Auditing algorithms and assessing their impact using (synthetic) **network data**

Lisette Espín-Noboa, PhD.

2025-02-05 | Topics in Digital Humanities @ CEU

@lespin

Who am I …

… and why am I here?

Who am I …

… and why am I here?

Guayaquil
Ecuador

Saarbrücken
Germany

Cologne
Germany

Vienna
Austria

Engineer
Computer Science
ESPOL

Master
Computer Science
Uni. Saarland & MPI-SWS

Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

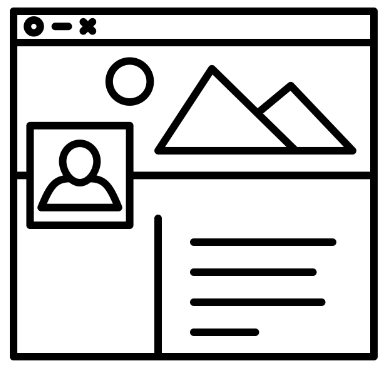
Postdoc
Computational Social Science
Complexity Science Hub, Central European University

Who am I …

… and why am I here?

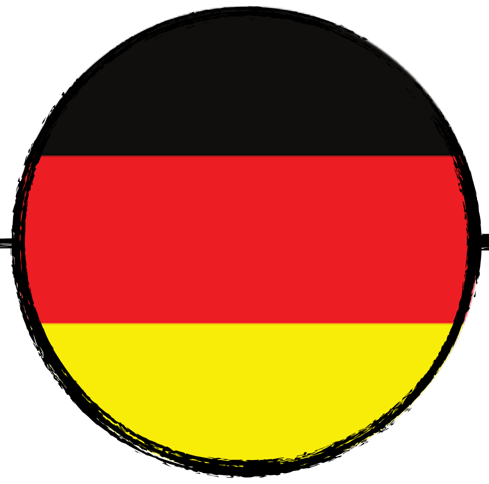
Guayaquil
Ecuador

Engineer
Computer
Science
ESPOL



Social Network
Web Development

Saarbrücken
Germany



Master
Computer
Science
Uni. Saarland & MPI-SWS

Cologne
Germany

Dr. rer. nat.
Computational
Social Science
Uni. Koblenz & GESIS

Vienna
Austria

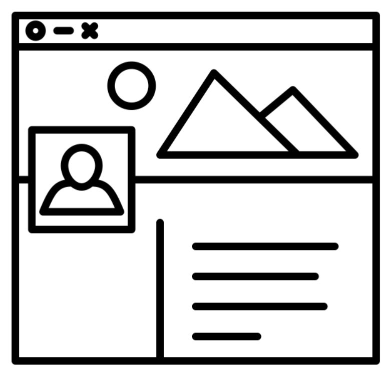
Postdoc
Computational
Social Science
Complexity Science Hub,
Central European University

Who am I …

… and why am I here?

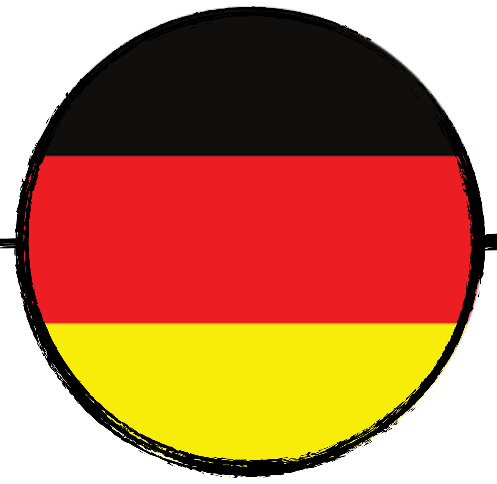
Guayaquil
Ecuador

Engineer
Computer Science
ESPOL

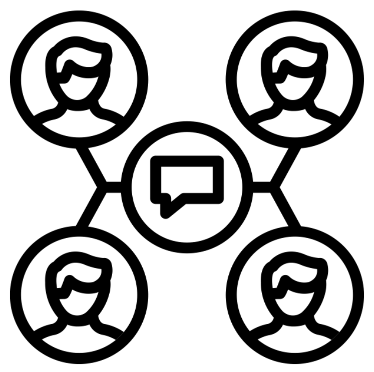


Social Network Web Development

Saarbrücken
Germany



Master
Computer Science
Uni. Saarland & MPI-SWS



Social Network Analysis

Cologne
Germany

Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

Vienna
Austria

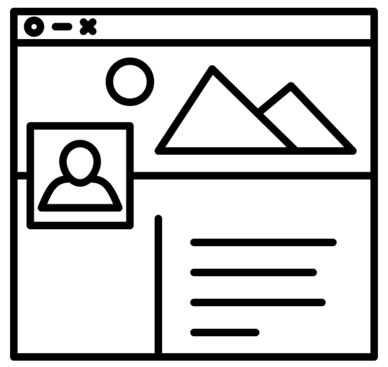
Postdoc
Computational Social Science
Complexity Science Hub, Central European University

Who am I …

… and why am I here?

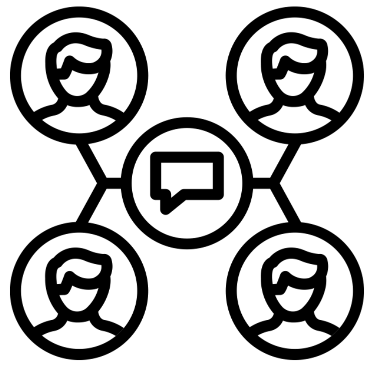
Guayaquil
Ecuador

Engineer
Computer Science
ESPOL



Saarbrücken
Germany

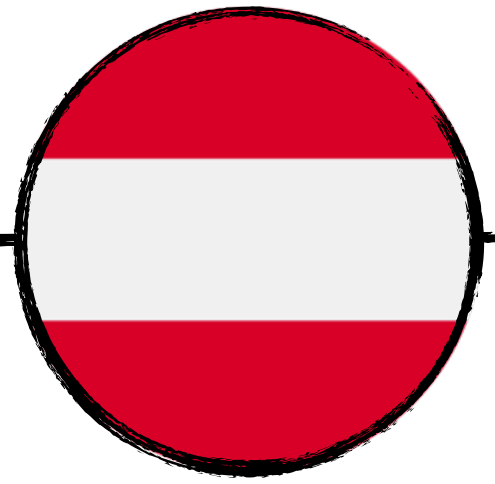
Master
Computer Science
Uni. Saarland & MPI-SWS



Cologne
Germany

Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

Vienna
Austria



Postdoc
Computational Social Science
Complexity Science Hub, Central European University

Social Network Web Development

Social Network Analysis

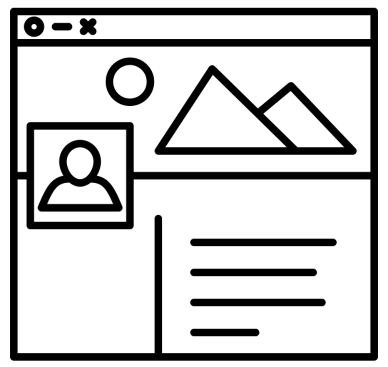
Algorithm Auditing & Impact Assessment for Social Impact

Who am I …

… and why am I here?

Guayaquil
Ecuador

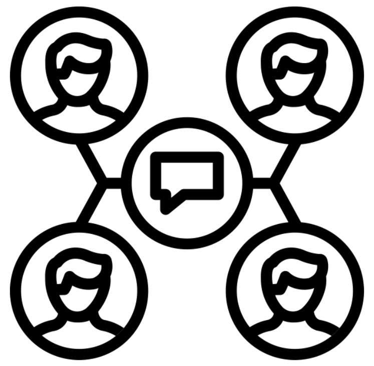
Engineer
Computer Science
ESPOL



Social Network Web Development

Saarbrücken
Germany

Master
Computer Science
Uni. Saarland & MPI-SWS



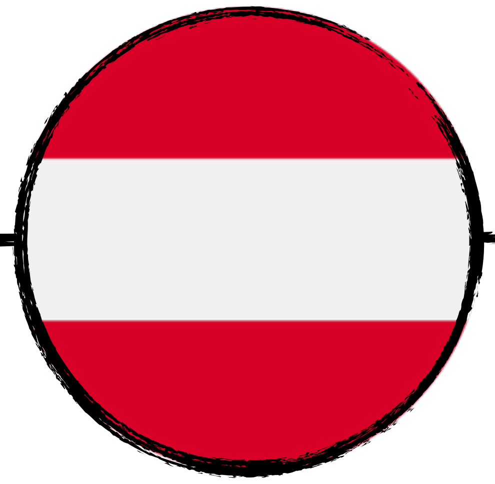
Social Network Analysis

Cologne
Germany

Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

Network **Inequality**

Vienna
Austria



Postdoc
Computational Social Science
Complexity Science Hub, Central European University

Algorithm Auditing & Impact Assessment for Social Impact

Who am I …

… and why am I here?

Guayaquil
Ecuador

Saarbrücken
Germany

Cologne
Germany

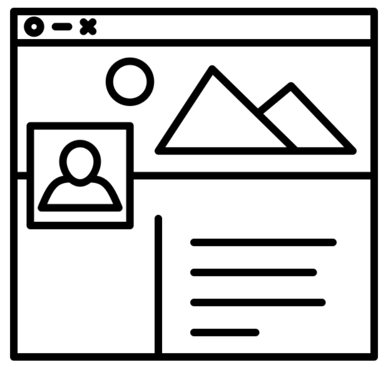
Vienna
Austria

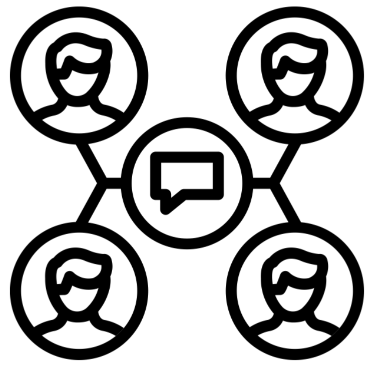
Engineer
Computer Science
ESPOL

Master
Computer Science
Uni. Saarland & MPI-SWS

Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

Postdoc
Computational Social Science
Complexity Science Hub, Central European University





Network **Inequality**

Network **Fairness**, Inferring **Poverty**

Social Network Web Development

Social Network Analysis

Algorithm Auditing & Impact Assessment for Social Impact

Who am I …

… and why am I here?

Guayaquil
Ecuador
Engineer
Computer Science
ESPOL

Saarbrücken
Germany
Master
Computer Science
Uni. Saarland & MPI-SWS

Cologne
Germany
Dr. rer. nat.
Computational Social Science
Uni. Koblenz & GESIS

Vienna
Austria
Postdoc
Computational Social Science
Complexity Science Hub, Central European University

Graz
Austria
Visiting Professor
CSS
Graz University of Technology

Network **Inequality**

Network **Fairness**, Inferring **Poverty**

Social Impact

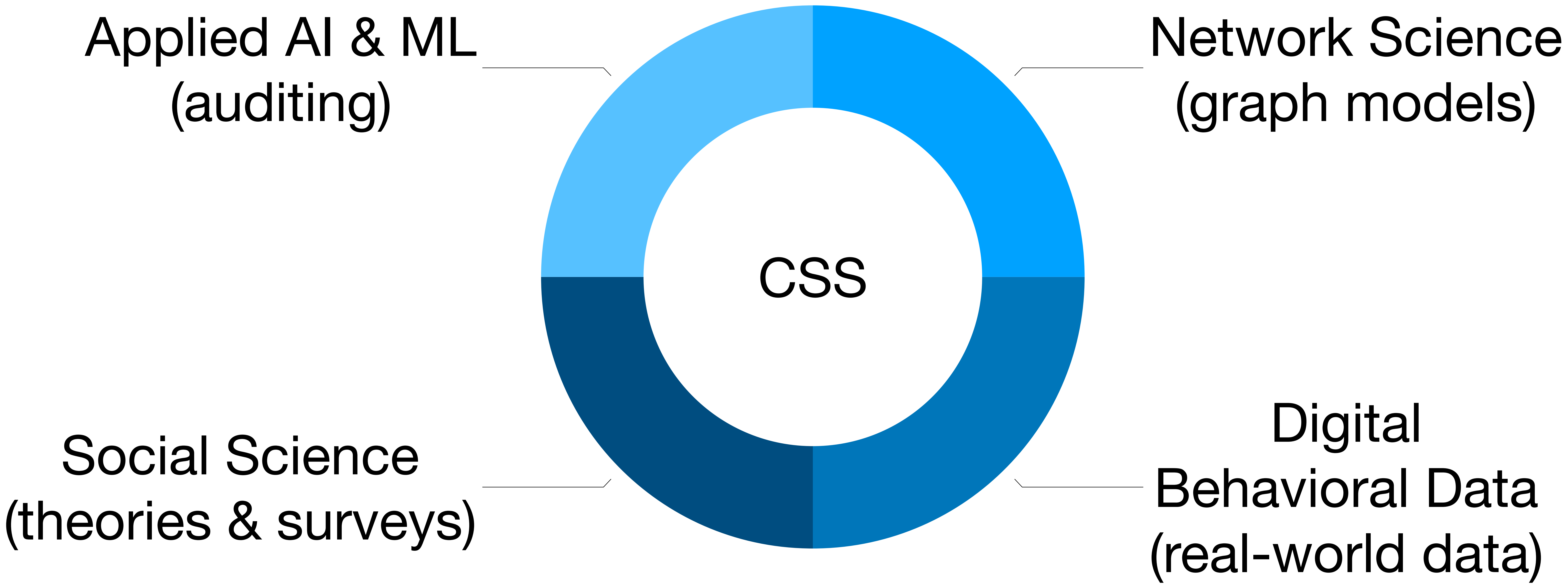
Social Network Web Development

Social Network Analysis

Algorithm Auditing & Impact Assessment for Social Impact

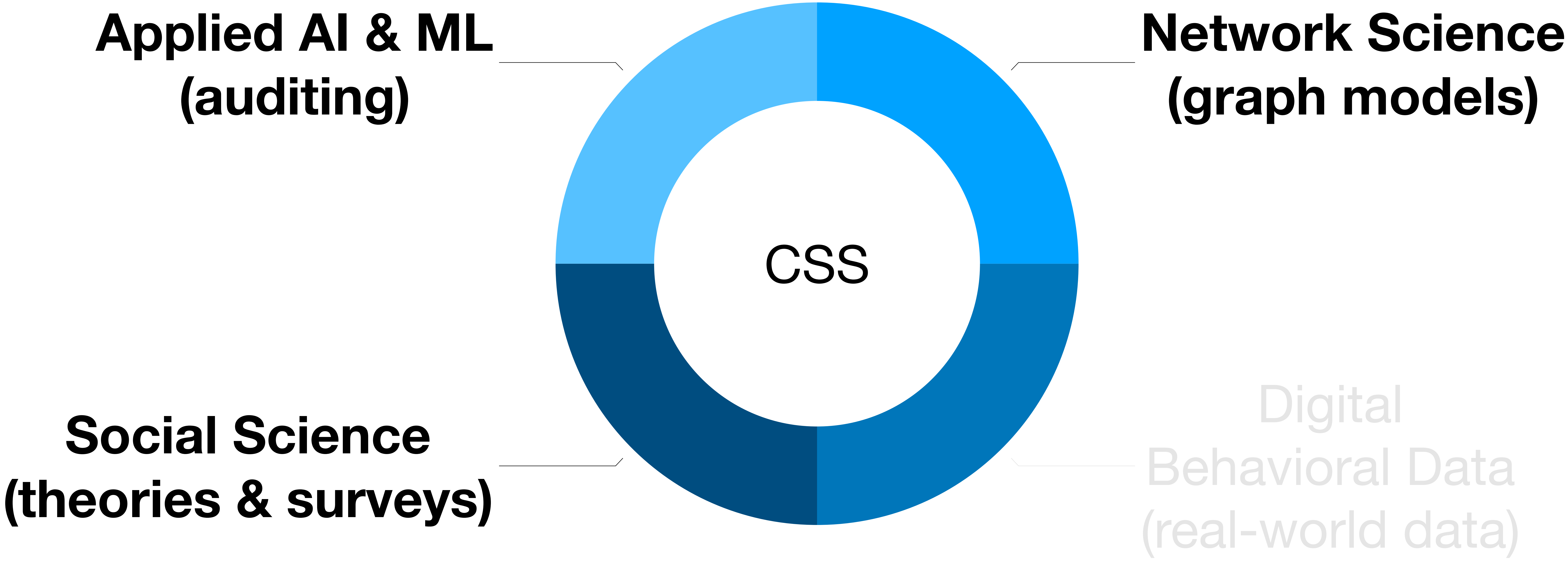
Computational Social Science

research areas



Computational Social Science

research areas



Why do we need algorithm auditing?

How do you reach out to people …

… in academic contexts?

How do you reach out to people …

… in academic contexts?

Word-of-mouth
Friend-of-friend

Google
Scholar

Google
Search

How do you reach out to people …

… in academic contexts?

Word-of-mouth
Friend-of-friend

Google
Scholar

Google
Search

They all suffer from popularity bias

The “rich-gets-richer” and “poor-gets-poorer”effect

How do you reach out to people …

… in academic contexts?

Google Scholar

They all suffer from popularity bias

The “rich-gets-richer” and “poor-gets-poorer”effect

How do you reach out to people …

… in academic contexts?

Google Scholar

They all suffer from popularity bias

The “rich-gets-richer” and “poor-gets-poorer”effect

Searching for: all “network scientists”

How do you reach out to people …

… in academic contexts?

Google Scholar

They all suffer from popularity bias

The “rich-gets-richer” and “poor-gets-poorer”effect

Searching for: all “network scientists”

- 1565 results, 20%women (1 in top-10)

How do you reach out to people …

… in academic contexts?

Google Scholar

They all suffer from popularity bias

The "rich-gets-richer" and "poor-gets-poorer" effect

Searching for: all "network scientists"

- 1565 results, 20% women (1 in top-10)
- 1565 results, 0.64% Black (0 in top-10)
  Browse until page 150 of 157 to find the top-10 most cited

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Google Scholar

They all suffer from popularity bias

ch-gets-richer” and “poor-gets-poorer”effect

Because people don’t update their profiles

“network scientists”

- 1565 results, **20%women** (1 in top-10)
- 1565 results, **0.64% Black** (0 in top-10)
  Browse until page 150 of 157 to find the top-10 most cited

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Google Scholar

Problem 2.
A biased ranking

They all suffer from popularity bias

The "rich-gets-richer" and "poor-gets-poorer"effect

Searching for: all "network scientists"

- 1565 results, 20%women (**1 in top-10**)
- 1565 results, 0.64% Black (**0 in top-10**)
  Browse until page 150 of 157 to find the top-10 most cited

Because results are ranked by citations

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Google Scholar

Problem 2.
A biased ranking

They all suffer from popularity bias
The "rich-gets-richer" and "poor-gets-poorer"effect

Searching for: all "network scientists"

- 1565 results, 20%women (**1 in top-10**)
- 1565 results, 0.64% Black (**0 in top-10**)
  Browse until page 150 of 157 to find the to

Because results are ranked by citations

Kong, H., Martin-Gutierrez, S., & Karimi, F. (2022). **Influence of the first-mover advantage on the gender disparities in physics citations**. Communications Physics, 5(1), 243.

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Google Scholar

Problem 2.
A biased ranking

They all suffer from popularity bias
The "rich-gets-richer" and "poor-gets-poorer"effect

Searching for: all "network scientists"

- 1565 results, 20%women (1 in top-10)
- 1565 results, 0.64% Black (0 in top-10)
  Browse until page 150 of 157 to find the top-10 most cited

We live in “algorithmically infused societies”

that influence our own decisions

Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021).
Measuring **algorithmically infused societies**. Nature, 595(7866), 197-204.

We live in “algorithmically infused societies”

that influence our own decisions

Criminal courts’ artificial intelligence: the way it reinforces bias and discrimination

Md. Abdul Malek

Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021).
Measuring **algorithmically infused societies**. Nature, 595(7866), 197-204.

We live in “algorithmically infused societies”

that influence our own decisions

Criminal courts’ artificial intelligence: the way it reinforces bias and discrimination

Md. Abdul Malek

Algorithmic Political Bias in Artificial Intelligence Systems

Uwe Peters

Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021).
Measuring **algorithmically infused societies**. Nature, 595(7866), 197-204.

We live in “algorithmically infused societies”

that influence our own decisions

Criminal courts’ artificial intelligence: the way it reinforces bias and discrimination

Md. Abdul Malek

Algorithmic Political Bias in Artificial Intelligence Systems

Uwe Peters

The unfairness of popularity bias in music recommendation: A reproducibility study

Dominik Kowald*, Markus Schedl, Elisabeth Lex

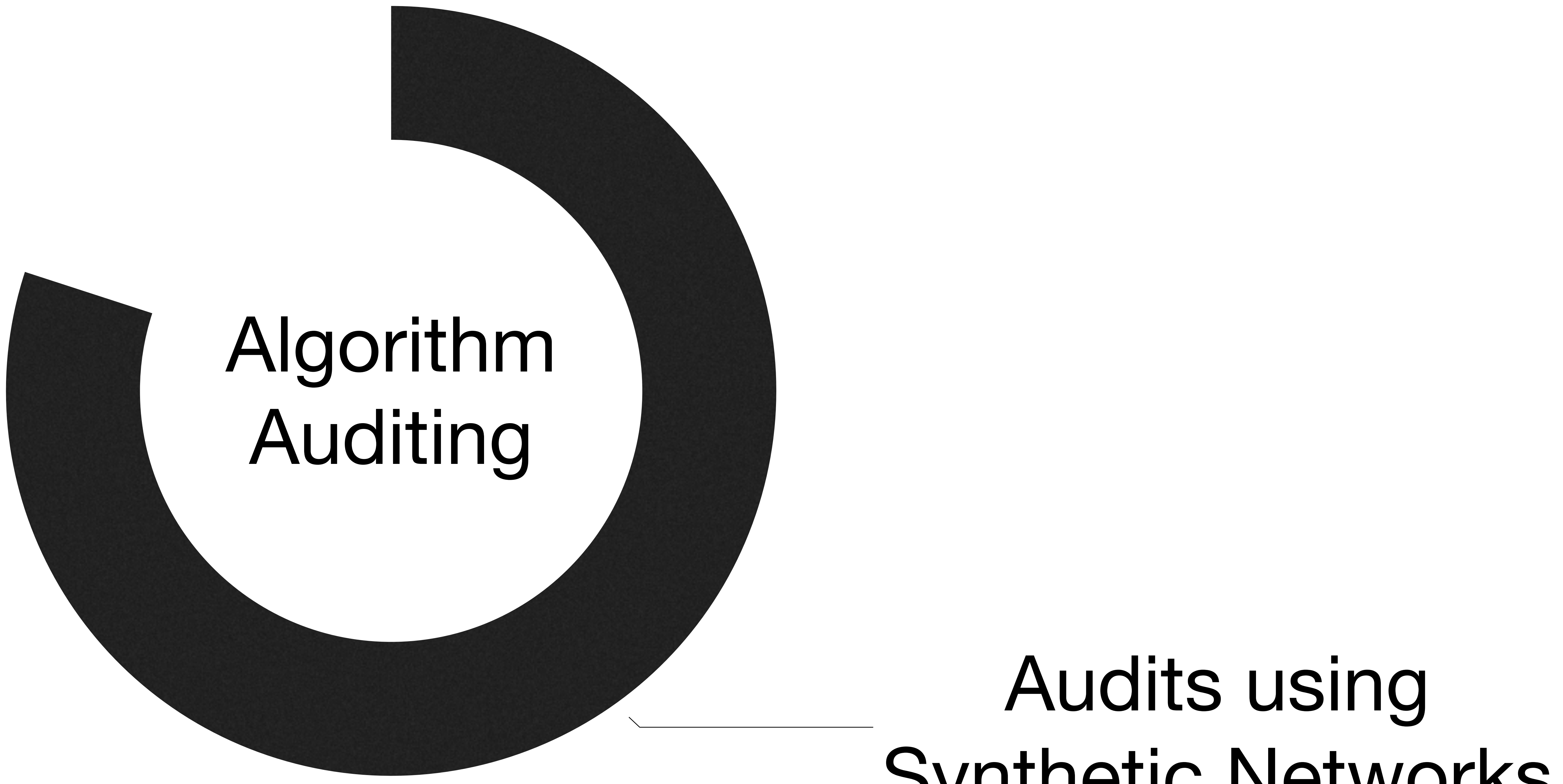
Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021).
Measuring **algorithmically infused societies**. Nature, 595(7866), 197-204.

We need **auditing** to spot
when algorithms fail,
and understand **why they fail**.

Algorithm Auditing

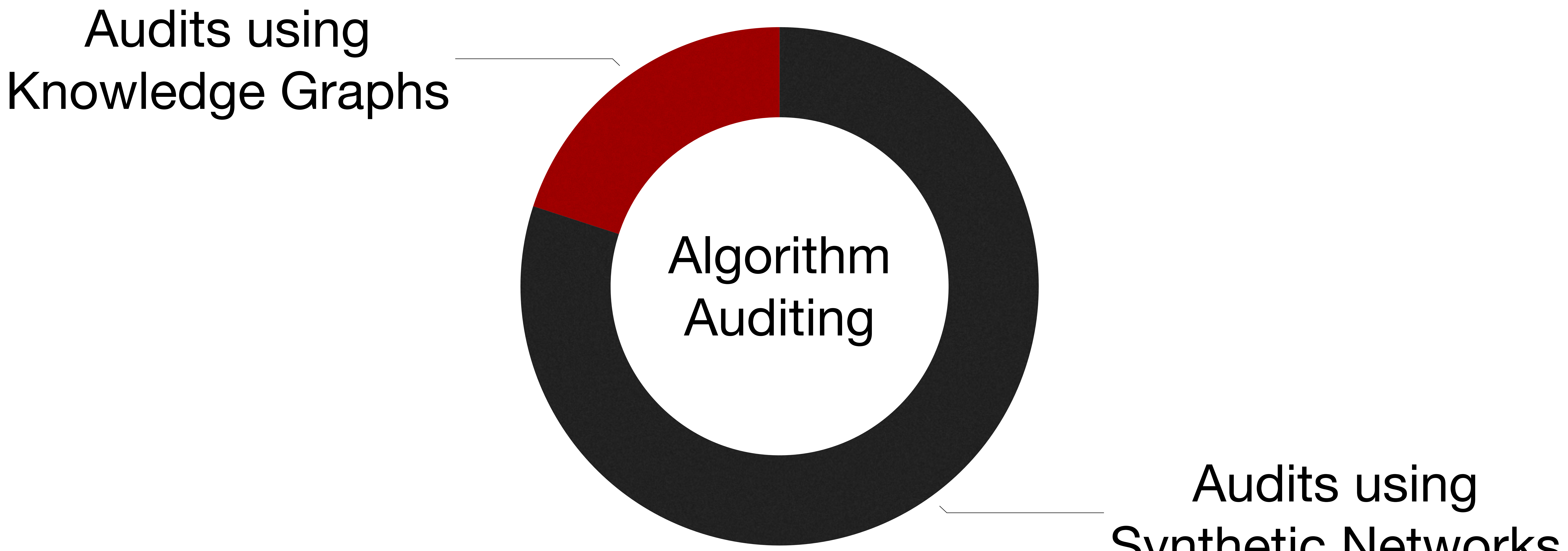
using network data

Algorithm Auditing
using network data



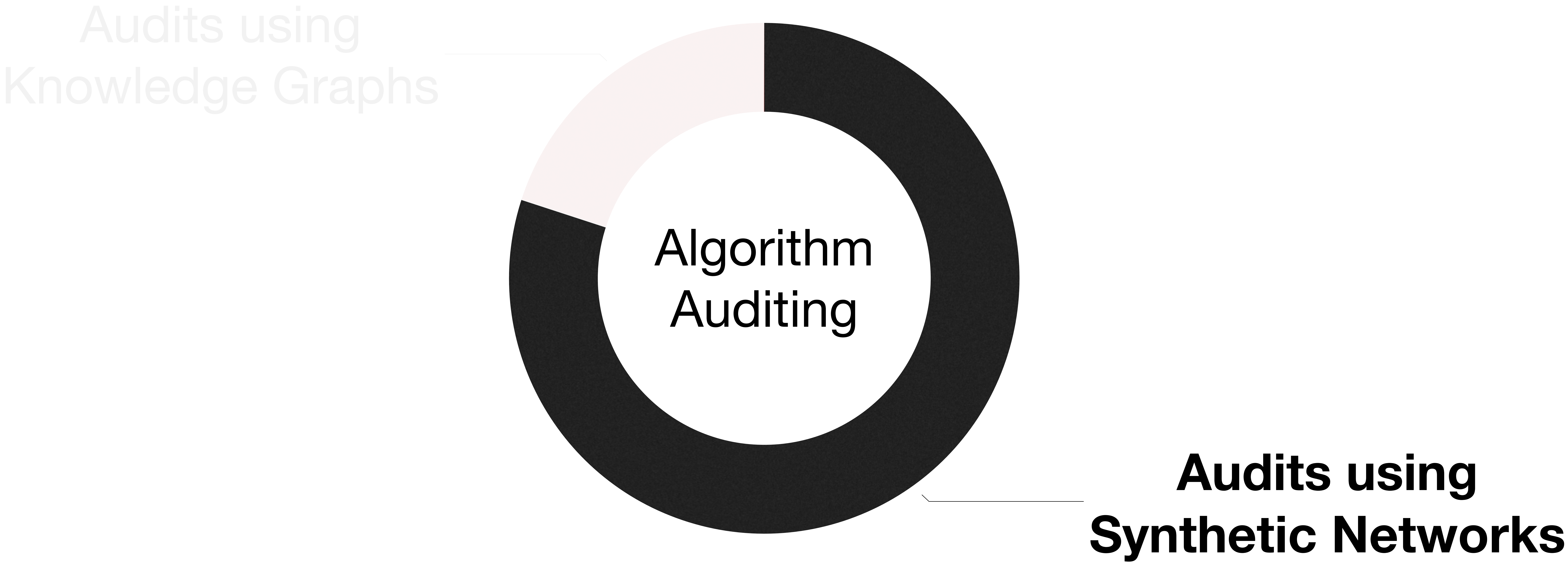
Algorithm Auditing

using network data



Algorithm Auditing

using network data

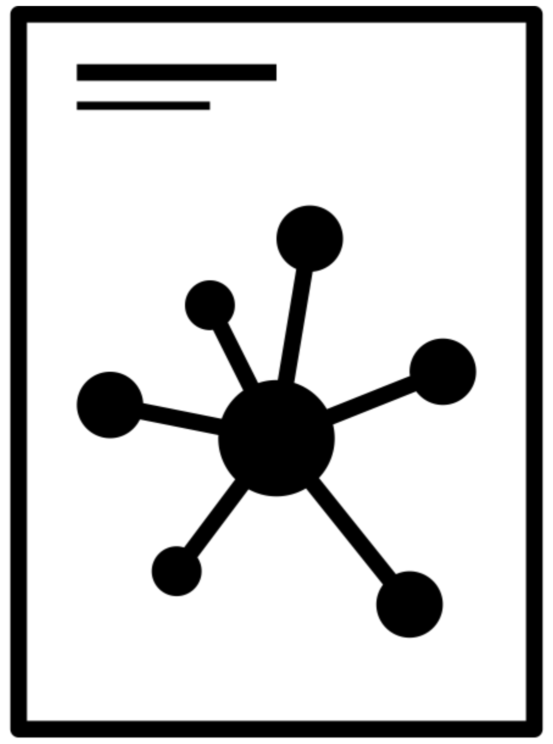


The CS approach …

… evaluating performance on real-world network data

The CS approach …

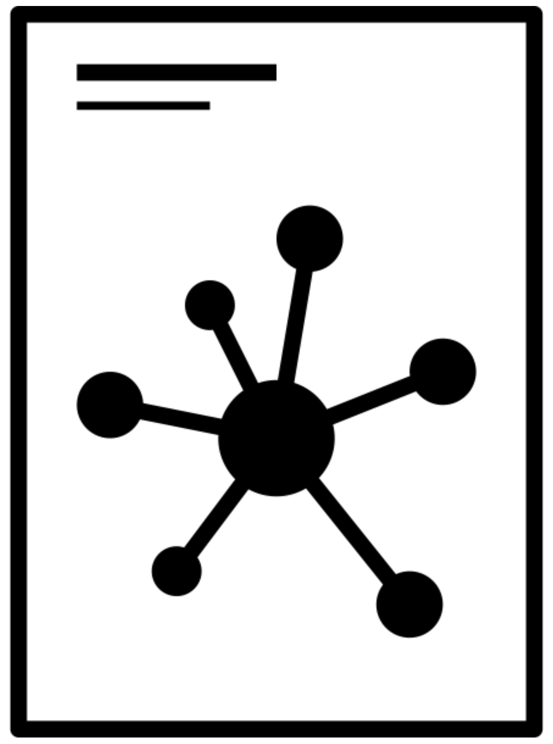
… evaluating performance on real-world network data



