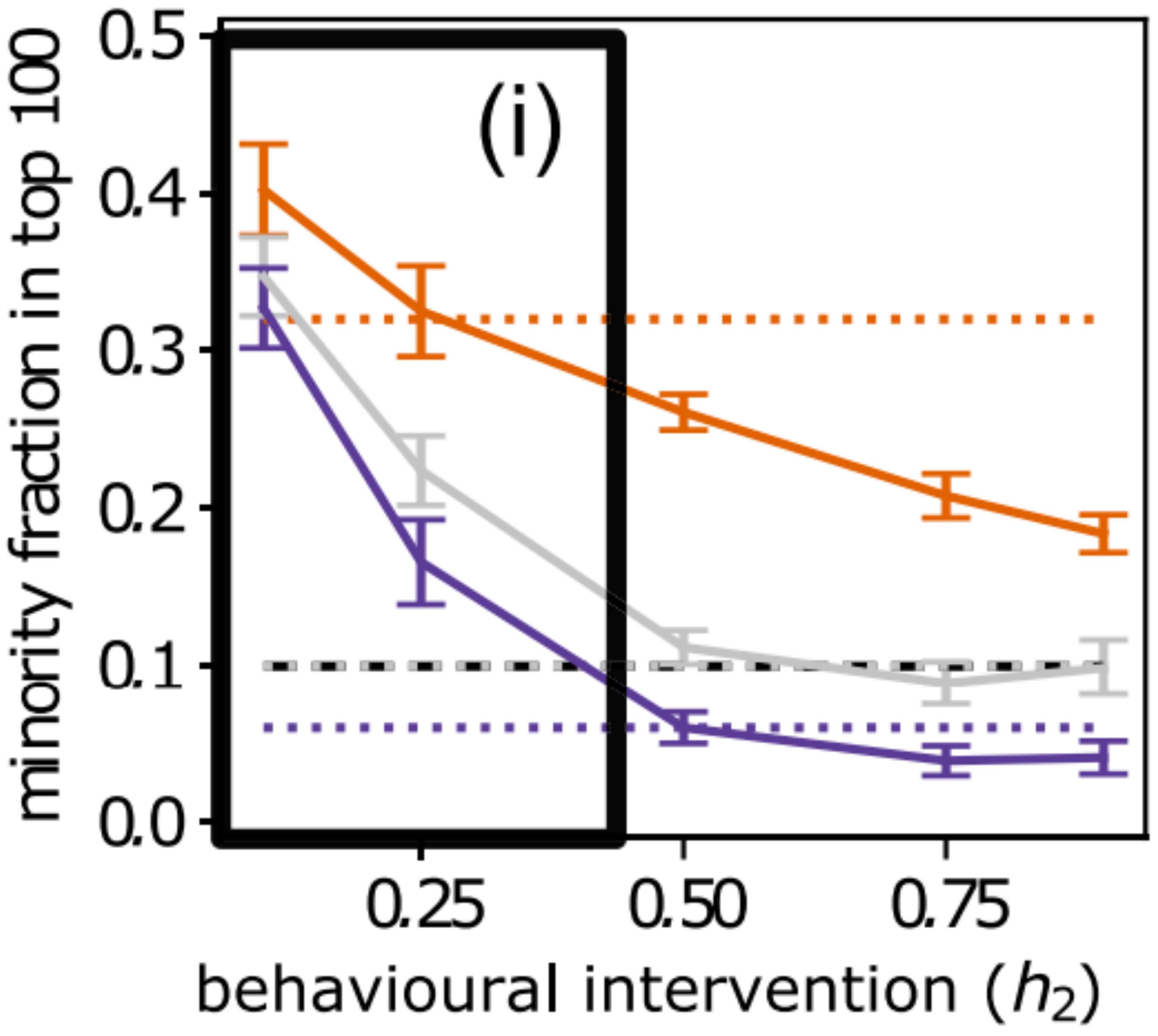
Neuhäuser, et al. 2023

The effects of changing homophily on minority representation in rankings

$h_1 = 0.1$ $h_1 = 0.5$ $h_1 = 0.9$

Heterophilic m1 = m2 = 0.1 Homophilic

Strategy 2: Changing h1 to h2

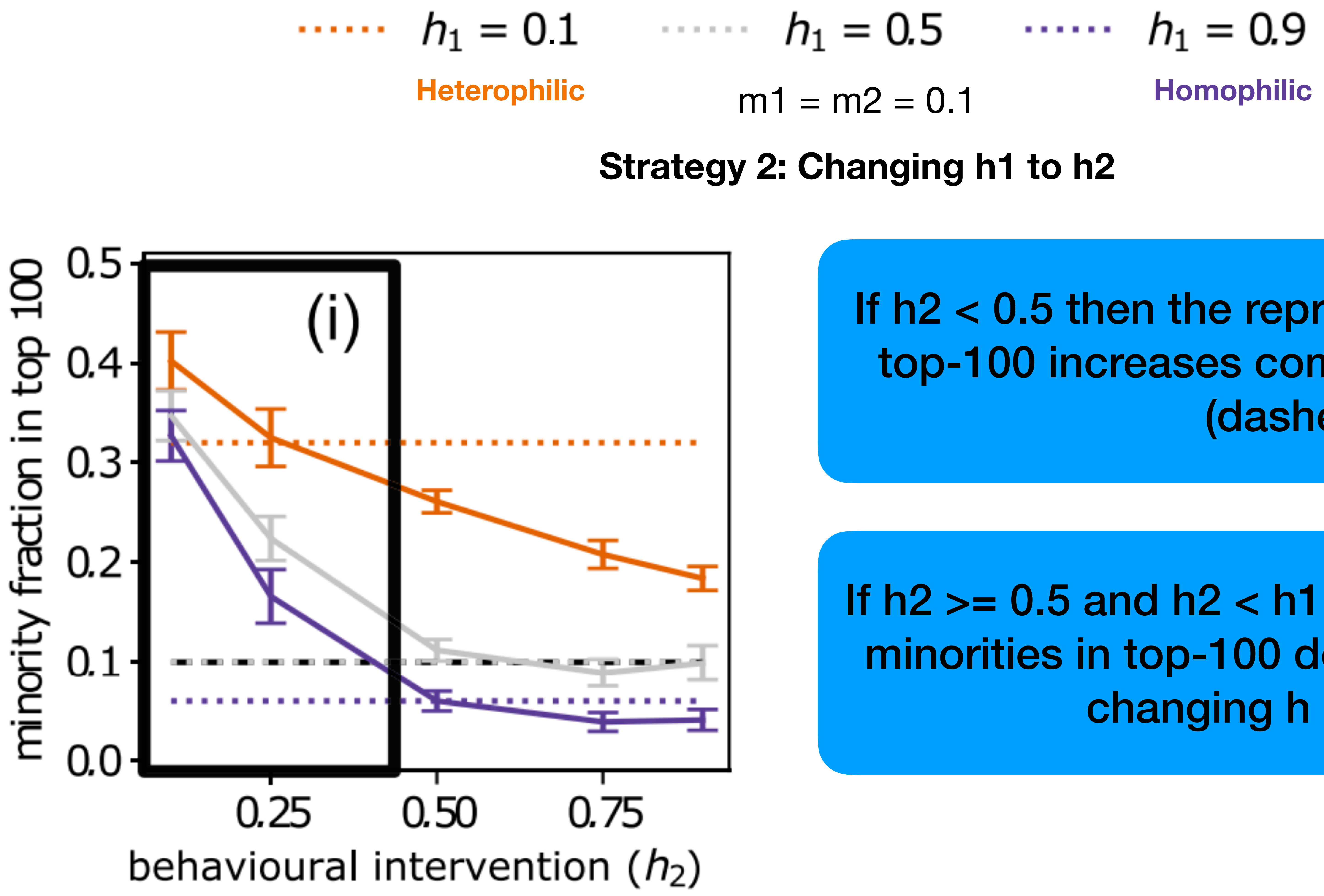


If h2 < 0.5 then the representation of minorities in top-100 increases compared to no change in h (dashed lines).

--- Parity (expectation, ideal)

Neuhäuser, et al. 2023

The effects of changing homophily on minority representation in rankings



If h2 < 0.5 then the representation of minorities in top-100 increases compared to no change in h (dashed lines).

If h2 >= 0.5 and h2 < h1 then the representation of minorities in top-100 decreases compared to no changing h (dashed lines).

--- Parity (expectation, ideal)

Neuhäuser, et al. 2023

The effects of "quotas" on minority representation in rankings

$h_1 = 0.1$ Heterophilic

$h_1 = 0.5$ h1 = h2

$h_1 = 0.9$ Homophilic

Strategy 1: Changing m1=0.1 to m2

Neuhäuser, et al. 2023

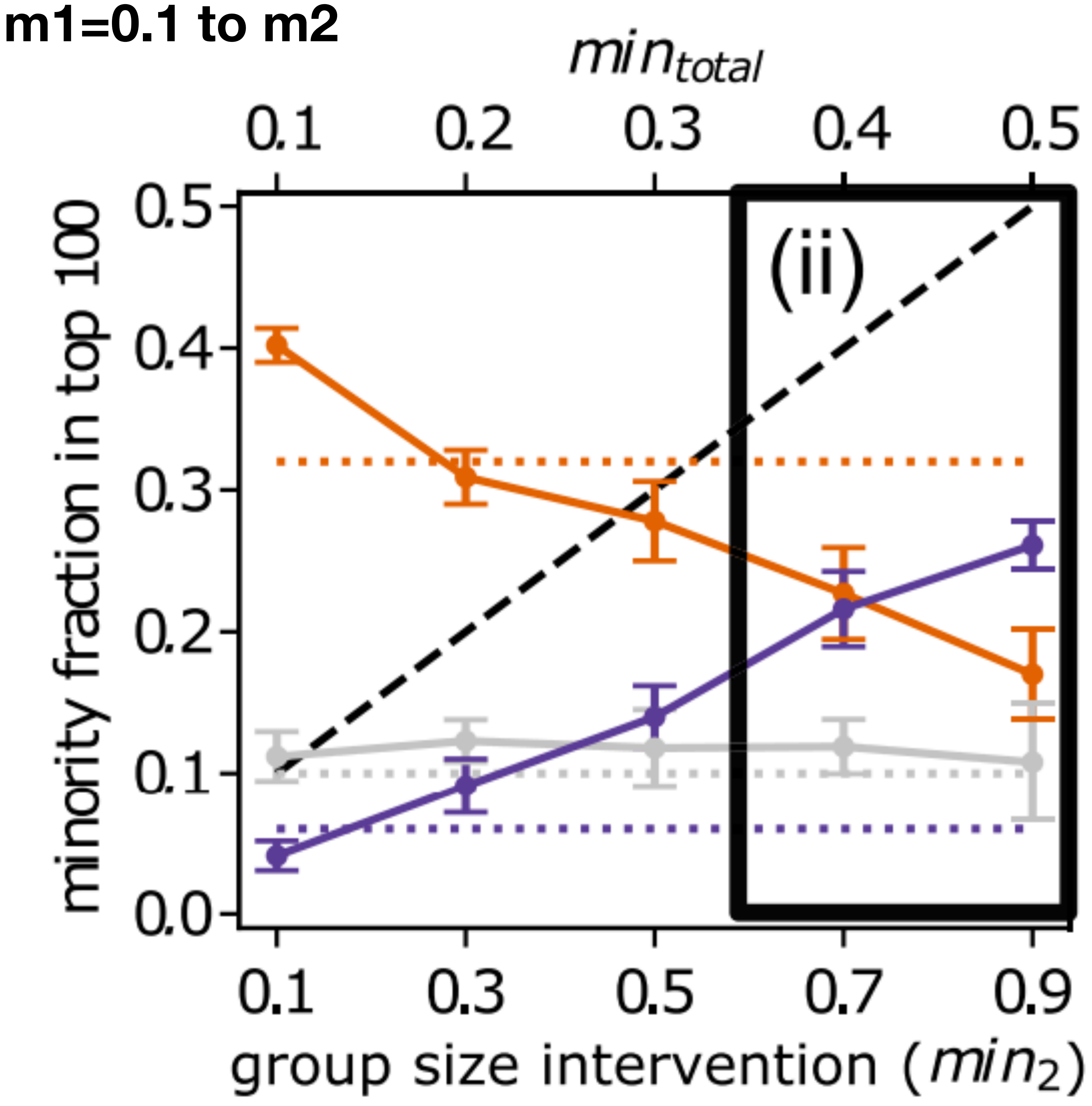
Parity (expectation, ideal)

min_{total} = min fraction in top100

The effects of “quotas” on minority representation in rankings

$h_1 = 0.1$ — Heterophilic
$h_1 = 0.5$ — h1 = h2
$h_1 = 0.9$ — Homophilic

Strategy 1: Changing m1=0.1 to m2



Parity (expectation, ideal)
mintotal = min fraction in top100

Neuhäuser, et al. 2023

The effects of "quotas" on minority representation in rankings

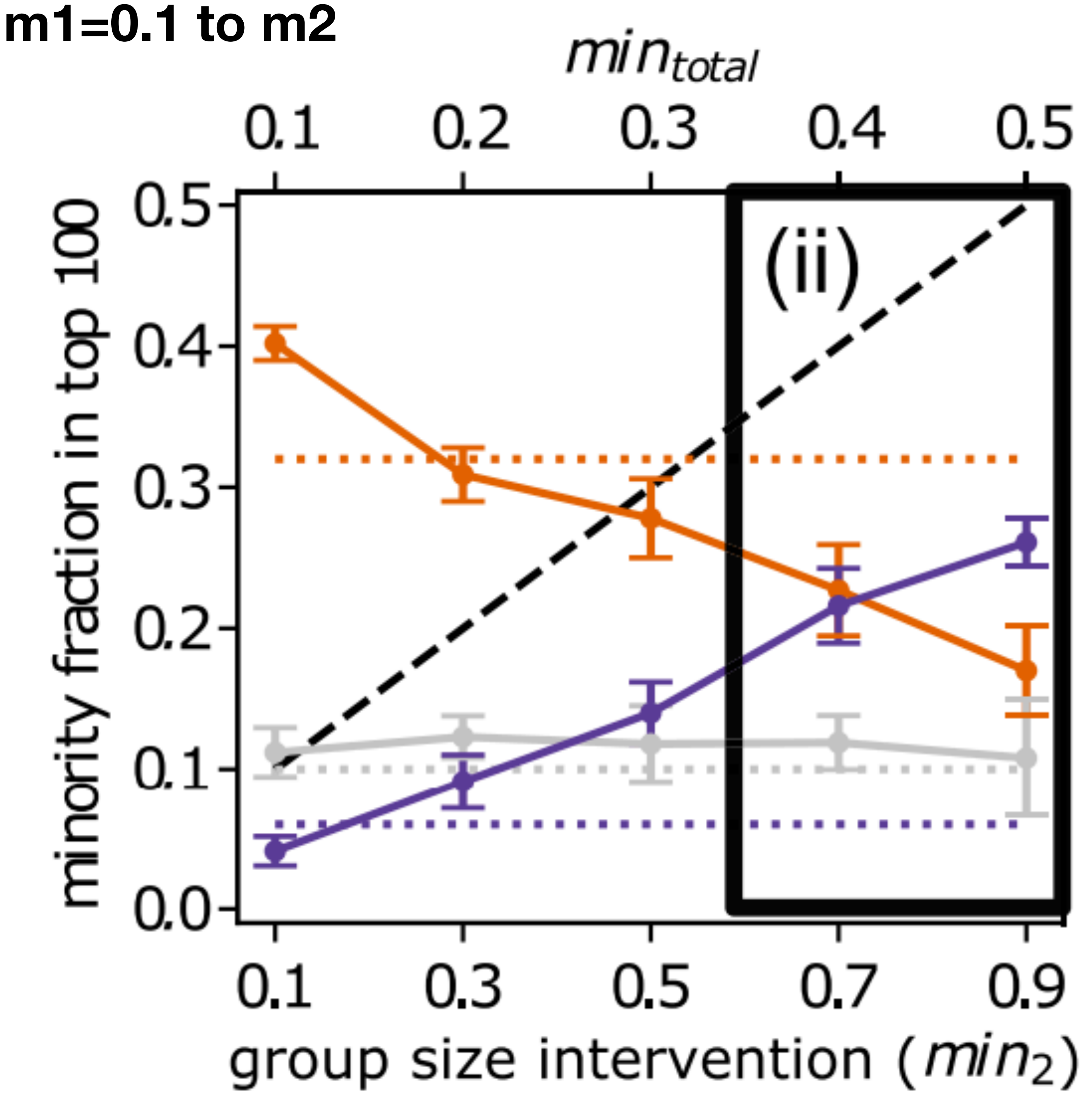
$h_1 = 0.1$ — Heterophilic

$h_1 = 0.5$ — h1 = h2

$h_1 = 0.9$ — Homophilic

Strategy 1: Changing m1=0.1 to m2

In the homophilic setting (purple h=0.9), **the minority representation grows** with the final minority size (min_{total}) and quota m2.



Parity (expectation, ideal)

min_{total} = min fraction in top100

Neuhäuser, et al. 2023

The effects of “quotas” on minority representation in rankings

$h_1 = 0.1$ — Heterophilic

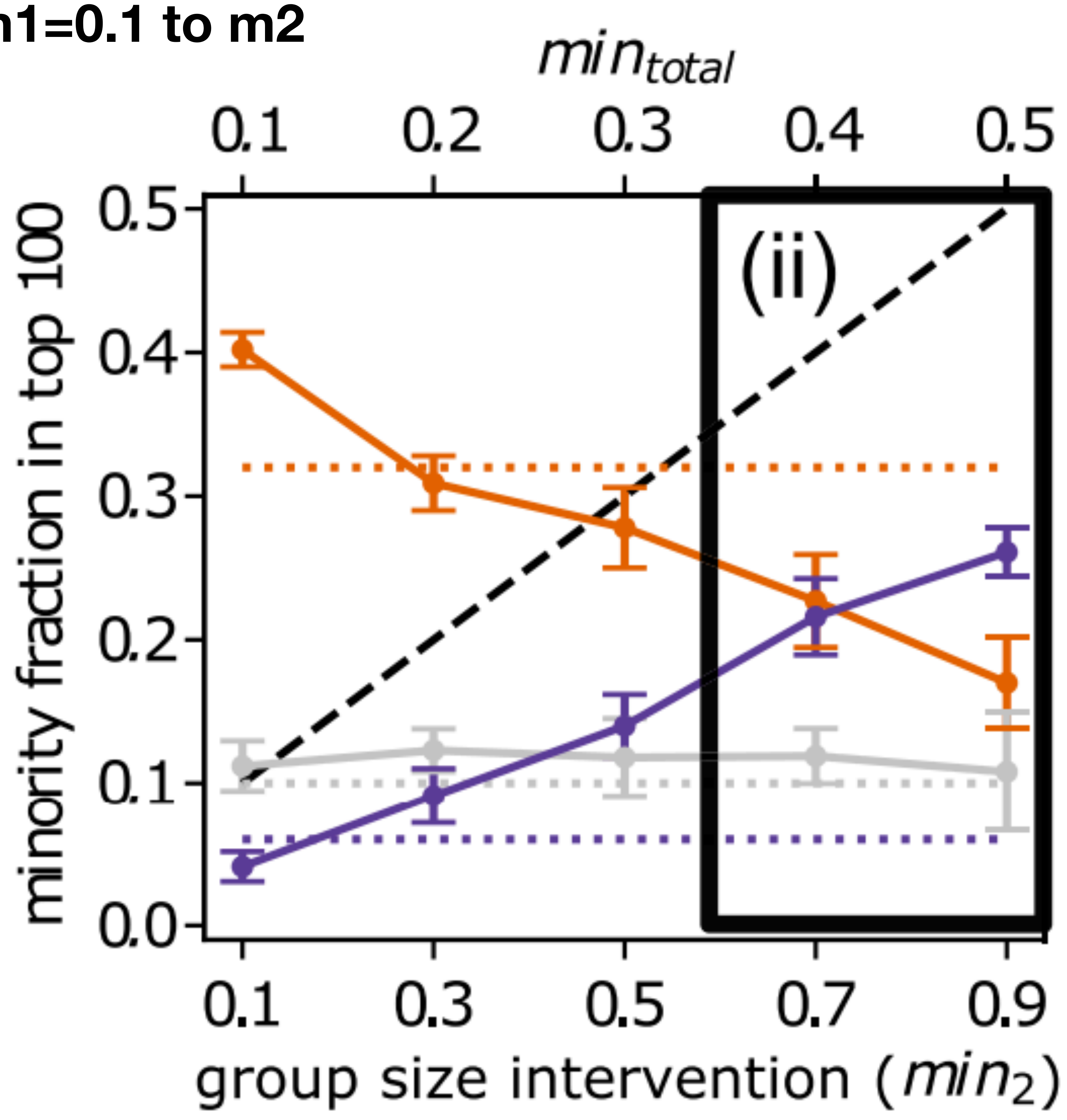
$h_1 = 0.5$ — h1 = h2

$h_1 = 0.9$ — Homophilic

Strategy 1: Changing m1=0.1 to m2

In the homophilic setting (purple h=0.9), **the minority representation grows** with the final minority size (min_{total}) and quota m2.

In the heterophilic setting (orange h=0.1), **the minority representation decreases** with the final minority size (min_{total}) and quota m2.



Parity (expectation, ideal)

min_{total} = min fraction in top100

Neuhäuser, et al. 2023

The effects of “quotas” on minority representation in rankings

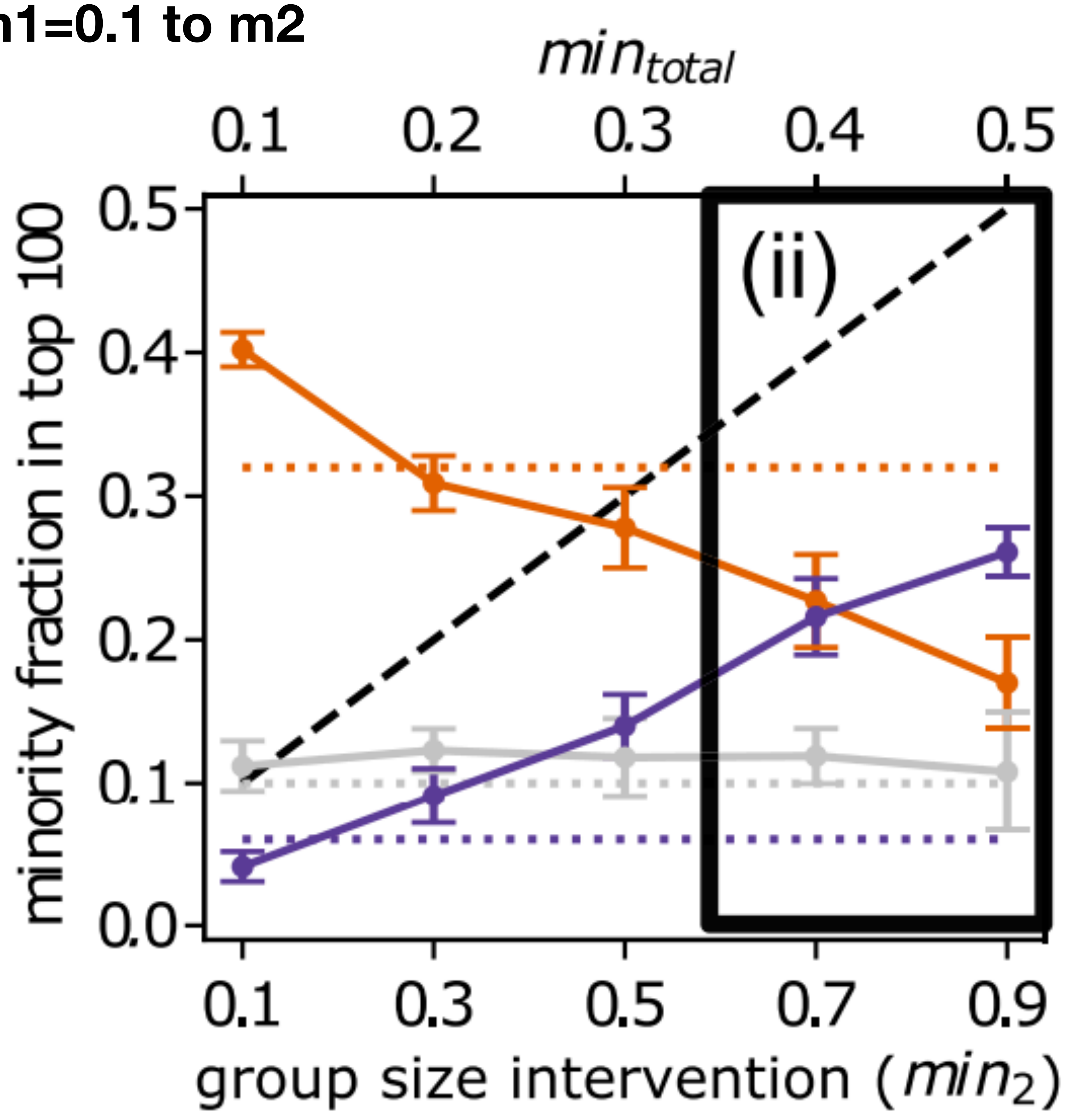
$h_1 = 0.1$ **Heterophilic** $h_1 = 0.5$ h1 = h2 $h_1 = 0.9$ **Homophilic**

Strategy 1: Changing m1=0.1 to m2

In the homophilic setting (purple h=0.9), **the minority representation grows** with the final minority size (min_{total}) and quota m2.

In the heterophilic setting (orange h=0.1), **the minority representation decreases** with the final minority size (min_{total}) and quota m2.

In the neutral case (grey h=0.5), we see no impact of a quota at all



Parity (expectation, ideal)
min_{total} = min fraction in top100

Neuhäuser, et al. 2023

Recap!

Inequalities on networks

Inequalities on networks

- Networks are biased because their nodes and edges are biased!

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
    - Groups are unbalanced

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
  - Groups are unbalanced
  - Mechanisms of edge formation are biased towards:

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
  - Groups are unbalanced
  - Mechanisms of edge formation are biased towards:
    - Popular nodes

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
  - Groups are unbalanced
  - Mechanisms of edge formation are biased towards:
    - Popular nodes
    - Similar nodes

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
  - Groups are unbalanced
  - Mechanisms of edge formation are biased towards:
    - Popular nodes
    - Similar nodes
  - Multiple mechanisms of edge formation can occur at the same time in one network

Inequalities on networks

- Networks are biased because their nodes and edges are biased!
  - Groups are unbalanced
  - Mechanisms of edge formation are biased towards:
    - Popular nodes
    - Similar nodes
  - Multiple mechanisms of edge formation can occur at the same time in one network
  - The **`netin`** python package offers multiple network models with homophily and preferential attachment

Inequalities on algorithms

Inequalities on algorithms

- Algorithms can be biased for multiple reasons

Inequalities on algorithms

- Algorithms can be biased for multiple reasons
  - The data is biased

Inequalities on algorithms

- Algorithms can be biased for multiple reasons
    - The data is biased
    - The algorithm is biased

Inequalities on algorithms

- Algorithms can be biased for multiple reasons
  - The data is biased
  - The algorithm is biased
  - The output is biased

Inequalities on algorithms

- Algorithms can be biased for multiple reasons
  - The data is biased
  - The algorithm is biased
  - The output is biased
  - Human assessment is biased

Inequalities on network-based algorithms

Inequalities on network-based algorithms

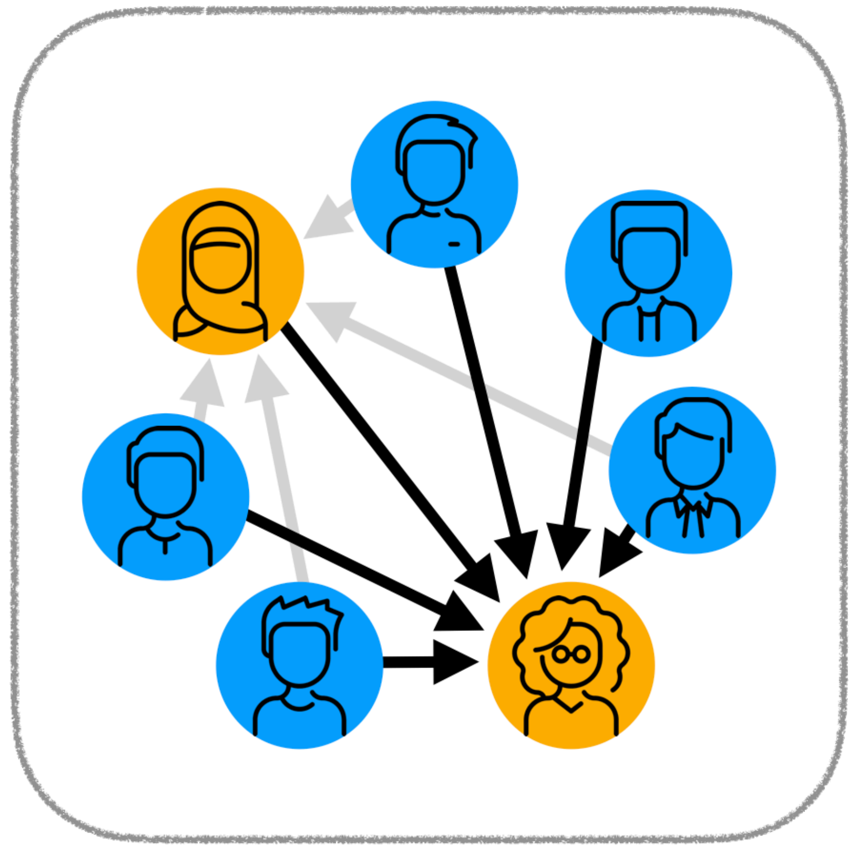
- Biases in network-based algorithms can be explained thanks to the structural nature of networks!

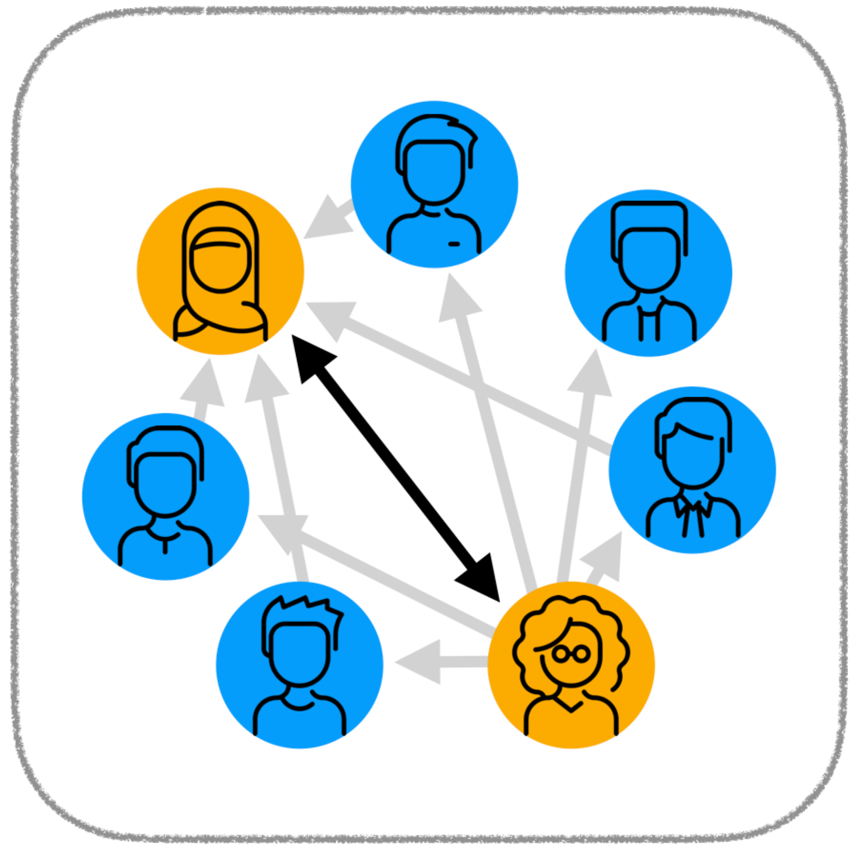
Inequalities on network-based algorithms

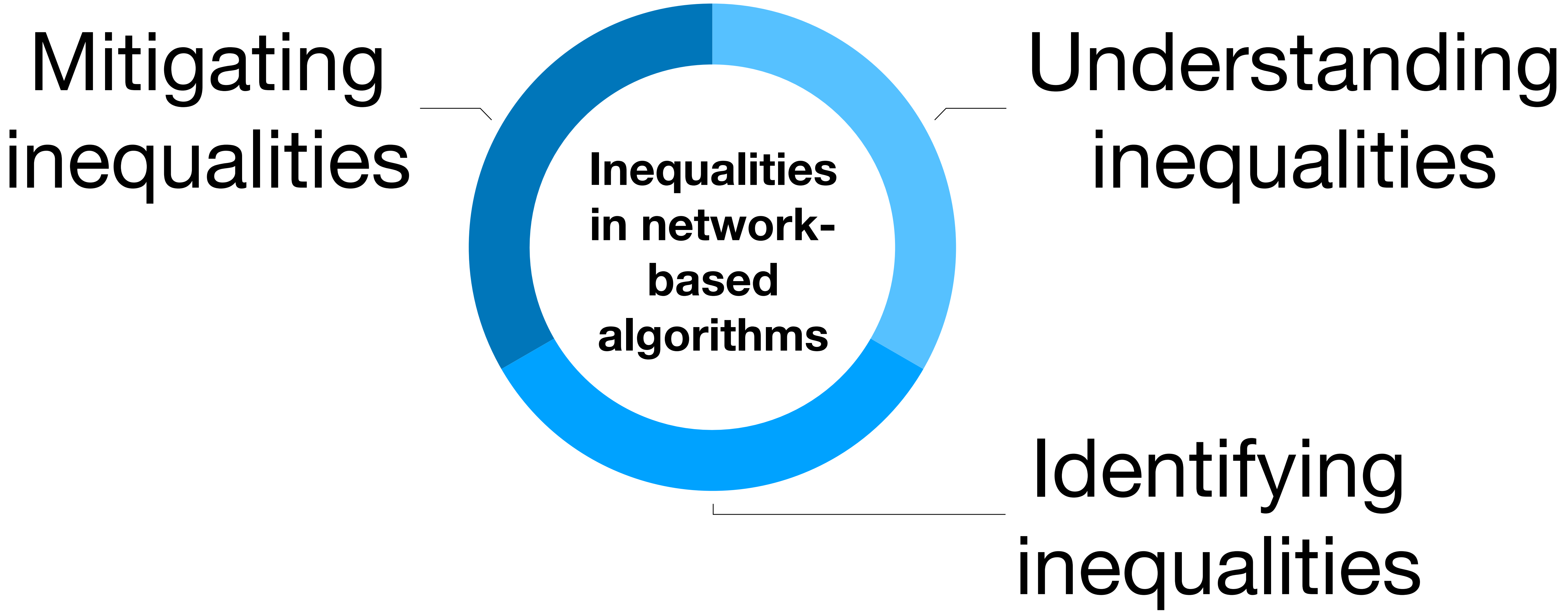
- Biases in network-based algorithms can be explained thanks to the structural nature of networks!
    - Certain mechanisms of edge formation can explain certain algorithmic biases

Inequalities on network-based algorithms

- Biases in network-based algorithms can be explained thanks to the structural nature of networks!
    - Certain mechanisms of edge formation can explain certain algorithmic biases
    - If we understand these correlations, we can propose better mitigation strategies to reduce inequalities







Inequality II.

Networks and Algorithms

Lisette Espín-Noboa | 31.May.2023 | TU Wien
espin@csh.ac.at
@lespin
www.lisetteespin.info

[Tutorial]
SOCIAL NETWORK MODELING AND APPLICATIONS
https://bit.ly/snma2023

Backups

Important properties

of social networks

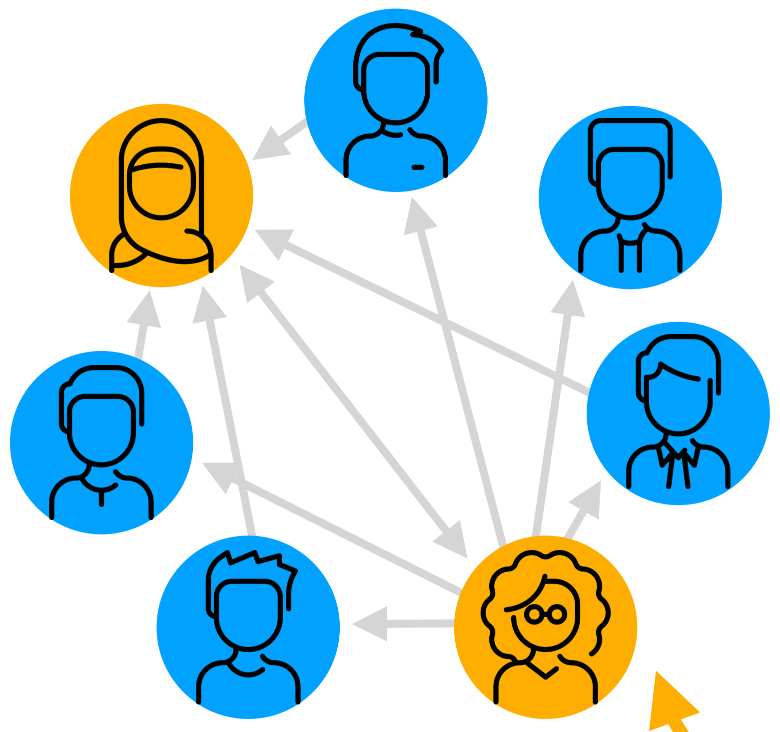
#	Graph	Node	Edge
1	Average and minimum degree (note: average degree does not make sense if you have power-law degree distribution)	Degree (in/out)	Weights
2	Degree distribution (the probability that a randomly selected node in the network has degree k)	Centrality (e.g,. PageRank)	Shortest path
3	Adjacency matrix / Mixing matrix	Clustering coefficient	Homophilous type (MM, Mm, mm, mM)
4	Sparsity (out of the total possible number of edges, how many actually exist)	Activity (high activity = high outdegree)	
5	Diameter		
6	Average Path Length		
7	Connected components		

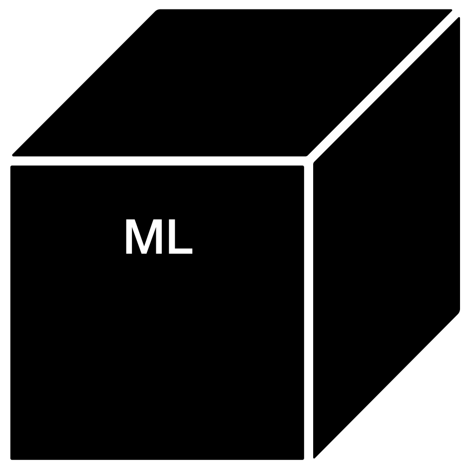
Evaluation benchmarks in “ranking inequalities”

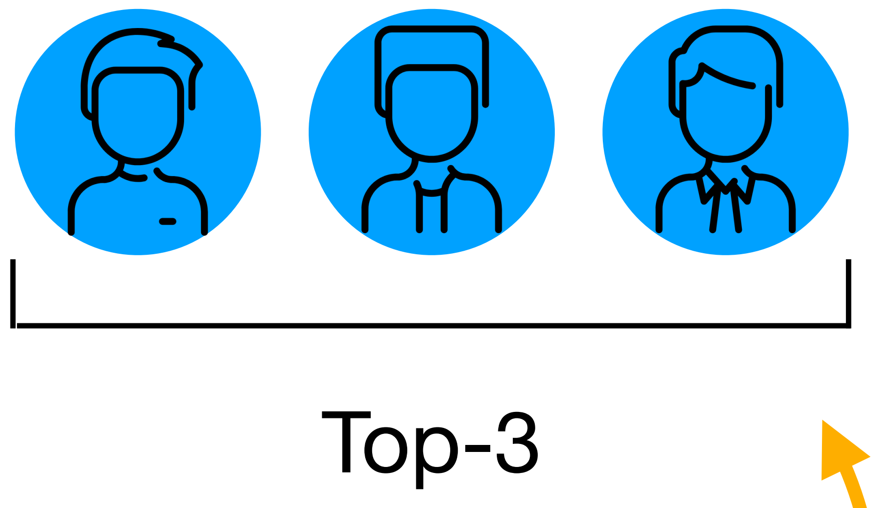
INPUT
directed
attributed graph

RANKING
of nodes

TOP-K
of nodes







How does **network structure** influence the top-k **rank**?