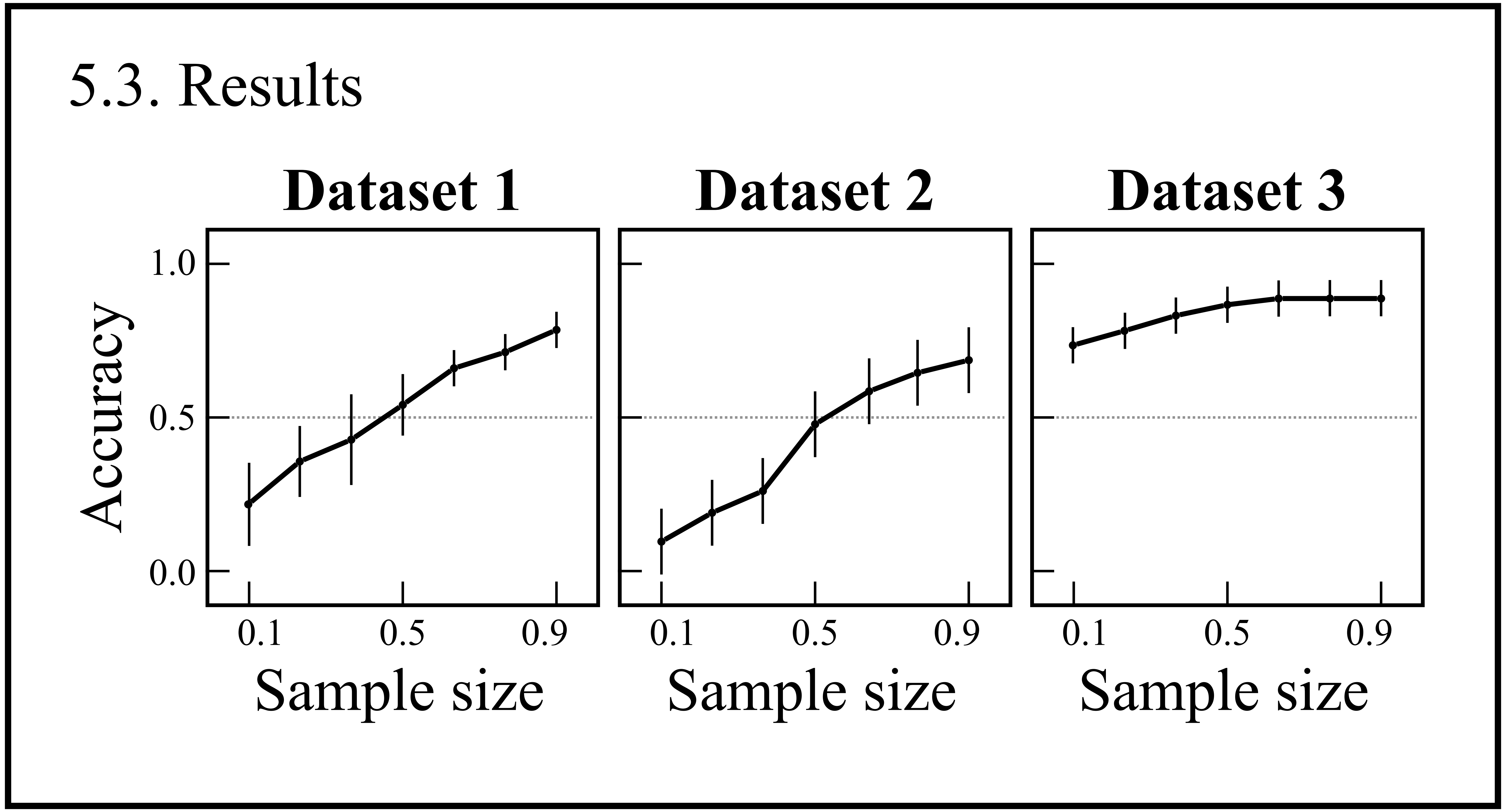
Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).

The CS approach …

… evaluating performance on real-world network data

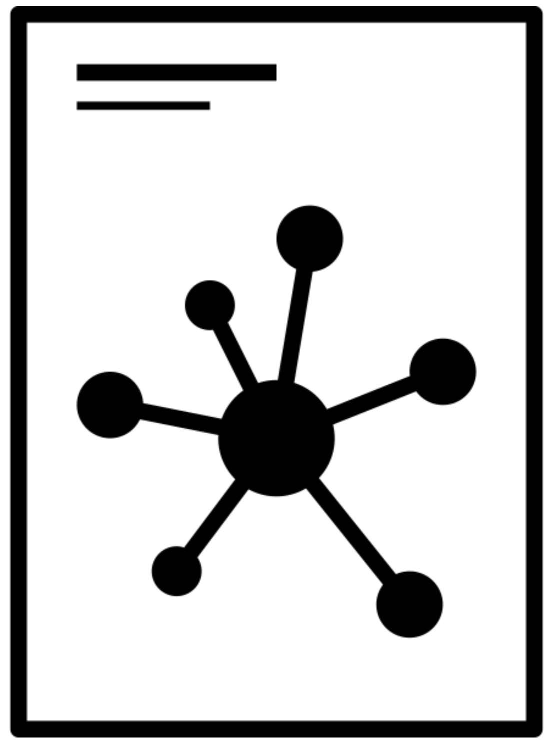


Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).

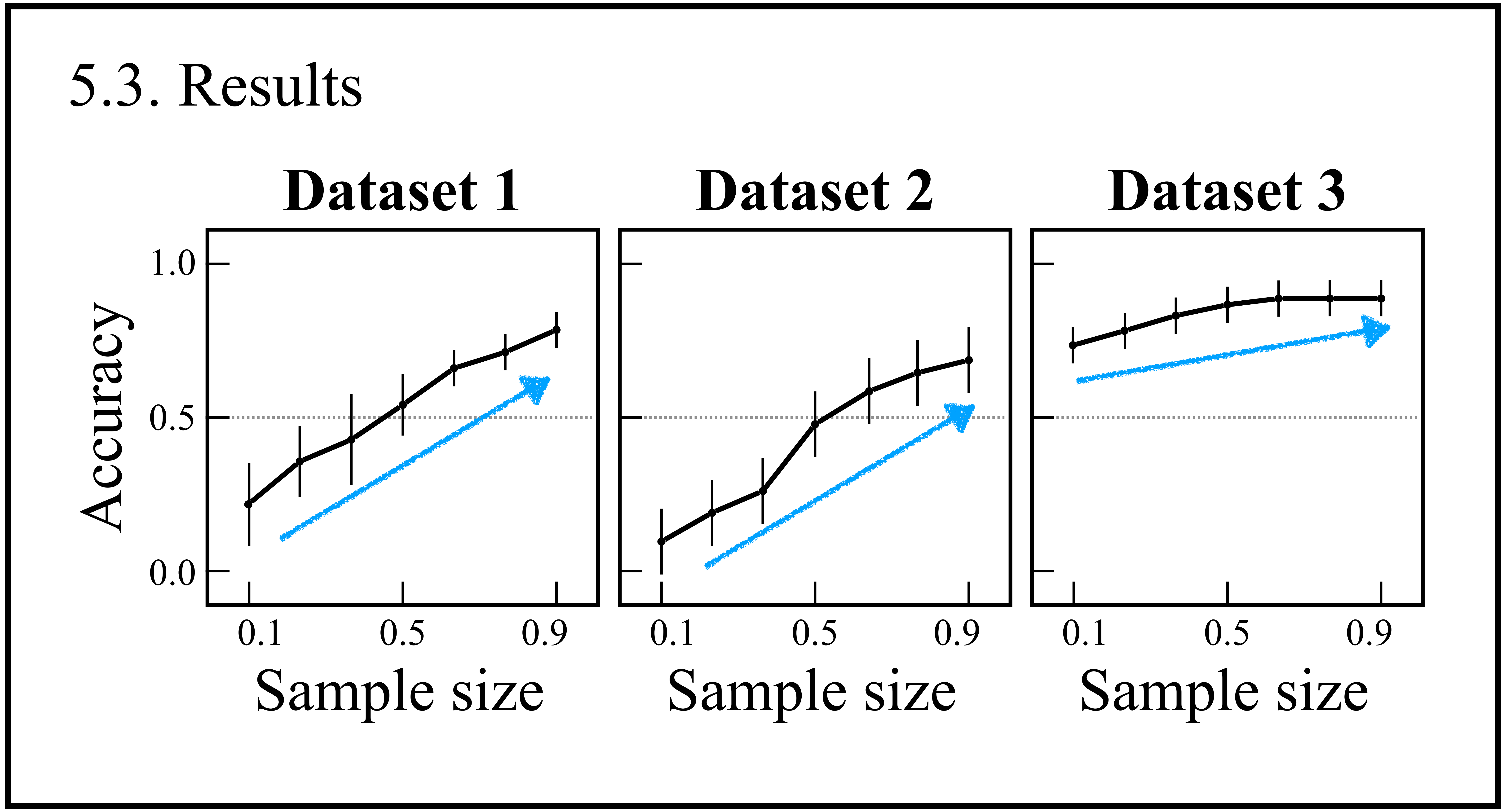


The CS approach …

… evaluating performance on real-world network data



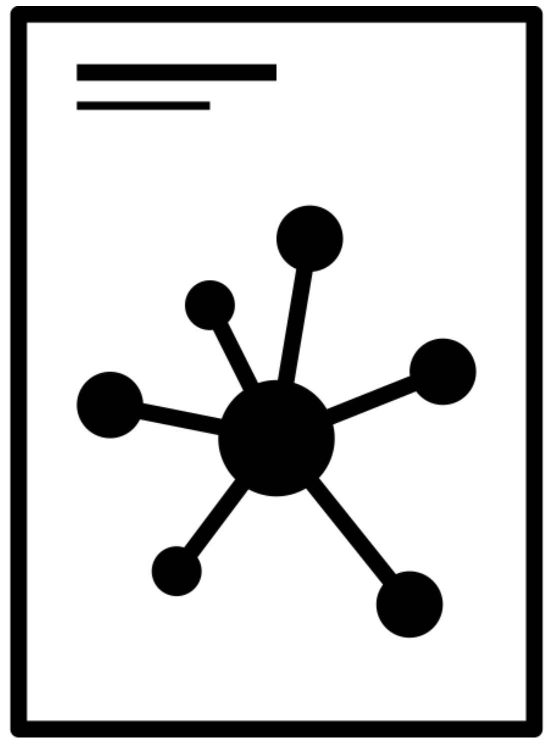
Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).



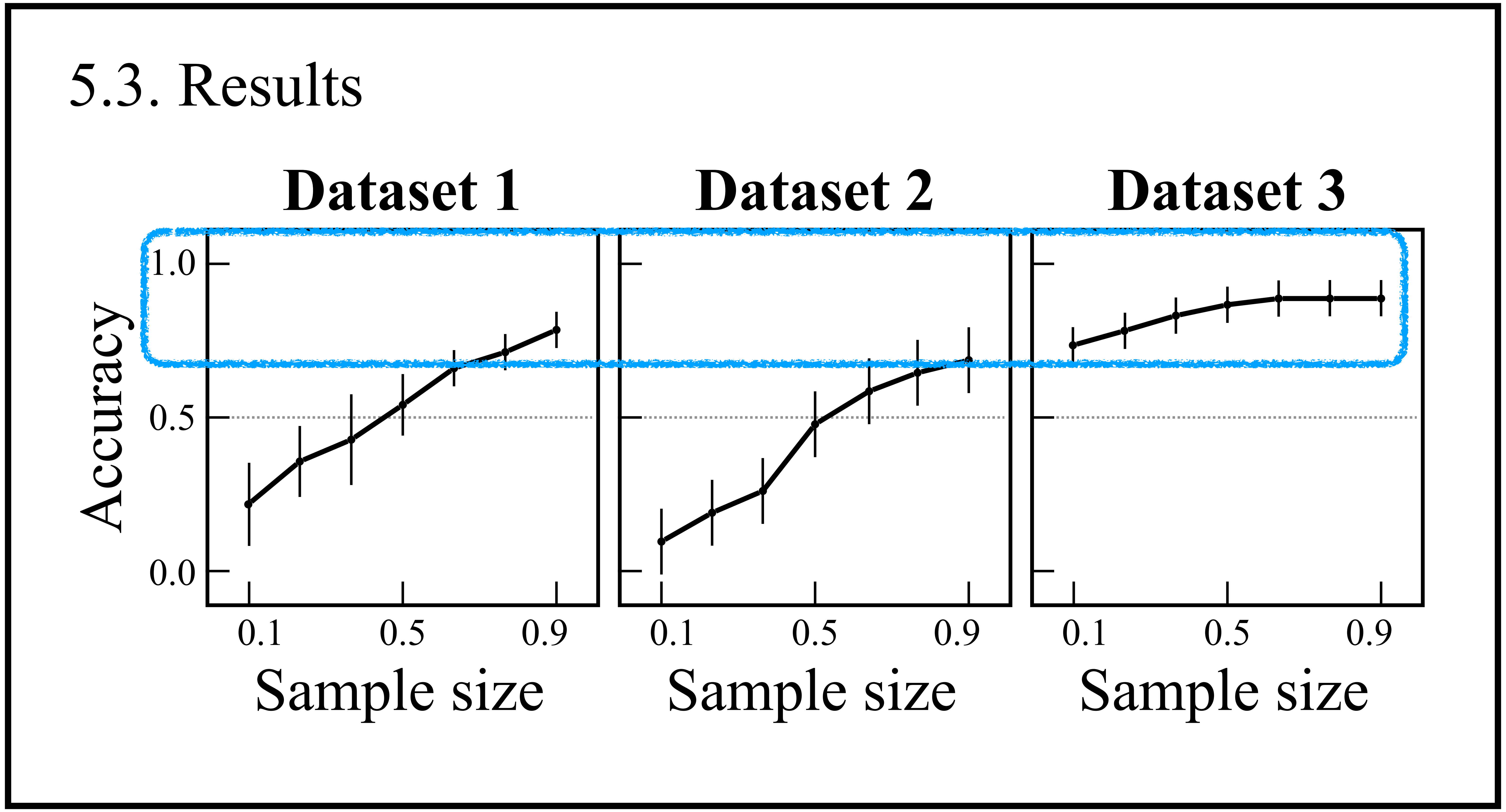
1 The larger the training sample, the better the accuracy

The CS approach …

… evaluating performance on real-world network data



Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).

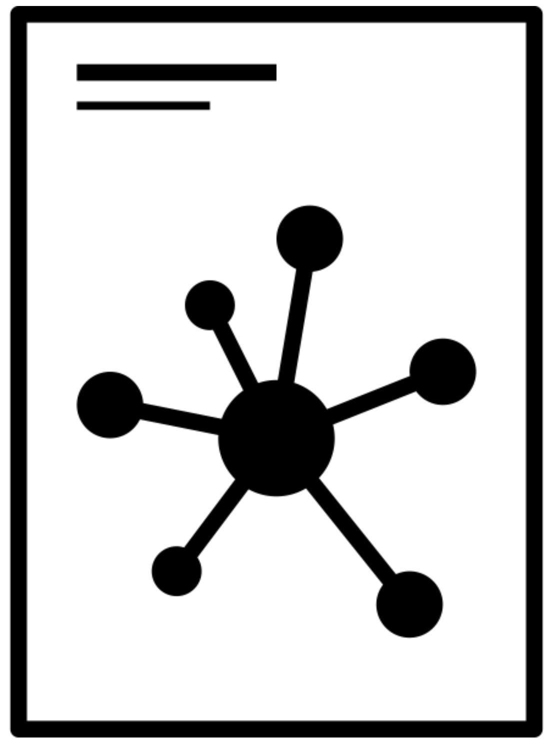


1 The larger the training sample, the better the accuracy

2 Accuracy “seems” to correlate w/ net. structure

The CS approach …

… evaluating performance on real-world network data



Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).

5.1. Datasets

Dataset	Nodes	Edges	Directed	Assortativity
Dataset 1	500	10K	Yes	0.1
Dataset 2	2K	345K	No	0.1
Dataset 3	2.5K	433K	Yes	0.9

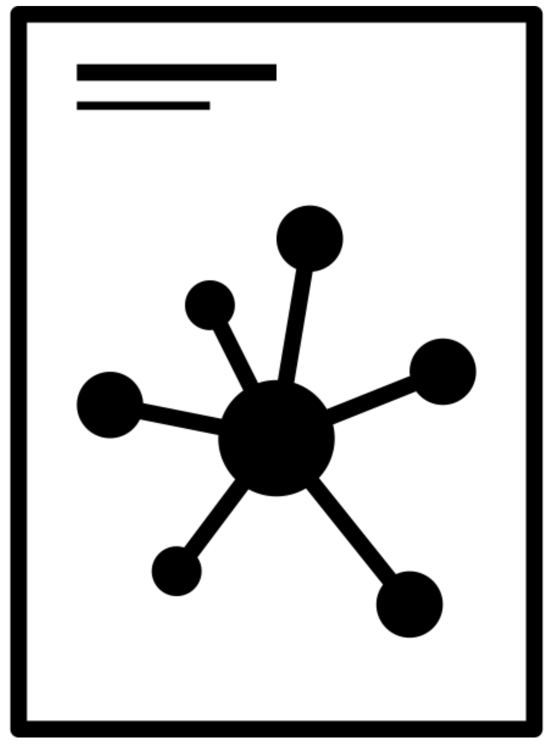
1 The larger the training sample, the better the accuracy

2 Accuracy "seems" to correlate w/ net. structure

3 It "seems" to work best for assortative & directed net.

The CS approach …

… evaluating performance on real-world network data



Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).

5.1. Datasets

Dataset	Nodes	Edges	Directed	Assortativity
Dataset 1	500	10K	Yes	0.1
Dataset 2	2K	345K	No	0.1
Dataset 3	2.5K	433K	Yes	0.9
Your Dataset	**10K**	**1M**	**No**	**0.9**

1 The larger the training sample, the better the accuracy

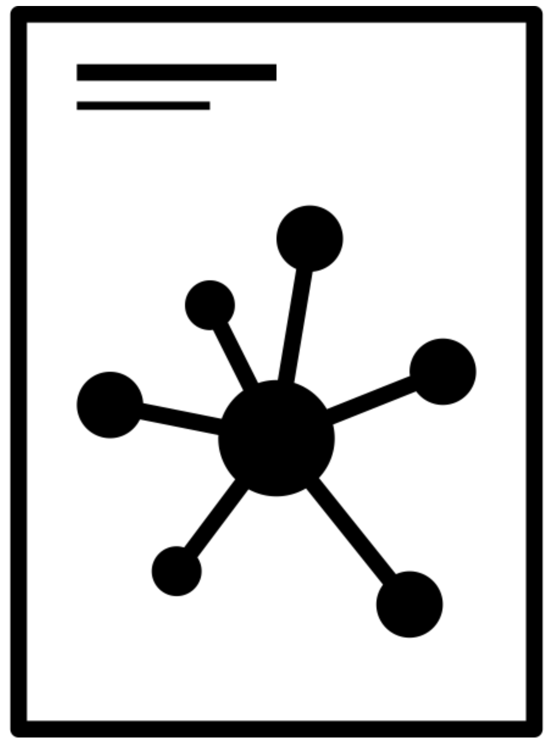
2 Accuracy “seems” to correlate w/ net. structure

3 It “seems” to work best for assortative & directed net.

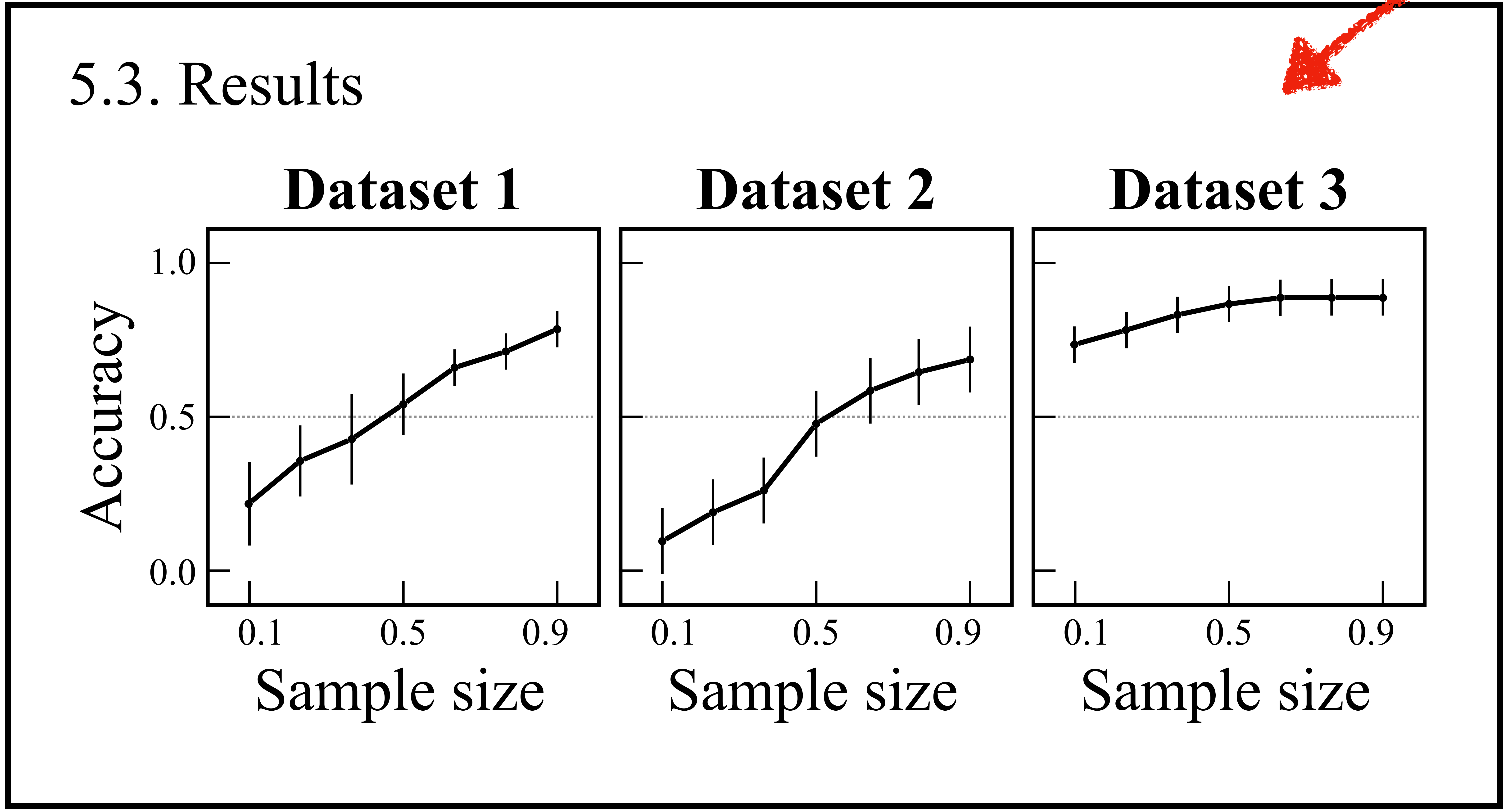
4 What about other types of networks?

The CS approach …

… evaluating performance on real-world network data



Smith et al.
Novel **node classification algorithm** outperforms state-of-the-art algorithm X. Top-tier Venue (2024).



This is not enough to understand the "WHY" of an algorithm's outcomes!

Can **synthetic data** help audit algorithms based on **network data**?

Modeling (social) networks

to generate synthetic data

Modeling (social) networks

to generate synthetic data

Real data can represent only **one realization** of the social structure.

Steinbacher, M., et al. (2021). Advances in the agent-based modeling of economic and social behavior. SN Business & Economics, 1(7), 99.

Modeling (social) networks

to generate synthetic data

Real data can represent only **one realization** of the social structure.

Realistic synthetic networks can represent **multiple realizations,** including cases we have never seen (what-if scenarios).

Steinbacher, M., et al. (2021). Advances in the agent-based modeling of economic and social behavior. SN Business & Economics, 1(7), 99.

Modeling (social) networks

to generate synthetic data

Real data can represent only **one realization** of the social structure.

Realistic synthetic networks can represent **multiple realizations,** including cases we have never seen (what-if scenarios).

Network models can generate synthetic networks by assuming one or more **mechanisms of edge formation.**

Steinbacher, M., et al. (2021). Advances in the agent-based modeling of economic and social behavior. SN Business & Economics, 1(7), 99.

Holme, P., & Liljeros, F. (2015). Mechanistic models in computational social science. Frontiers in Physics, 78.

Modeling (social) networks

to generate synthetic data

Real data can represent only **one realization** of the social structure.

Realistic synthetic networks can represent **multiple realizations,** including cases we have never seen (what-if scenarios).

Network models can generate synthetic networks by assuming one or more **mechanisms of edge formation.**

Steinbacher, M., et al. (2021). Advances in the agent-based modeling of economic and social behavior. SN Business & Economics, 1(7), 99.

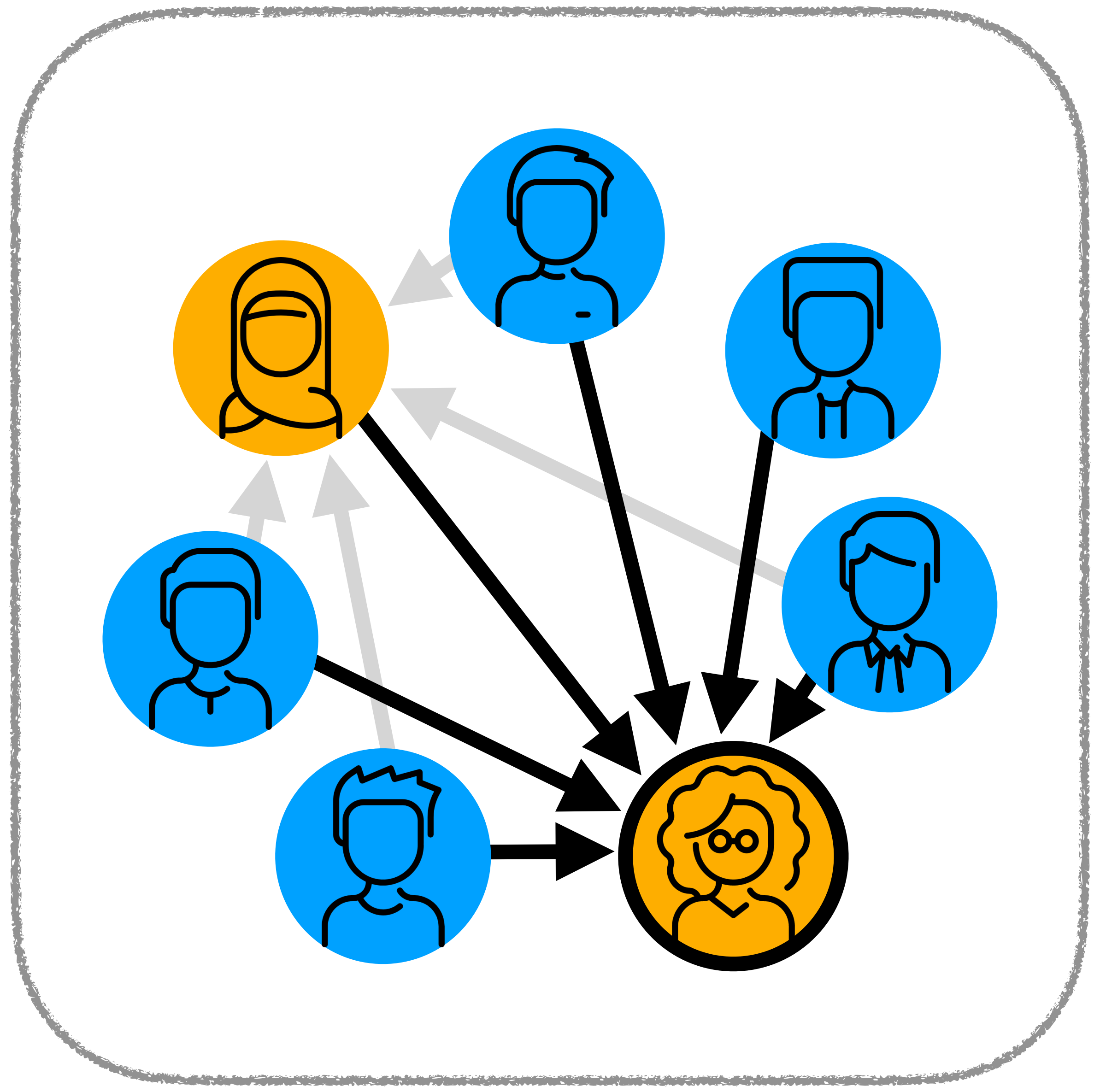
Holme, P., & Liljeros, F. (2015). Mechanistic models in computational social science. Frontiers in Physics, 78.