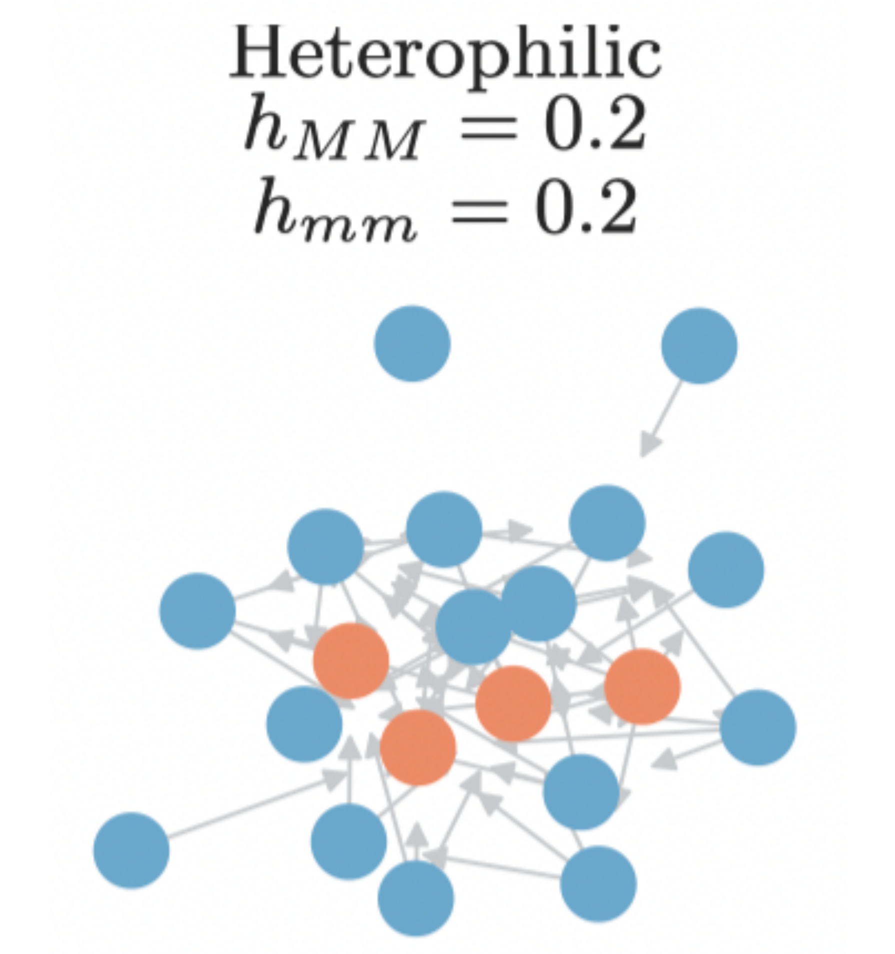
Espín-Noboa et al. SciRep’2022

Inequality, inequity, and disparity

Espín-Noboa et al. 2022

Inequality, inequity, and disparity

Given a network,

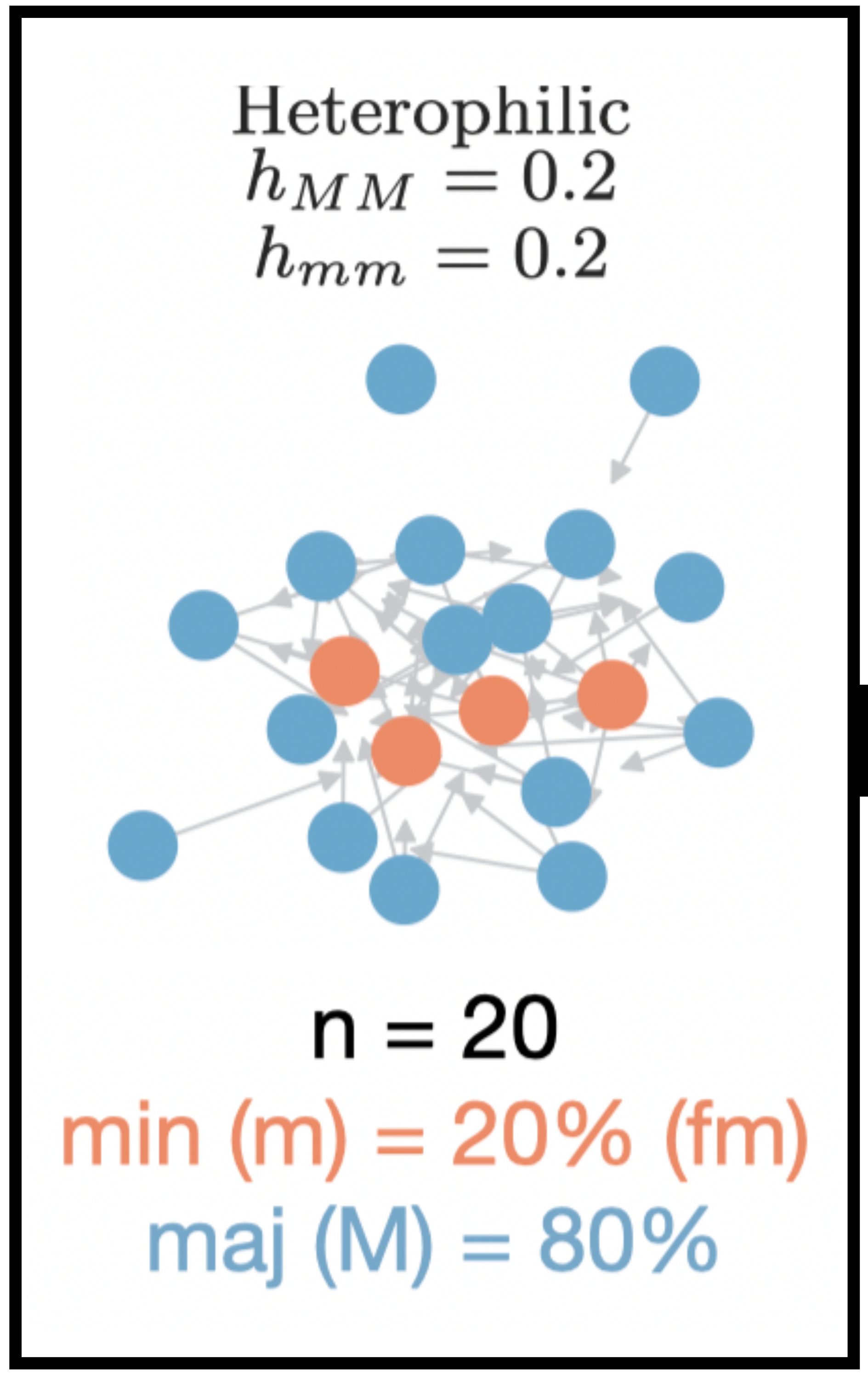


n = 20
min (m) = 20% (fm)
maj (M) = 80%

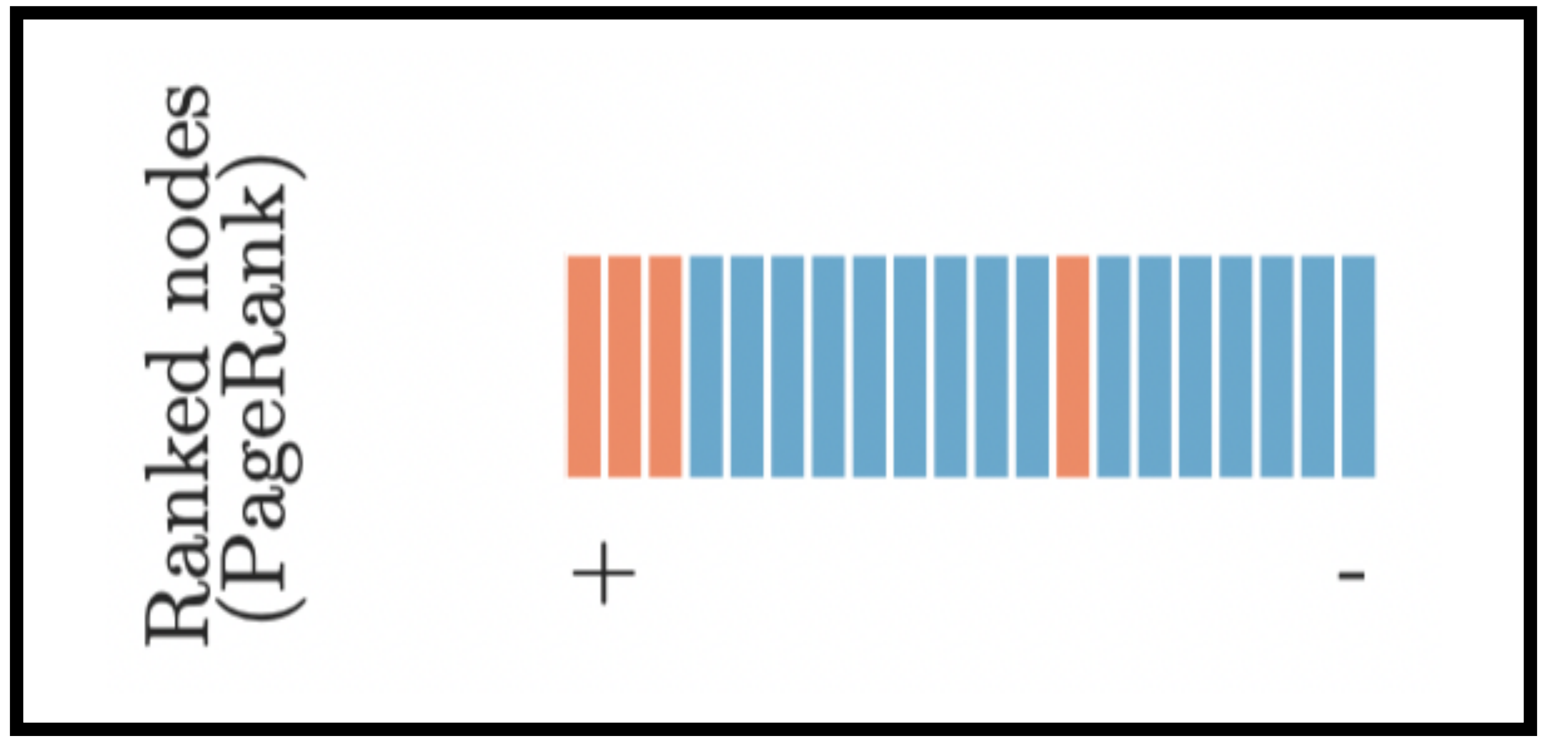
Espín-Noboa et al. 2022

Inequality, inequity, and disparity

Given a network,



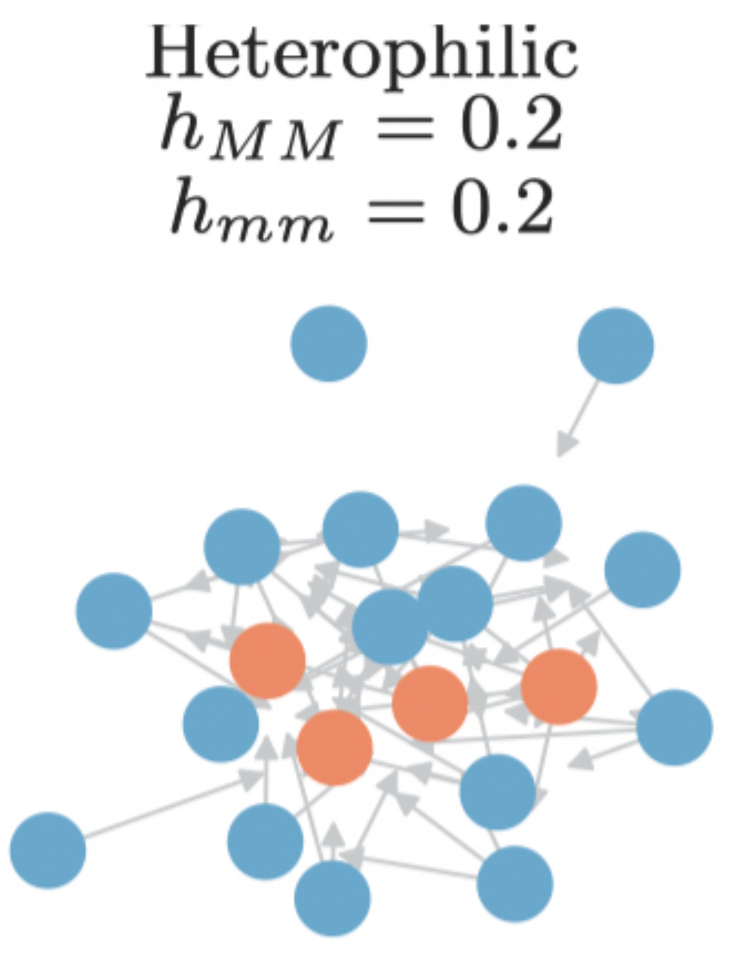
… and a ranking of its nodes



Espín-Noboa et al. 2022

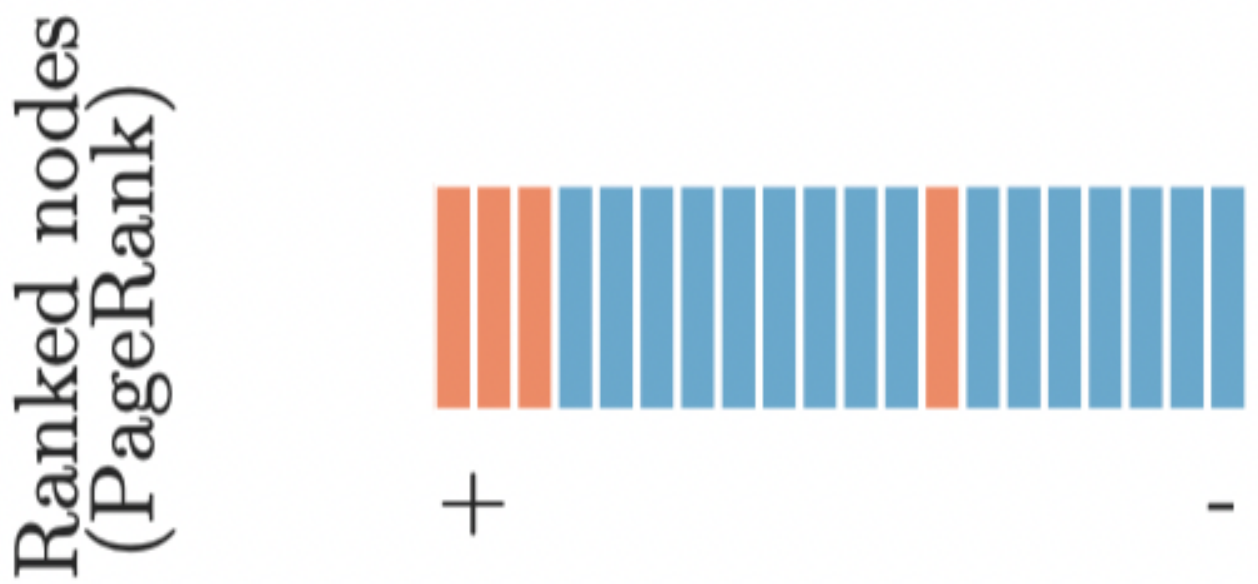
Inequality, inequity, and disparity

Given a network,

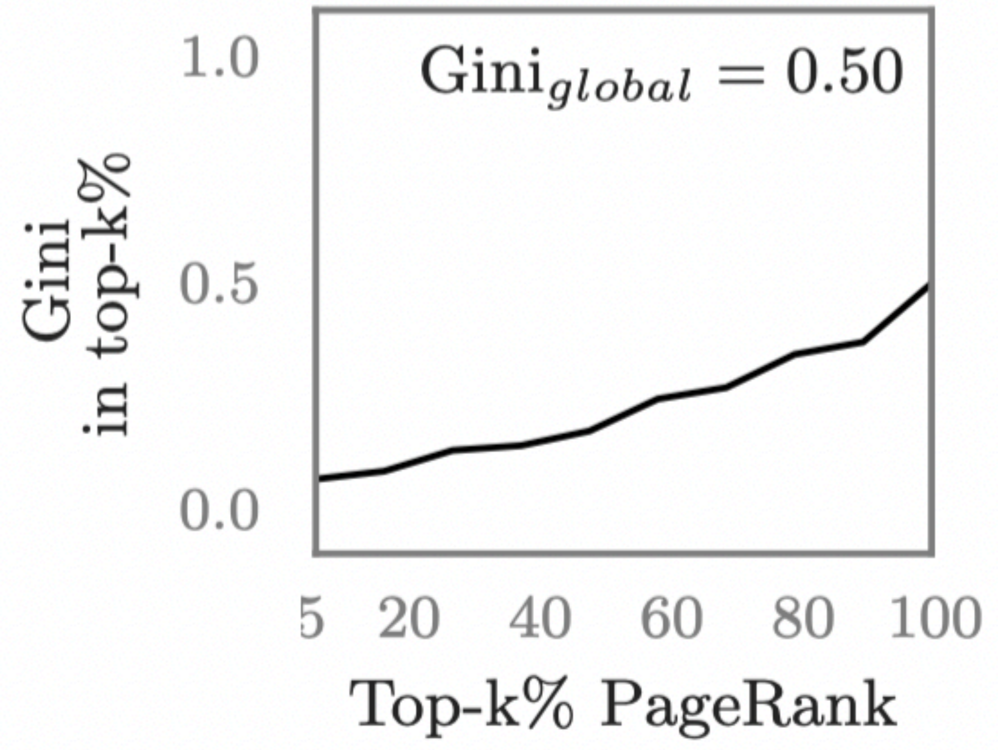


n = 20
min (m) = 20% (fm)
maj (M) = 80%

… and a ranking of its nodes



Measure:

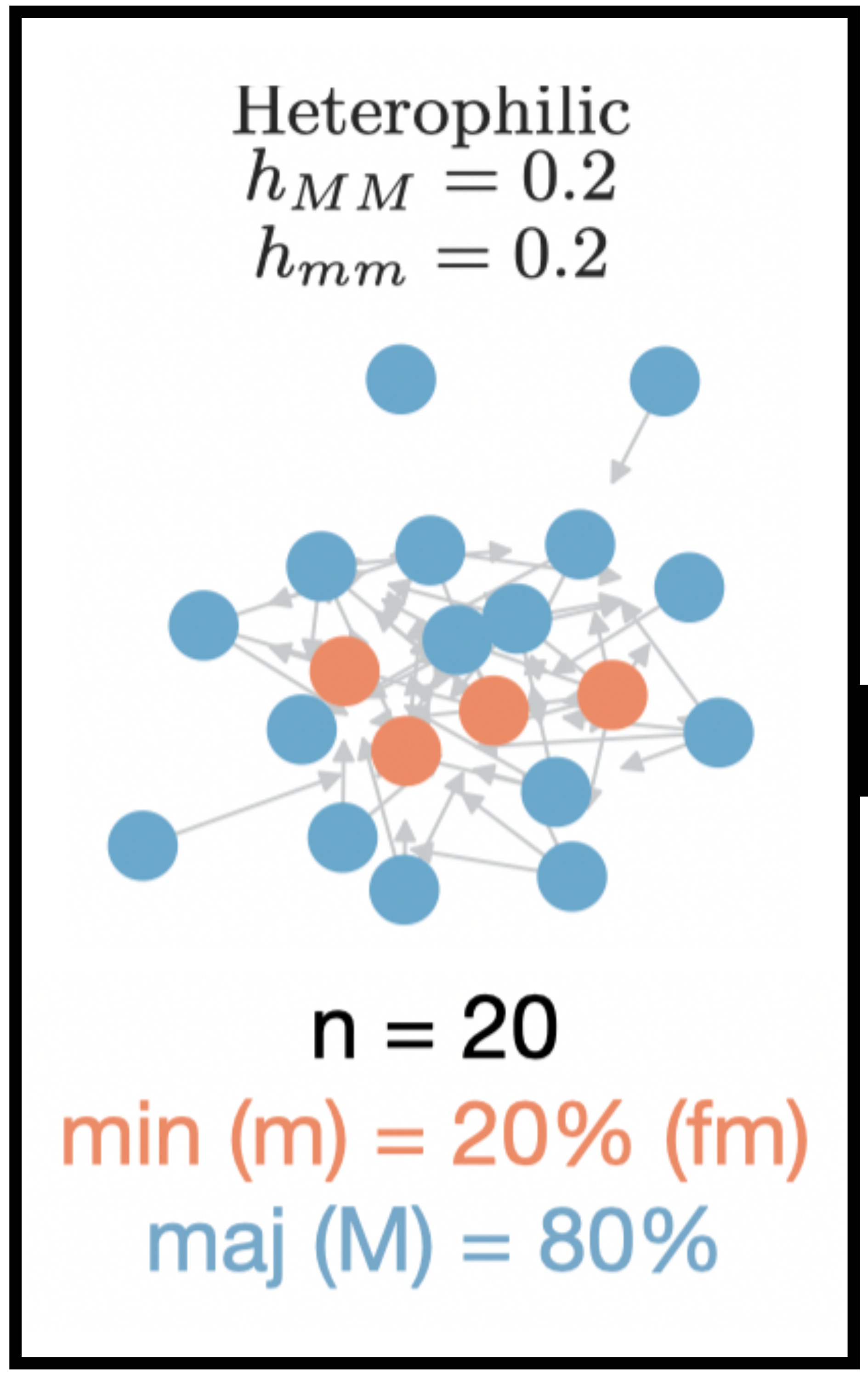


Inequality

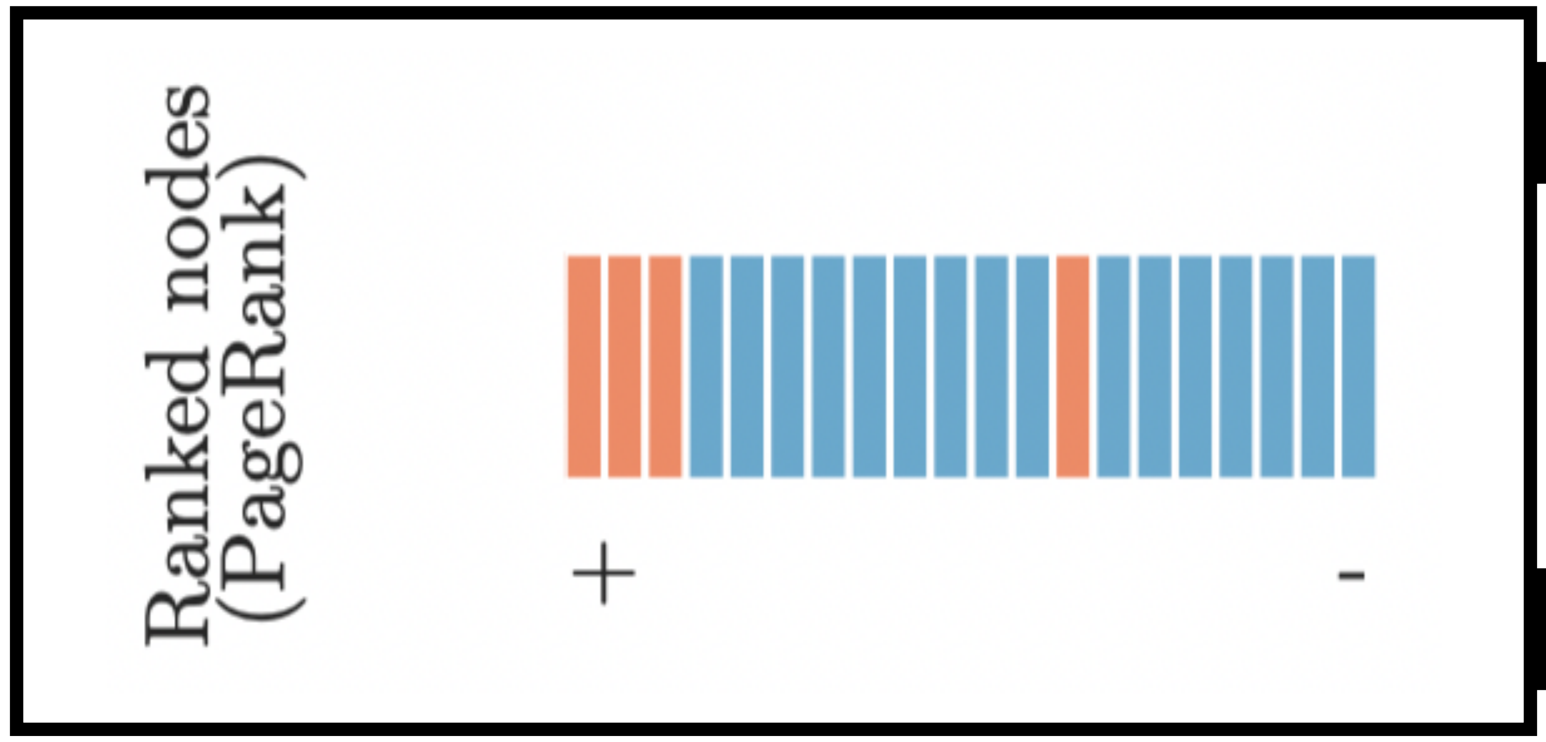
Espín-Noboa et al. 2022

Inequality, inequity, and disparity

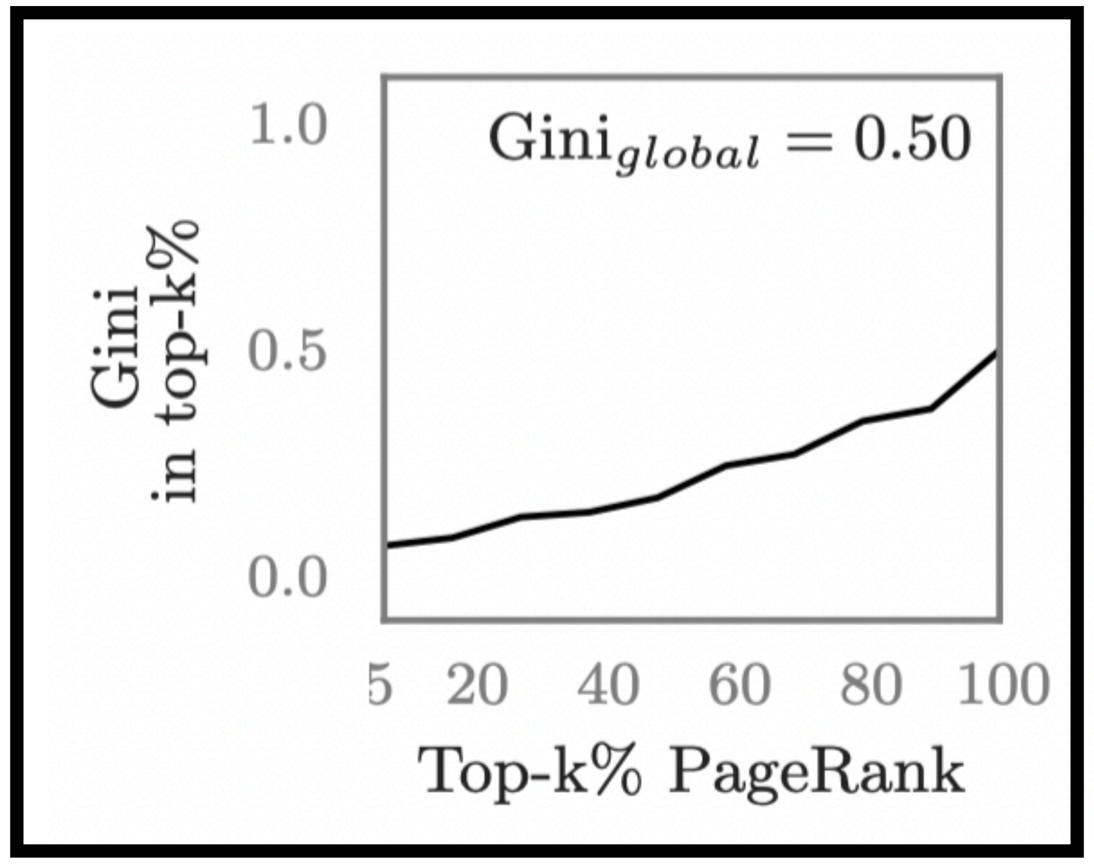
Given a network,



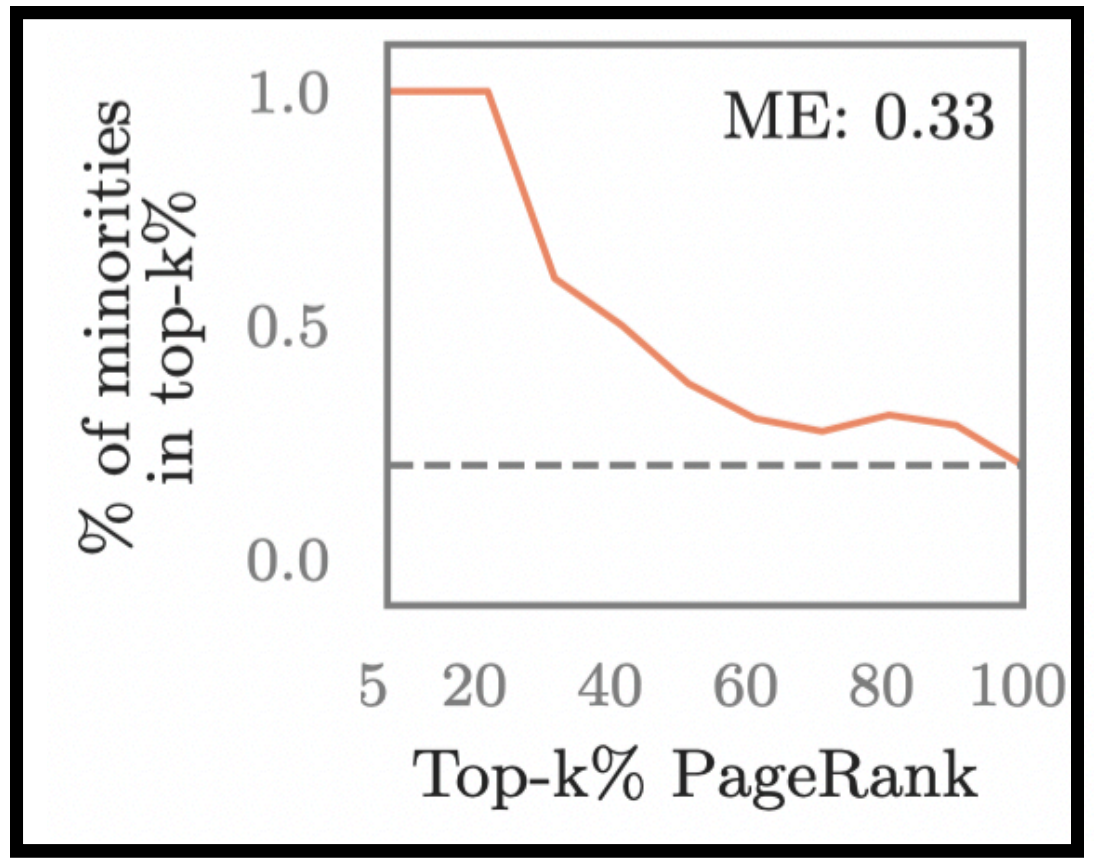
… and a ranking of its nodes



Measure:



Inequality

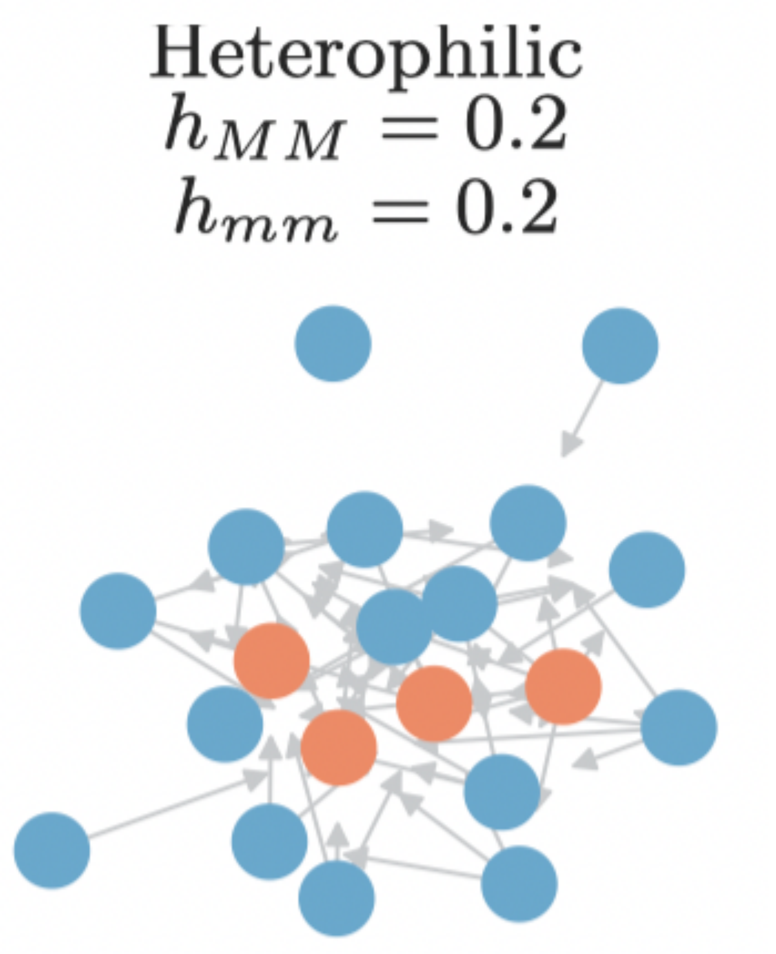


Inequity

Espín-Noboa et al. 2022

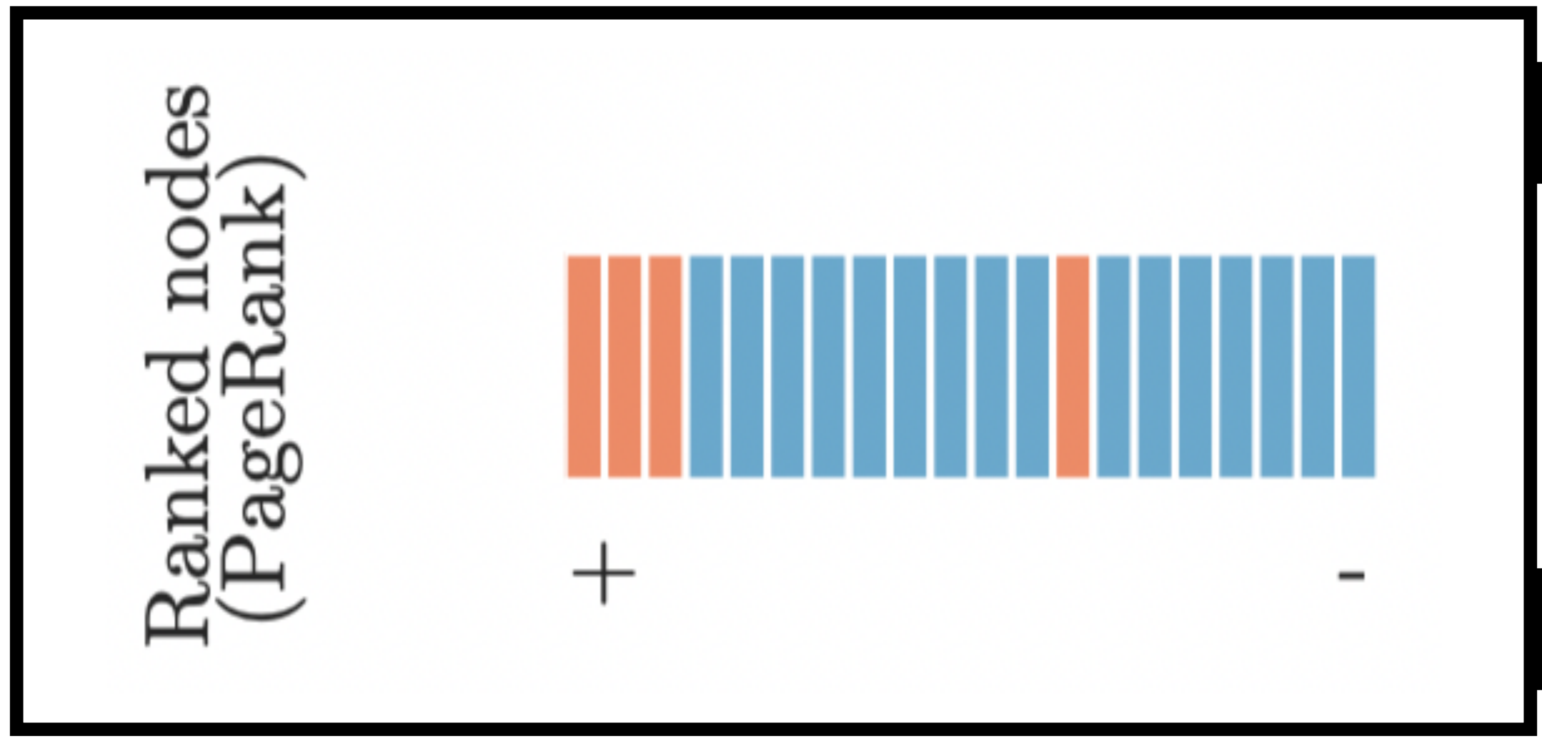
Inequality, inequity, and disparity

Given a network,

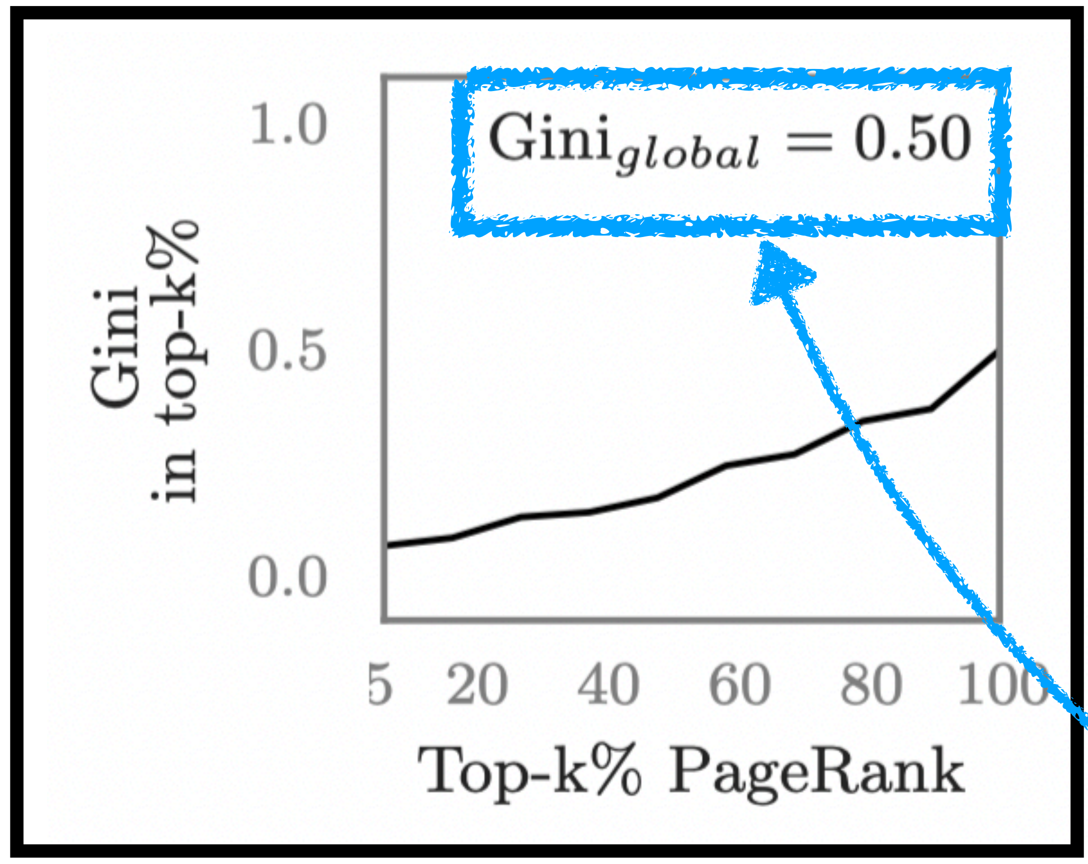


n = 20
min (m) = 20% (fm)
maj (M) = 80%

… and a ranking of its nodes

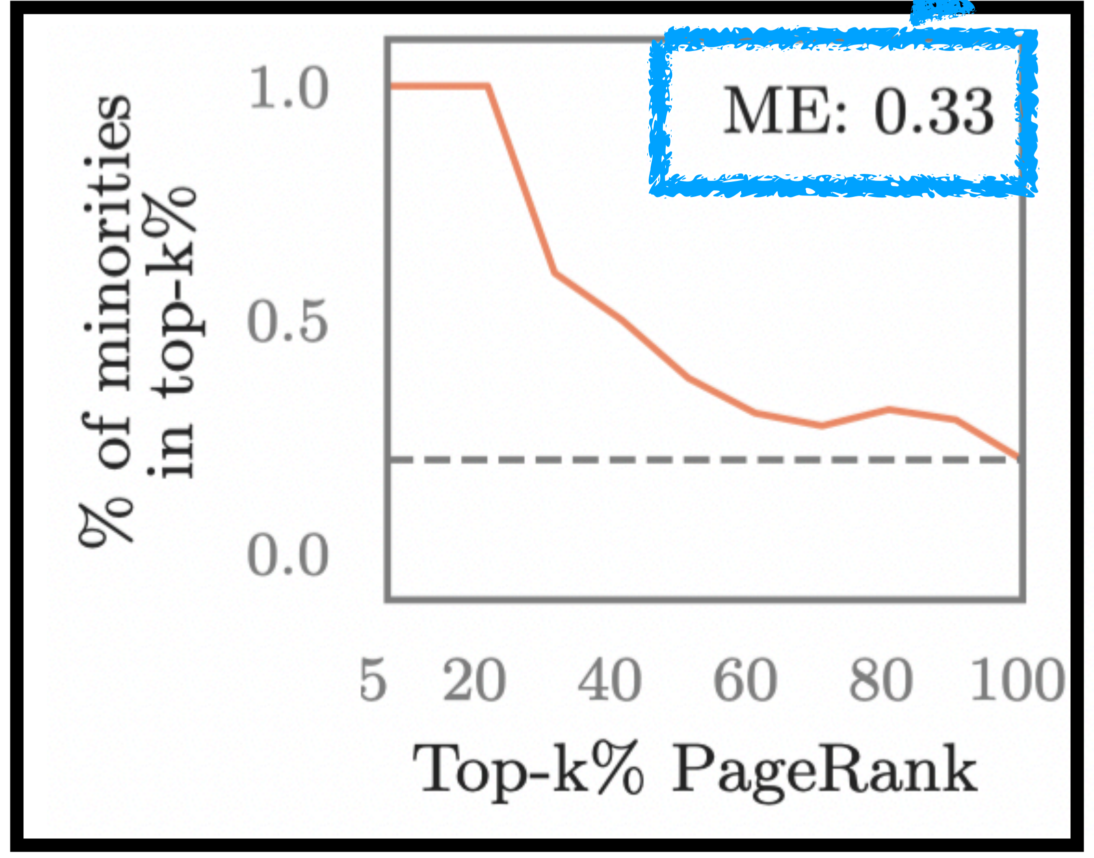


Measure:



Inequality

Disparity
ME vs. Gini_{global}



Inequity

Espín-Noboa et al. 2022

Inequality, inequity, and disparity

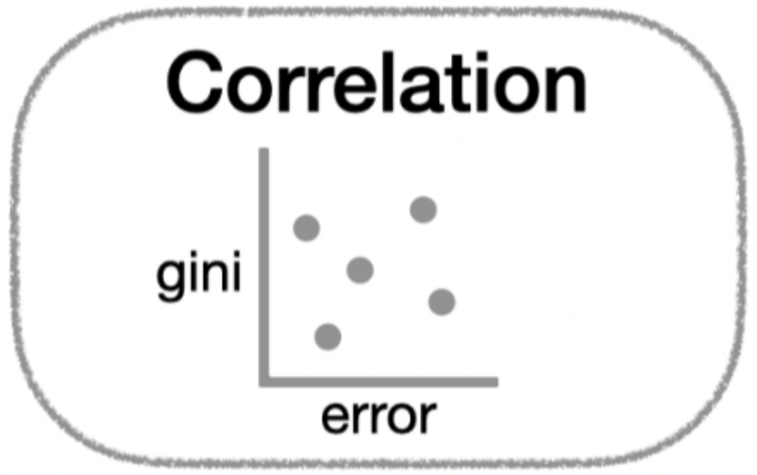
Ranking of nodes
based on PageRank
in descending order

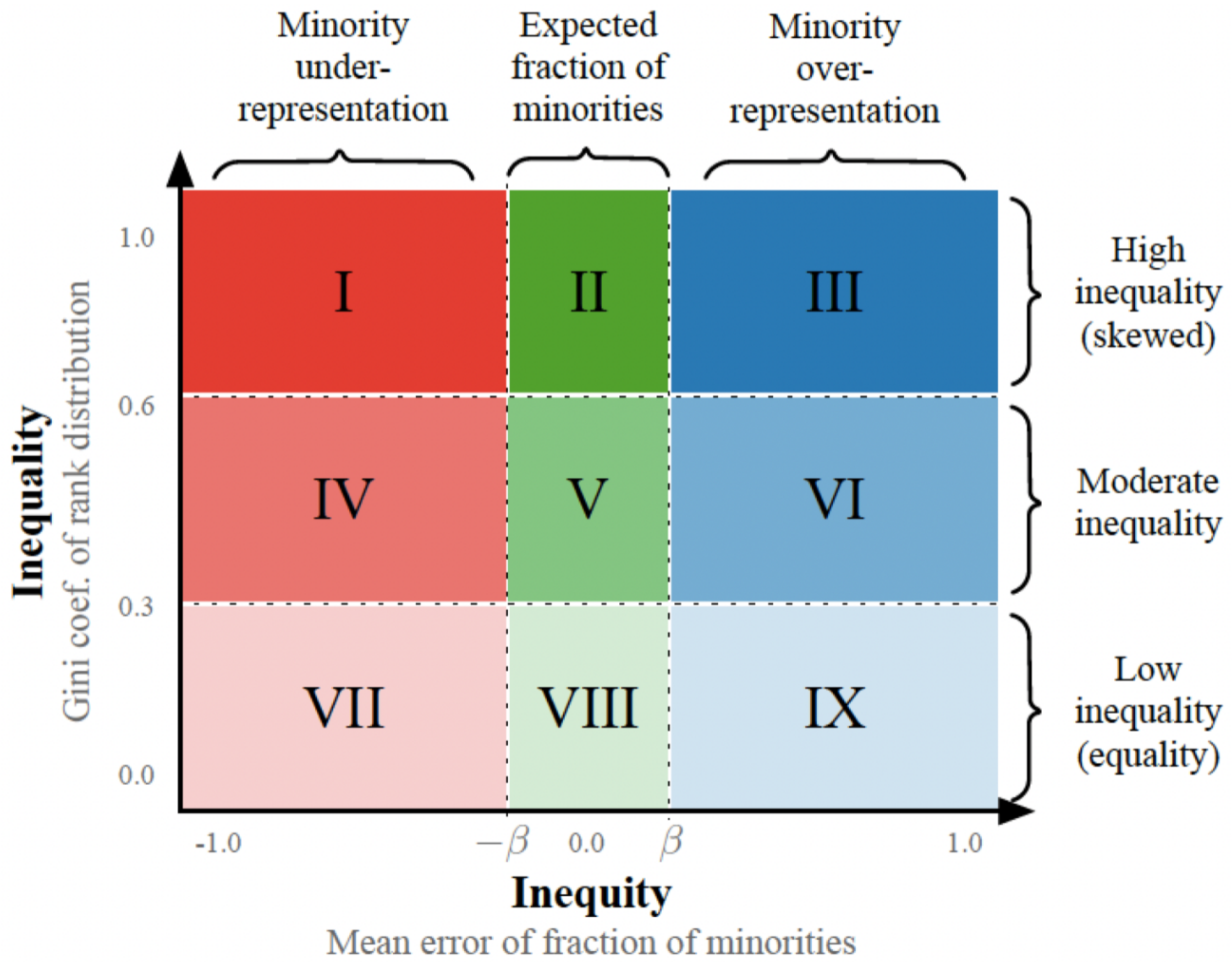
minority (**25%**)
majority (75%)

top-k%	PR		Gini	% min (error)	
20%	0.39	j	0.00	0% (-0.25)	Local
40%	0.14	o	0.23	50% (+0.25)	Local
60%	0.13	l	0.26	33% (+0.08)	Local
80%	0.09	k	0.31	**25% (0.00)**	Local
100%	0.06	i, m, n, p	0.37	**25% (0.00)**	Local
			0.37	(0.06)	Global