Social mechanisms

of edge formation

Social mechanisms

of edge formation

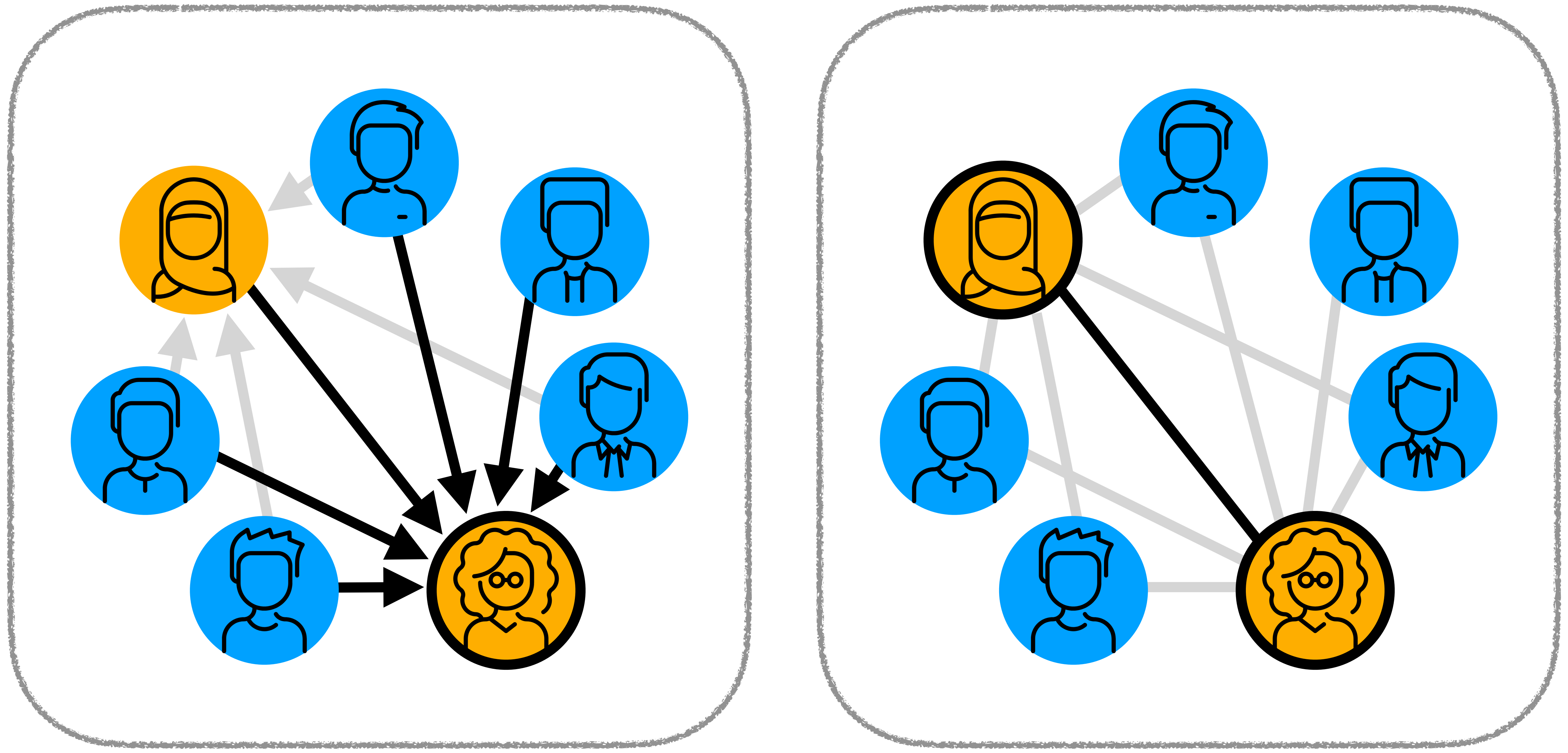


Preferential Attachment
(popularity)

Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56-63.
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. science, 286(5439), 509-512.

Social mechanisms

of edge formation



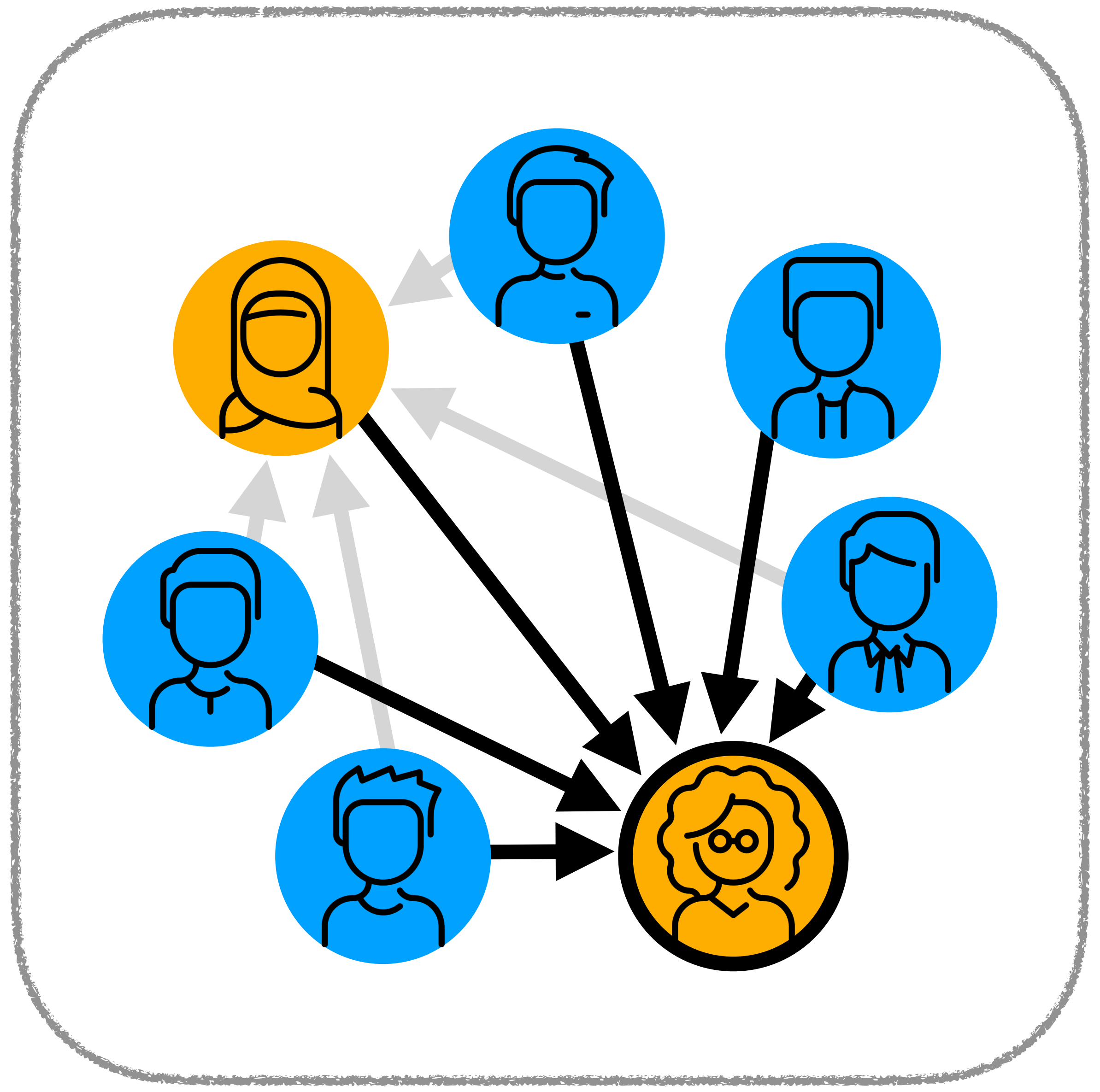
Preferential Attachment
(popularity)

Homophily
(similarity)

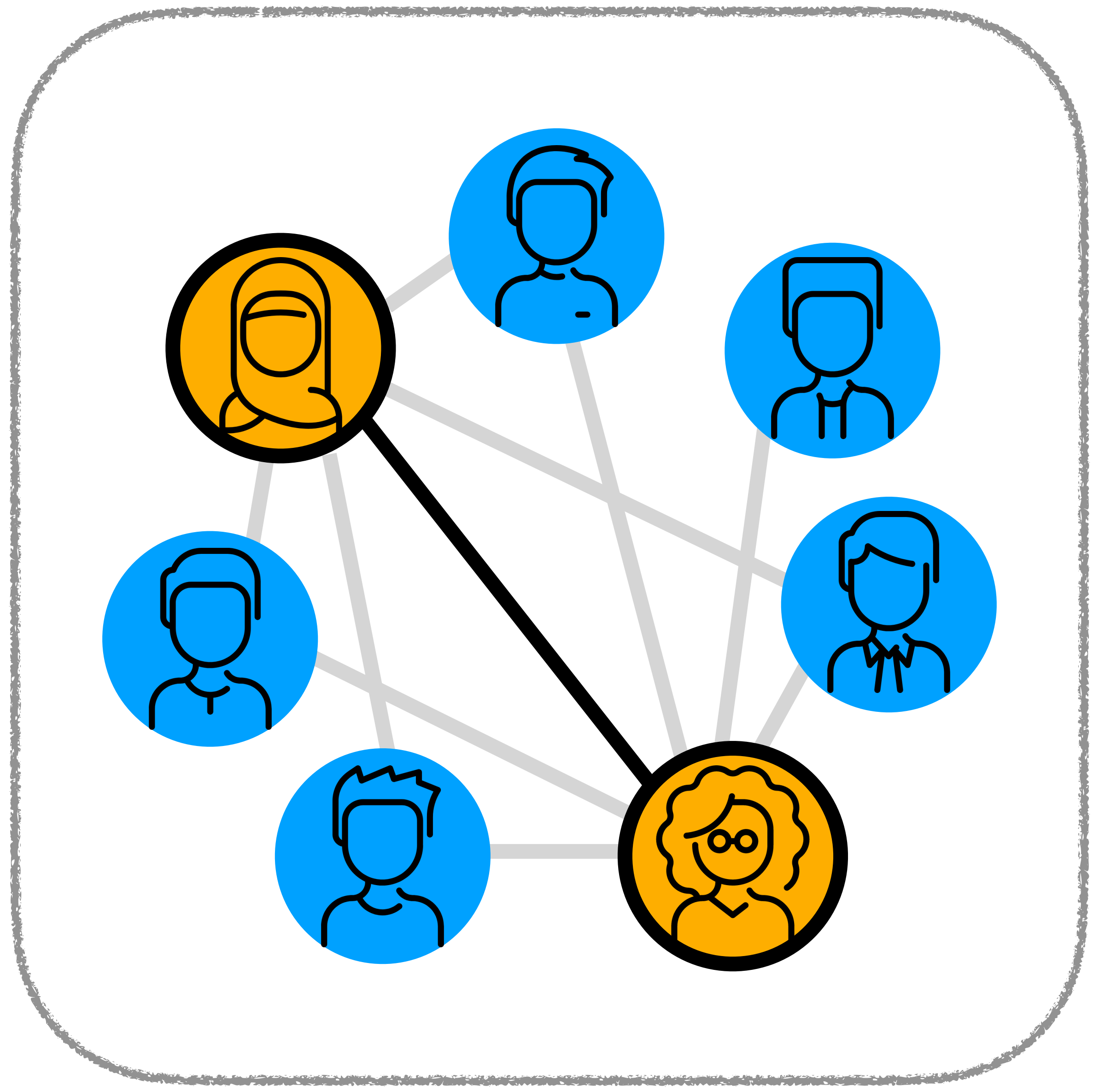
Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56-63.
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. science, 286(5439), 509-512.
McPherson, M., et al, (2001). Birds of a feather: Homophily in social networks. Annual review of sociology, 27(1), 415-444.
Newman, M. E. (2003). Mixing patterns in networks. Physical review E, 67(2), 026126.

Social mechanisms

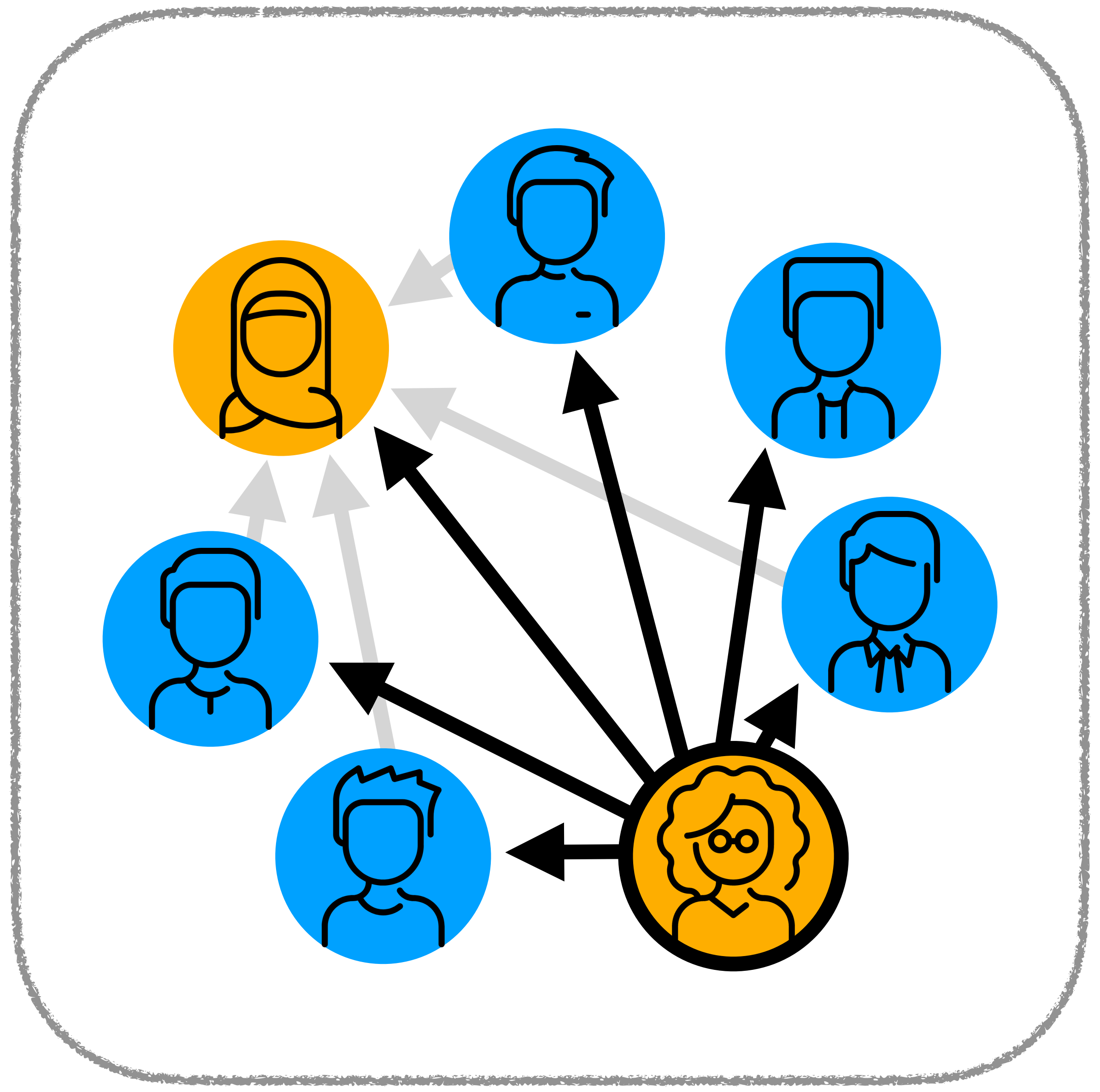
of edge formation



Preferential Attachment
(popularity)



Homophily
(similarity)



Activity
(outreach)

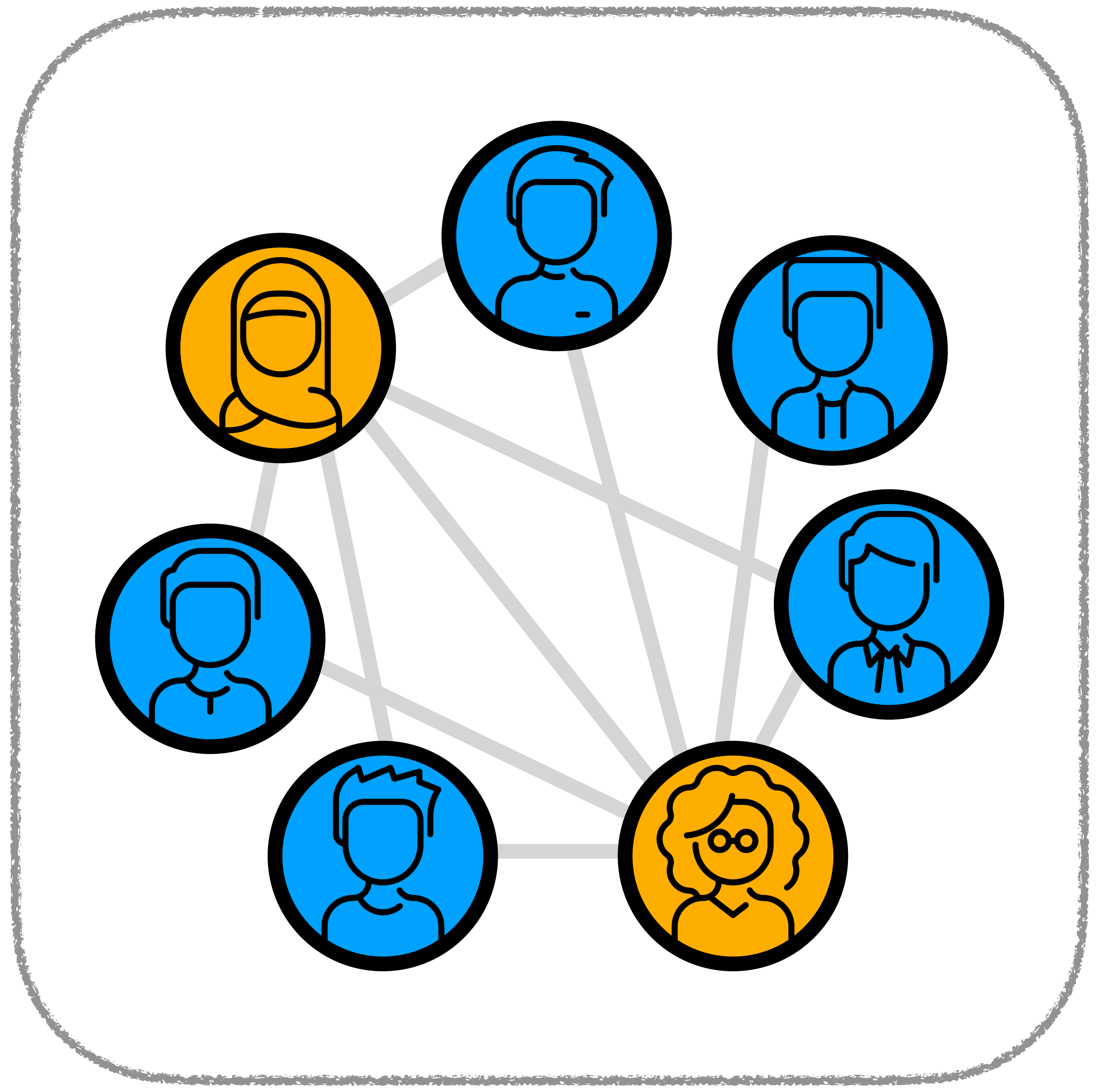
Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56-63.
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. science, 286(5439), 509-512.
McPherson, M., et al, (2001). Birds of a feather: Homophily in social networks. Annual review of sociology, 27(1), 415-444.
Newman, M. E. (2003). Mixing patterns in networks. Physical review E, 67(2), 026126.
Perra, N., et al. (2012). Activity driven modeling of time varying networks. Scientific reports, 2(1), 469.

Network structure

or properties

Network structure

or properties

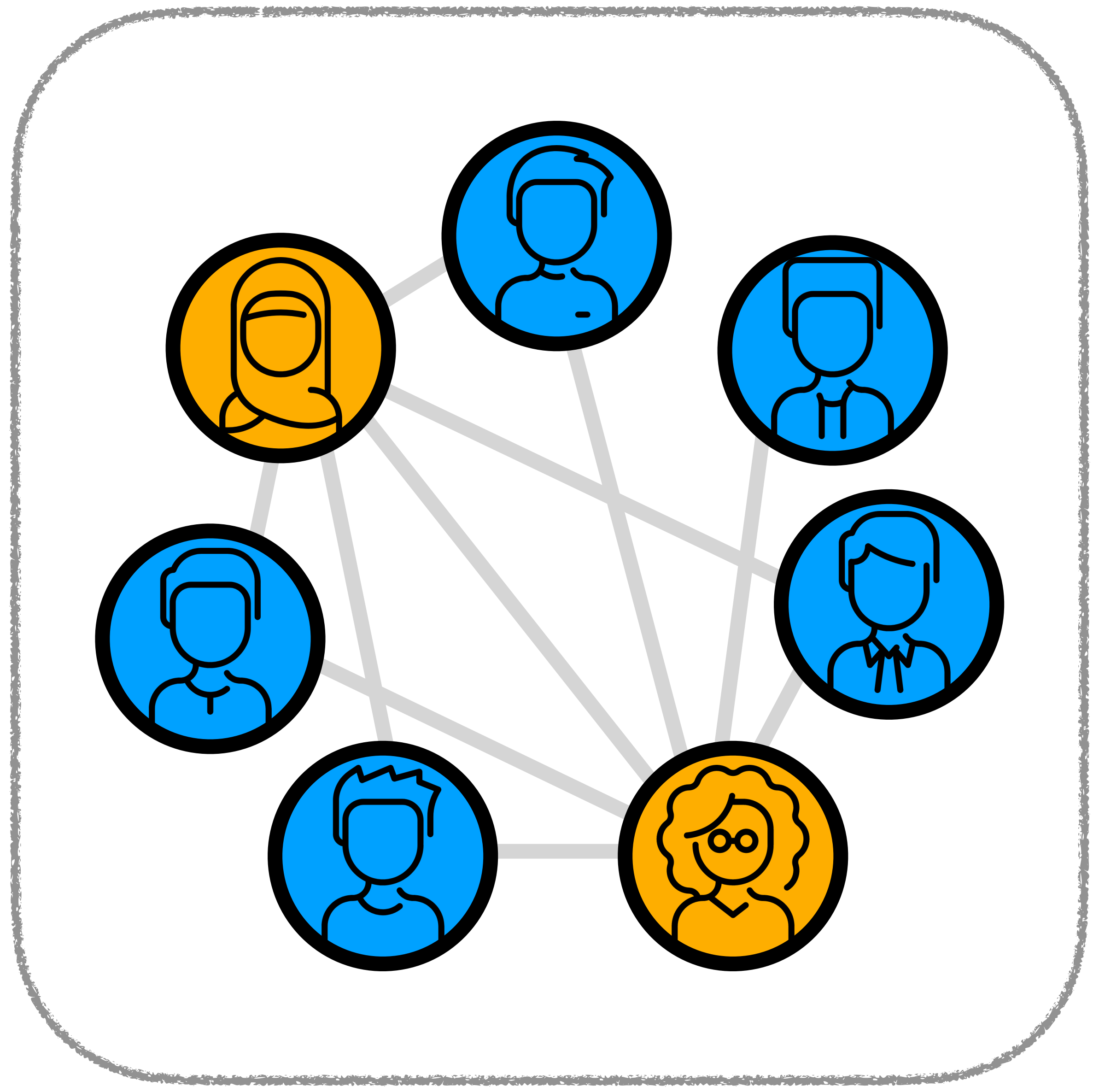


Network size

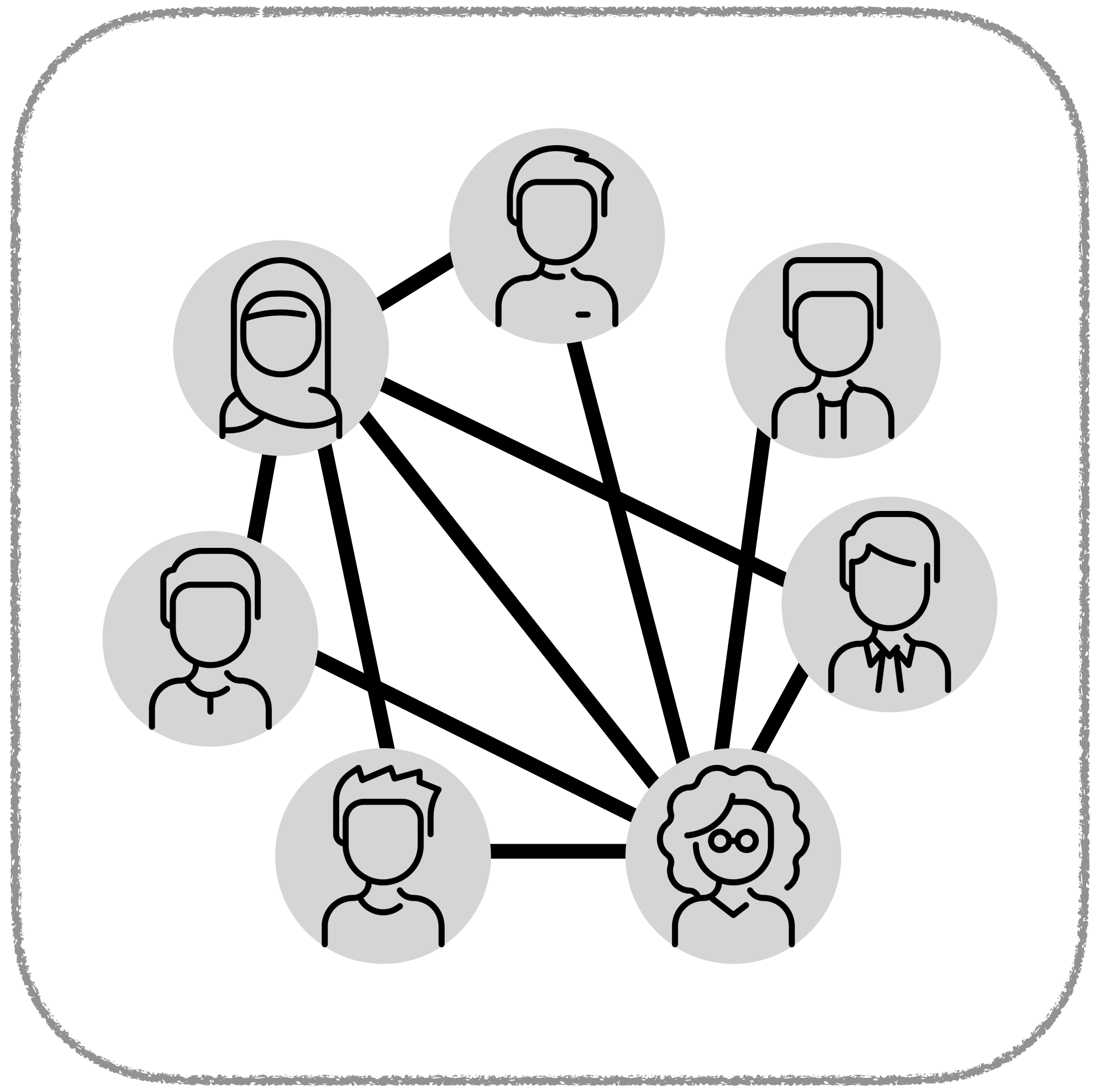
(number of nodes)

Network structure

or properties



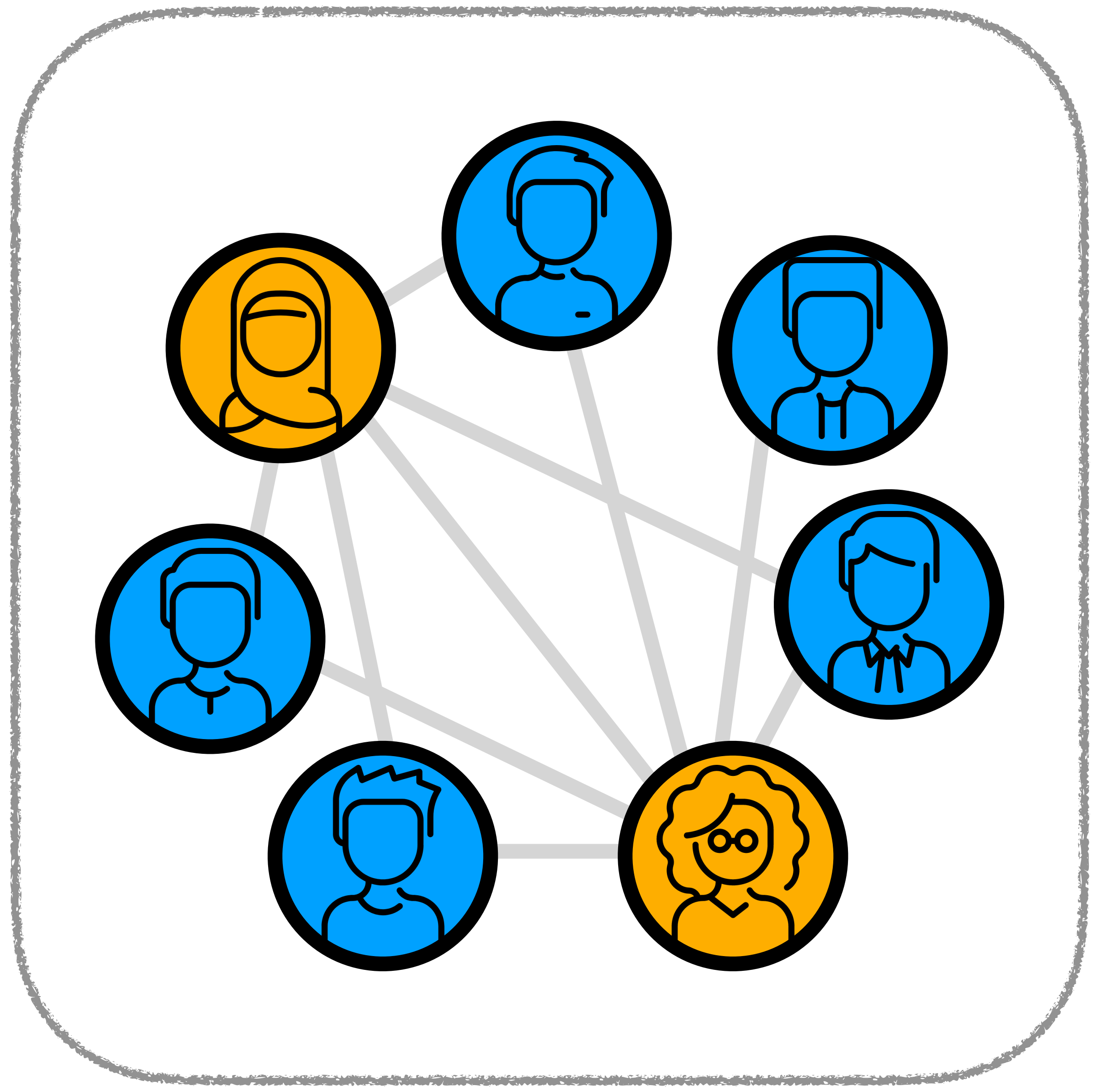
Network size
(number of nodes)