Espín-Noboa et al. 2022

Inequality vs. inequity

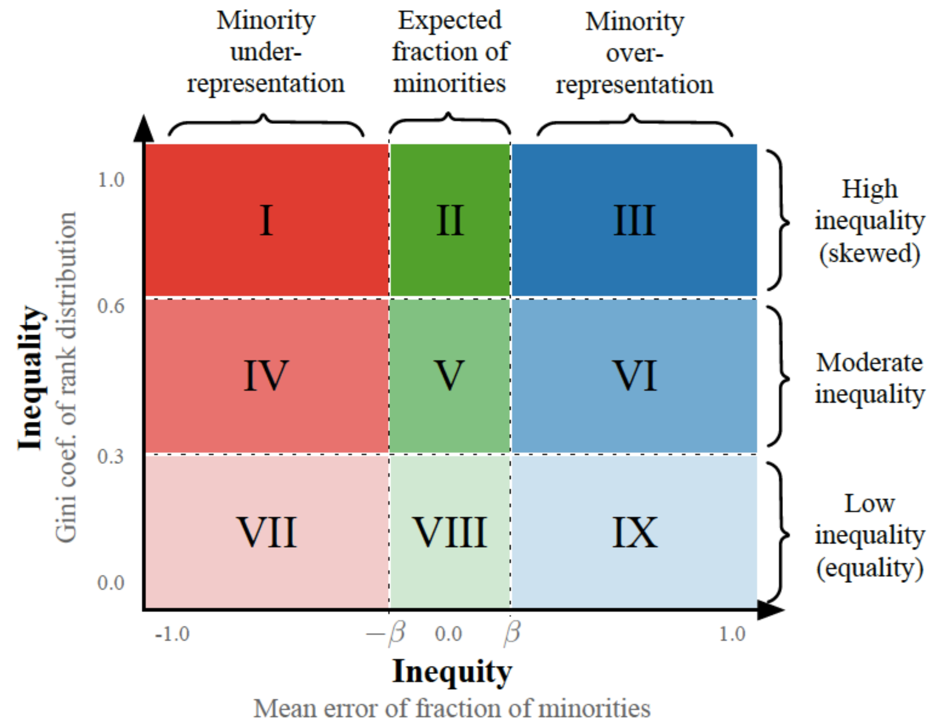


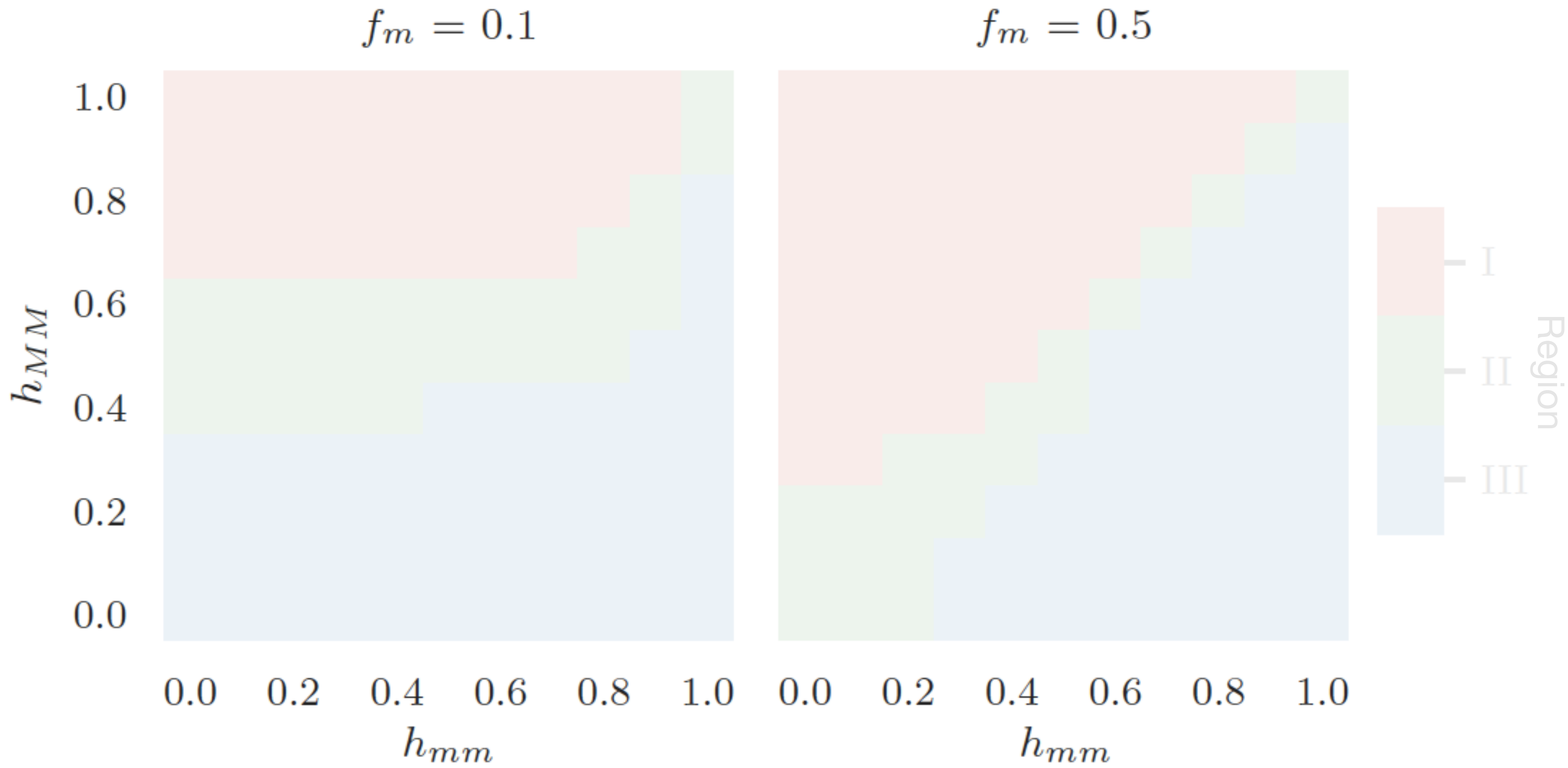


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

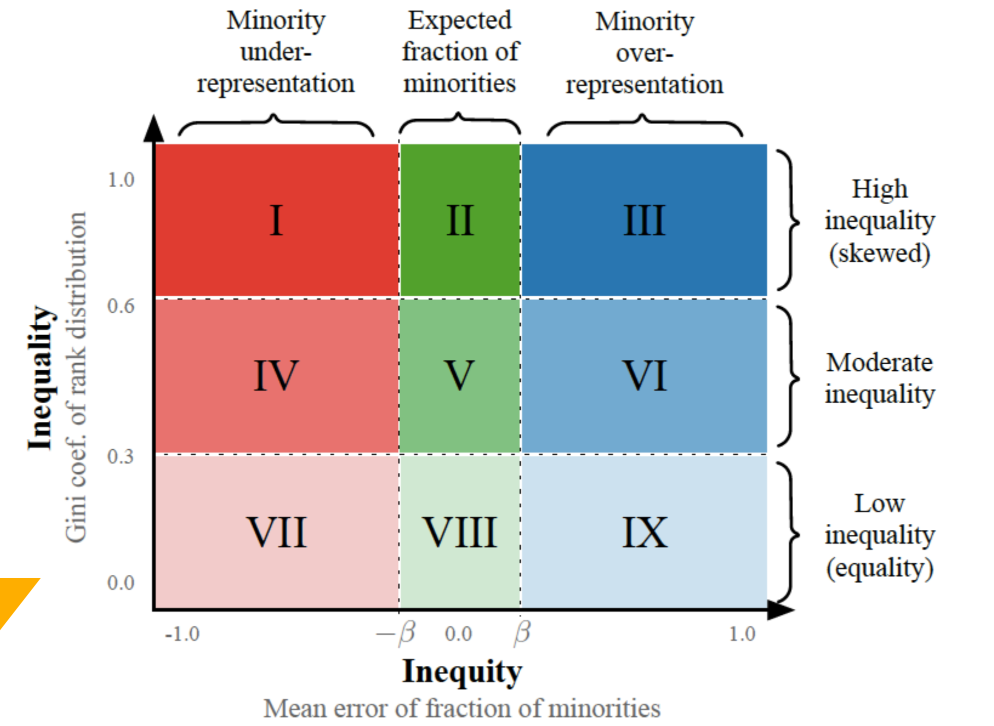




Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

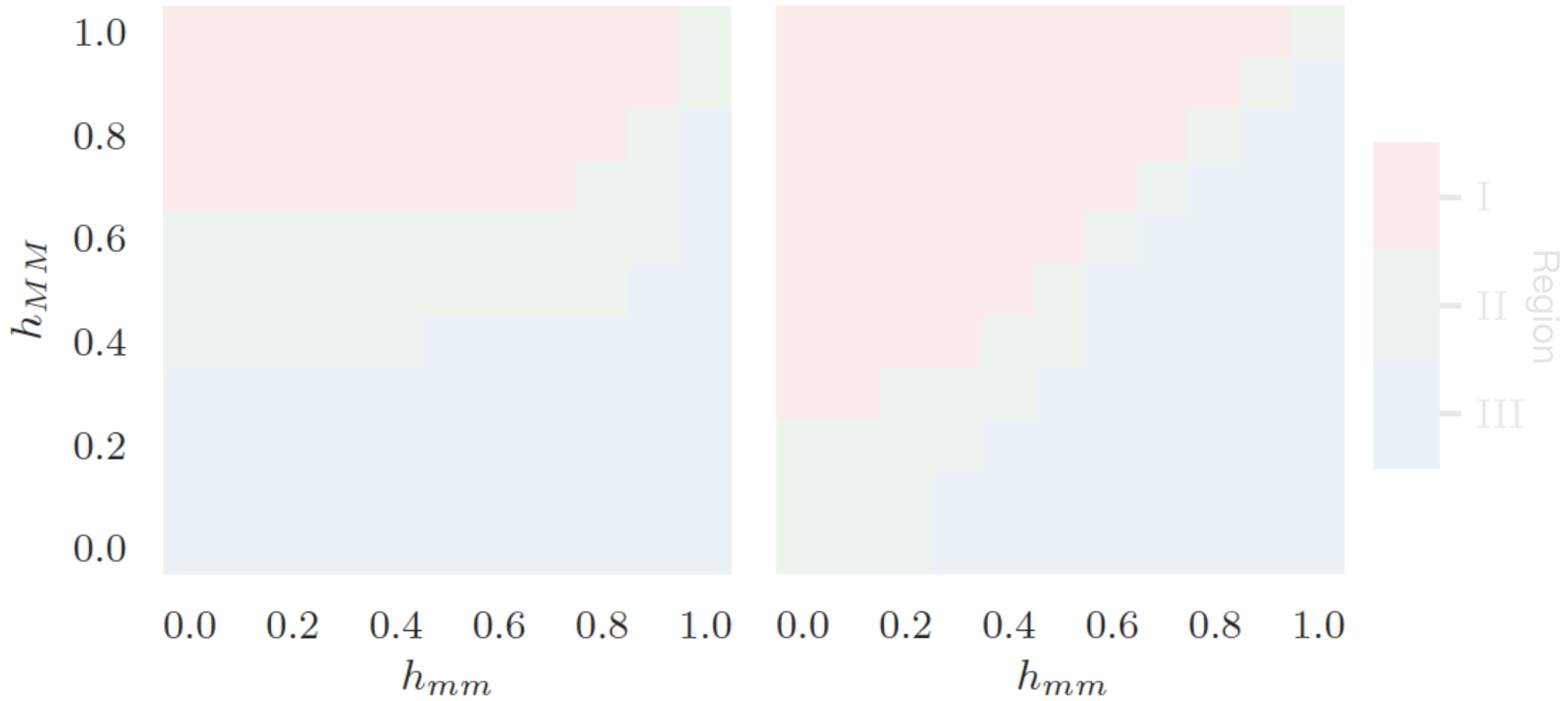
As a function of homophily and fraction of minorities



Columns
Fraction of minority group

unbalance
(minority 10%)

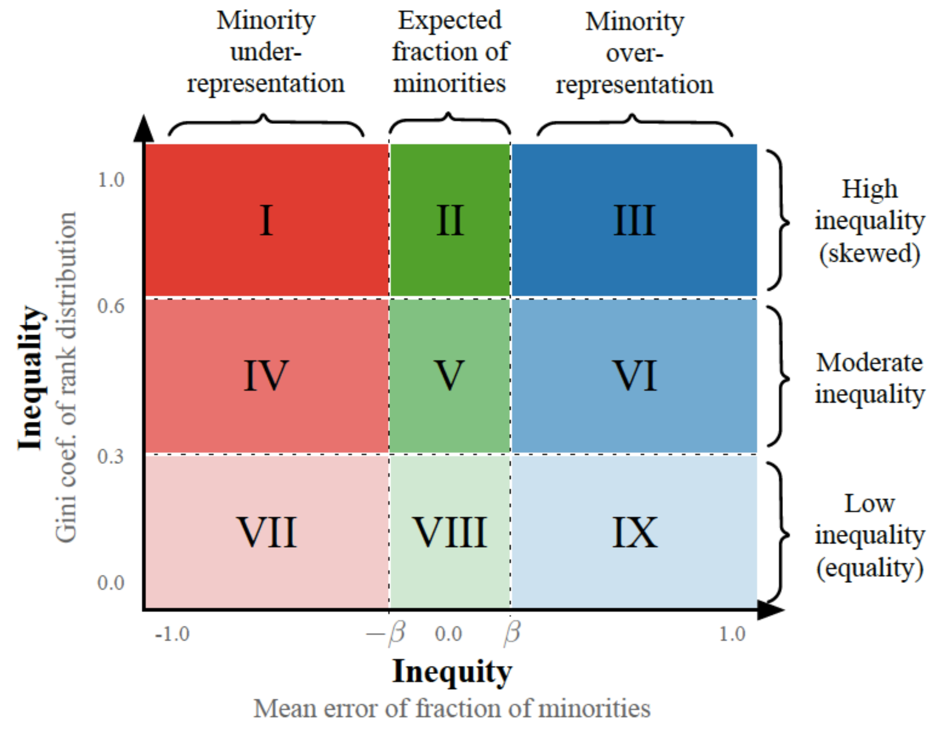
balance
(no minority)



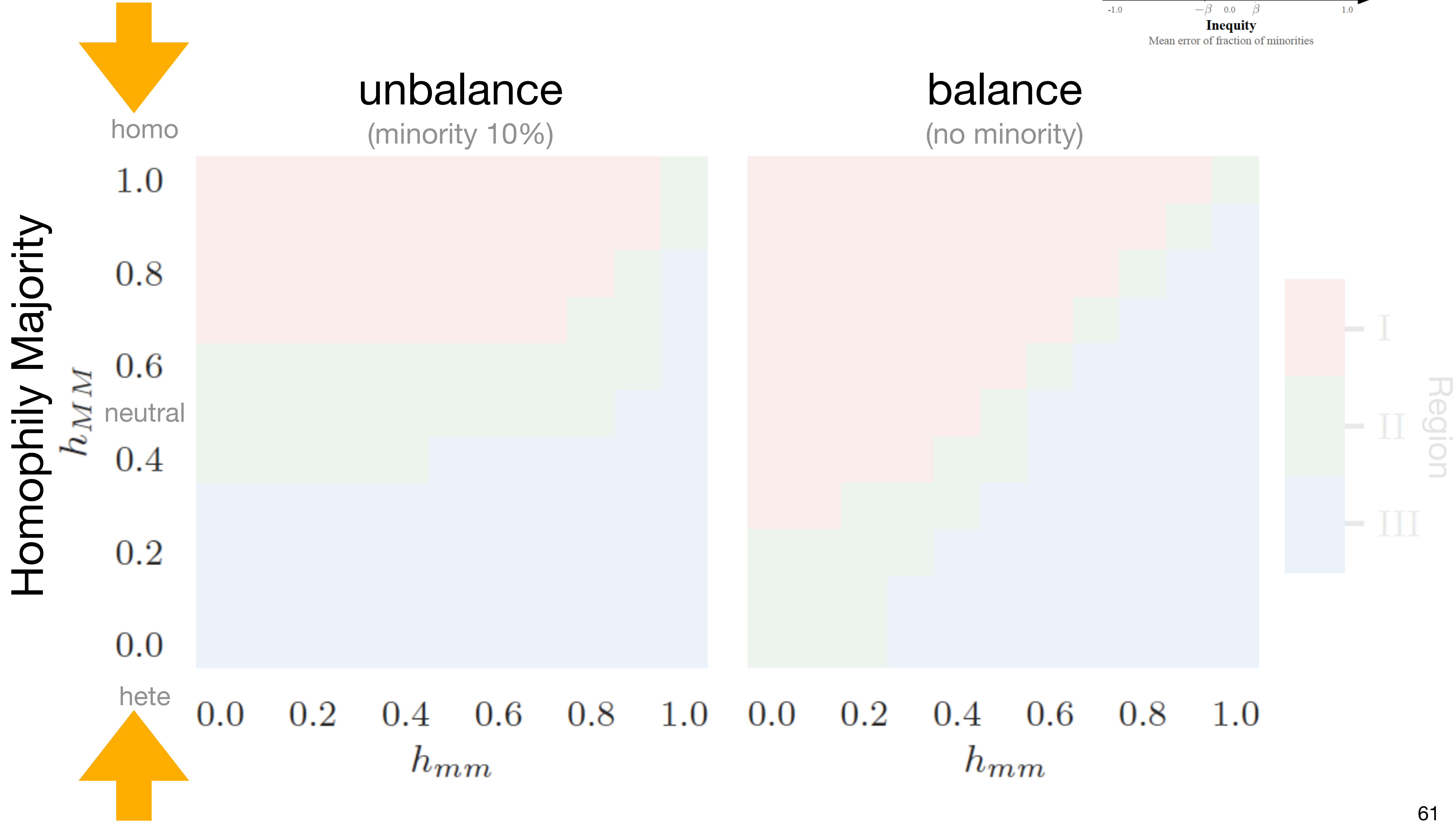
Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities



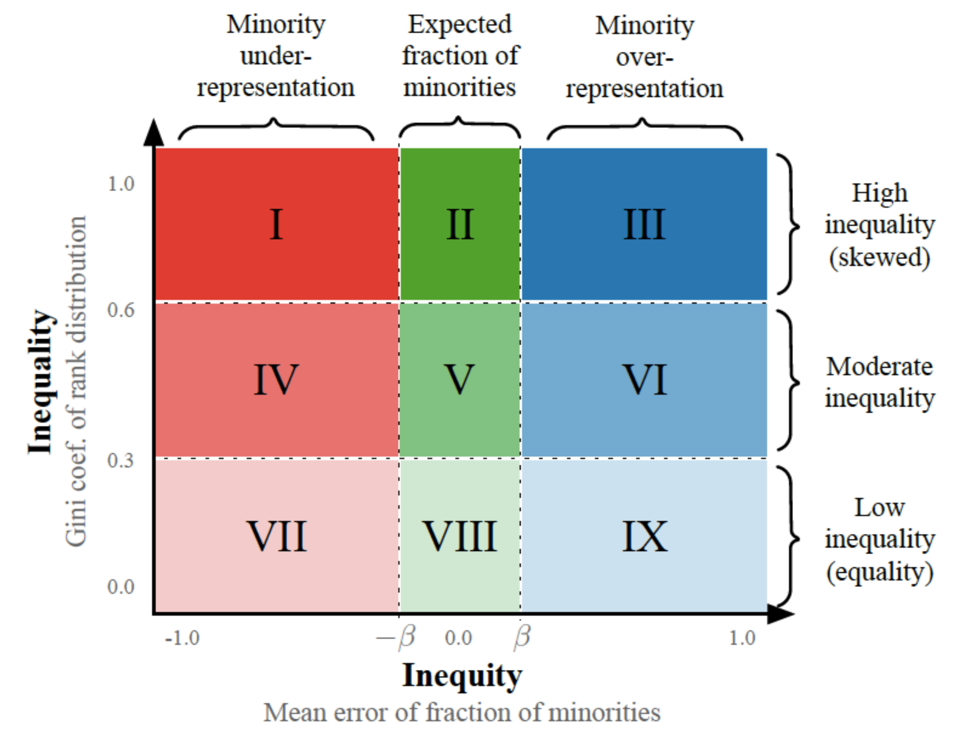
y-axis
Homophily within majority group

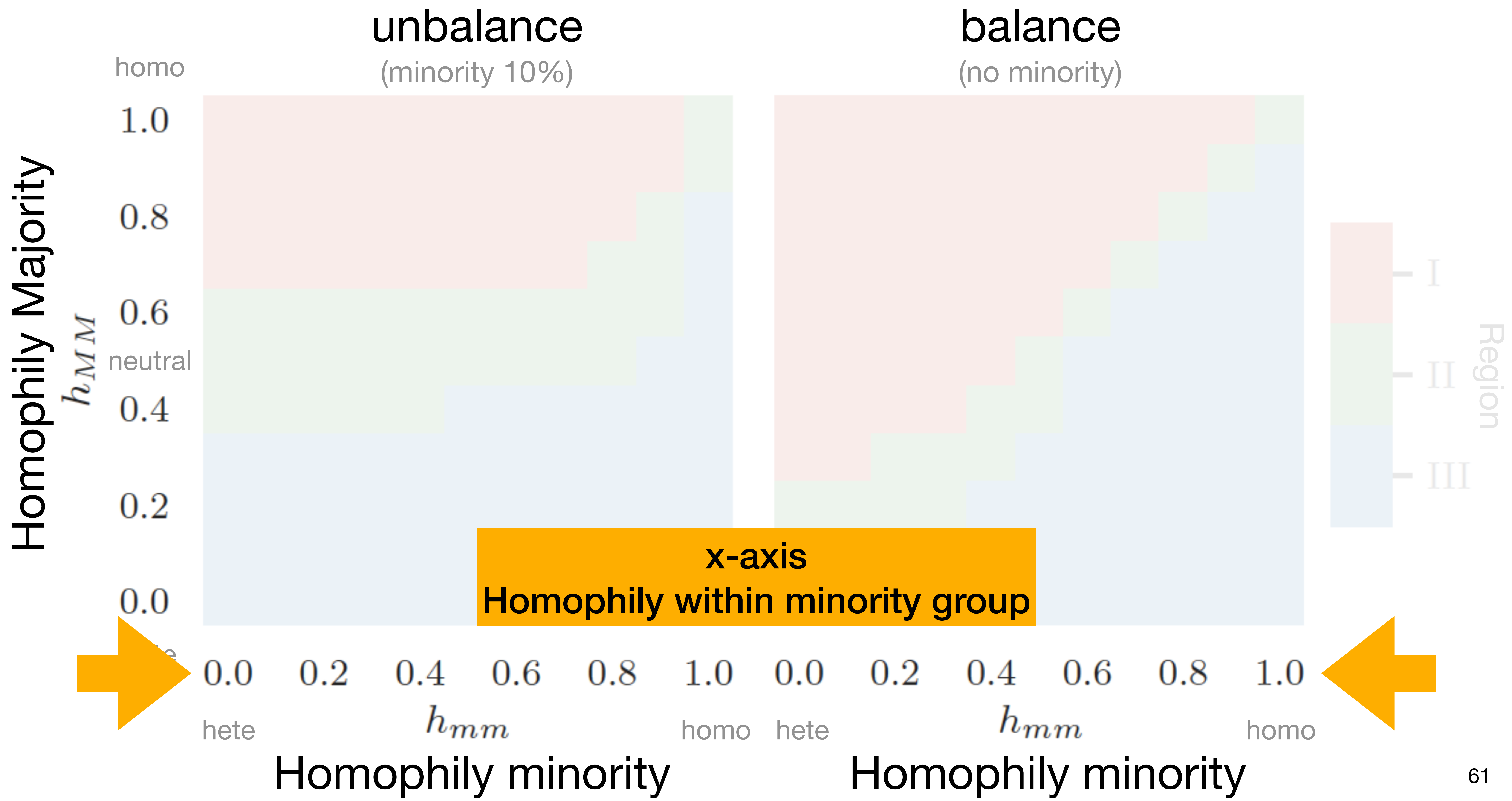


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

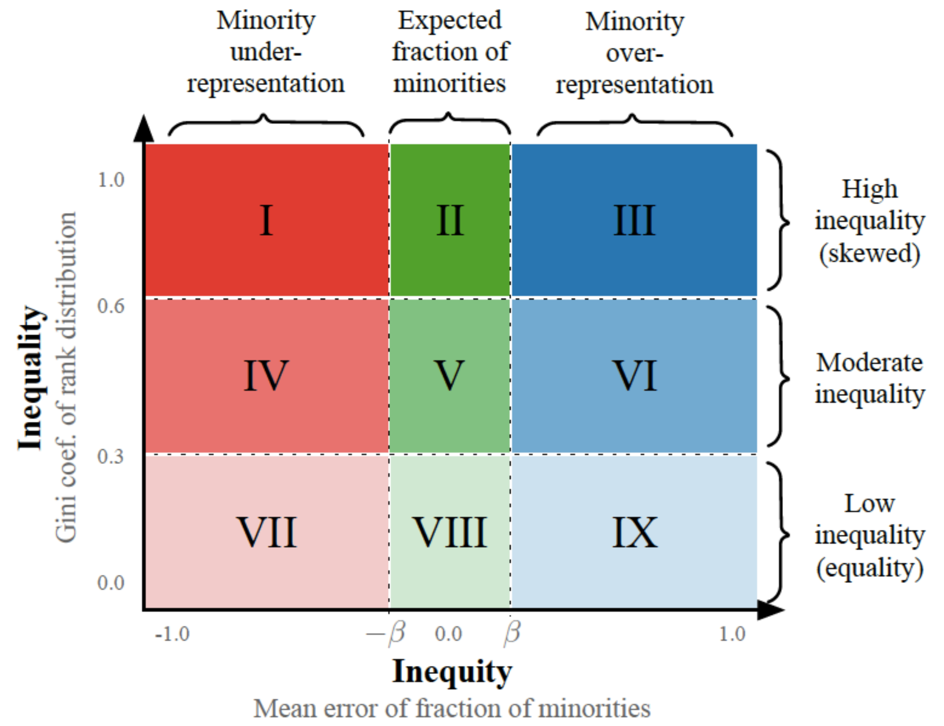


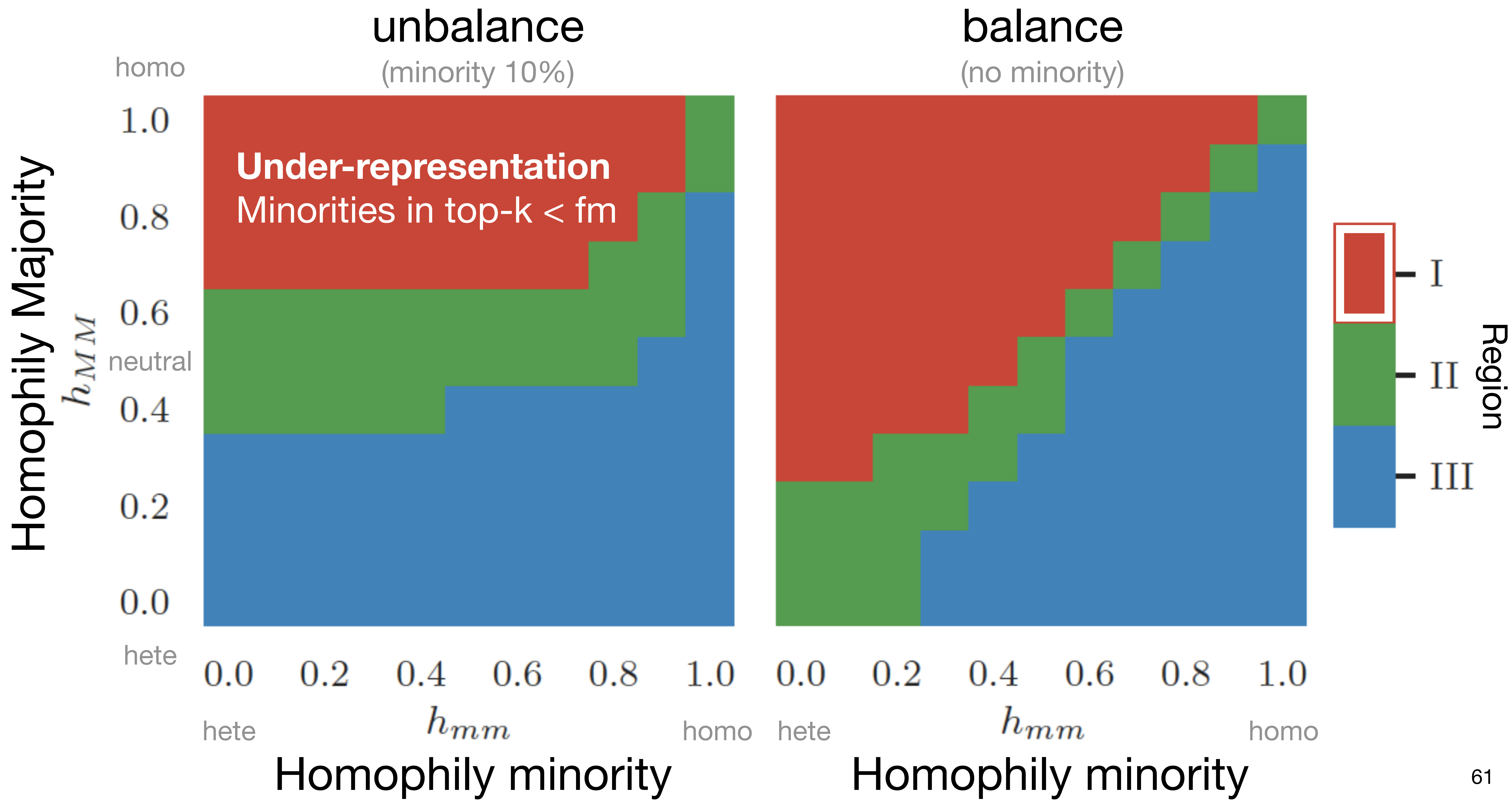


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

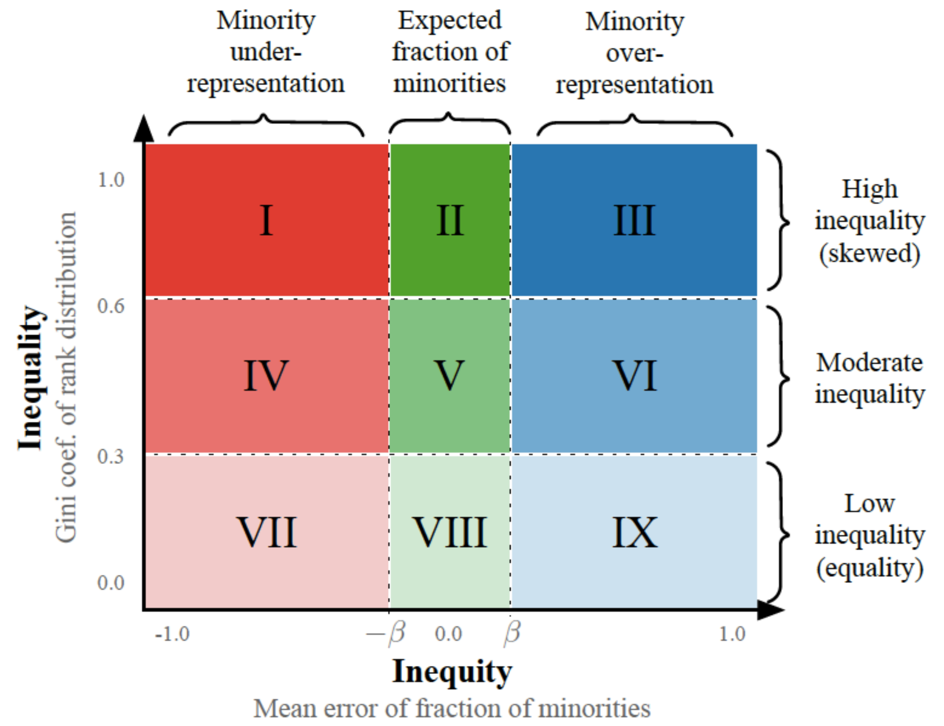


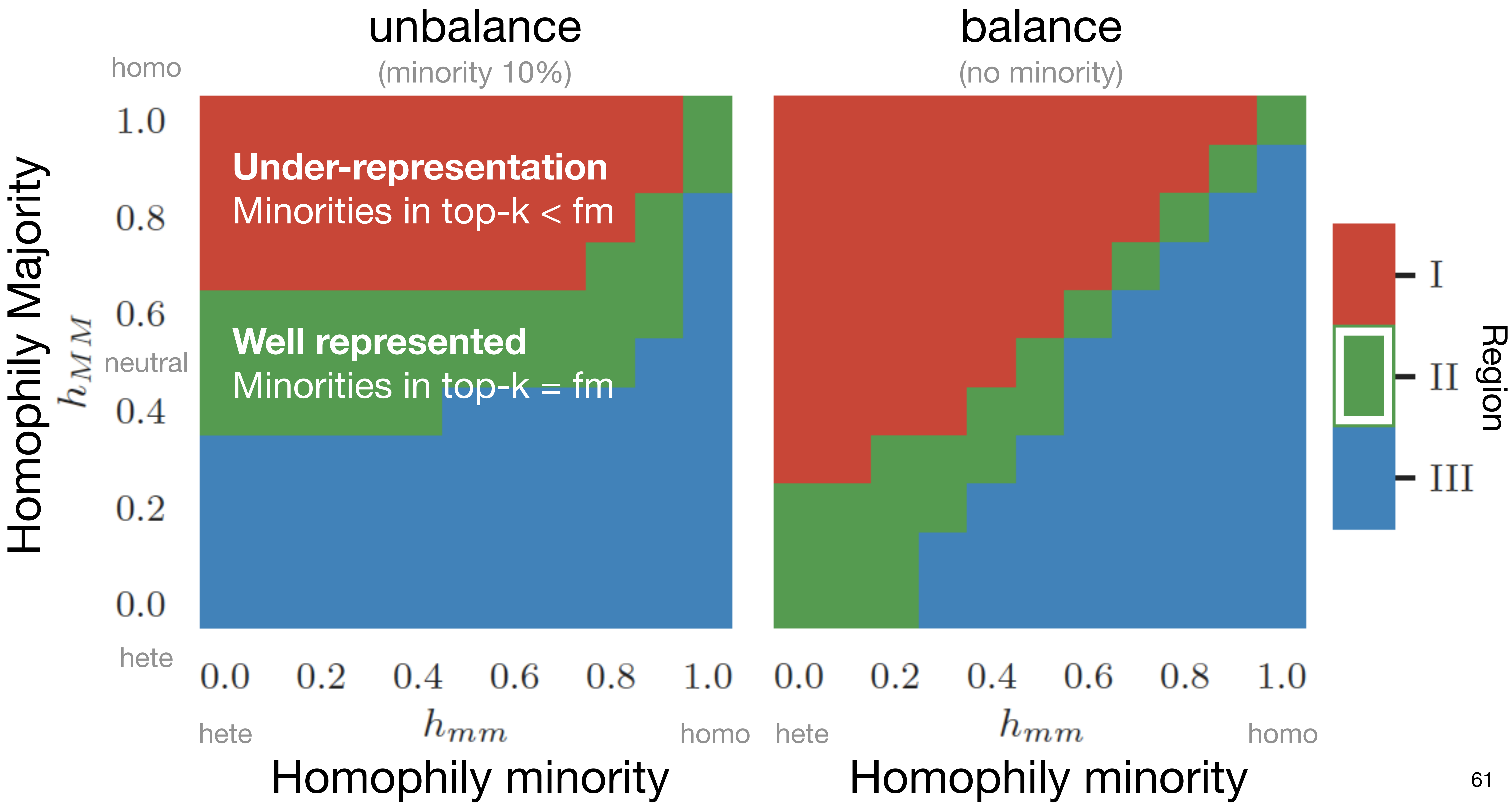


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

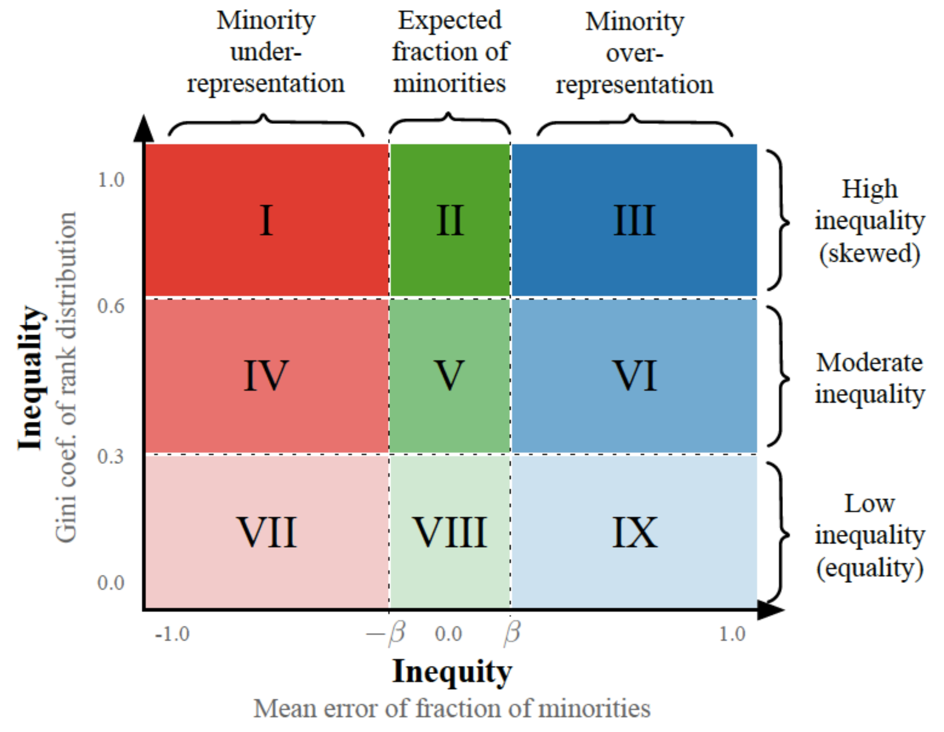


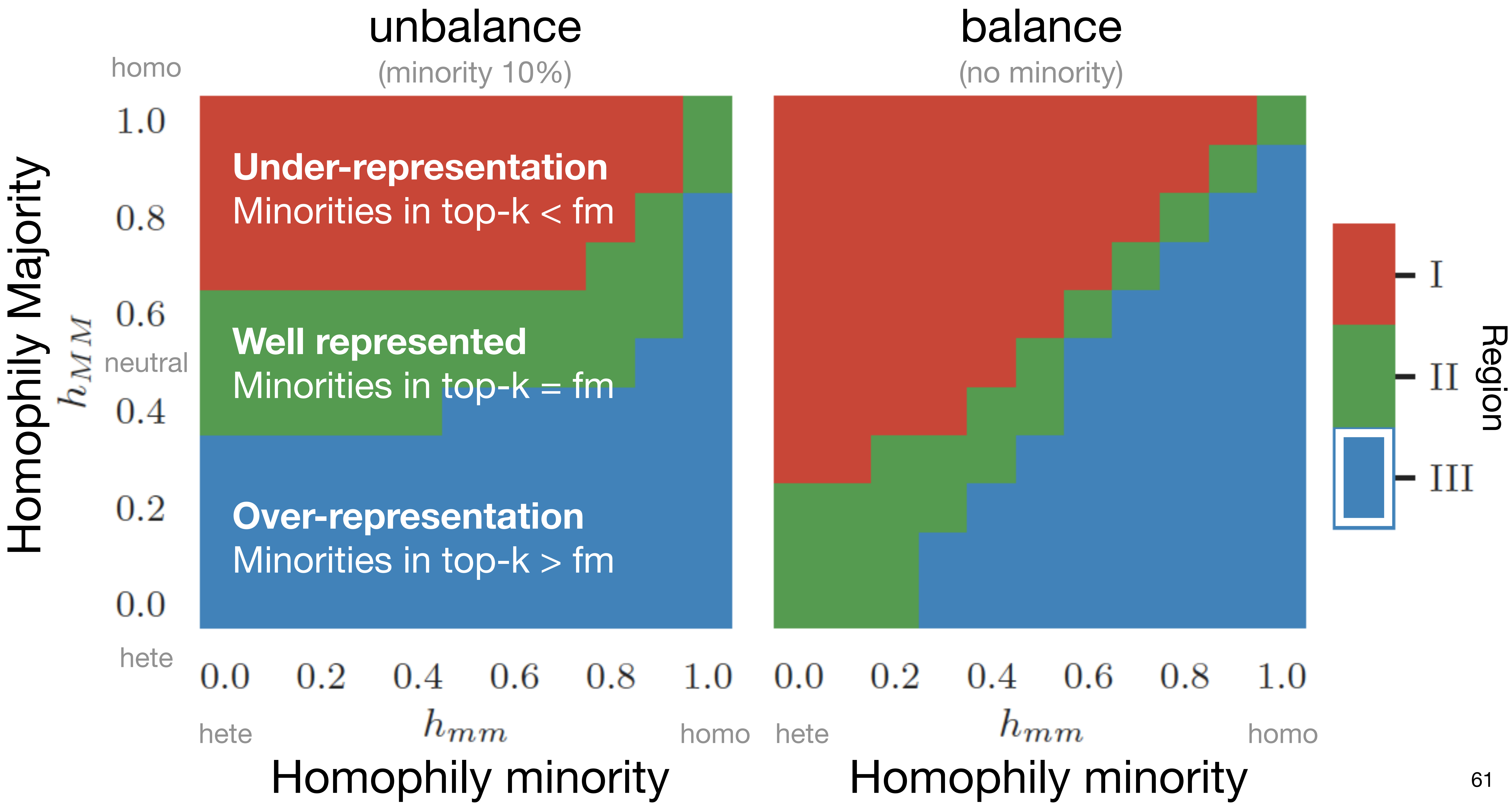


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

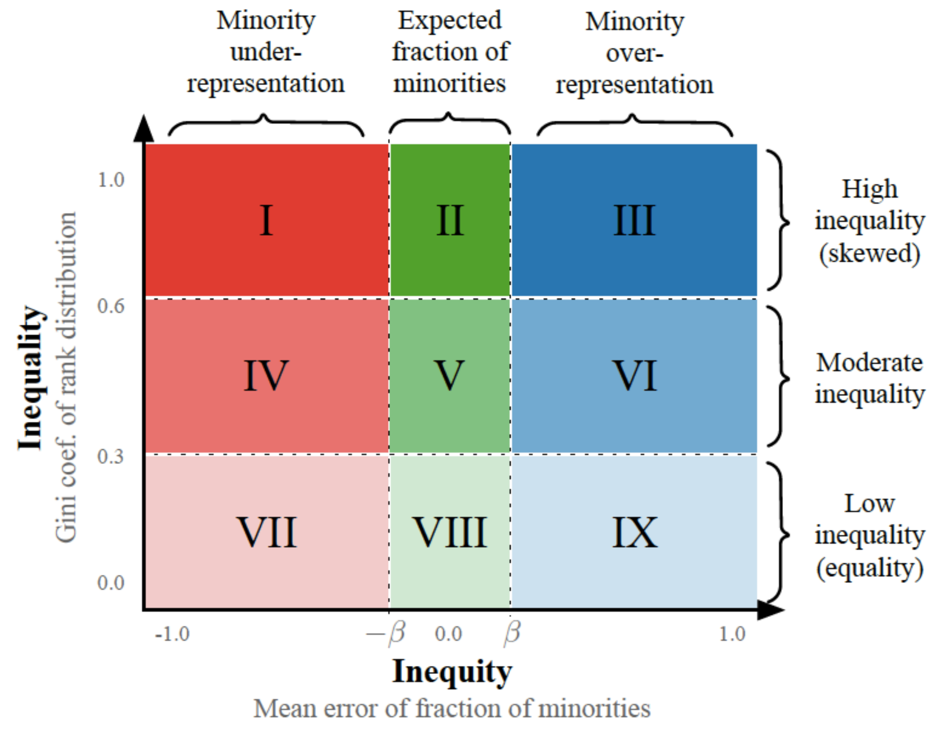


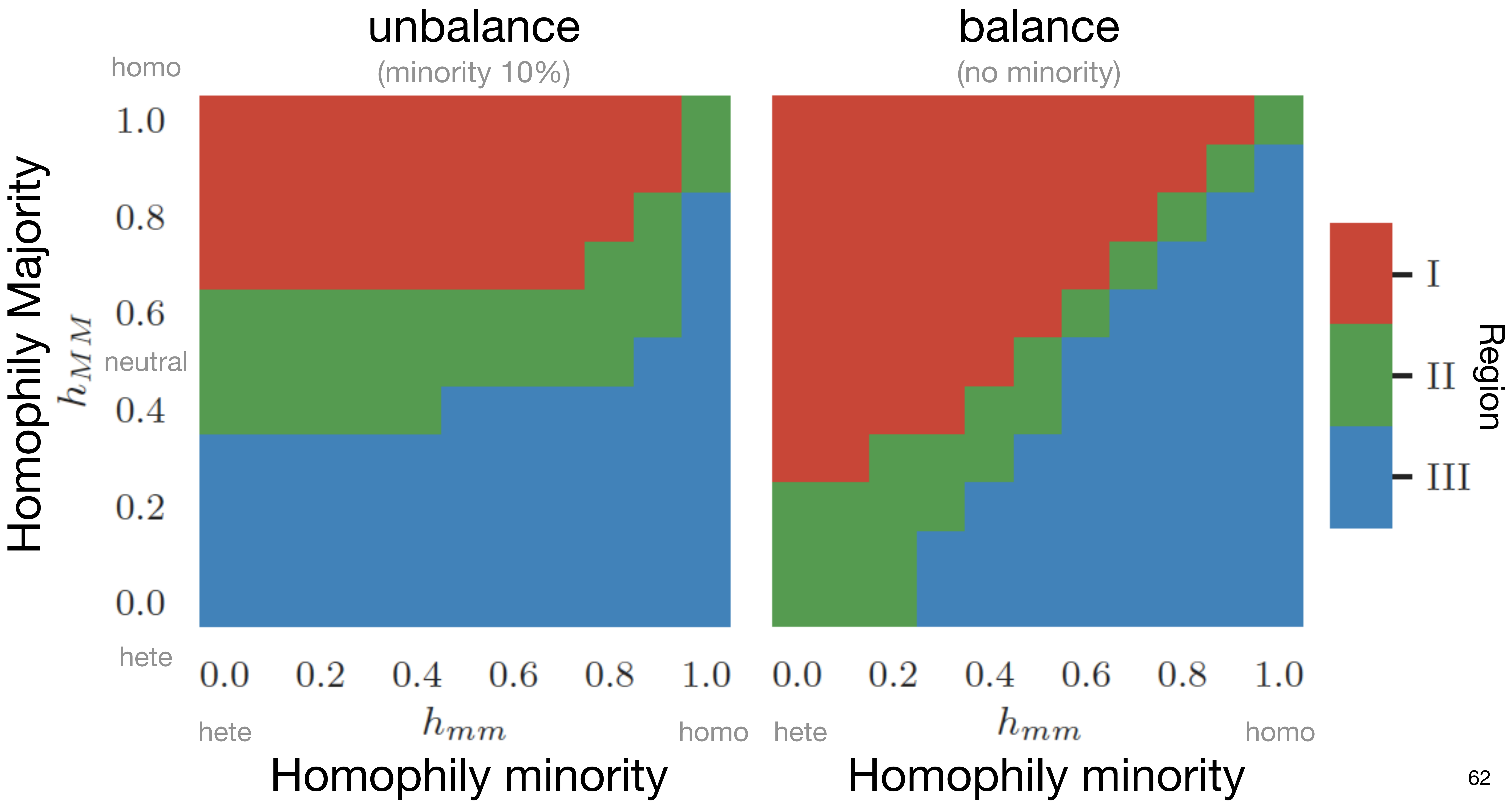


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities

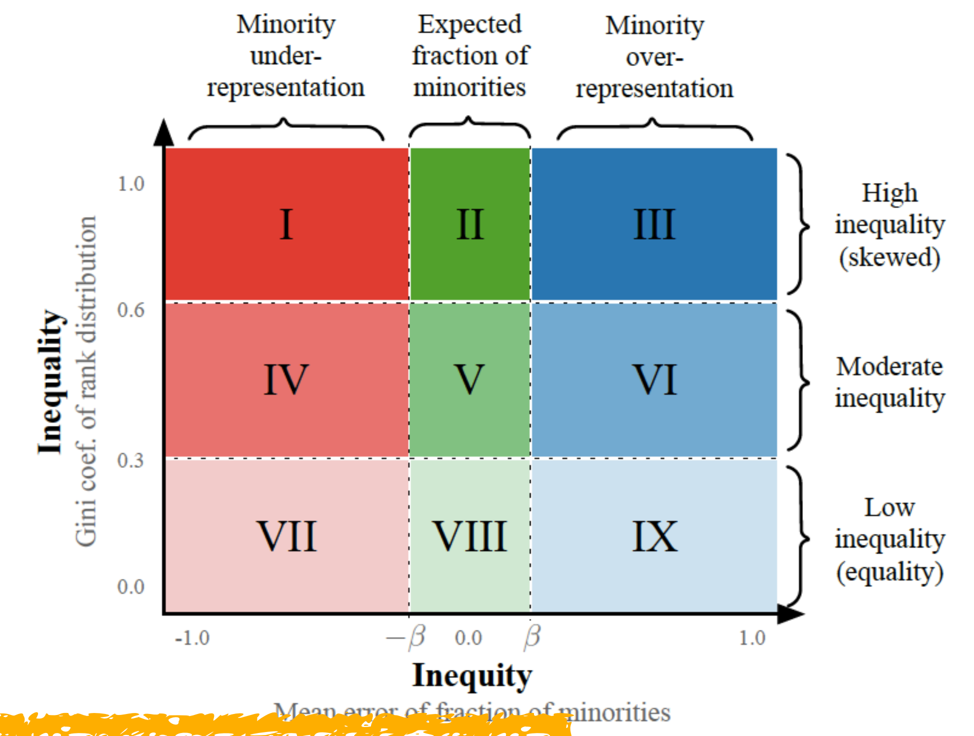




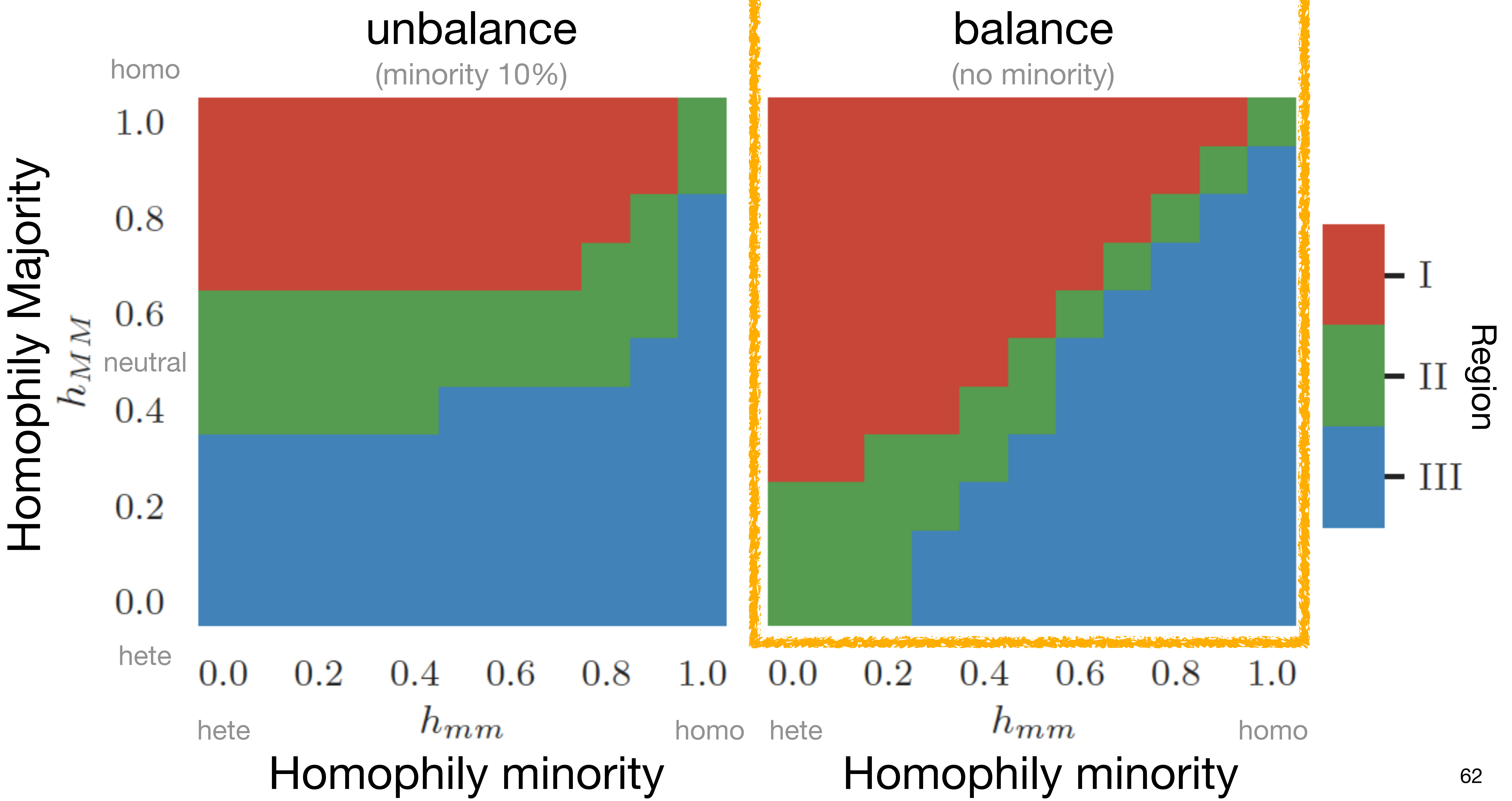
Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities



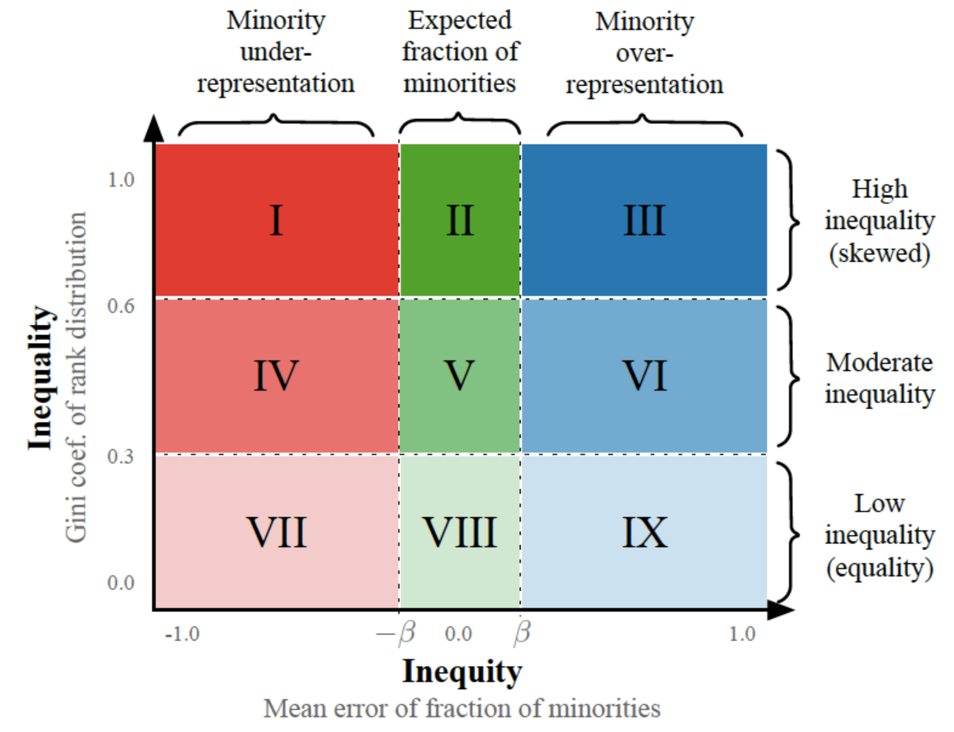
1. In balanced networks, both groups are well represented if hmm = hMM



Espín-Noboa et al. 2022