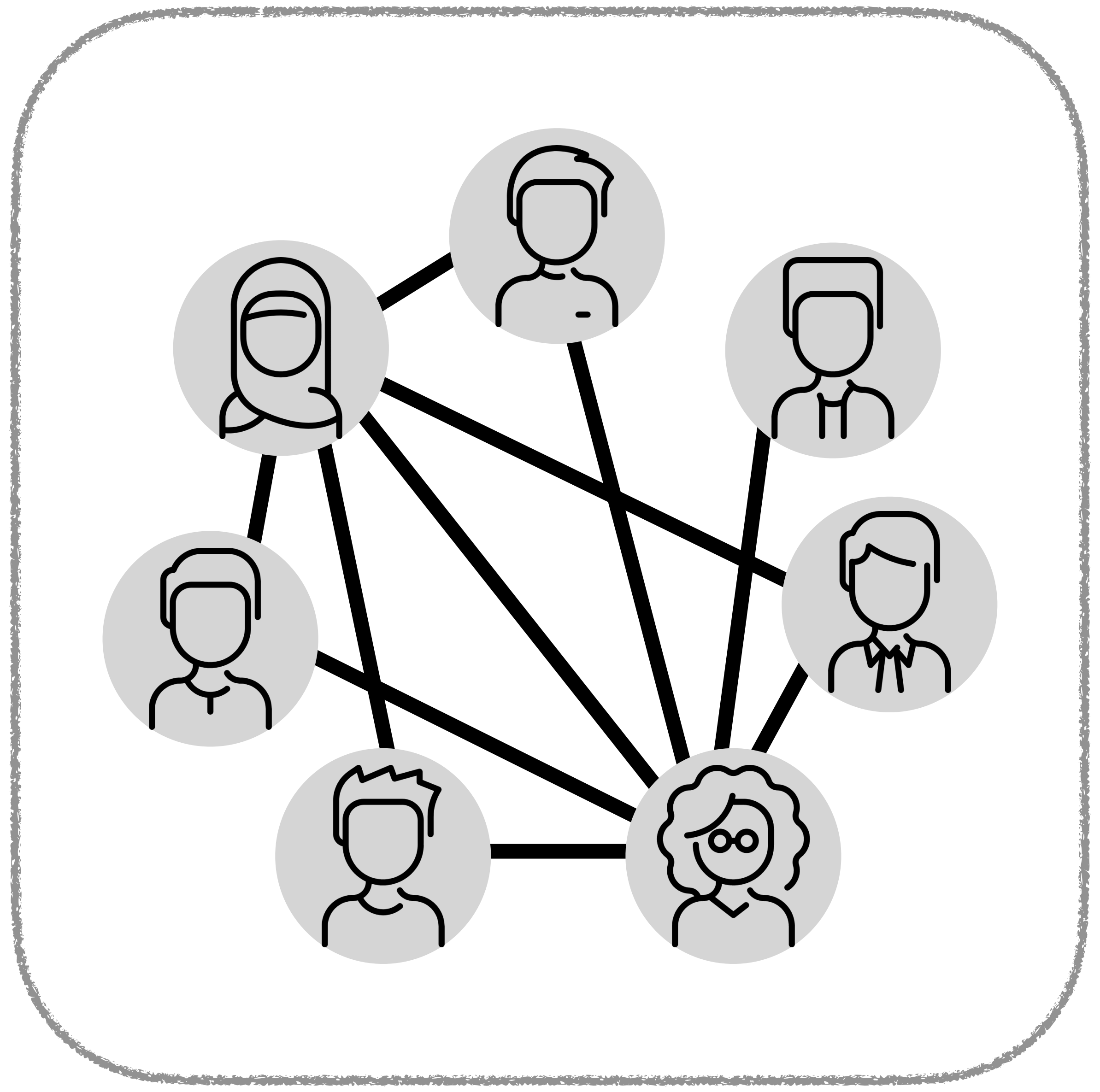
Density
(number of edges given number of nodes)

Network structure

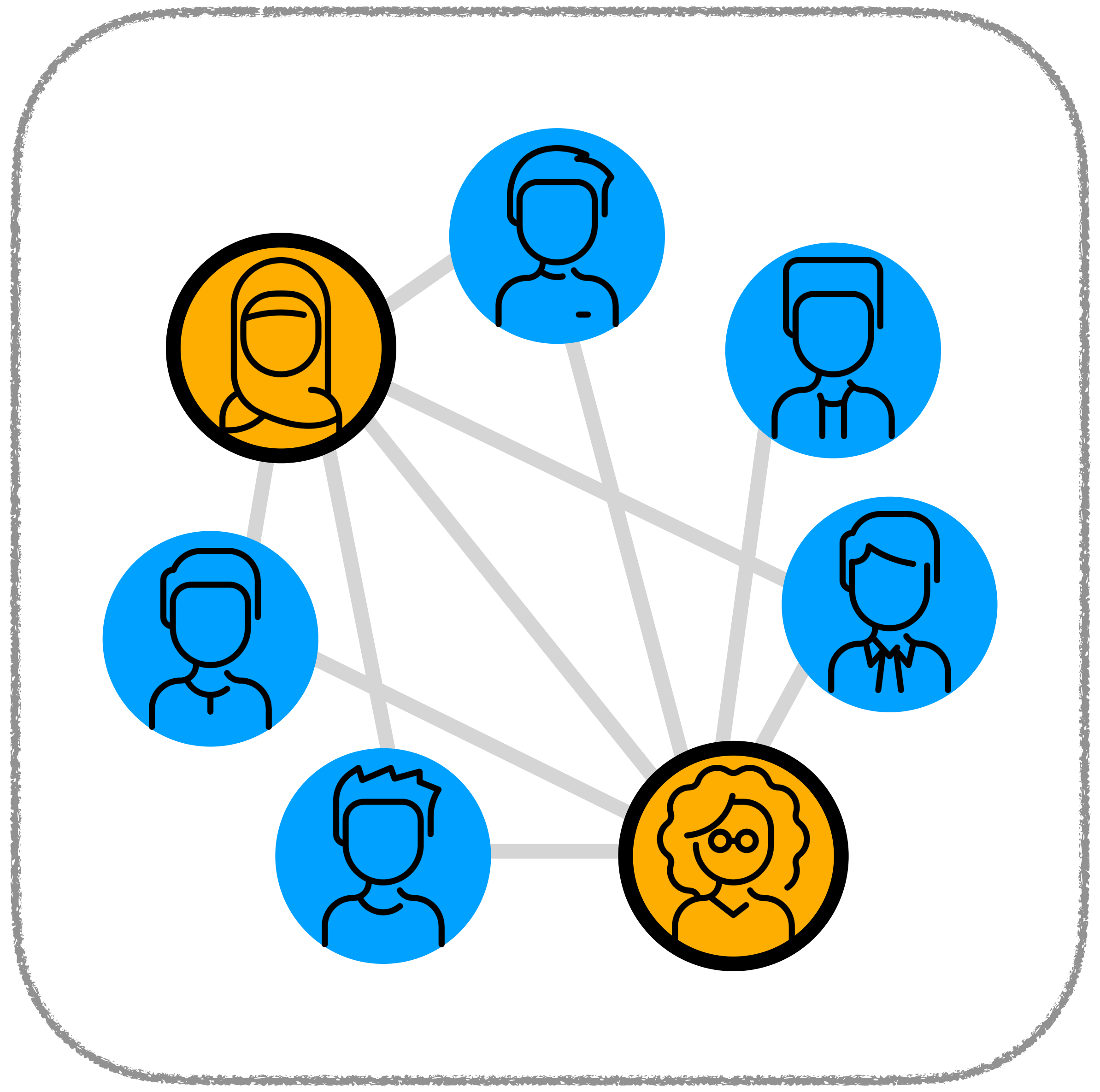
or properties



Network size
(number of nodes)



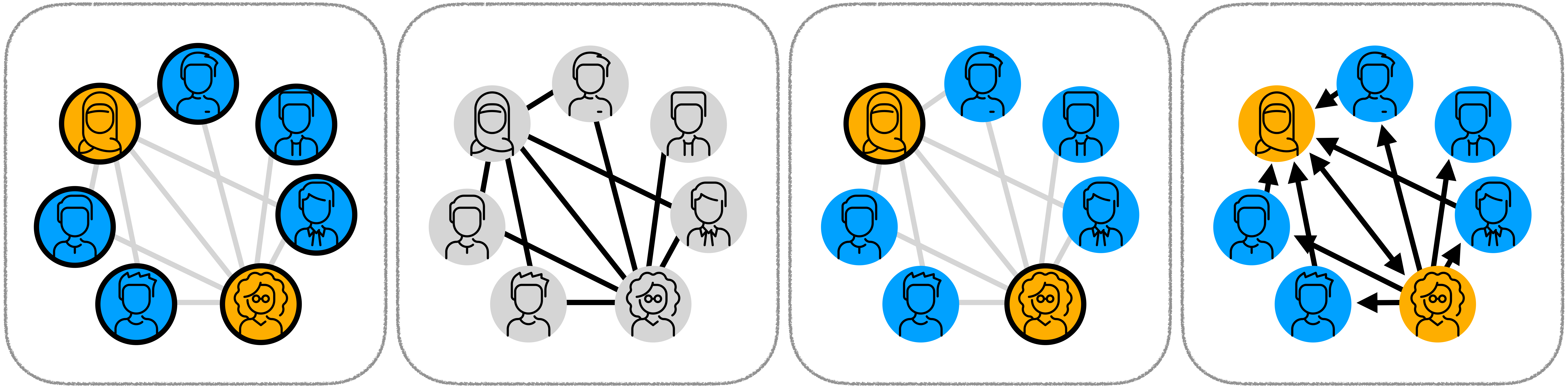
Density
(number of edges given number of nodes)



Fraction of minority
(class balance, aka. group size)

Network structure

or properties



Network size
(number of nodes)

Density
(number of edges given number of nodes)

Fraction of minority
(class balance, aka. group size)

Directionality
(directed edges)

Network models

in network science

Network models

in network science

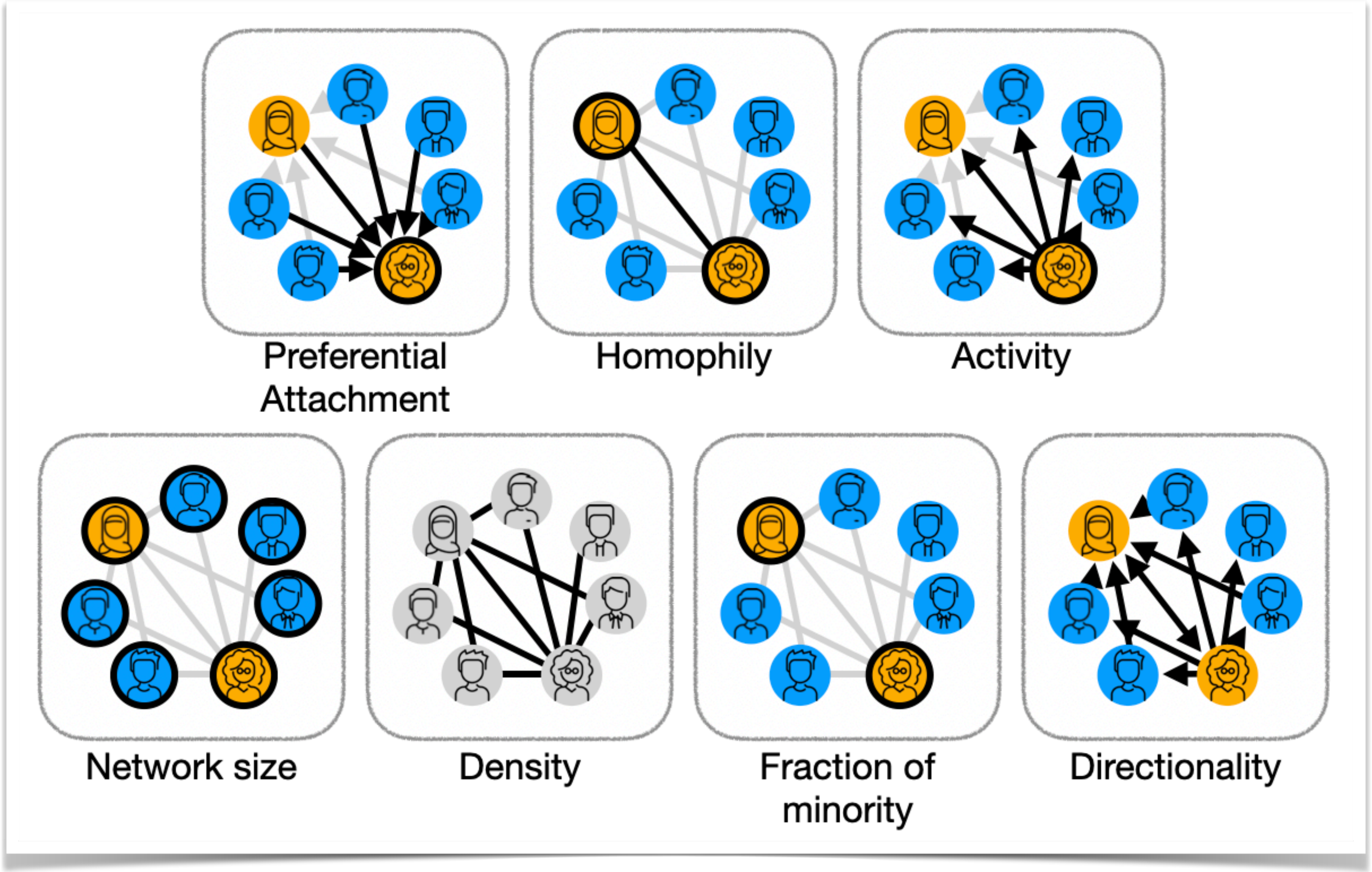
Model class	**Examples**
Static models	Erdos-Rényi Watts-Strogatz
Generative models	Configuration model Hidden parameter model Stochastic block model
Evolving network models	Barabási-Albert model Bianconi-Barabási model Initial Attractiveness model Internal links model Node deletion model Accelerated growth model Aging model

Network Science by Albert Lászlo Barabási: http://networksciencebook.com/chapter/6#summary6

Network models

in network science

Model class	**Examples**
Static models	Erdos-Rényi Watts-Strogatz
Generative models	Configuration model Hidden parameter model Stochastic block model
Evolving network models	Barabási-Albert model Bianconi-Barabási model Initial Attractiveness model Internal links model Node deletion model Accelerated growth model Aging model

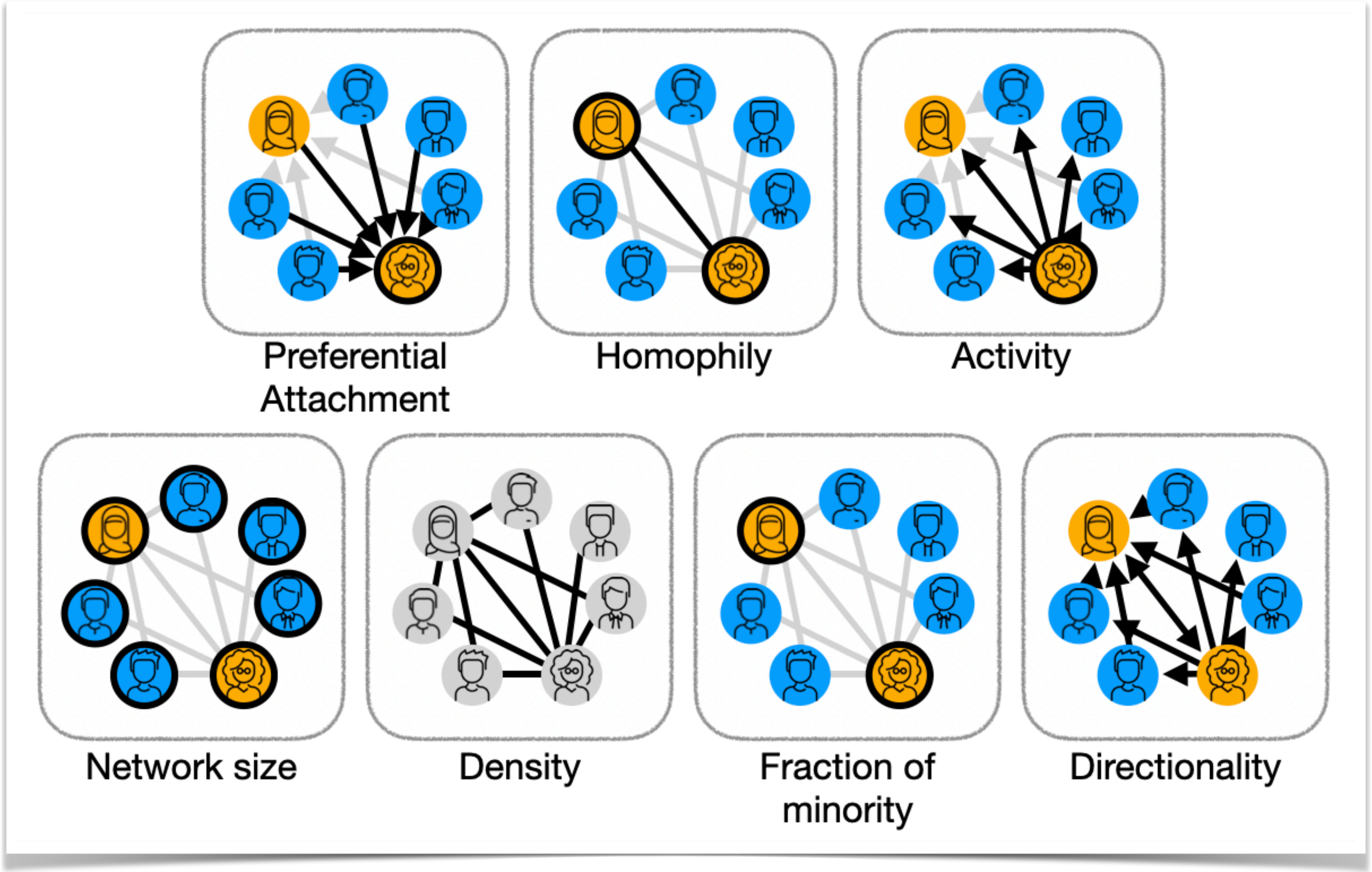


Network Science by Albert Lászlo Barabási: http://networksciencebook.com/chapter/6#summary6

Network models

in network science

Model class	Examples
Static models	Erdos-Rényi Watts-Strogatz
Generative models	Configuration model Hidden parameter model Stochastic block model
Evolving network models	Barabási-Albert model Bianconi-Barabási model Initial Attractiveness model Internal links model Node deletion model Accelerated growth model Aging model **PAH (Karimi et al. 2018)** **DPAH (Espín-Noboa et al. 2022)**



More realistic models with **P**referential **A**ttachment and **H**omophily and other properties.

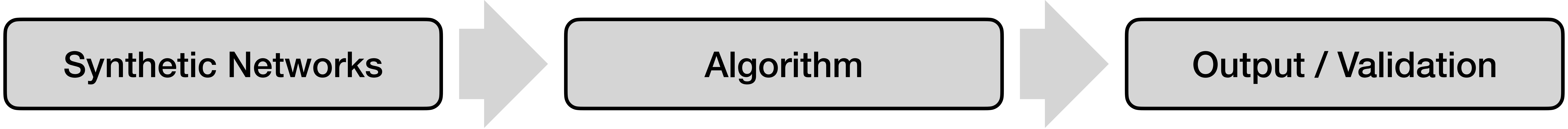
Network Science by Albert Lászlo Barabási: http://networksciencebook.com/chapter/6#summary6

The auditing pipeline

with synthetic networks

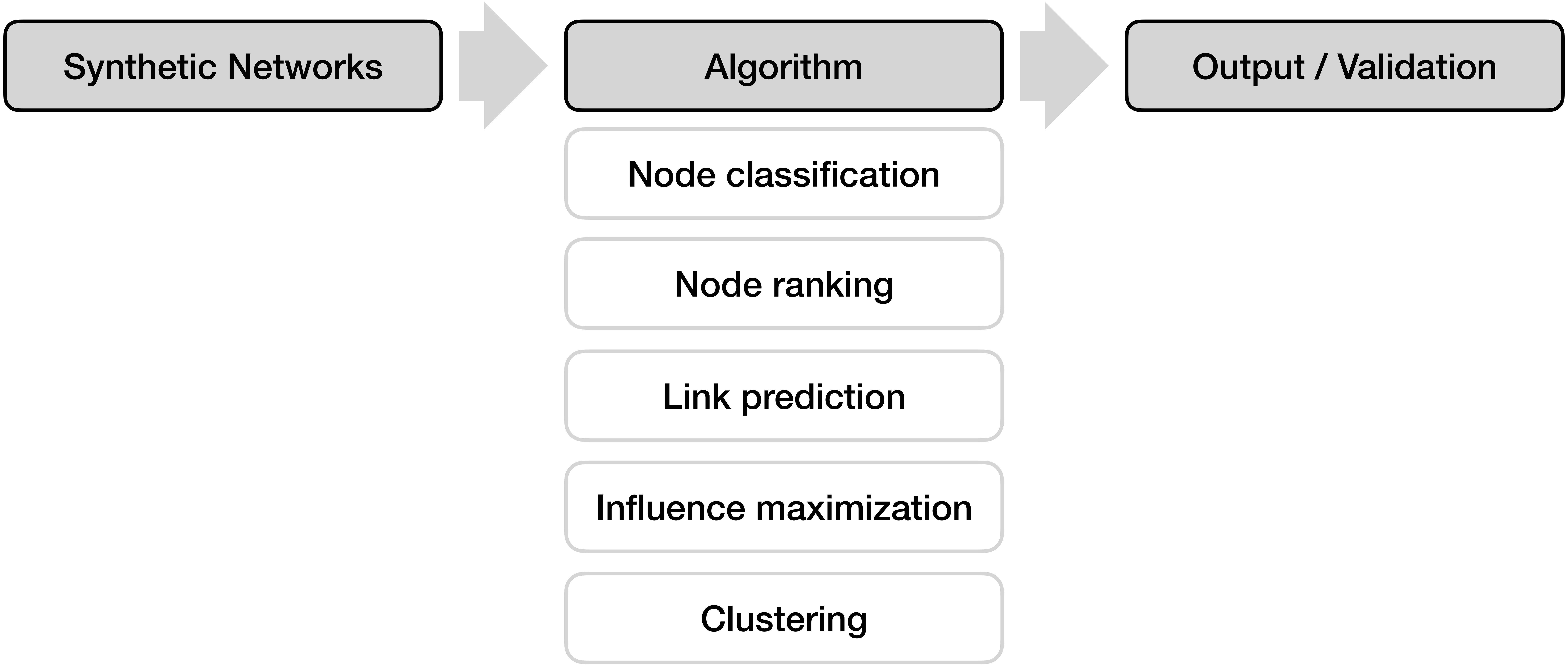
The auditing pipeline

with synthetic networks



The auditing pipeline

with synthetic networks



The auditing pipeline

with synthetic networks

Synthetic Networks	Algorithm	Output / Validation
	Node classification	Accuracy
	Node ranking	Bias (inequality)
	Link prediction	Fairness (equity)
	Influence maximization	
	Clustering	

The auditing pipeline

with synthetic networks

Synthetic Networks	Algorithm	Output / Validation
G = **model**(parameters)	Node classification	Accuracy
	Node ranking	Bias (inequality)
	Link prediction	Fairness (equity)
	Influence maximization	
	Clustering	

The auditing pipeline

with synthetic networks

Synthetic Networks

G = **model**(parameters)

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)

The auditing pipeline

with synthetic networks

Synthetic Networks

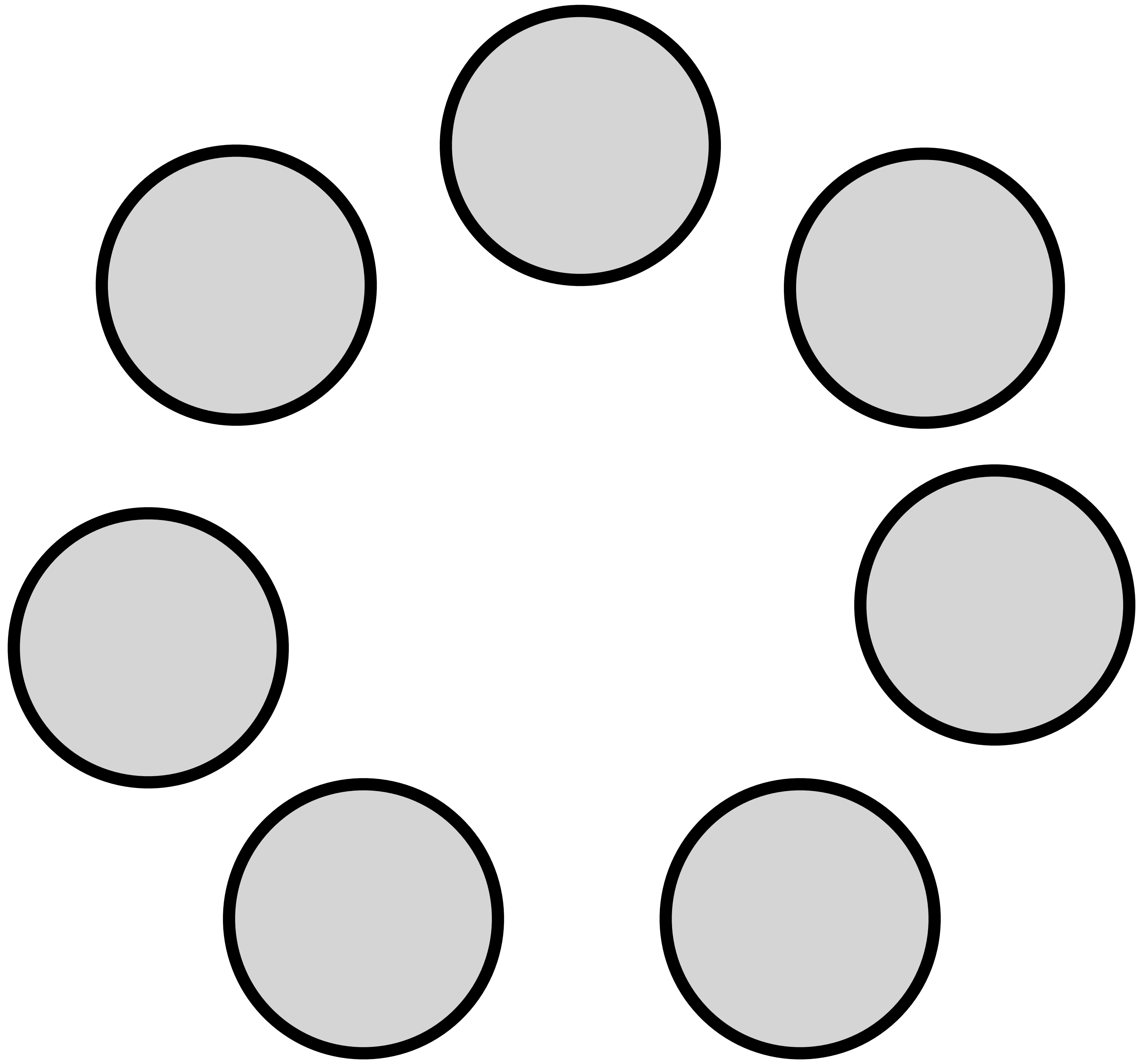
G = **model**(parameters)

Number of nodes

n

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)



$n = 7$

The auditing pipeline

with synthetic networks

Synthetic Networks

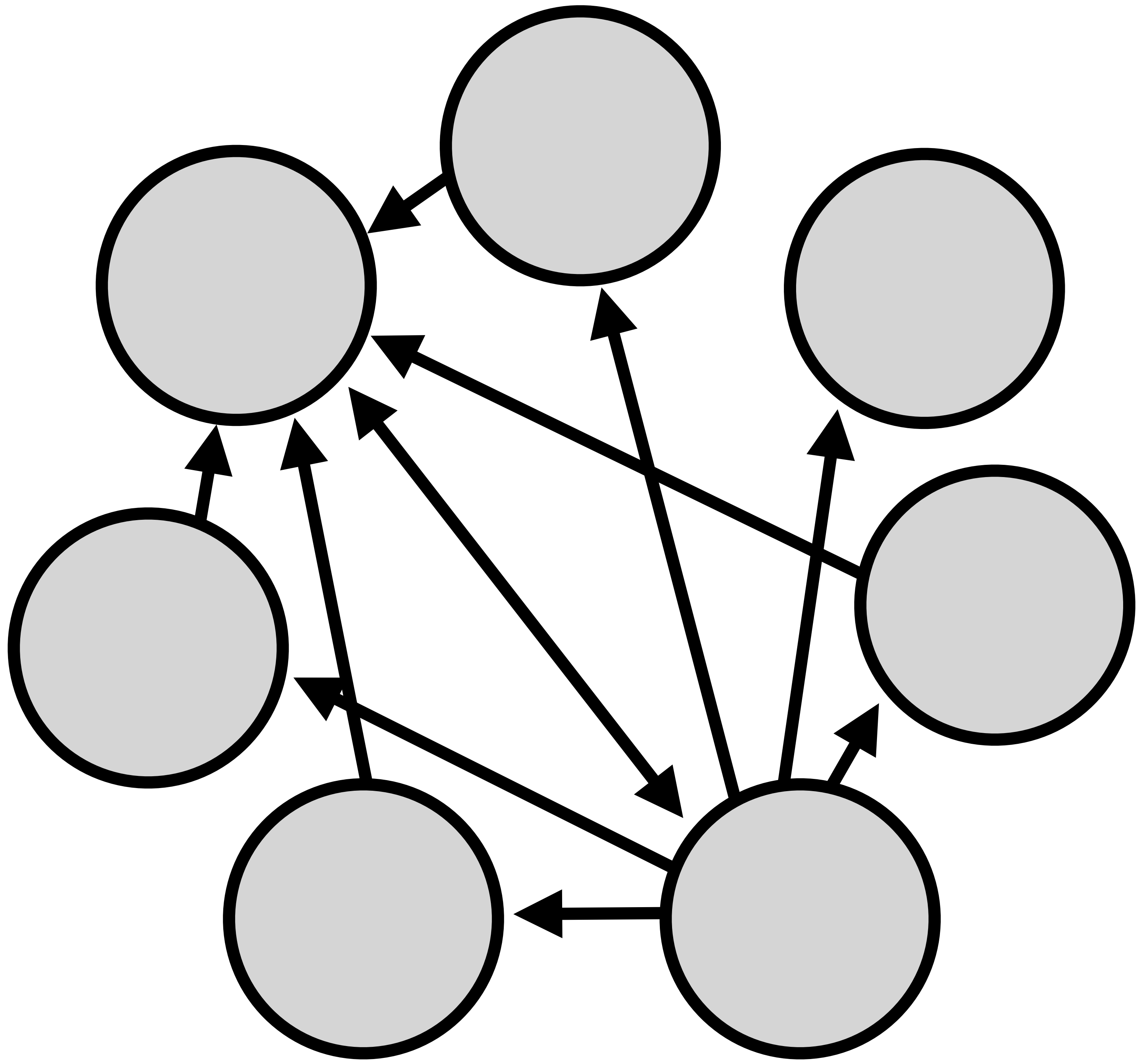
G = **model**(parameters)

Density

n **d**

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)



$$d = \frac{10}{7 * 7} = 0.02$$

The auditing pipeline

with synthetic networks

Synthetic Networks

G = **model**(parameters)