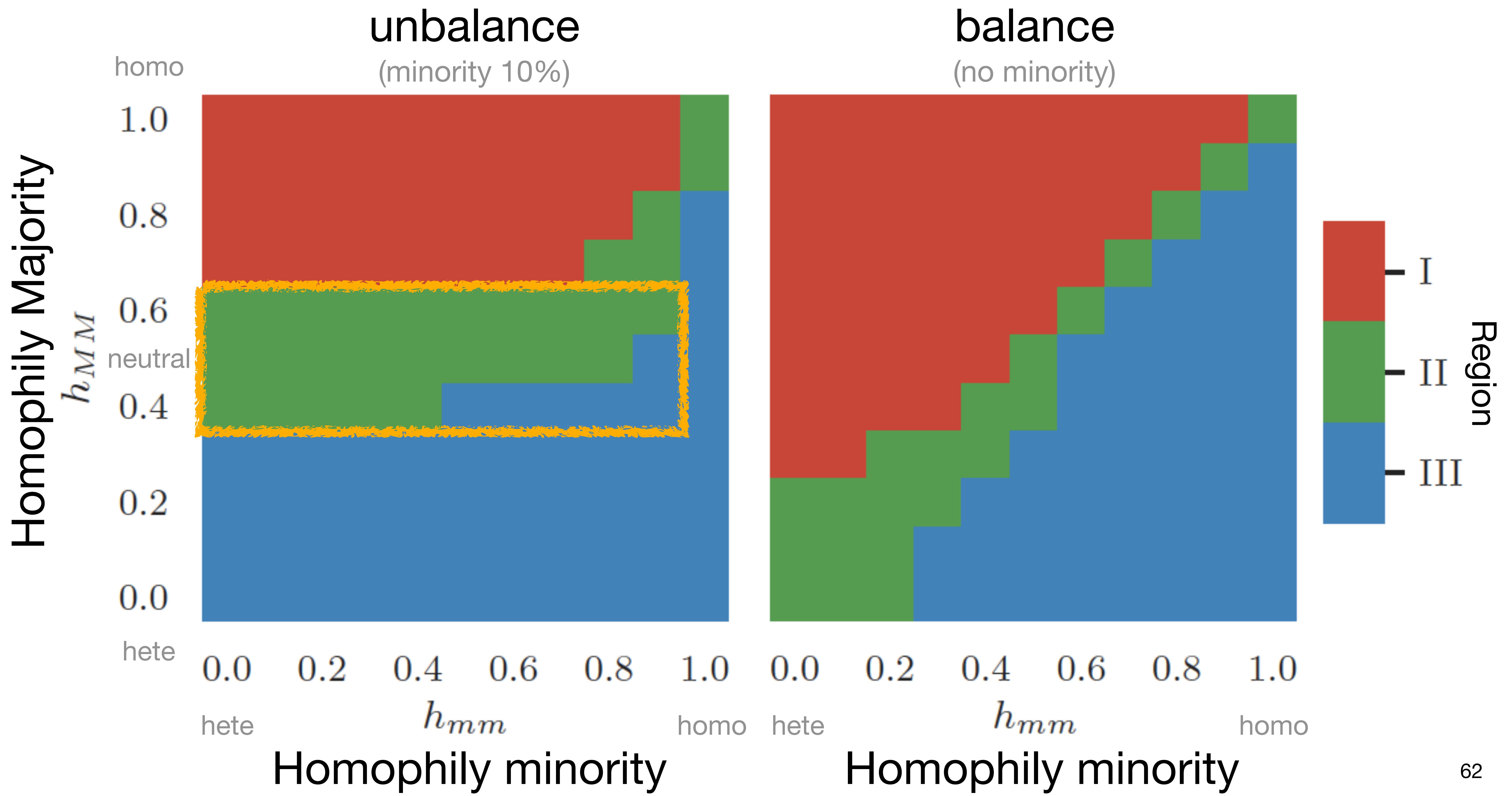
Inequality vs. inequity in PageRank

As a function of homophily and fraction of minorities



1. In **balanced networks**, both groups are well represented if hmm = hMM

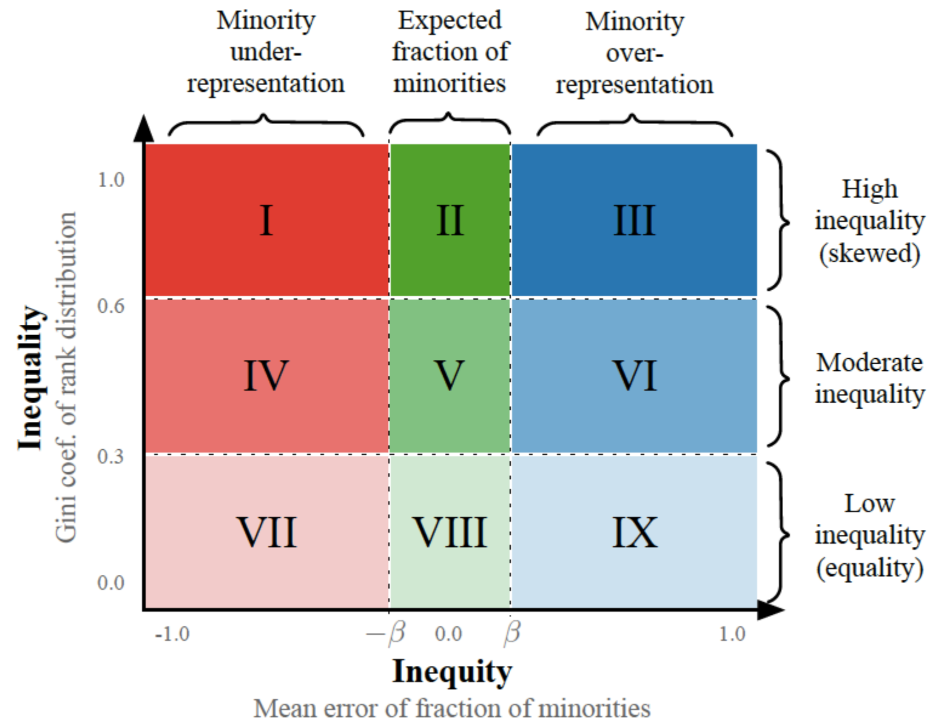
2. In **unbalanced networks**, minorities are well represented when majority is neutral and the minority is not too homophilic.



Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

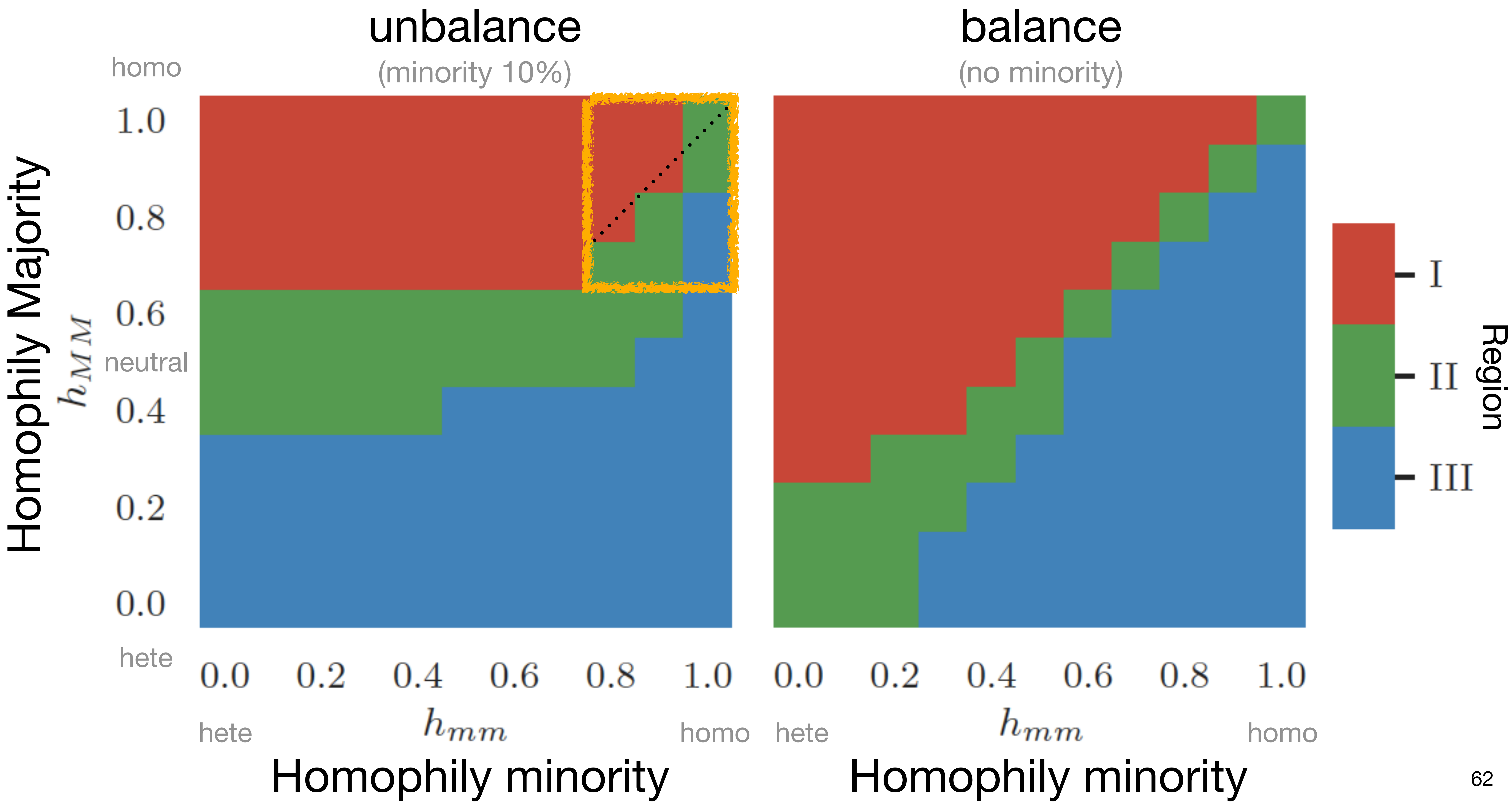
As a function of homophily and fraction of minorities



1. In **balanced networks**, both groups are well represented if hmm = hMM

2. In **unbalanced networks**, minorities are well represented when majority is neutral and the minority is not too homophilic.

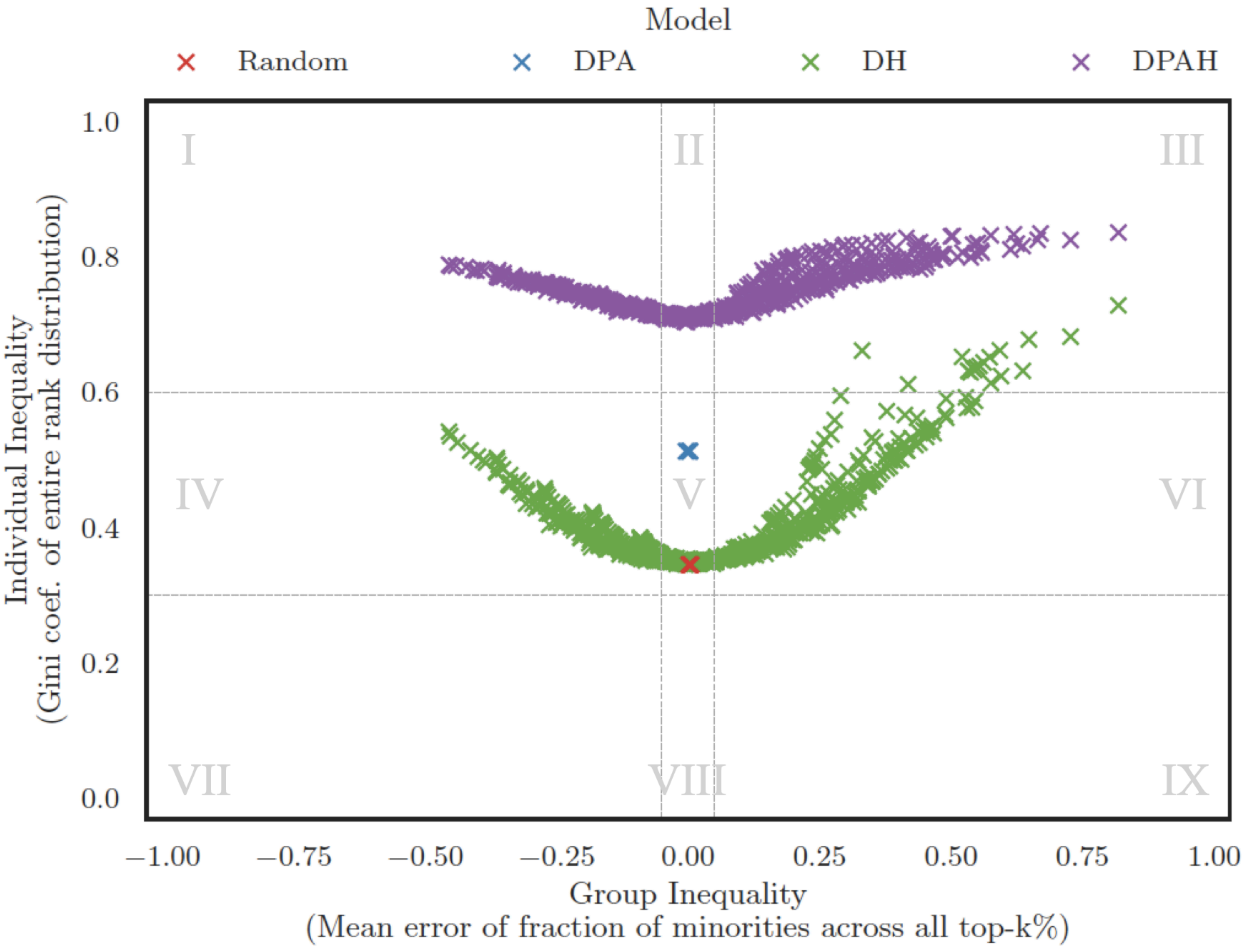
3. In **unbalanced and homophilic networks**, minorities are well represented when hmm > hMM.



Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

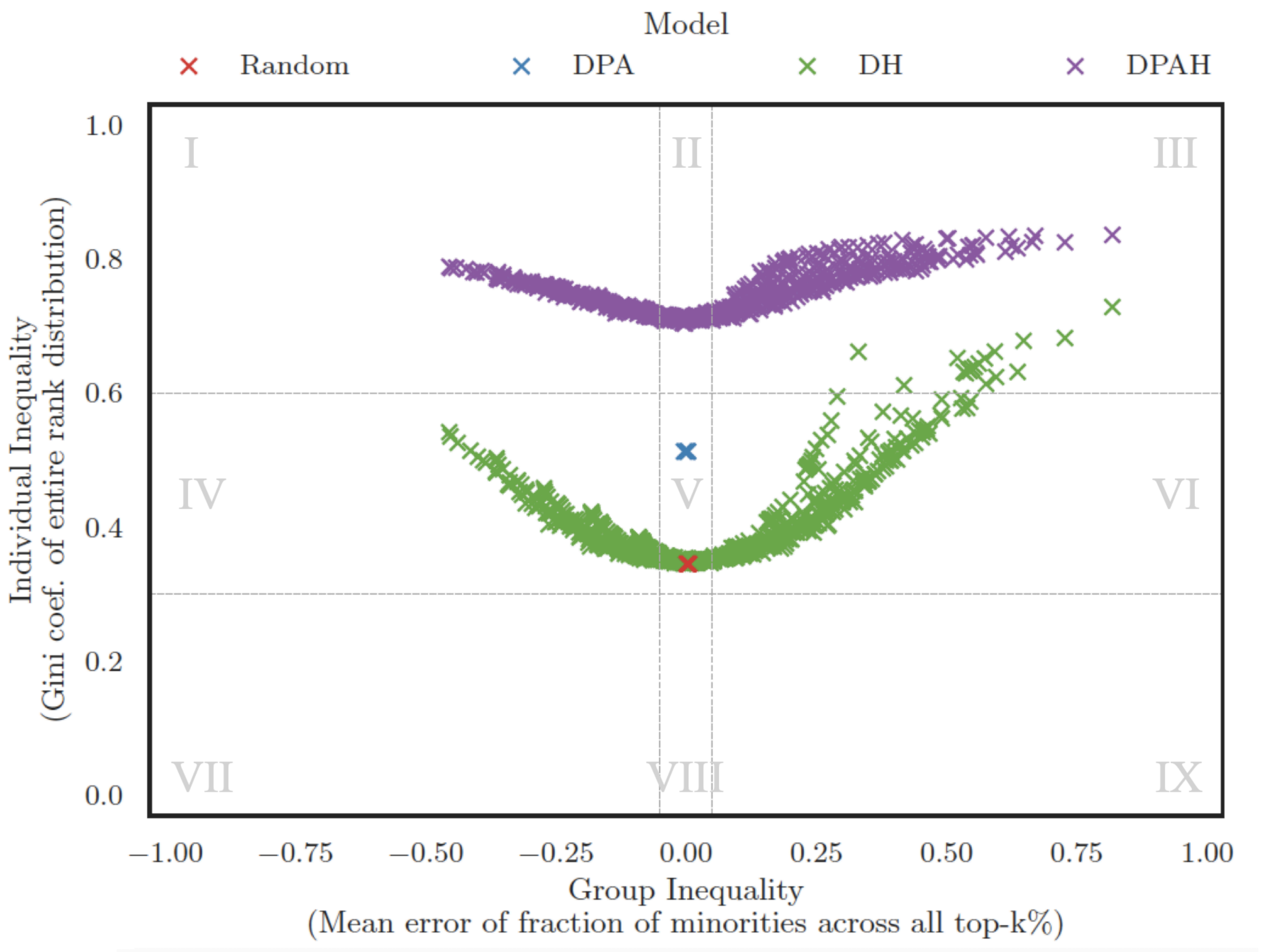
What mechanism of edge formation contributes to ranking inequality and inequity?