Fraction of minority

n d $\mathbf{f_m}$

D.PA.H(............)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)

$$f_m = \frac{2}{7} = 0.29$$

The auditing pipeline

with synthetic networks

Synthetic Networks

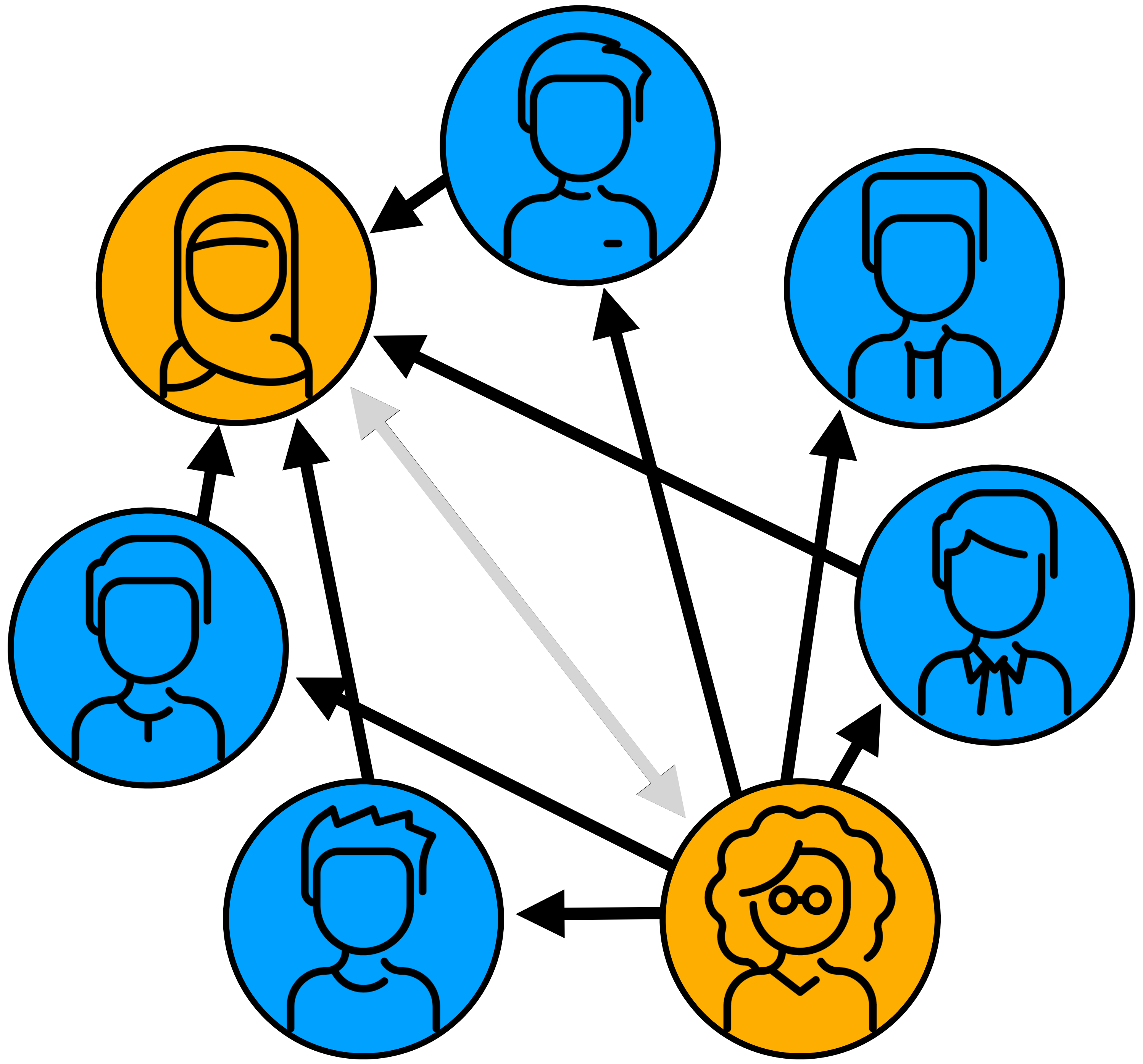
G = **model**(parameters)

Homophily (h_{min}, h_{maj})

n d f_m H

D.PA.H(............)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)



$$h_{min} = 0.2 \quad h_{maj} = 0.0$$

The auditing pipeline

with synthetic networks

Synthetic Networks

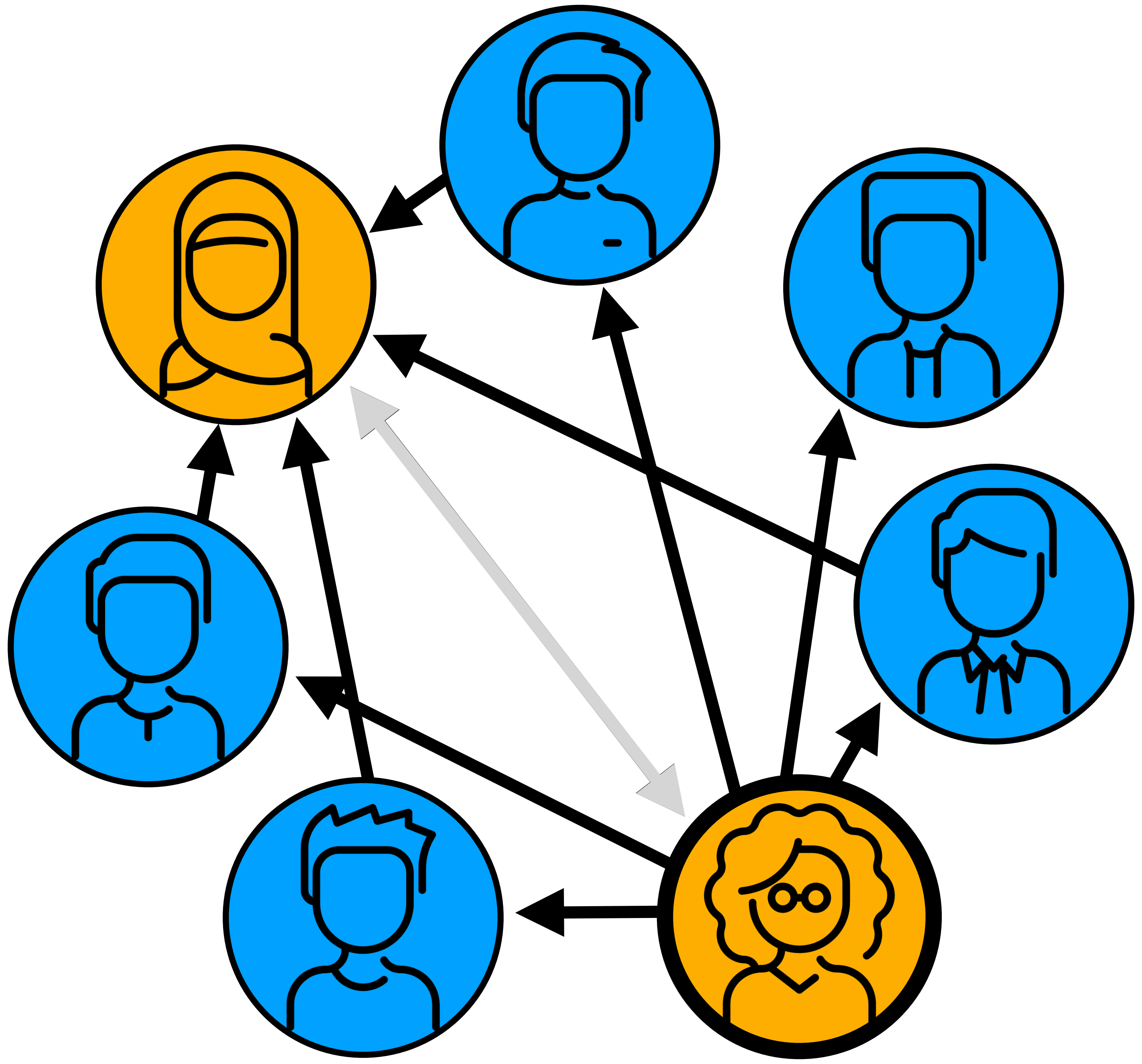
G = **model**(parameters)

Activity (γ_{min}, γ_{maj})

n d f_m H A

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)



$$\gamma_{min} = 1.5 \quad \gamma_{maj} = 3.0$$

The auditing pipeline

with synthetic networks

Synthetic Networks

G = **model**(parameters)

n **d** $\mathbf{f_m}$ **H** **A**

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)

The auditing pipeline
with synthetic networks

Synthetic Networks	Algorithm	Output / Validation
G = **model**(parameters)	Node classification	Accuracy
	Node ranking	Bias (inequality)
	Link prediction	Fairness (equity)
	Influence maximization	
	Clustering	

n d f_m H A

D.PA.H(…………)

Espín-Noboa et al. Inequality and inequity in network-based ranking and recommendation algorithms. Nature Scientific Reports (2022)

The auditing pipeline

with synthetic networks

Synthetic Networks	Algorithm	Output / Validation
G = **model**(parameters)	Node classification	Accuracy

How does **network structure** influence the outcomes of **algorithms**?

Assessing the plethora of outcomes

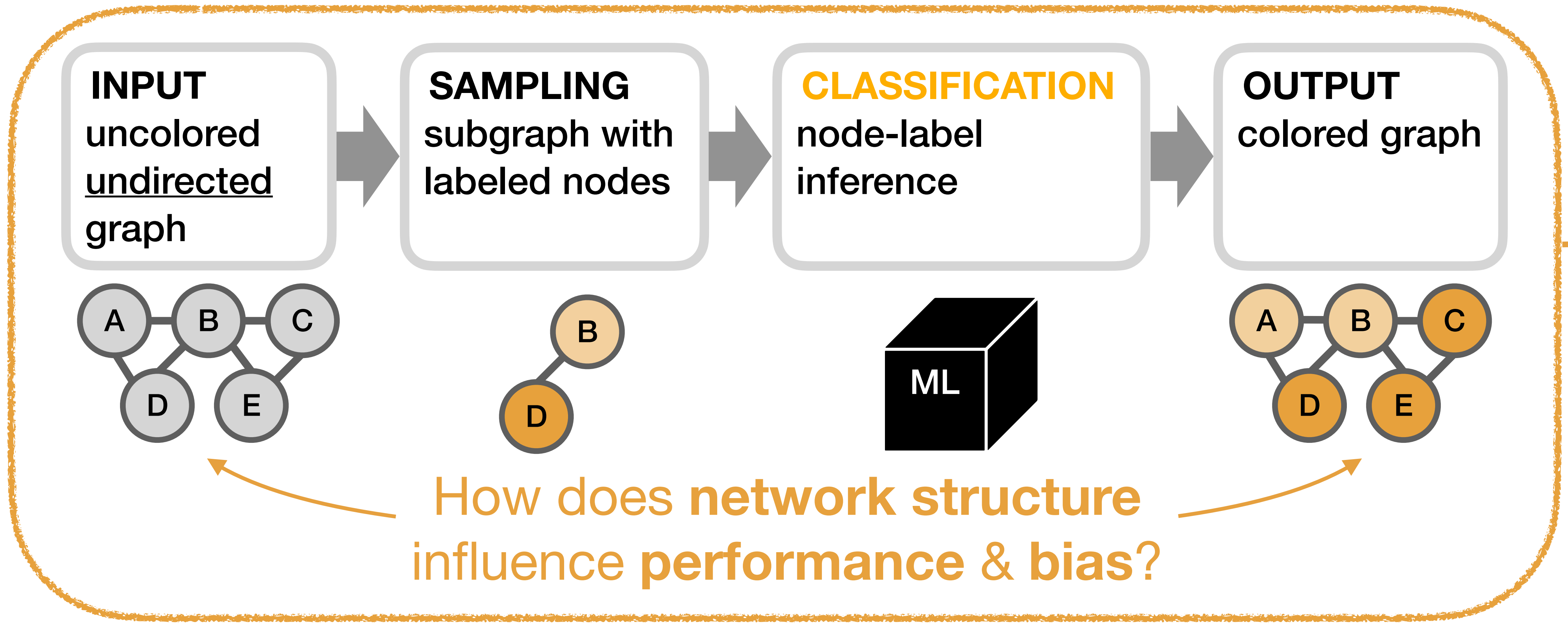
using synthetic networks

Assessing the plethora of outcomes

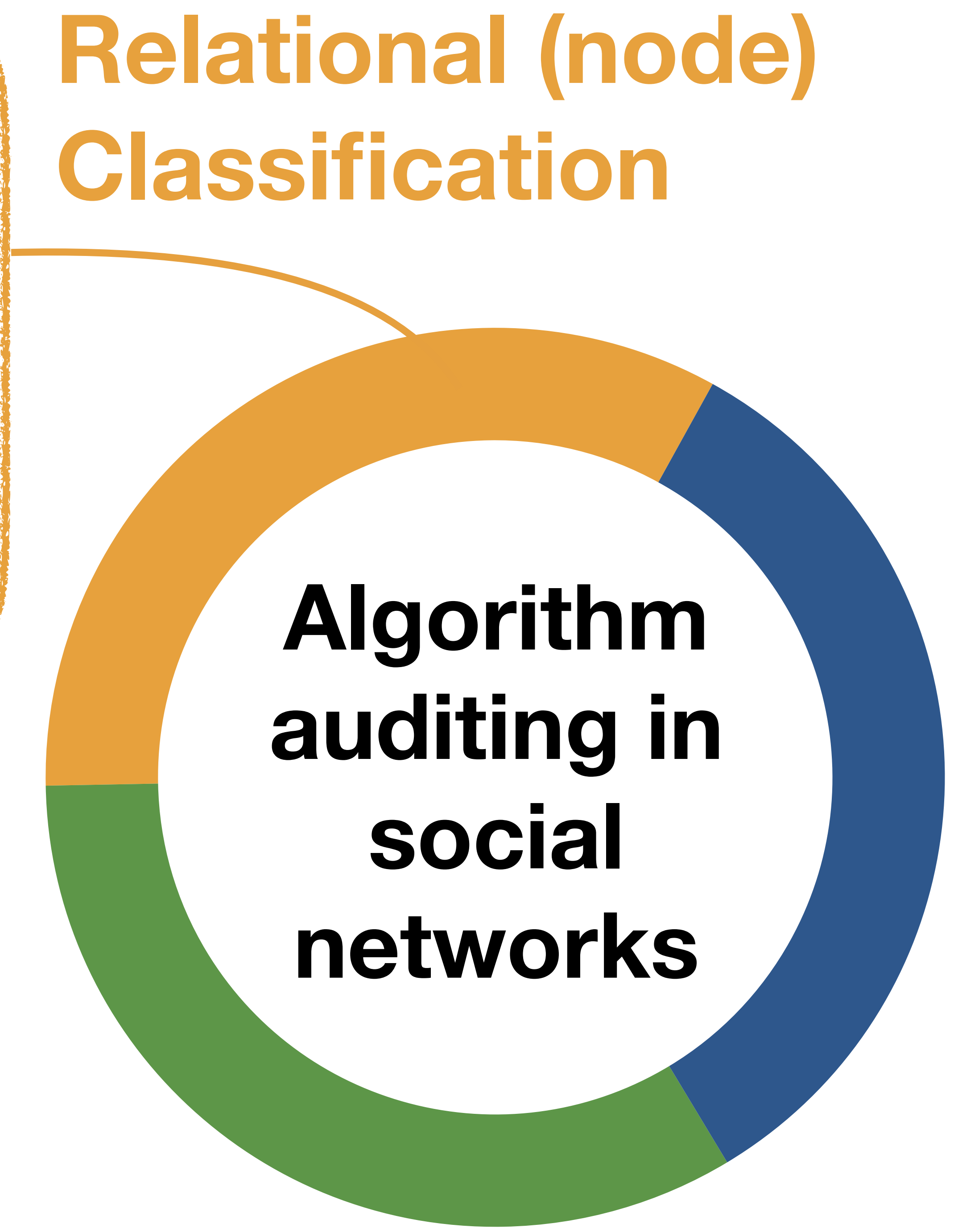
using synthetic networks

Assessing the plethora of outcomes

using synthetic networks

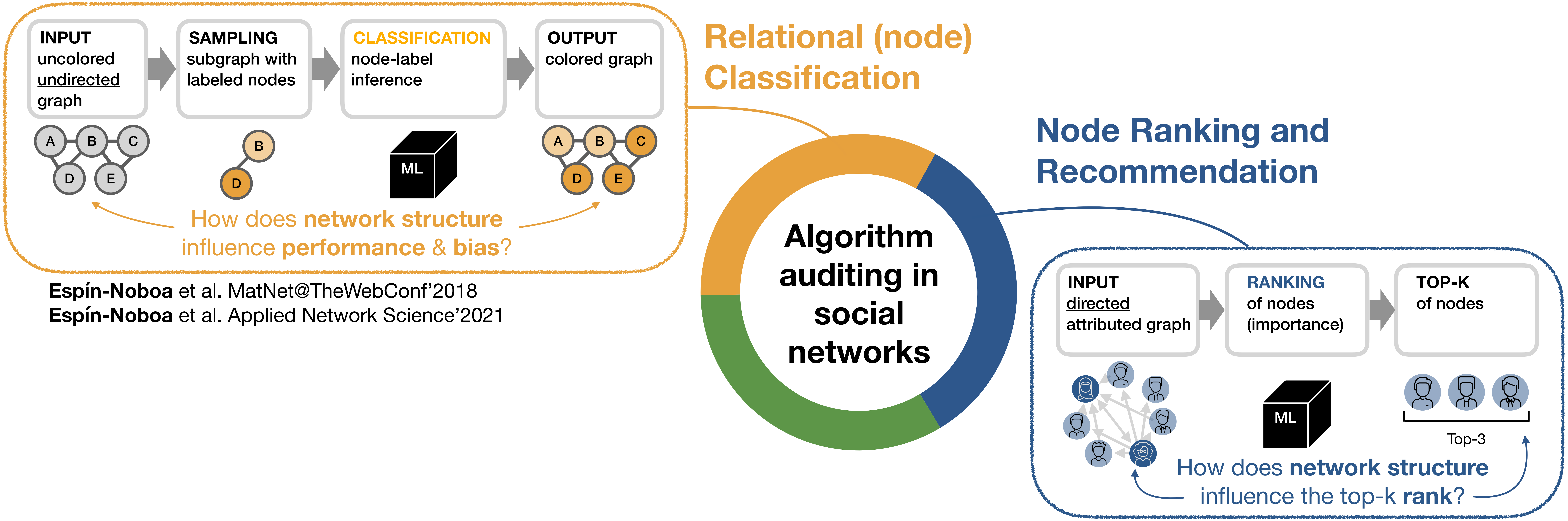


Espín-Noboa et al. MatNet@TheWebConf'2018
Espín-Noboa et al. Applied Network Science'2021



Assessing the plethora of outcomes

using synthetic networks

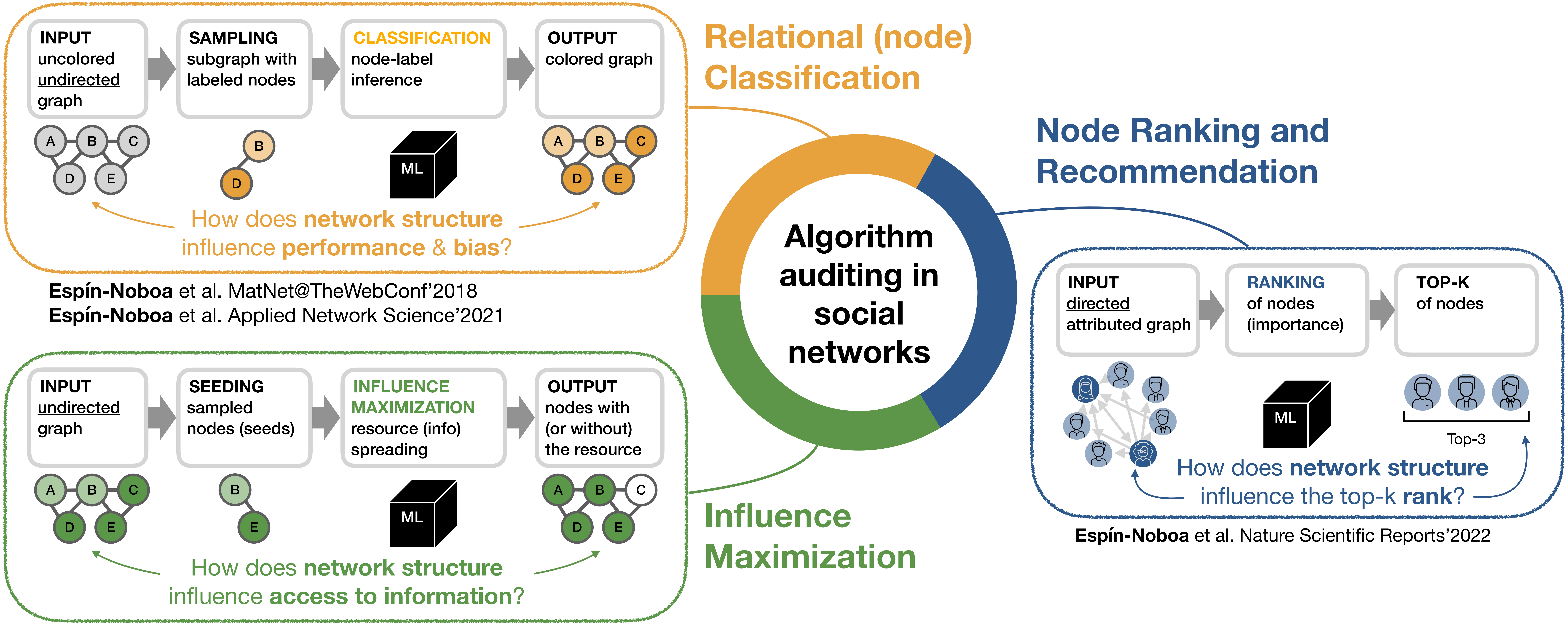


Espín-Noboa et al. MatNet@TheWebConf'2018
Espín-Noboa et al. Applied Network Science'2021

Espín-Noboa et al. Nature Scientific Reports'2022

Assessing the plethora of outcomes

using synthetic networks



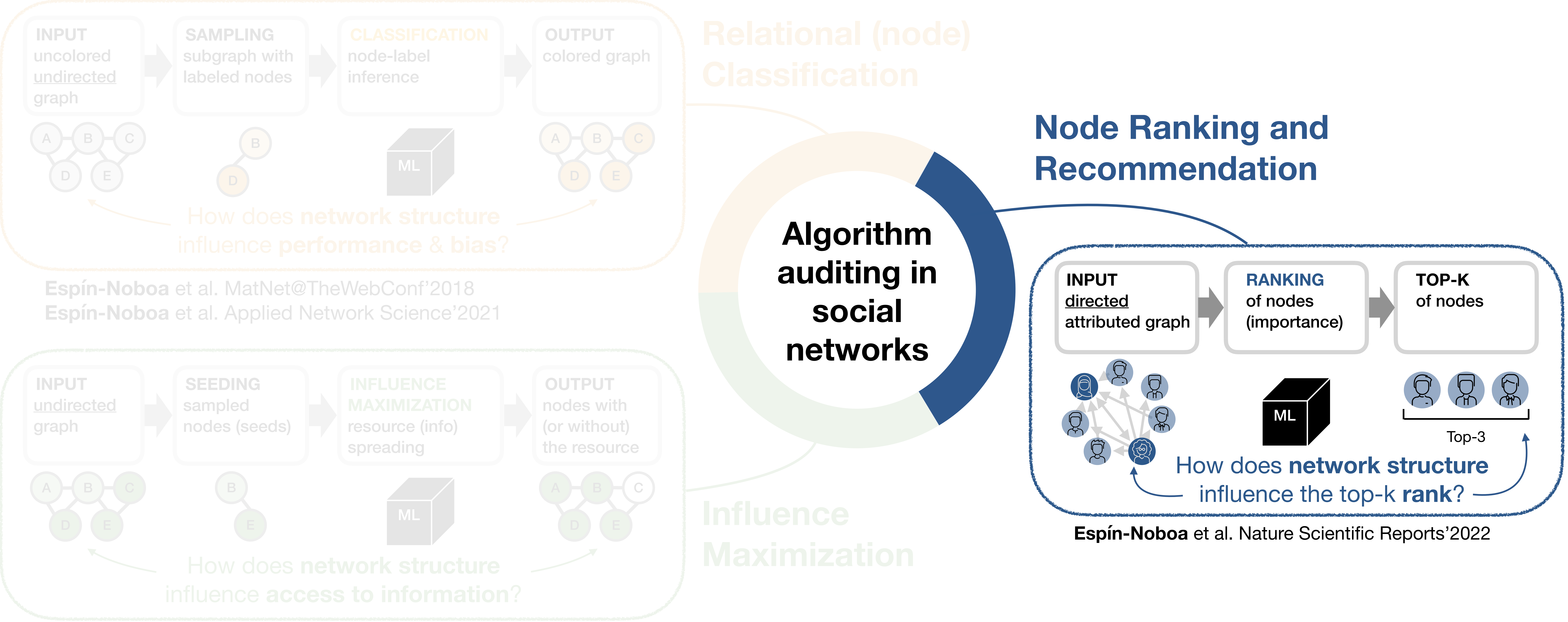
Espín-Noboa et al. MatNet@TheWebConf'2018
Espín-Noboa et al. Applied Network Science'2021

Espín-Noboa et al. Nature Scientific Reports'2022

Wang, Varol, and Eliassi-Rad, Applied Network Science'2022

Assessing the plethora of outcomes

using synthetic networks



Wang, Varol, and Eliassi-Rad, Applied Network Science'2022

How does **network structure** influence the representation of **minority** nodes in **top-k ranks**?


scientific reports

OPEN Inequality and inequity in network-based ranking and recommendation algorithms

Inequality and inequity in network-based ranking and recommendation algorithms.
Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Nature Scientific Reports, 12(1), 1-14.

Network characteristics:

Espín-Noboa et al. (2022)

Network characteristics:

n d f_m H A

D.PA.H(…………)

Espín-Noboa et al. (2022)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

f_m = 20% (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$

n d f_m H A

D.PA.H(…………)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$

Espín-Noboa et al. (2022)

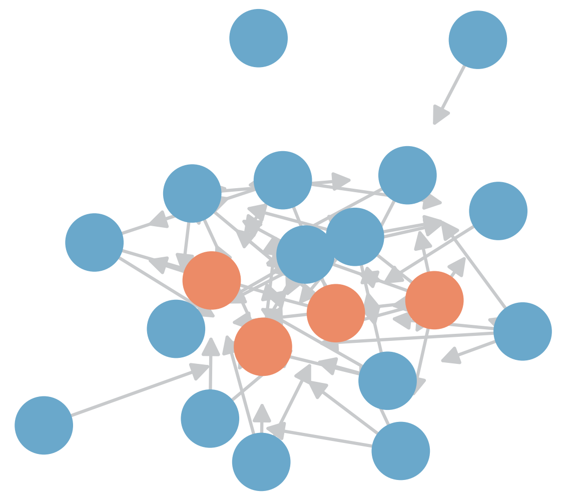
Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

f_m = 20% (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$

Graph vizualization

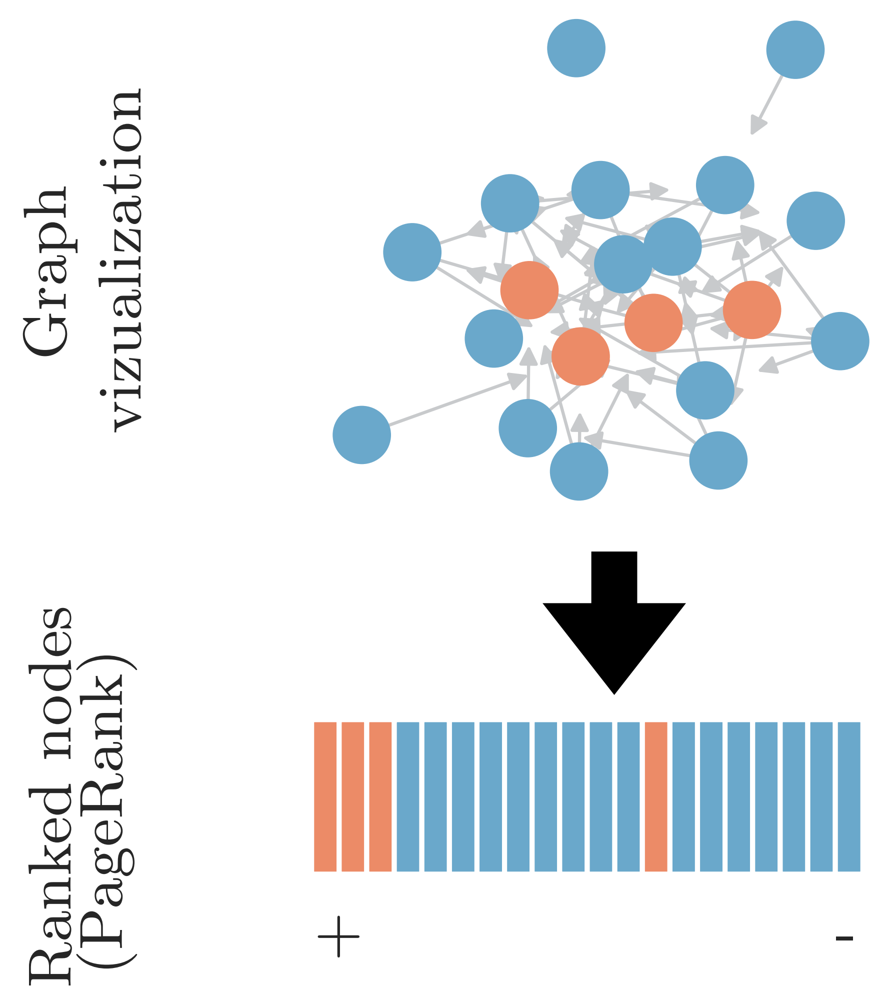


Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



Ranking of nodes

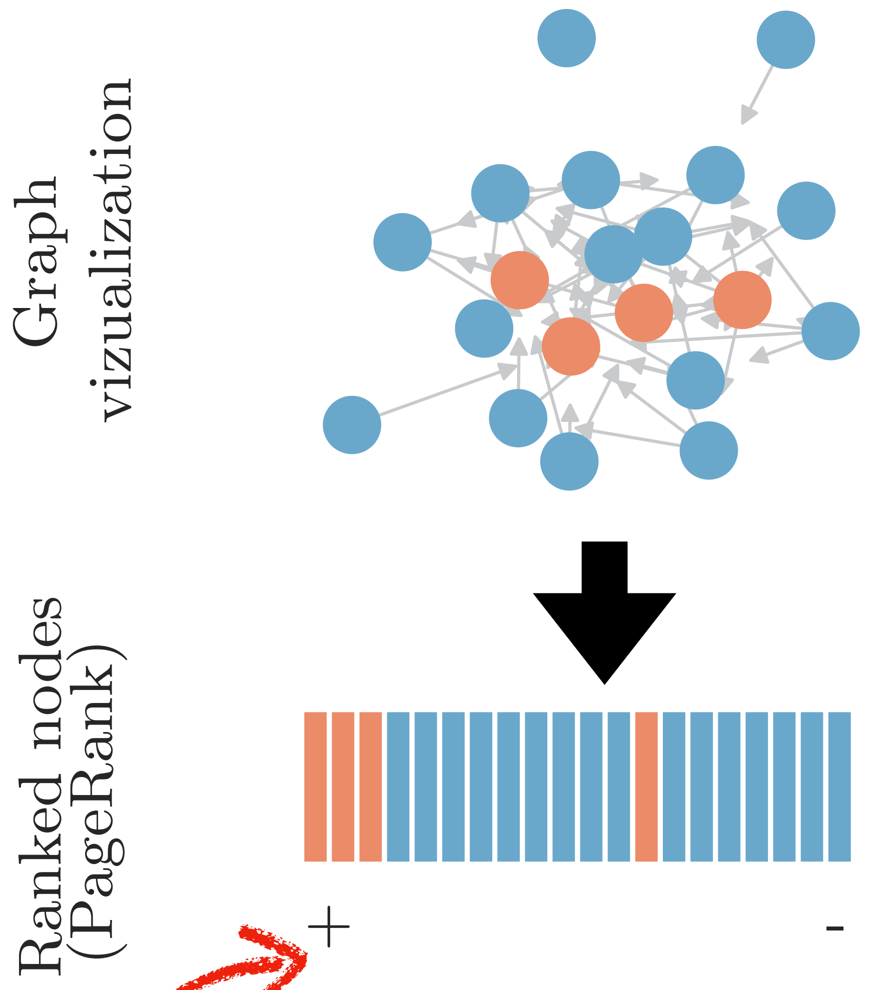
Espín-Noboa et al. (2022)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



Ranking of nodes

most important

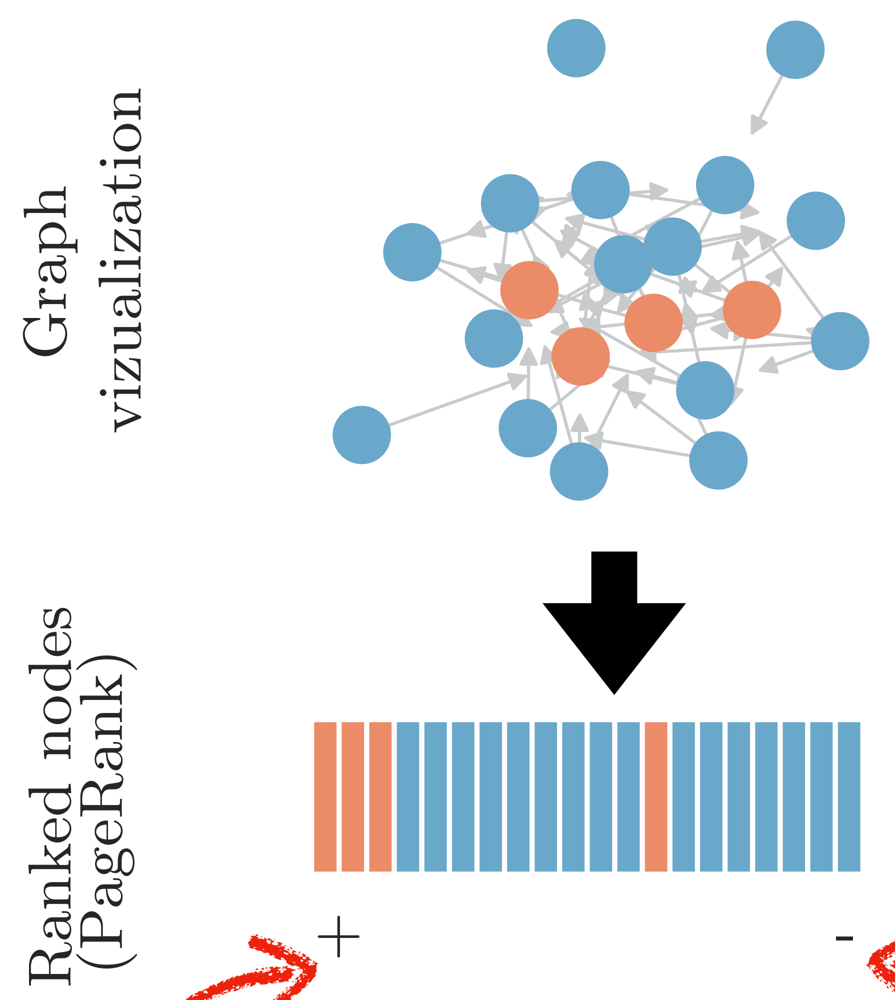
Espín-Noboa et al. (2022)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



Ranking of nodes

most important

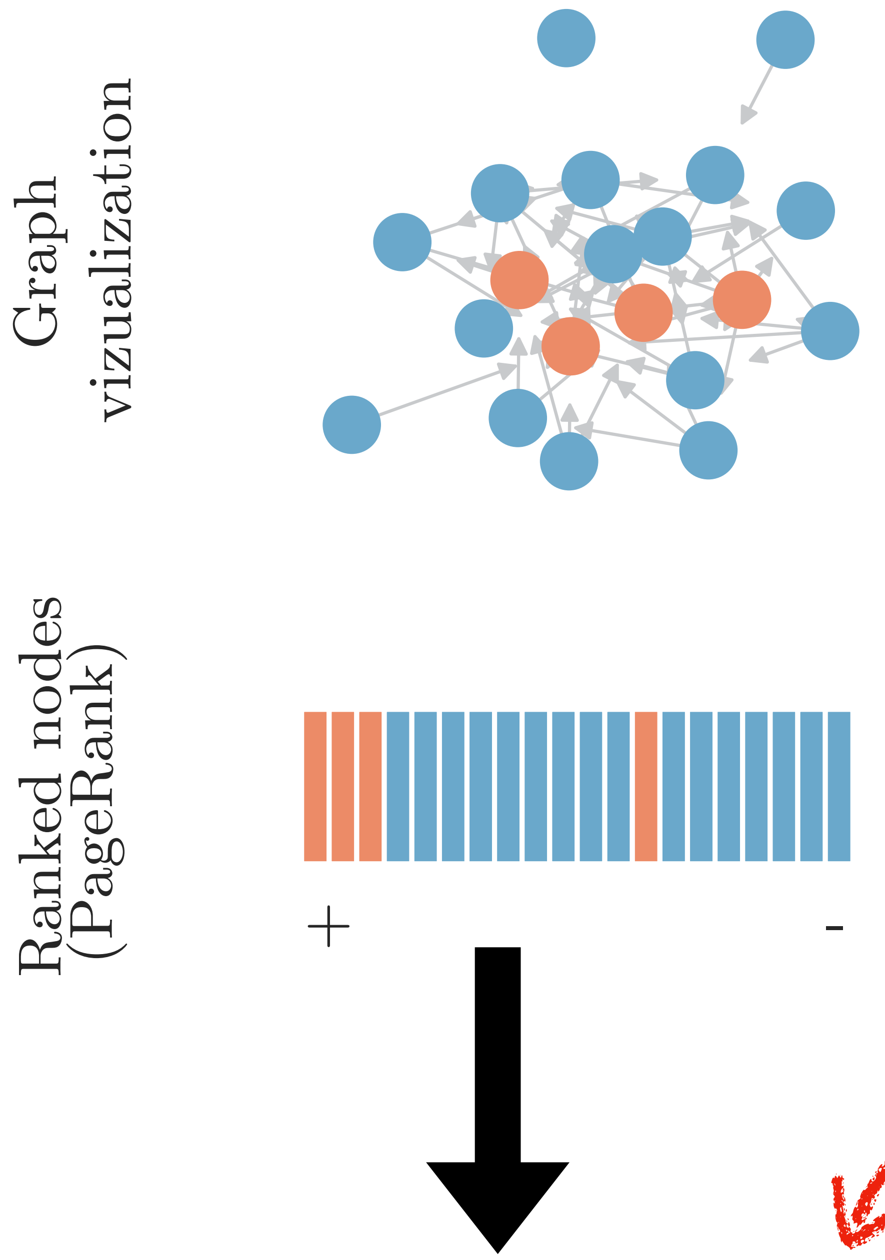
least important

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



Gini

$$Gini(X) = 0.5$$

Inequality

Skewness of the importance of nodes

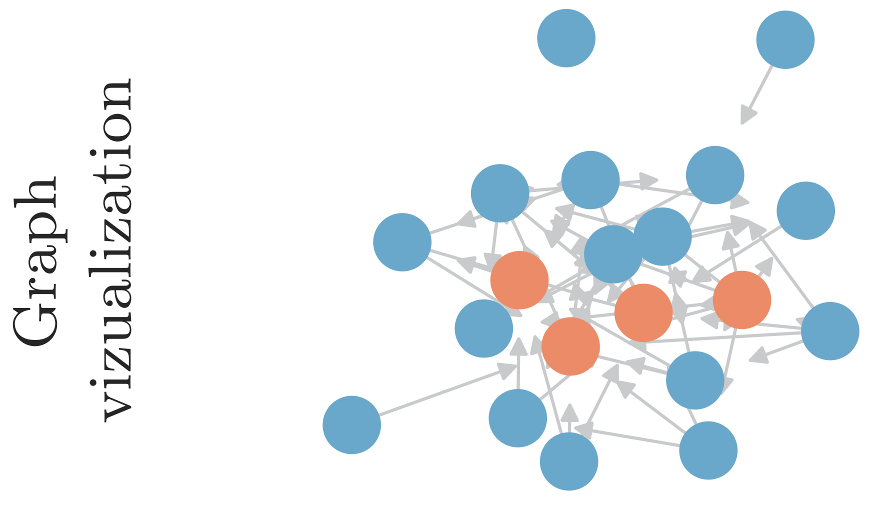
Espín-Noboa et al. (2022)

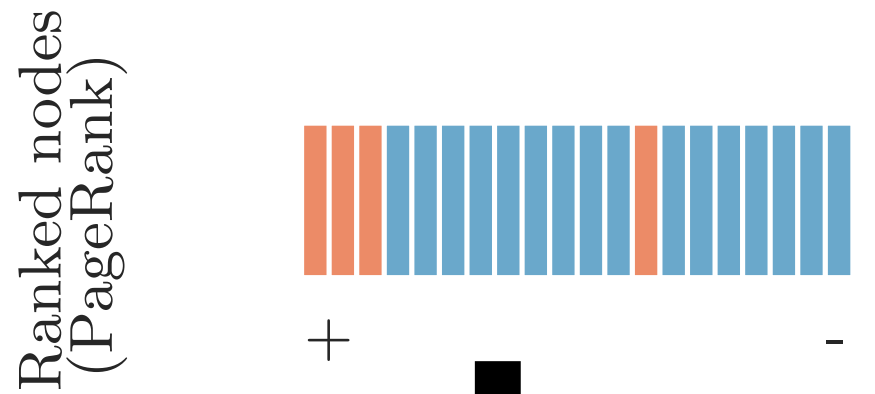
Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



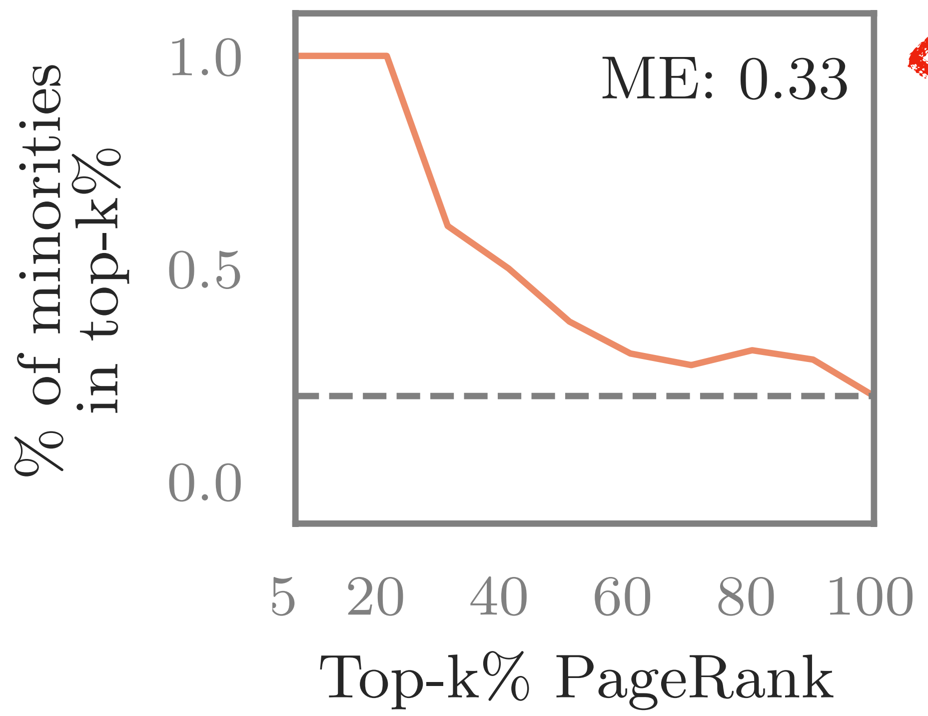


Gini

$$Gini(X) = 0.5$$

Inequality

Skewness of the importance of nodes



Inequity

Representation of minority nodes in top-k ranks compared to a baseline

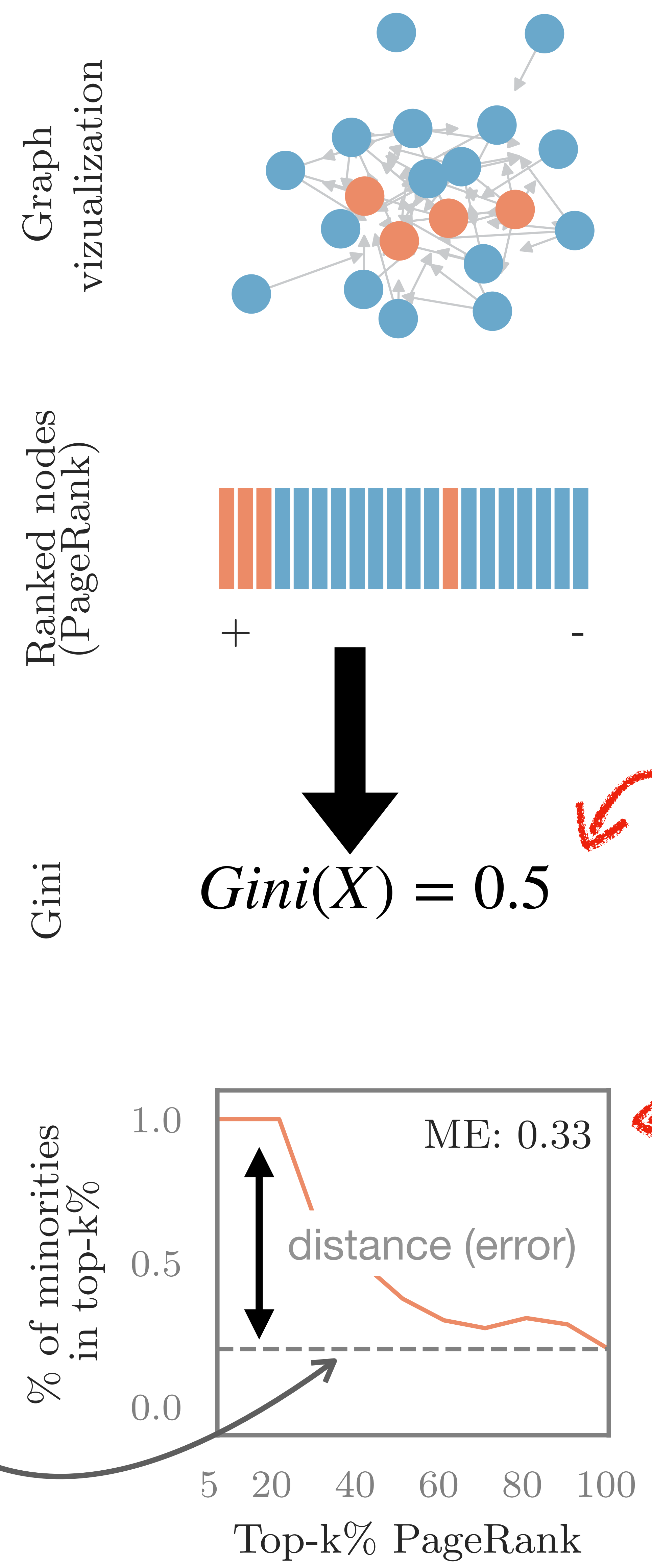
Espín-Noboa et al. (2022)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

f_m = 20% (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



Inequality
Skewness of the importance of nodes

Inequity
Representation of minority nodes in top-k ranks compared to a baseline

Statistical parity
When the representation of minorities in top-k = f_m

Dwork et al. 2012

Espín-Noboa et al. (2022)

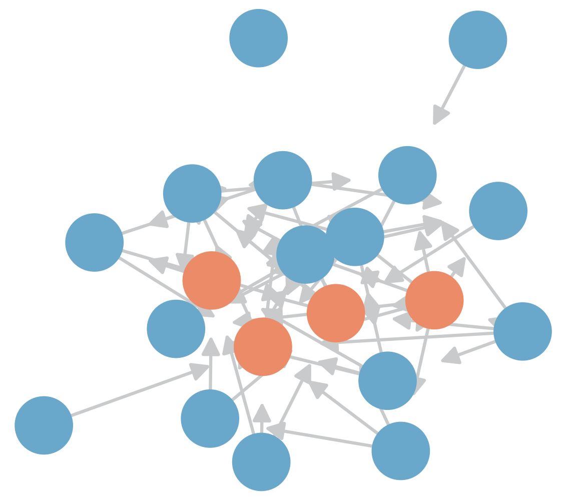
Network characteristics:

n = 20
d = 0.2
A = 3.0

f_m = 20% (min)
80% (maj)

h_{mm} = 0.2
h_{MM} = 0.2

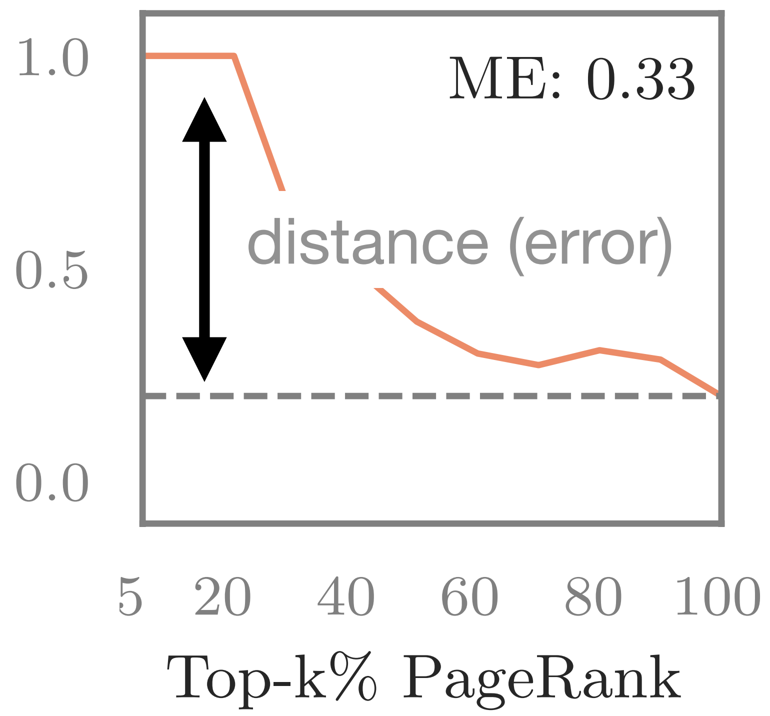
Graph vizualization



Ranked nodes (PageRank)

Gini

$$Gini(X) = 0.5$$

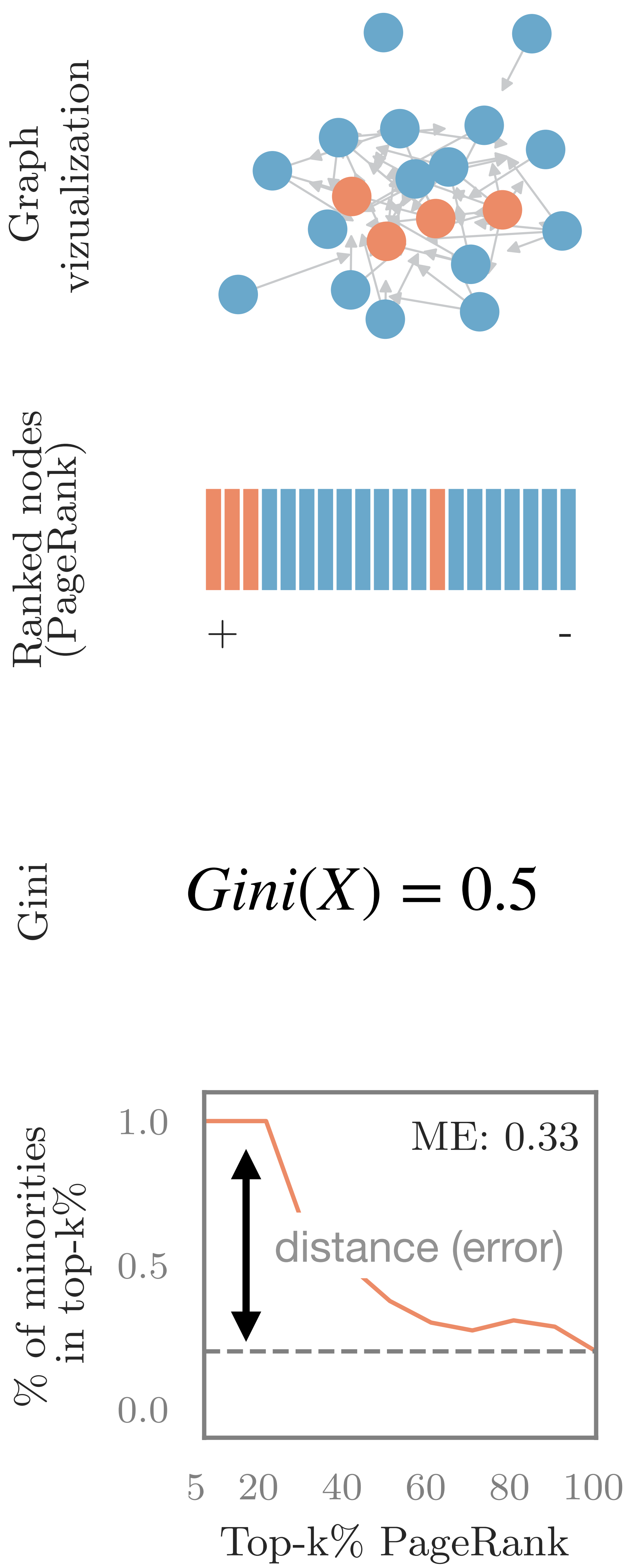


Network characteristics:

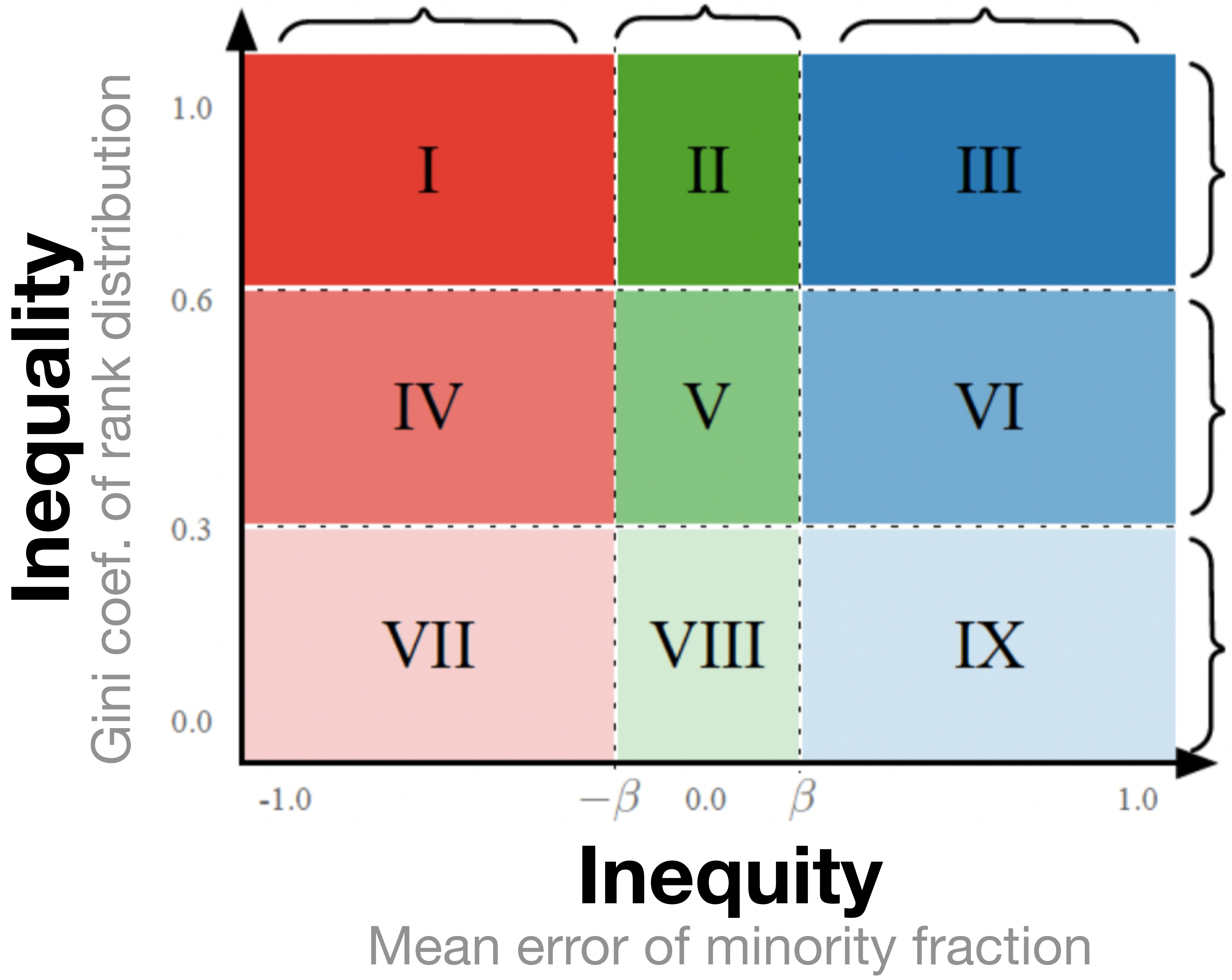
n = 20
d = 0.2
A = 3.0

f_m = 20% (min)
80% (maj)

h_{mm} = 0.2
h_{MM} = 0.2



9 regions of disparity



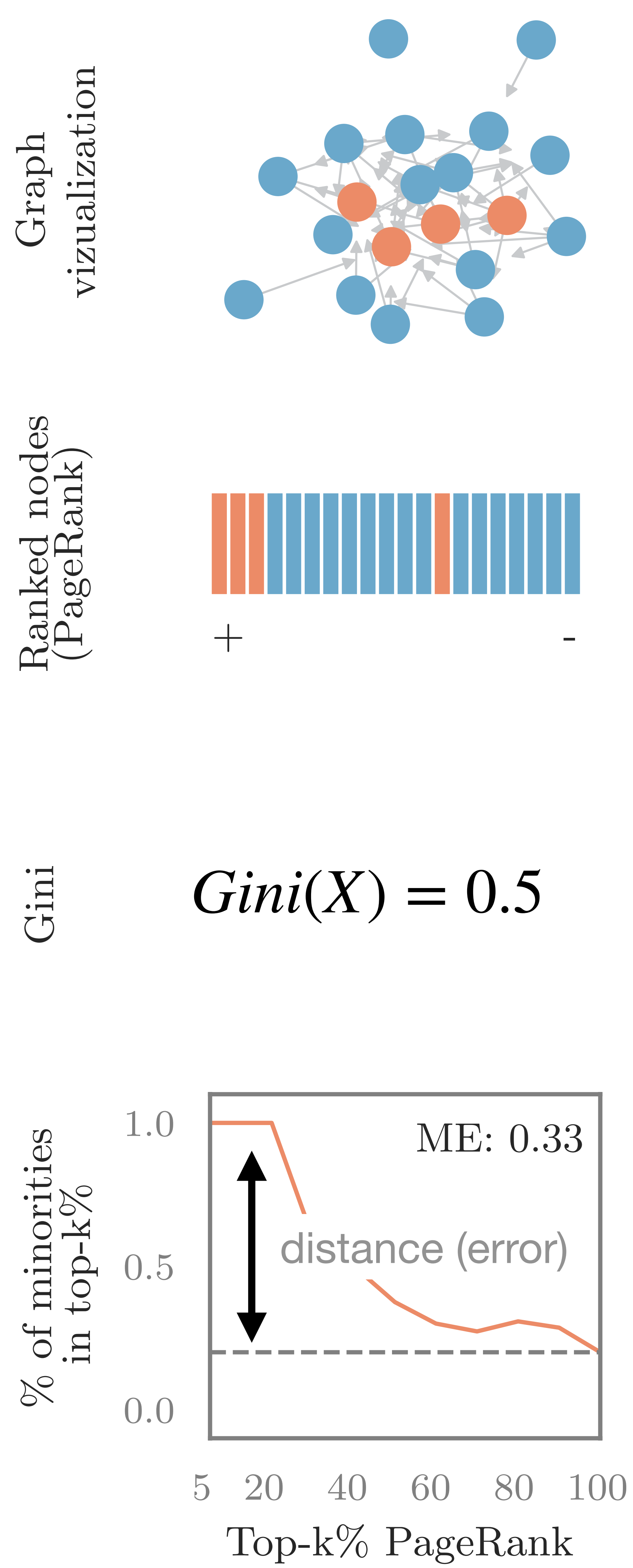
Espín-Noboa et al. (2022)

Network characteristics:

$n = 20$
$d = 0.2$
$A = 3.0$

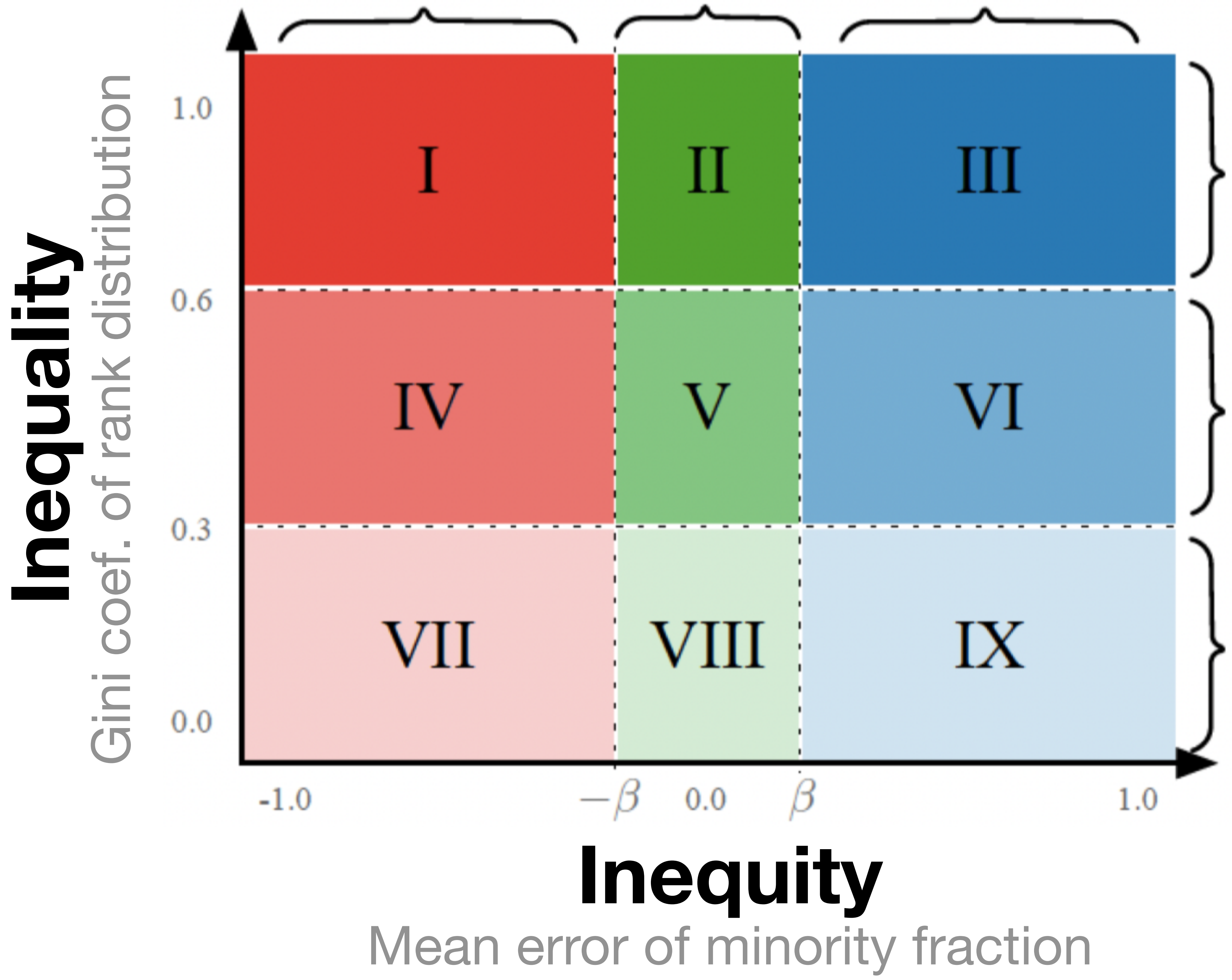
$f_m = 20\%$ (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