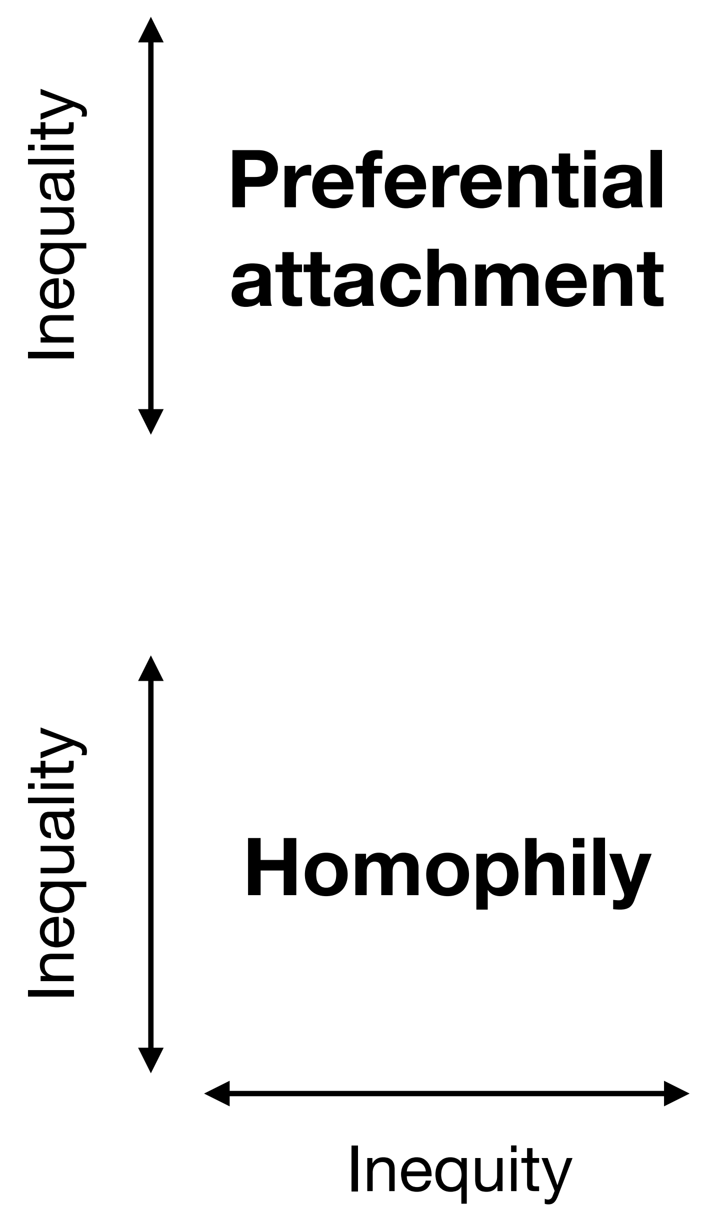


Espín-Noboa et al. 2022

Inequality vs. inequity in PageRank

What mechanism of edge formation contributes to ranking inequality and inequity?





Espín-Noboa et al. 2022

So what?

Real-world scenario

So what?

Real-world scenario

Real-world <u>social network</u>
(people with attributes)

Ranking / RecSys.
(PageRank or WTF)

So what?

Real-world scenario

Real-world social network
(people with attributes)

Ranking / RecSys.
(PageRank or WTF)

1. Identify network structure

Fraction min. $fm=0.3$

Node activity $\gamma M = \gamma m = 3$

Density $d=0.0015$

Homophily Maj. $H_{MM}=0.8$

Homophily min. $H_{mm}=0.4$

(Inequity is driven by homophily and fraction of minorities)

So what?

Real-world scenario

Real-world social network
(people with attributes)

Ranking / RecSys.
(PageRank or WTF)

1. Identify network structure

Fraction min.
fm=0.3

Node activity
γM = γm = 3

Density
d=0.0015

Homophily Maj.
H_{MM}=0.8

Homophily min.
H_{mm}=0.4

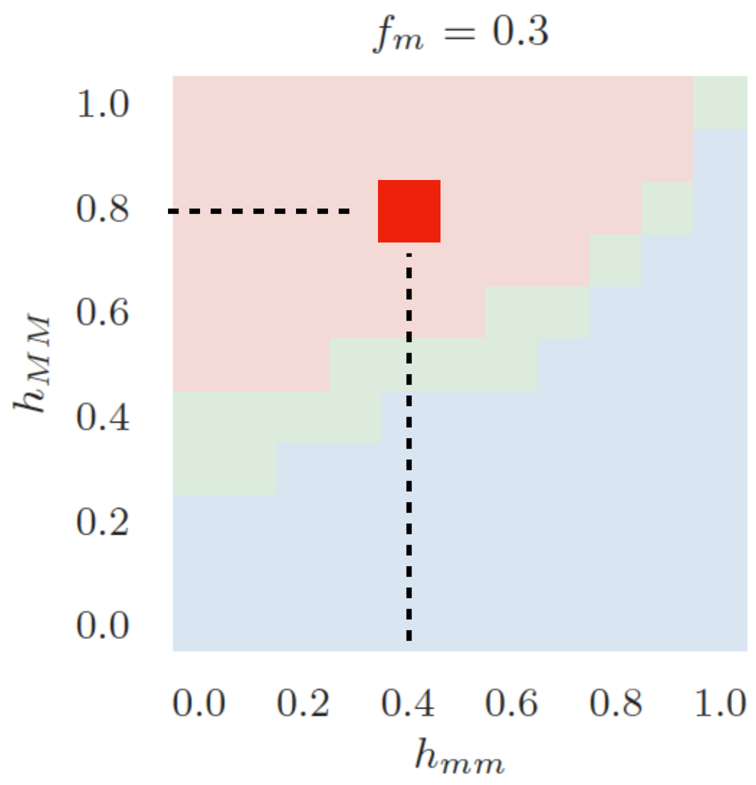
(Inequity is driven by homophily and fraction of minorities)

2. Identify inequality and inequity in ranking

OPEN

Inequality and inequity in network-based ranking and recommendation algorithms

Lisette Espín-Noboa[1,2,3], Claudia Wagner[1,4,5], Markus Strohmaier[1,4,6] & Fariba Karimi[1]



On average minorities are under-represented in top-k's
(Interventions needed)

Real-world (empirical) networks

Network properties

APS
Scientific citations

38% Classical statistical mechanics
62% Quantum statistical mechanics

Hate
Twitter users

11% haters
80% normal

Political Blogs
Hyper-links

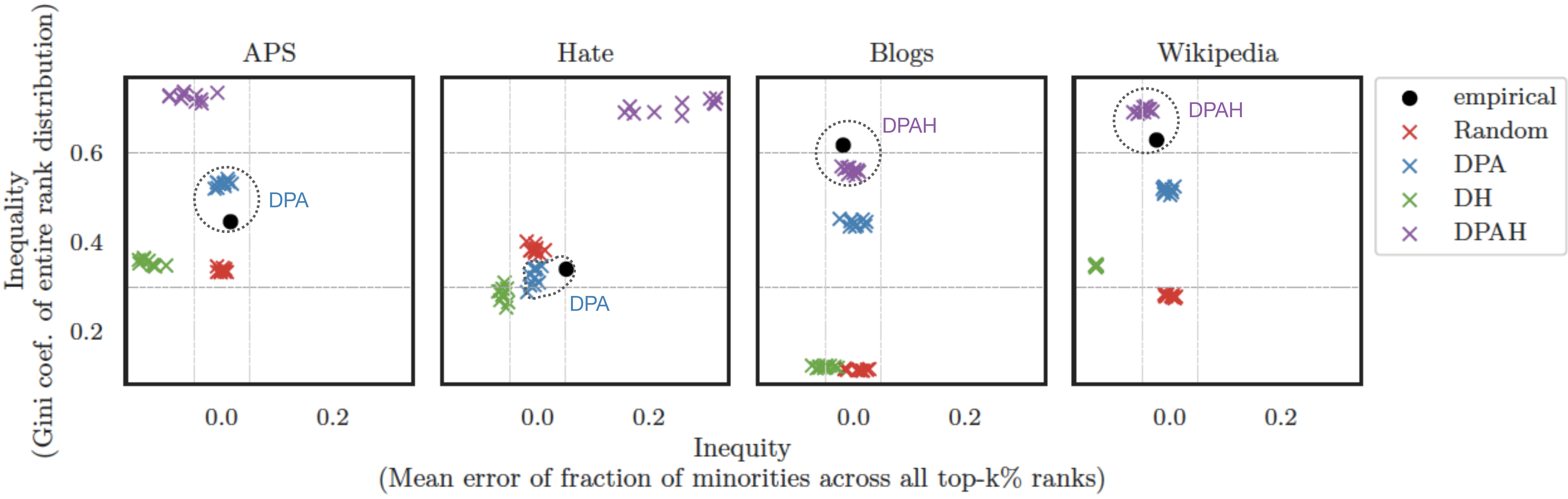
48% left-wing
52% right-wing

Politicians in Wikipedia
Hyper-links

15% women
85% men

Empirical networks

Model fitting (which model explains best the disparity?)



Disparities in vaccination (Agarwal 2021)

Social and Community Context

Political Ideology. To model political ideology, we used the share of votes cast for the Republican candidate (President Donald J. Trump) in the 2020 election collected from USA Today (34). We use Republican vote share because survey evidence suggests that political affiliation is an important predictor of attitudes toward COVID-19 and the COVID-19 vaccine in particular (35). This approach is also consistent with prior research that showed Democratic counties were more likely to reduce mobility in response to stay-at-home-orders for COVID-19 prevention, which used the vote margin in favor of President Trump as a predictor (36).

Segregation Index. We collected Black-White segregation index measures from the 2021 County Health Rankings (37). Based on data from ACS 2015 to 2019, the residential segregation index ranges from 0 (complete integration) to 100 (complete segregation). These values represent the percentage of either Black or White residents who would have to move to yield a distribution that is similar to that of the larger area. Segregation in communities has been associated with a number of important racial disparities in health (38).

Racial Bias. We modeled implicit racial bias using estimates published in prior research (39), which are based on responses collected from Project Implicit, a website where individuals voluntarily complete psychological attitude measures (40). We used the weighted measure of implicit bias, which accounts for the number of observations available for each county. Measures of implicit bias assess the automatic association between race (White and Black) and attitudes (good or bad), with larger values indicating greater bias against Blacks. Such measures of bias have previously been linked to racial disparities in school disciplinary actions (39).

Real-world (empirical) networks

Network properties

Table 1 Empirical networks: Structural properties of five real-world networks: Escorts, Swarthmore42, Caltech36, Wikipedia, and GitHub. In addition to the properties of interest, we report β, the power-law exponent of the degree distribution computed as described in [8]. N_{fit} and m_{fit} represent the number of nodes and minimum degree utilized to generate synthetic networks, respectively.

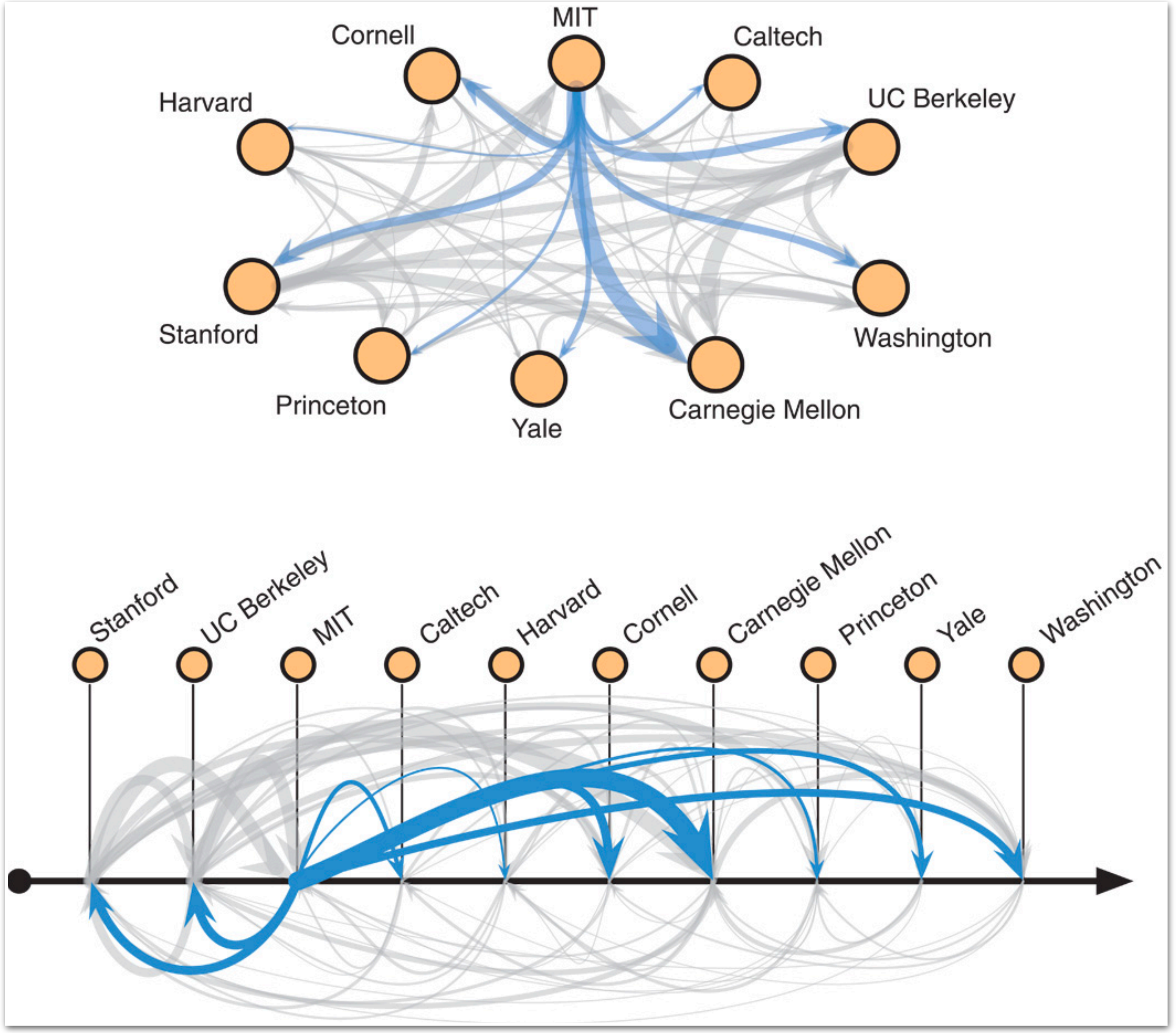
dataset	Escorts	Swarth.	Caltech	Wiki.	GitHub
N	16730	1519	701	2132	37700
m	1	1	1	1	1
class	role	gender	gender	gender	dev
minority	escort	2 (m)	1 (f)	female	1 (ML)
B	0.40	0.49	0.33	0.15	0.26
E	39044	53726	15464	3143	289003
d	0.0003	0.05	0.06	0.001	0.0004
β	2.87	5.50	4.90	2.87	2.54
H	0.00	0.52	0.54	0.64	0.84
N_{fit}	14338	208	179	2893	9830
m_{fit}	2	2	2	2	2

Inequalities in networks

Examples

Inequalities in networks

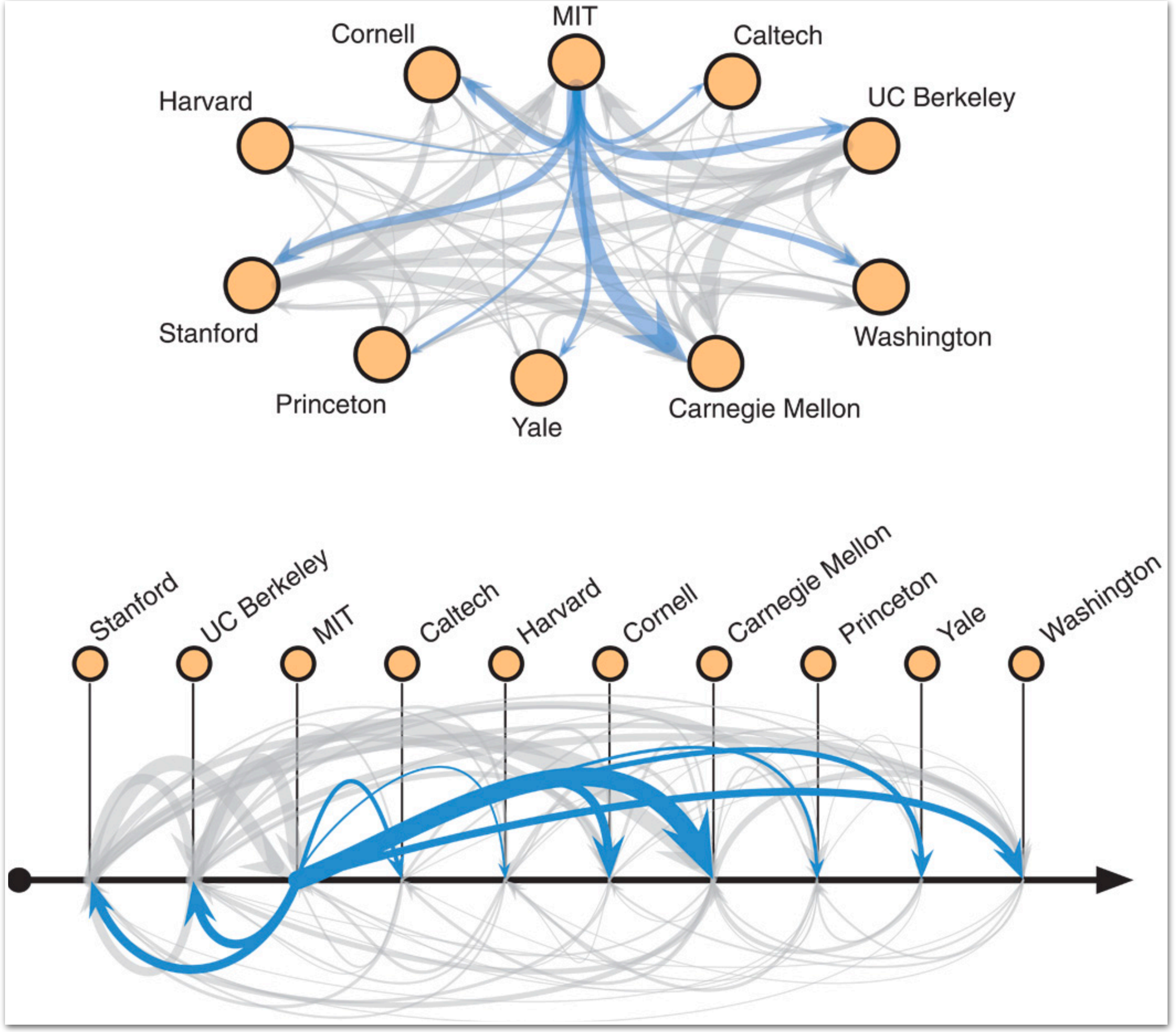
Examples



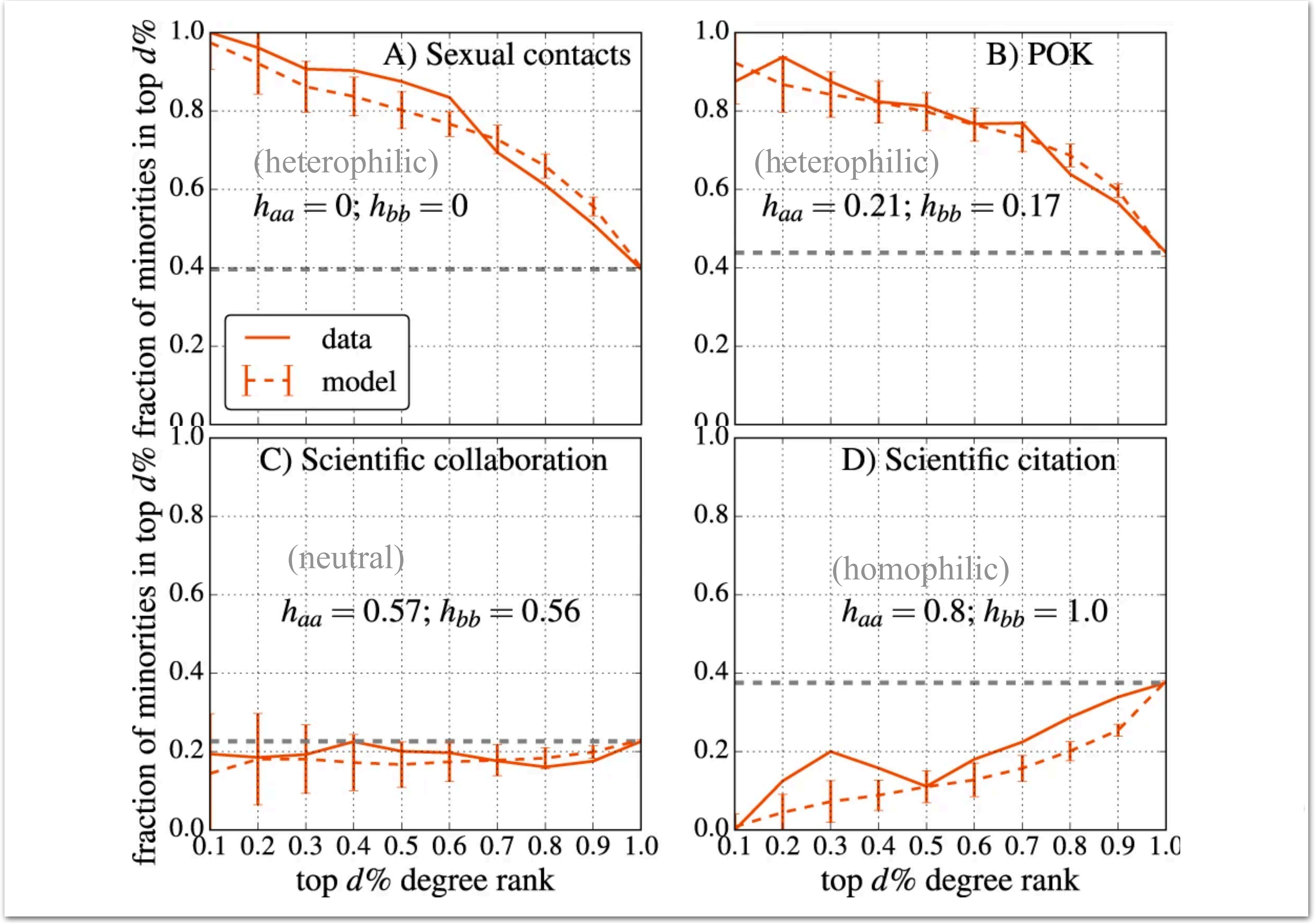
Systematic inequality and hierarchy in faculty hiring networks. Clauset et al. 2015.

Inequalities in networks

Examples



Systematic inequality and hierarchy in faculty hiring networks. Clauset et al. 2015.



Homophily influences ranking of minorities in social networks. Karimi et al. 2018.