9 regions of disparity

Under-
represented



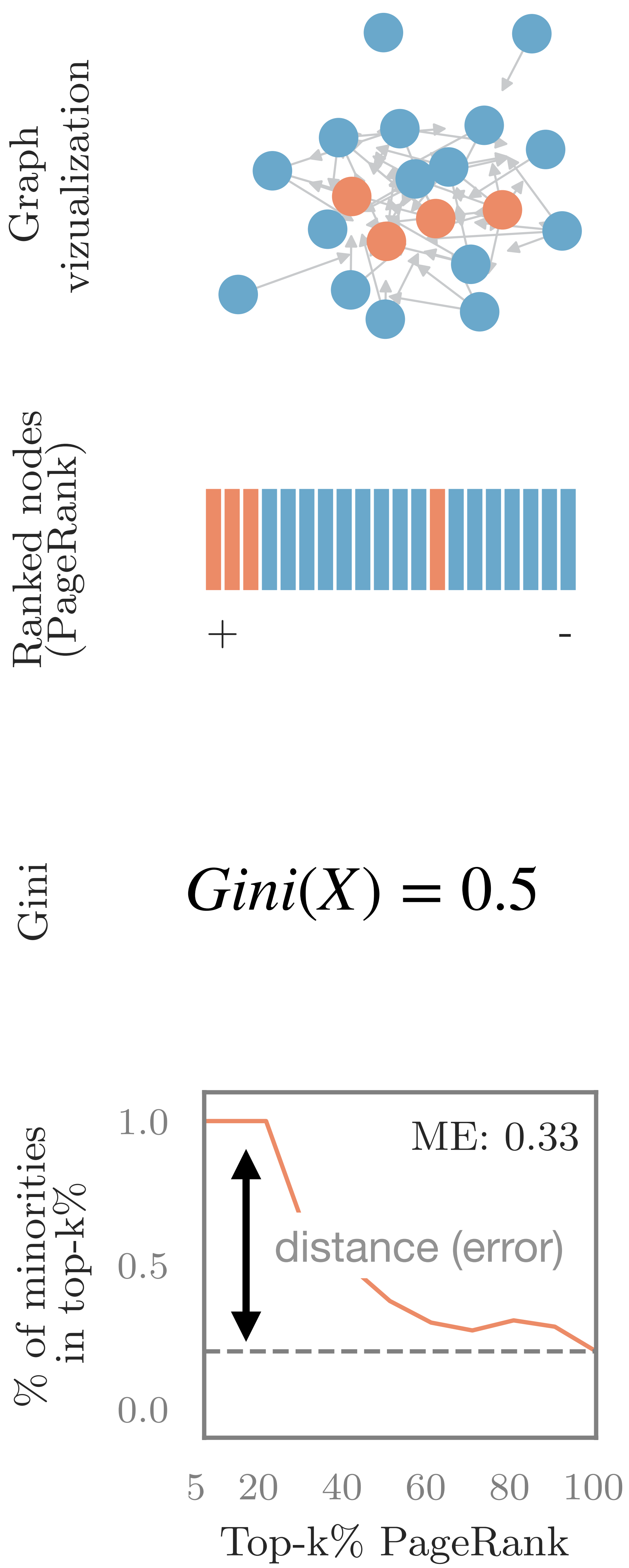
Espín-Noboa et al. (2022)

Network characteristics:

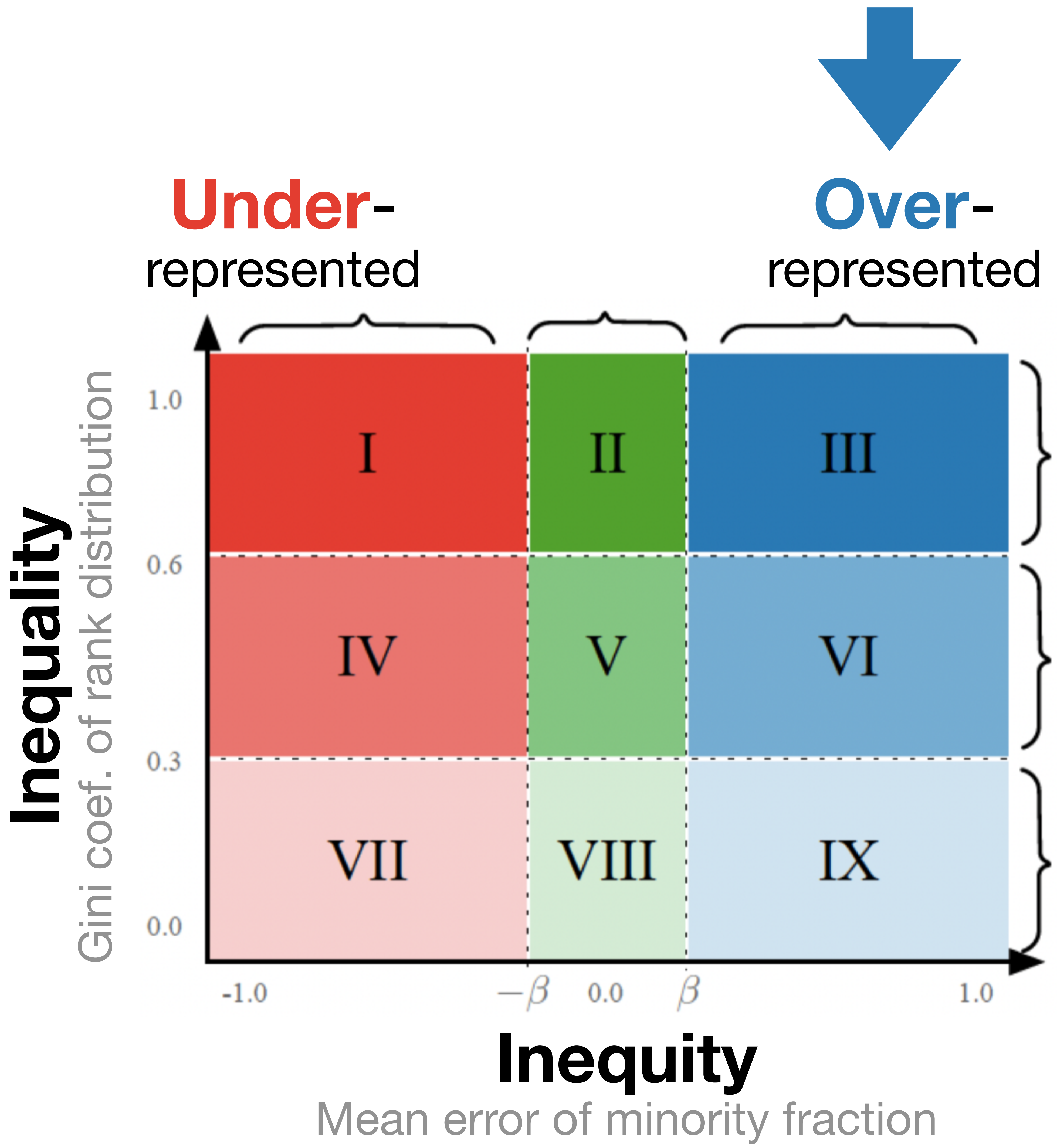
$n = 20$
$d = 0.2$
$A = 3.0$

f_m = 20% (min)
80% (maj)

$h_{mm} = 0.2$
$h_{MM} = 0.2$



9 regions of disparity



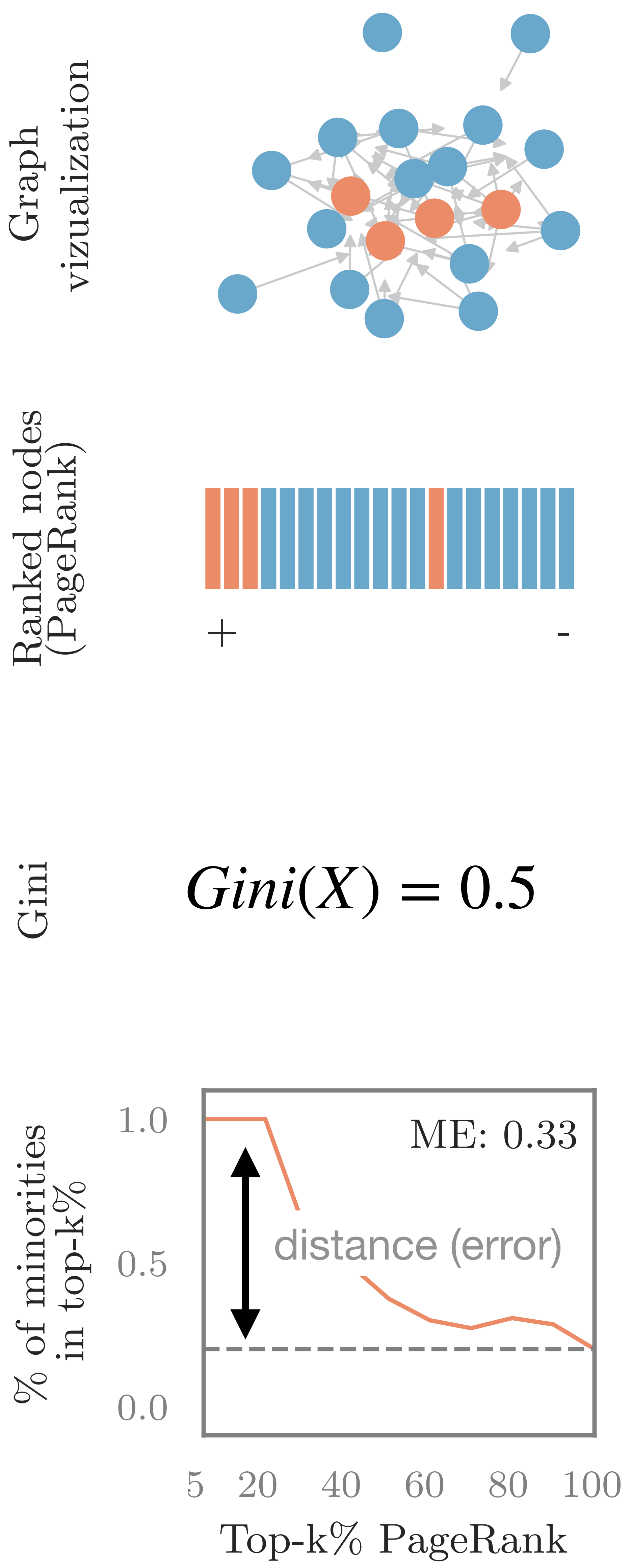
Espín-Noboa et al. (2022)

Network characteristics:

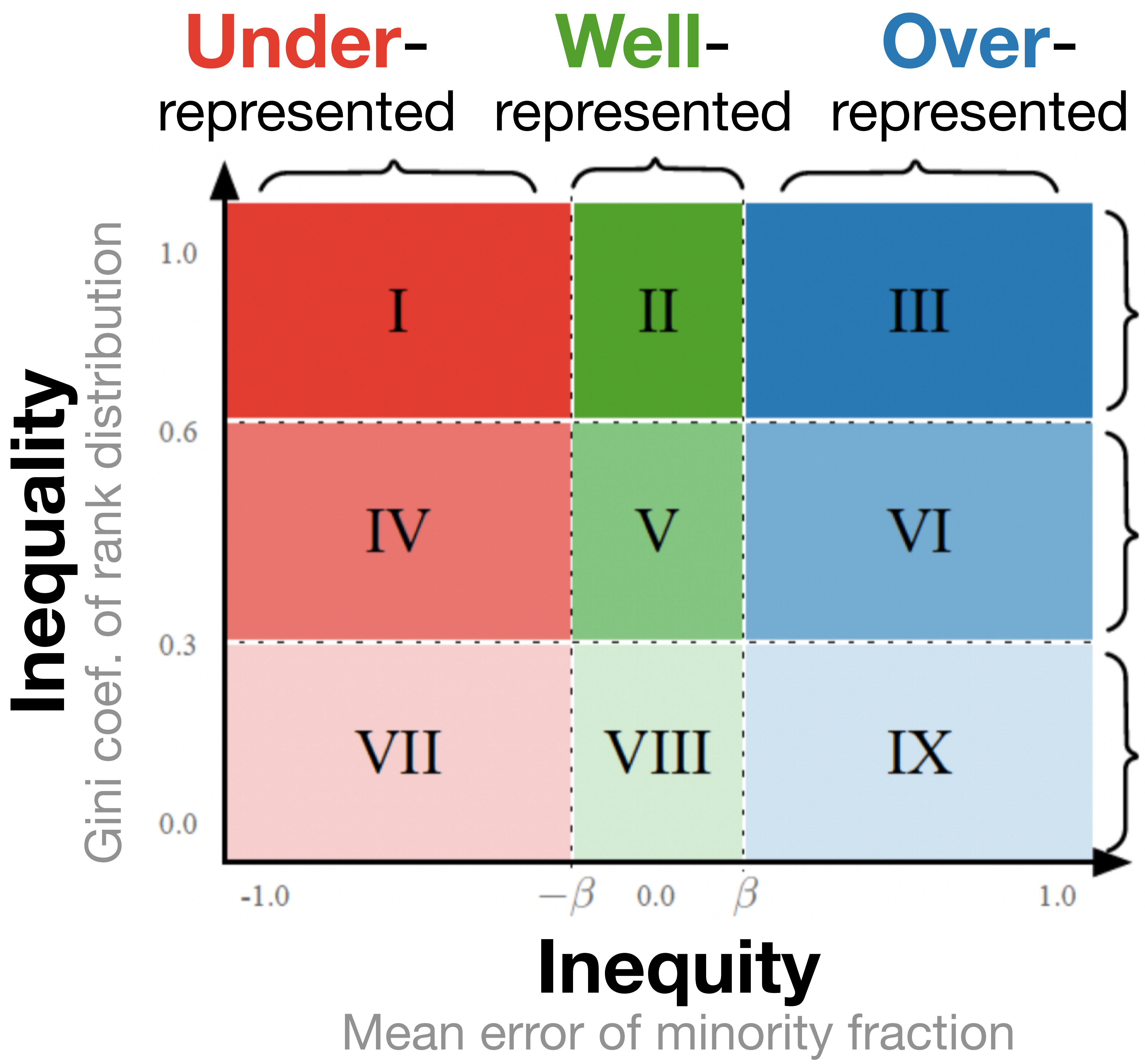
n = 20
d = 0.2
A = 3.0

f_m = 20% (min)
80% (maj)

h_{mm} = 0.2
h_{MM} = 0.2



9 regions of disparity



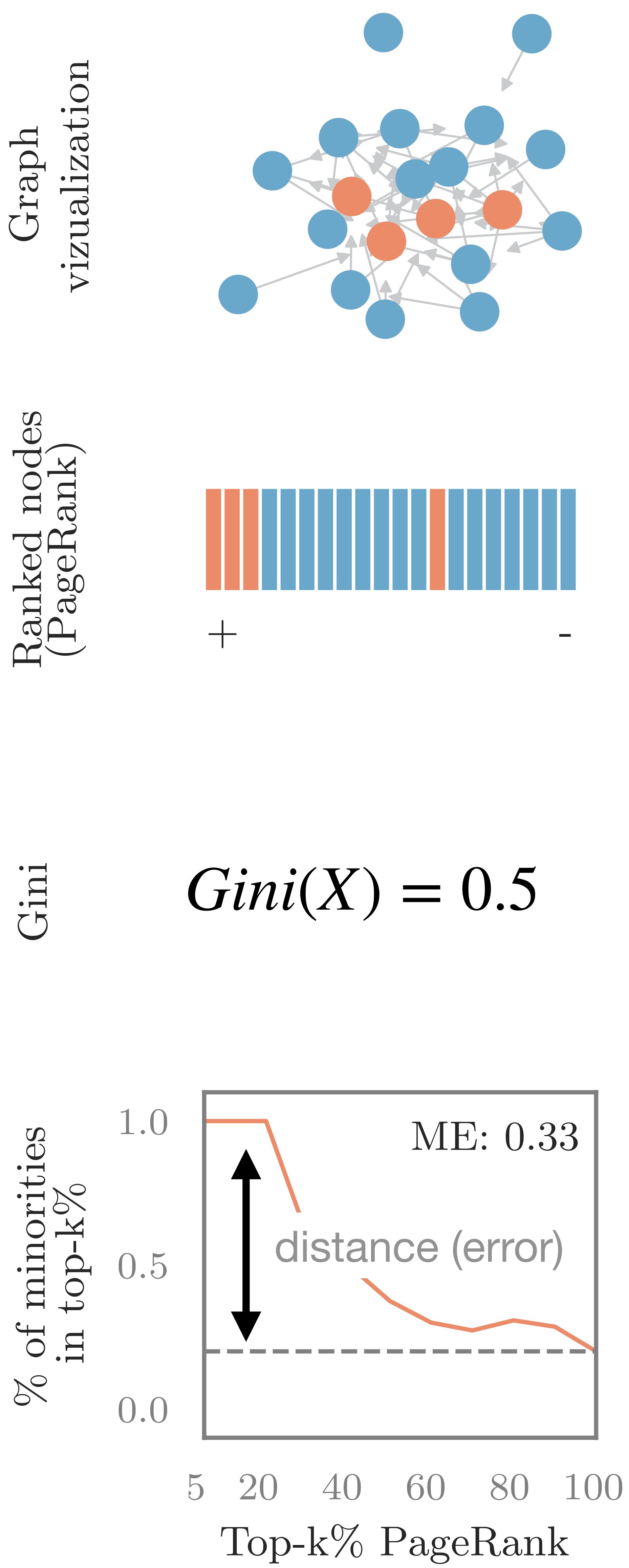
Espín-Noboa et al. (2022)

Network characteristics:

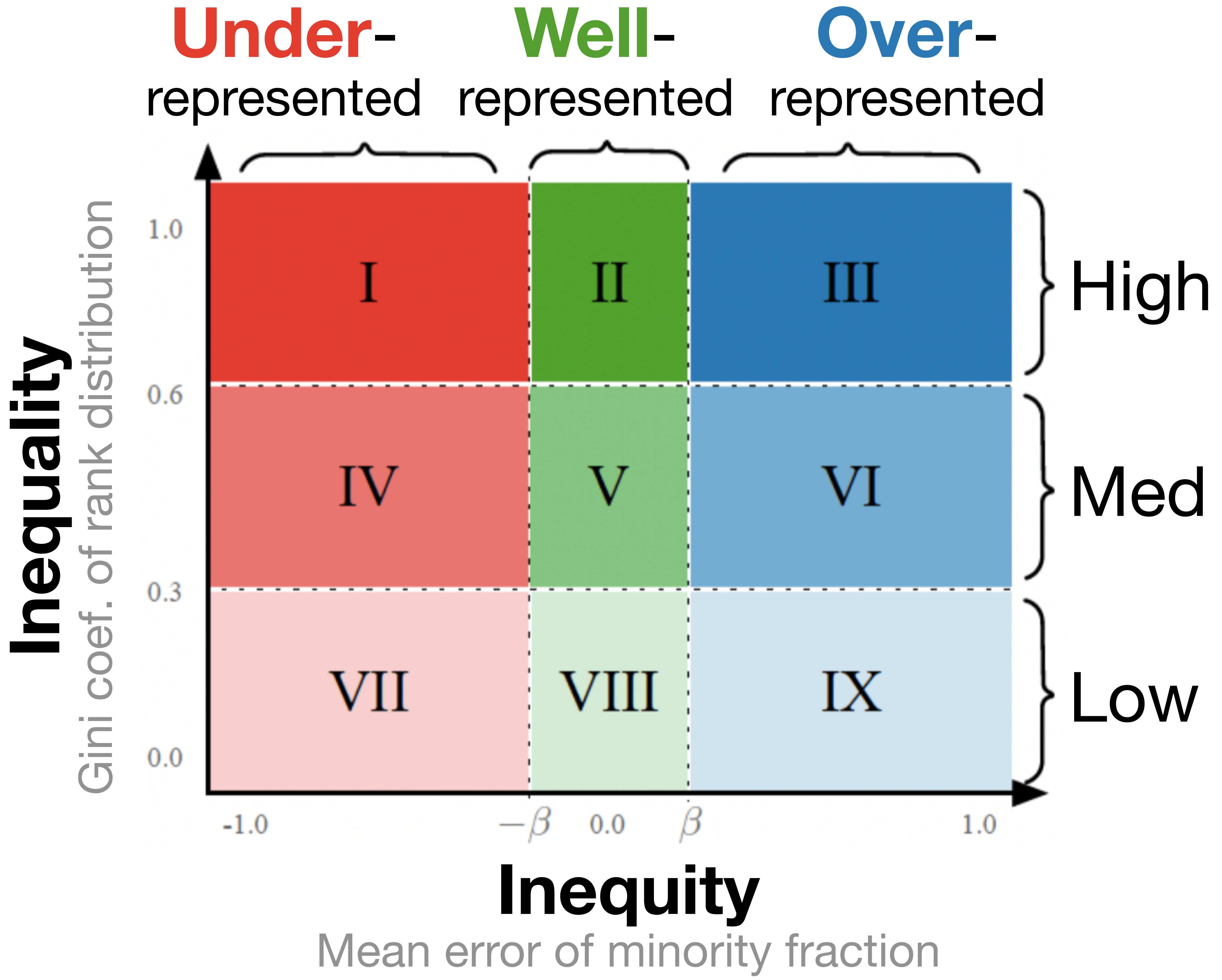
n = 20
d = 0.2
A = 3.0

f_m = 20% (min)
80% (maj)

h_{mm} = 0.2
h_{MM} = 0.2



9 regions of disparity



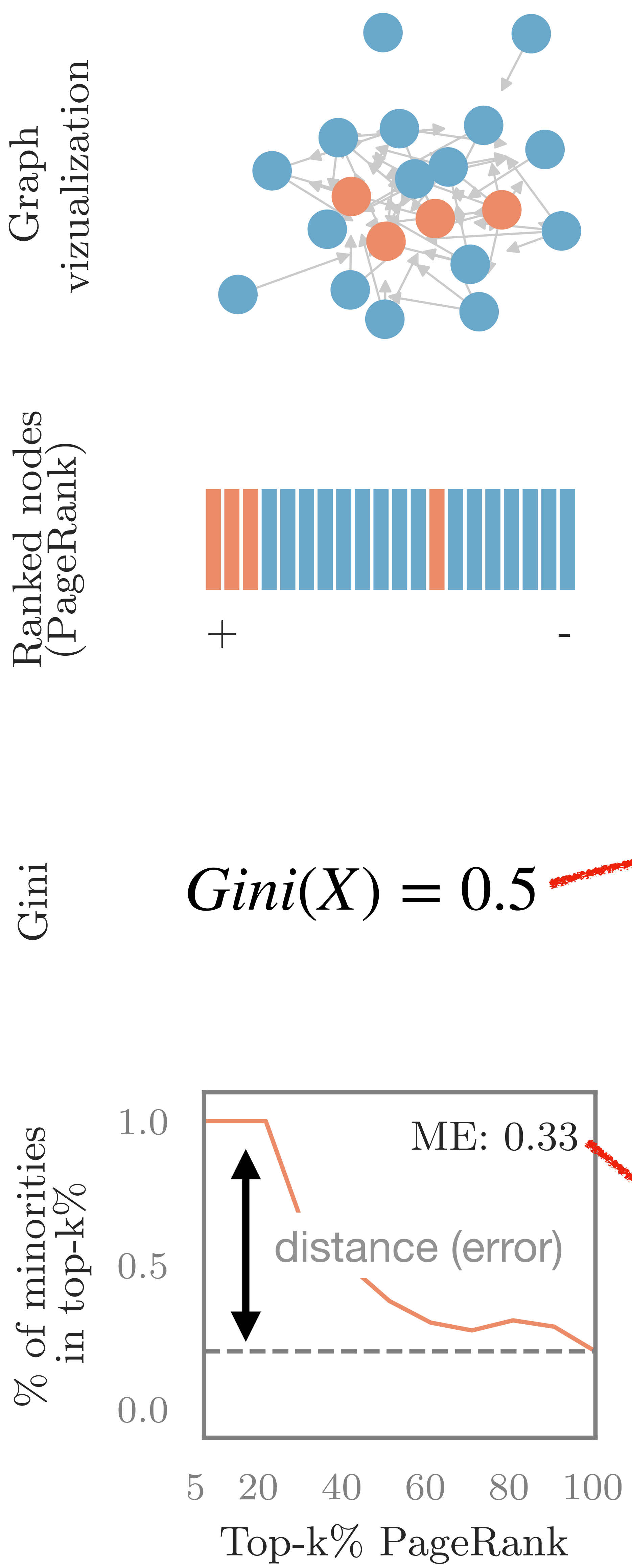
Espín-Noboa et al. (2022)

Network characteristics:

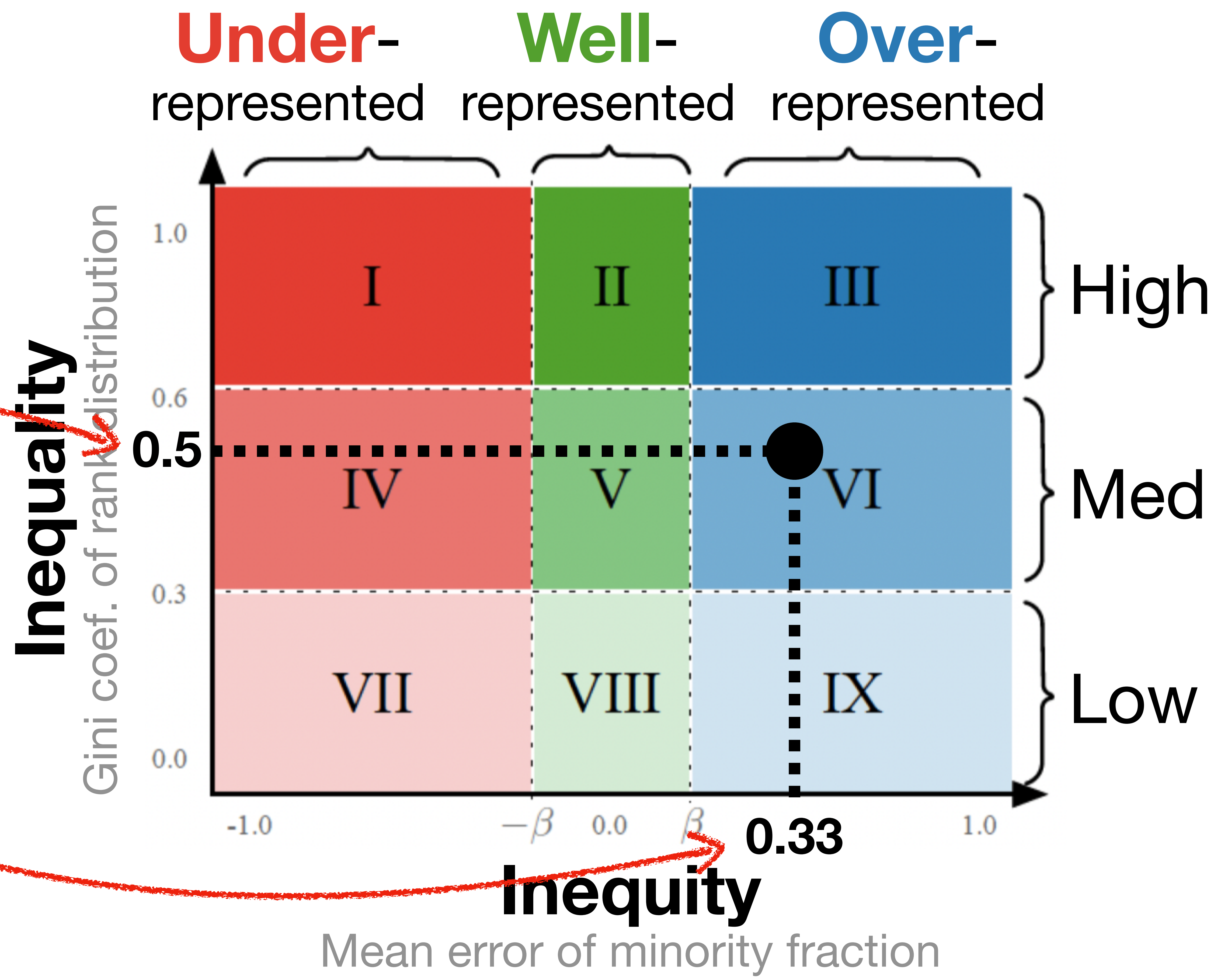
n = 20
d = 0.2
A = 3.0

f_m = 20% (min)
80% (maj)

h_{mm} = 0.2
h_{MM} = 0.2



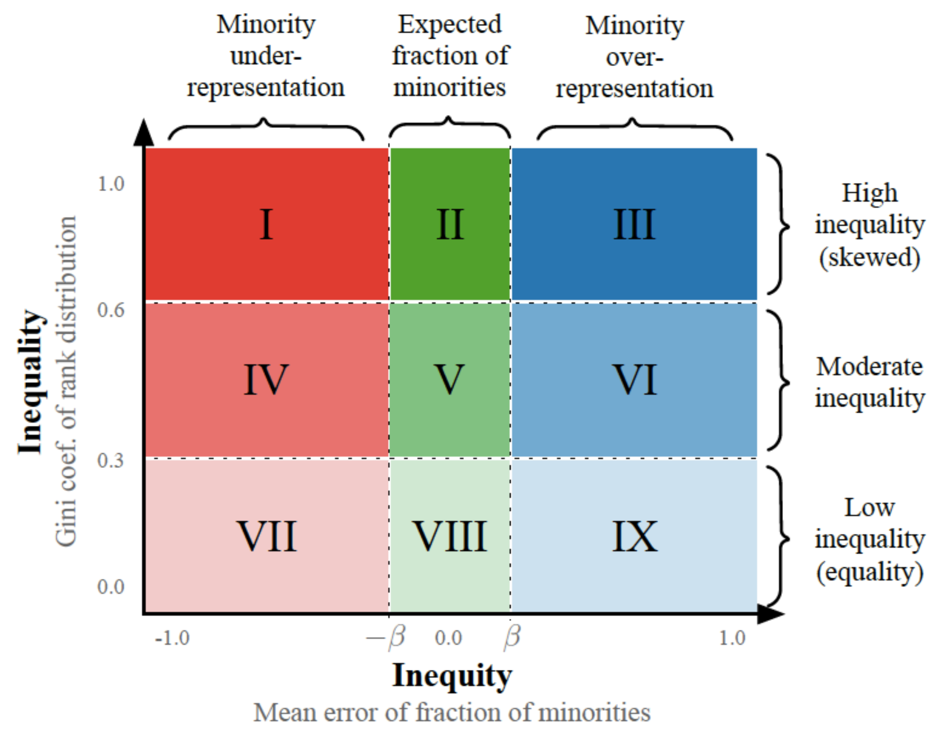
9 regions of disparity



Espín-Noboa et al. (2022)

Disparity in PageRank

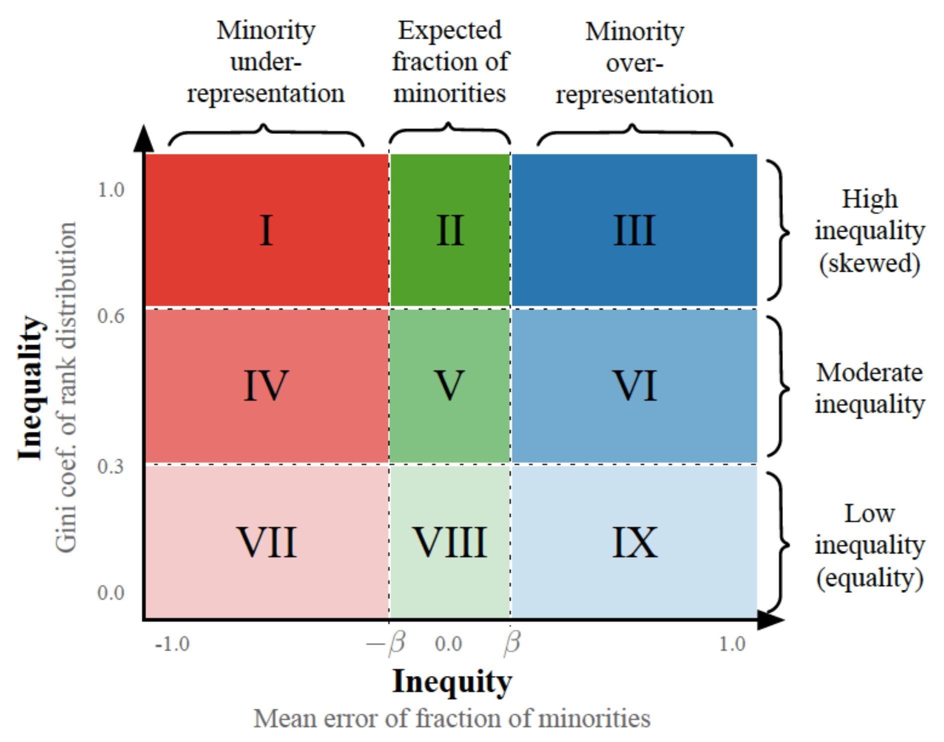
as a function of homophily and fraction of minority nodes

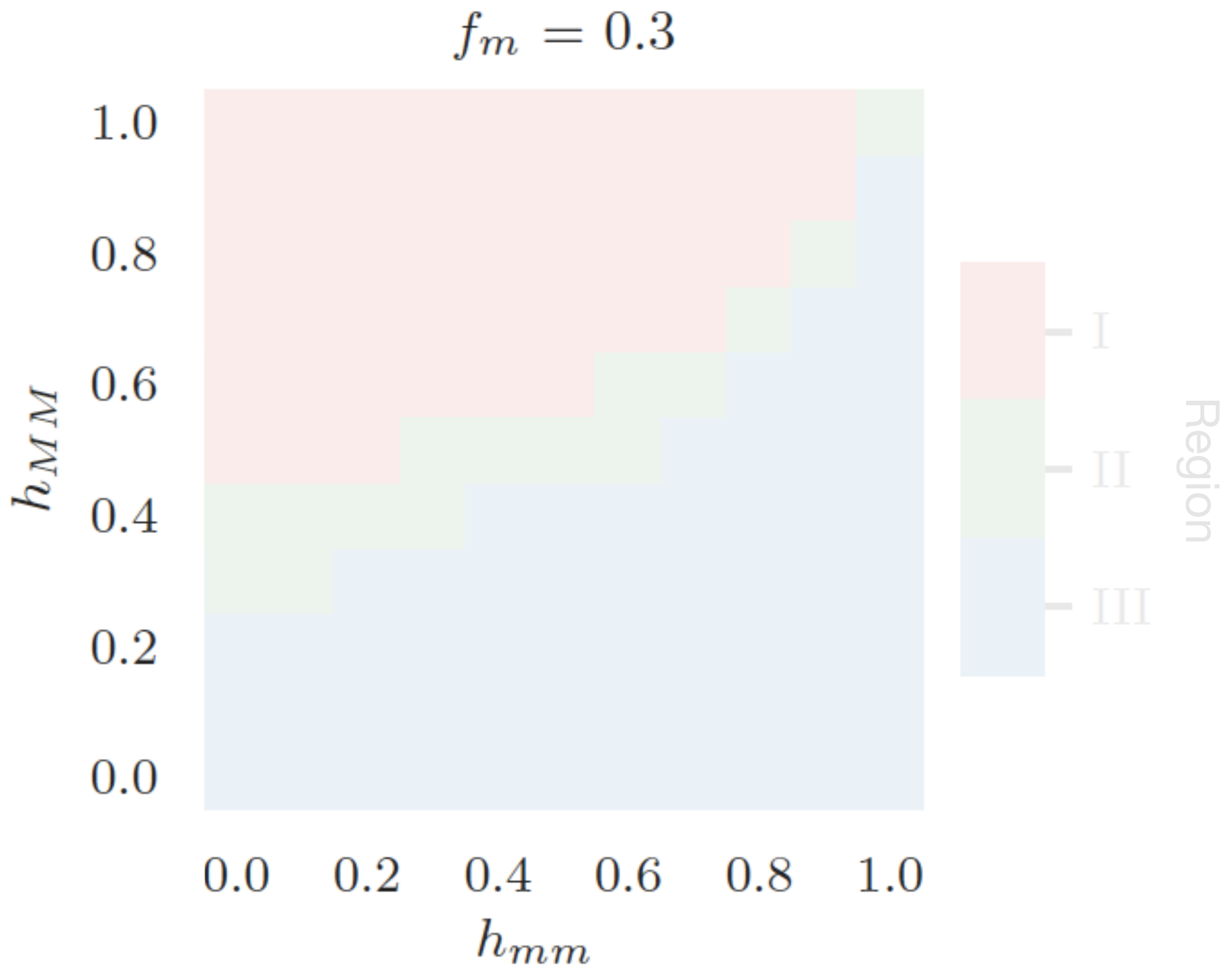


Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

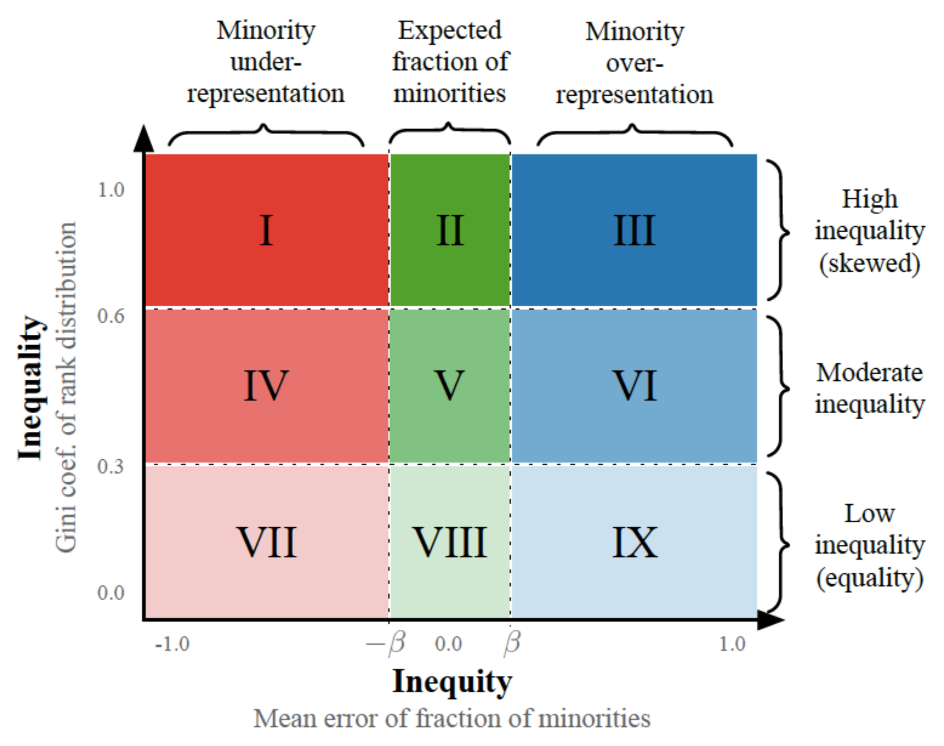




Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

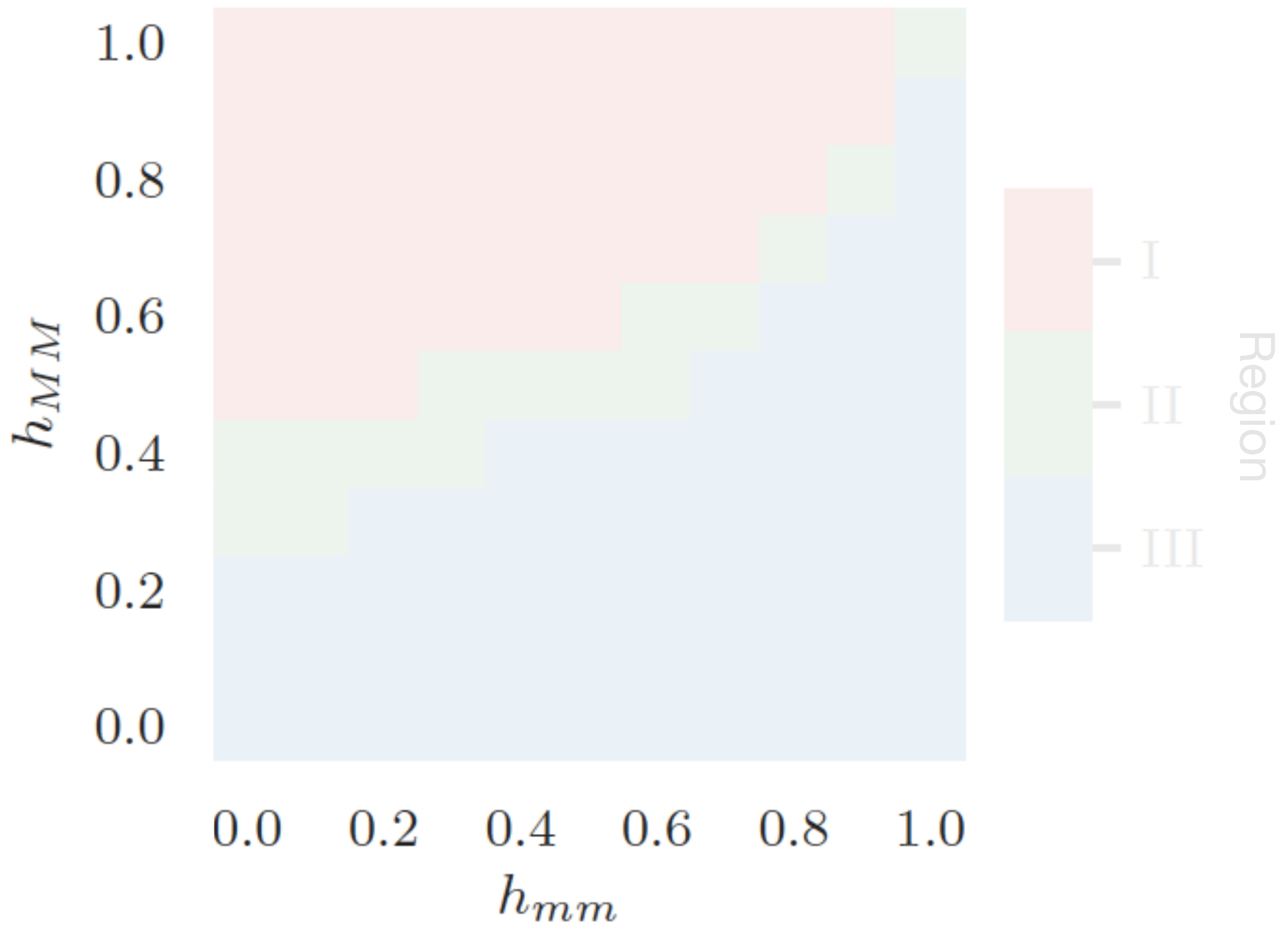
as a function of homophily and fraction of minority nodes



Column
Fraction of minority group

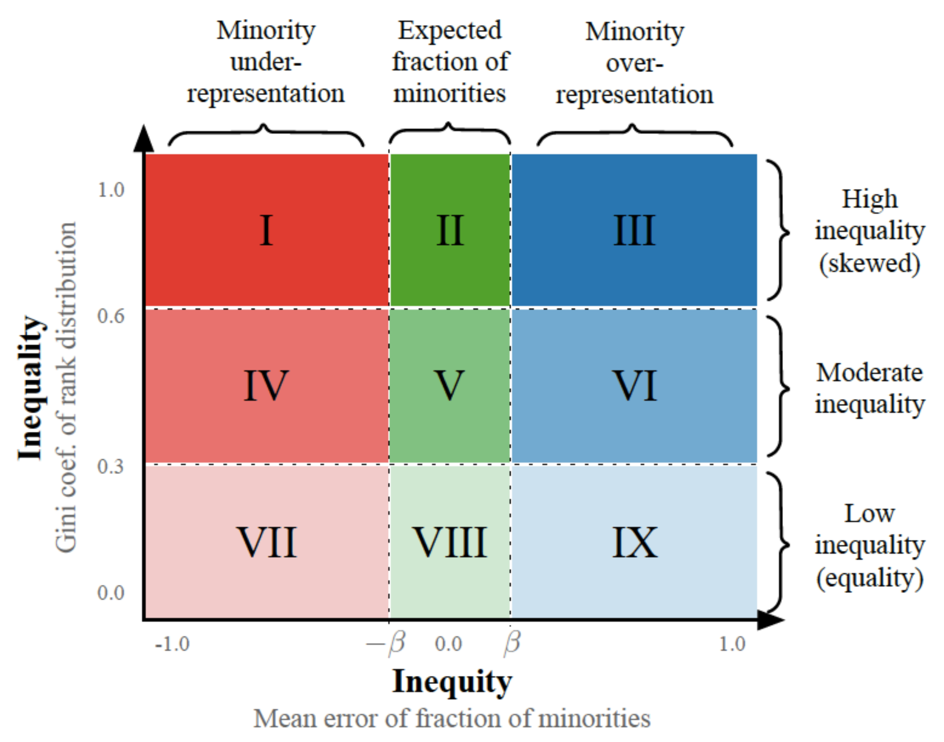
unbalance

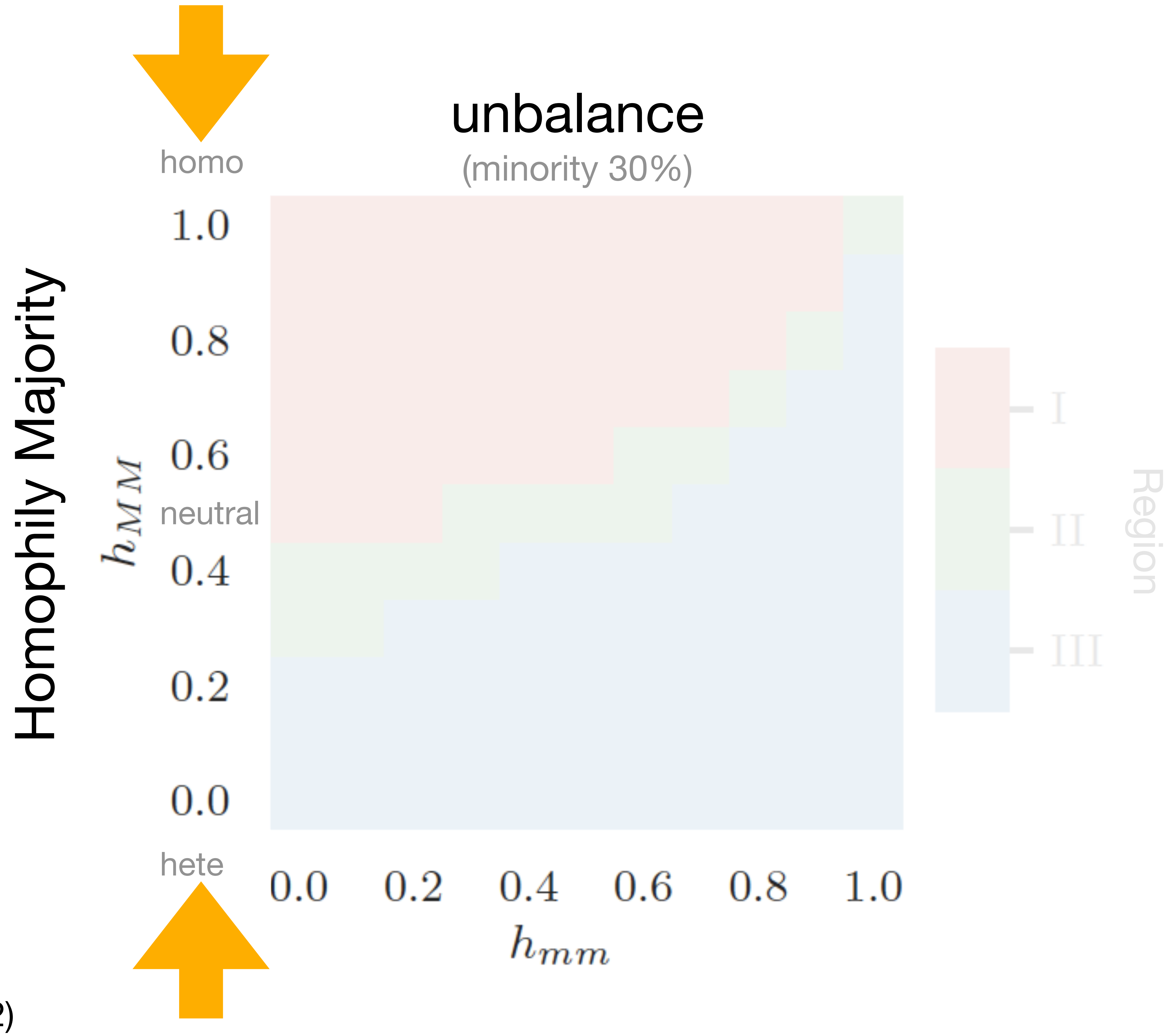
(minority 30%)



Disparity in PageRank

as a function of homophily and fraction of minority nodes

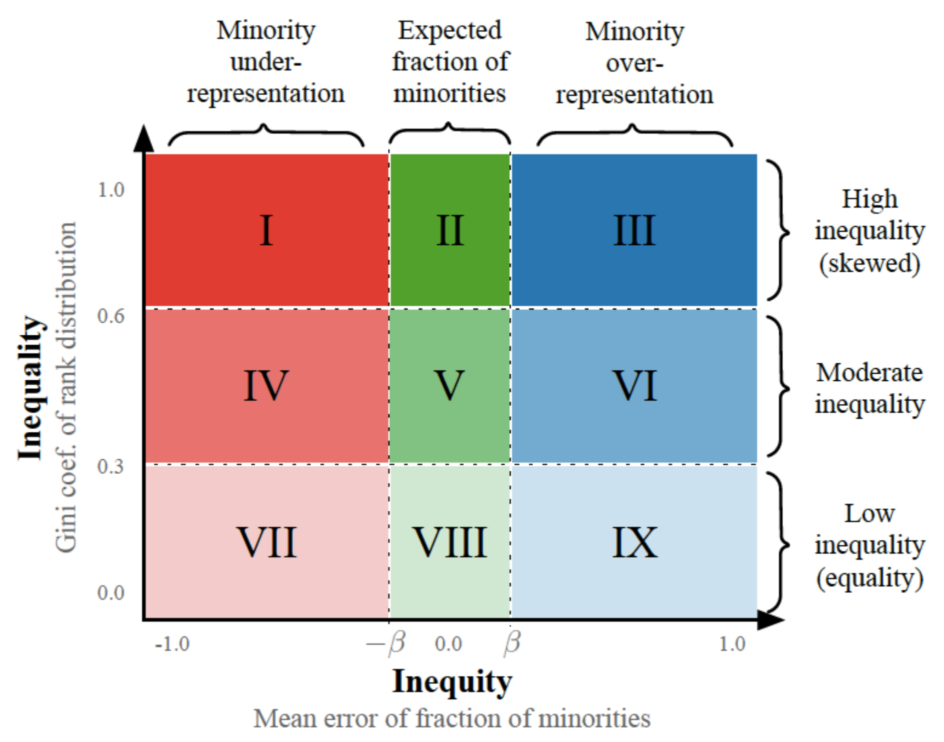


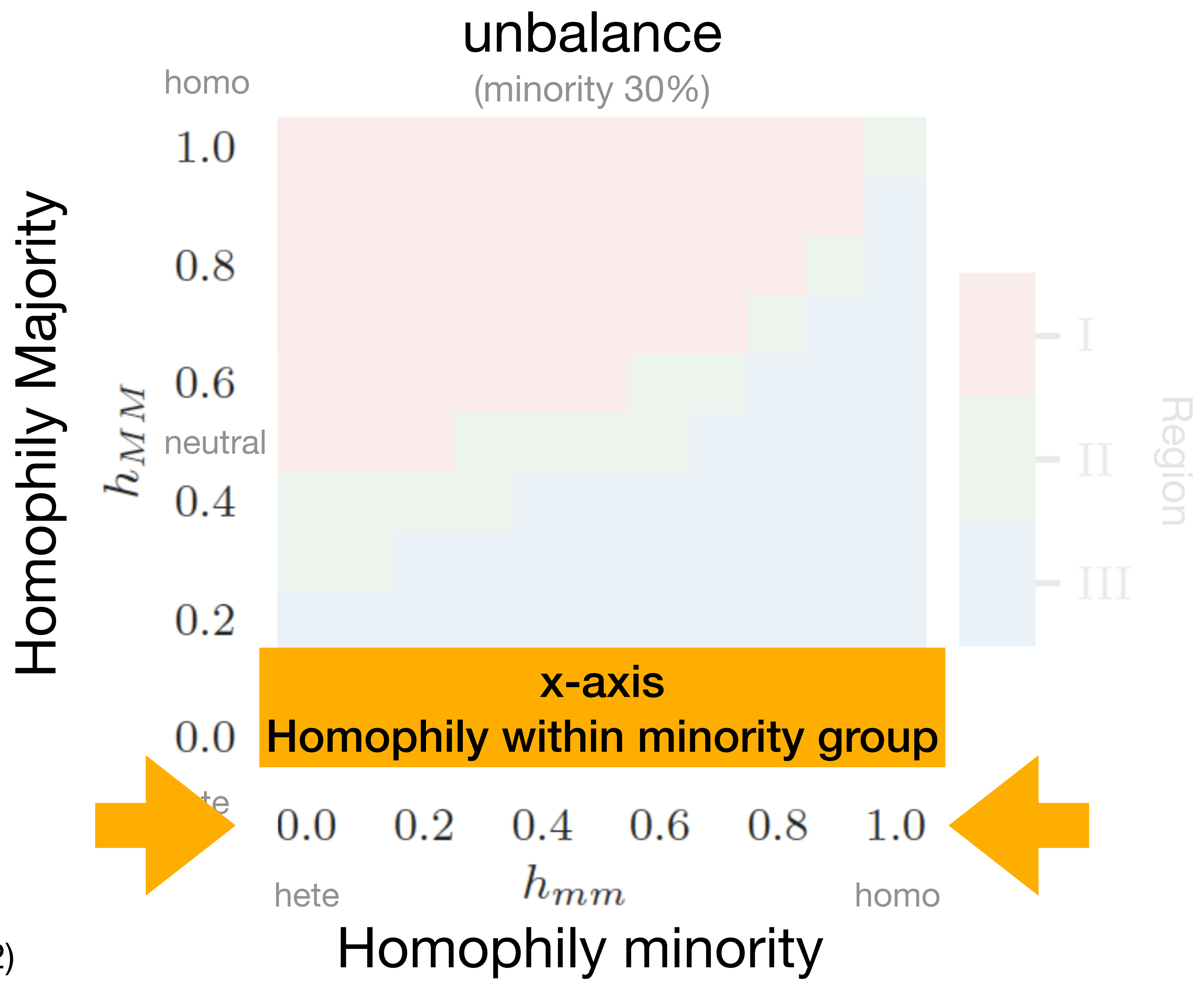


y-axis
Homophily within majority group

Disparity in PageRank

as a function of homophily and fraction of minority nodes

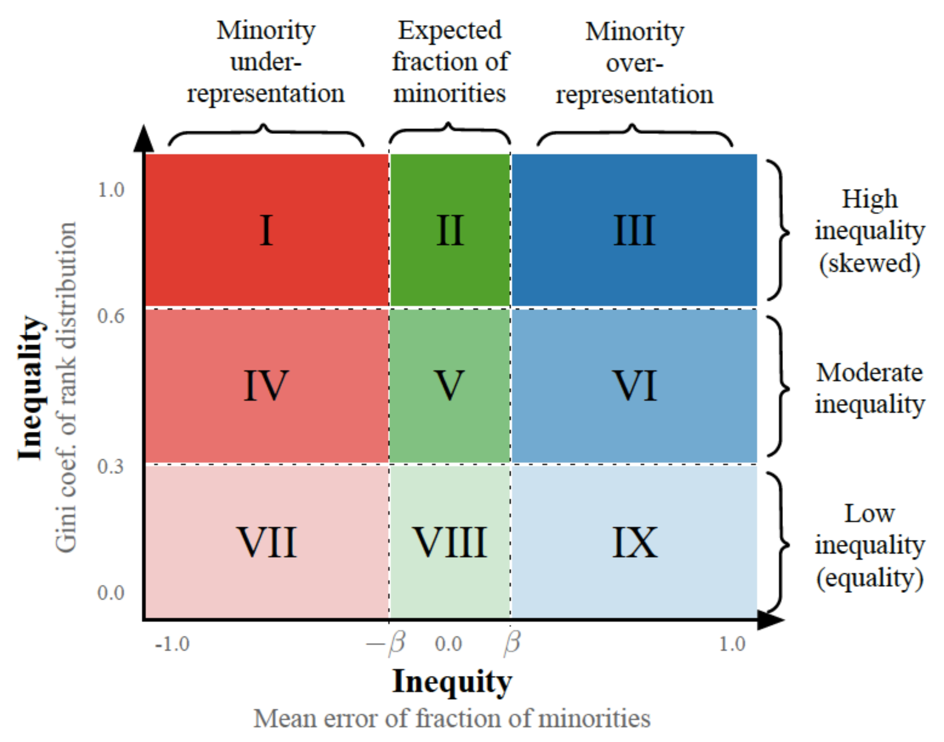


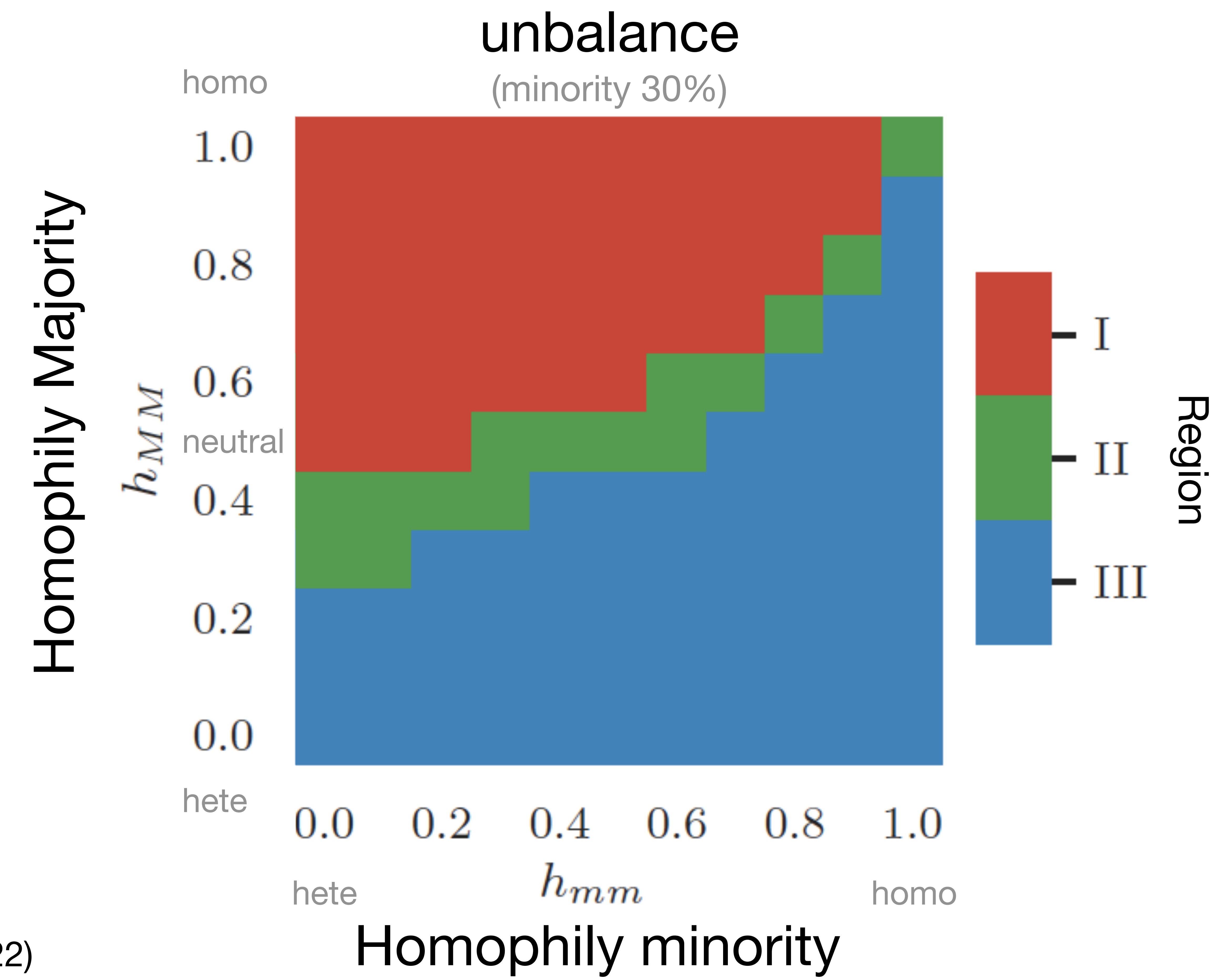


Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes





Under-representation
Less than 30% in top-k

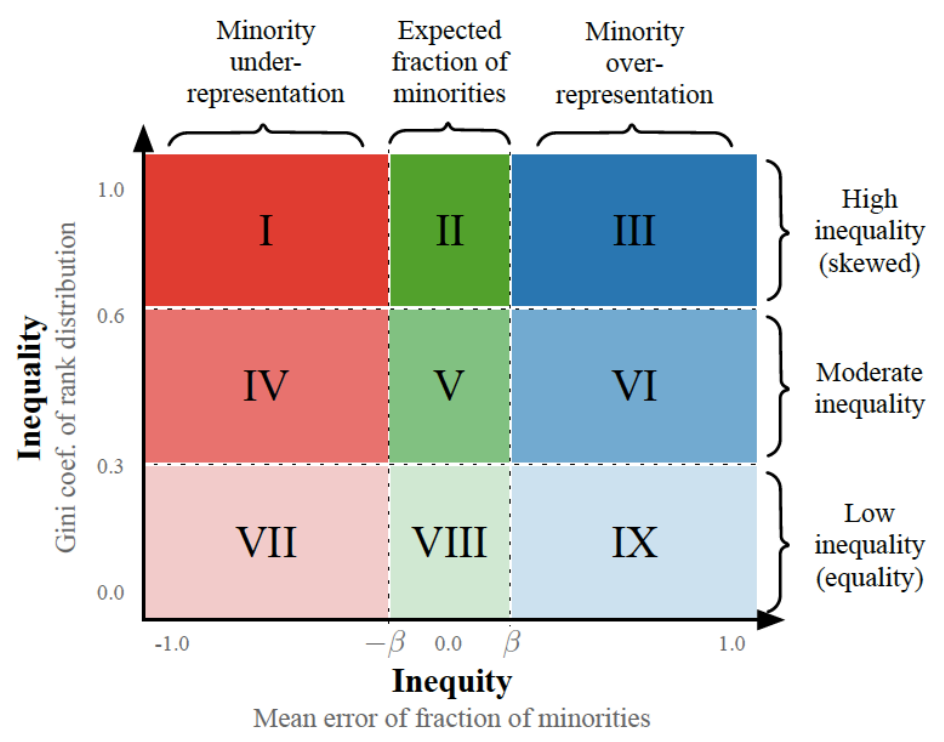
Well-representation
30% in top-k

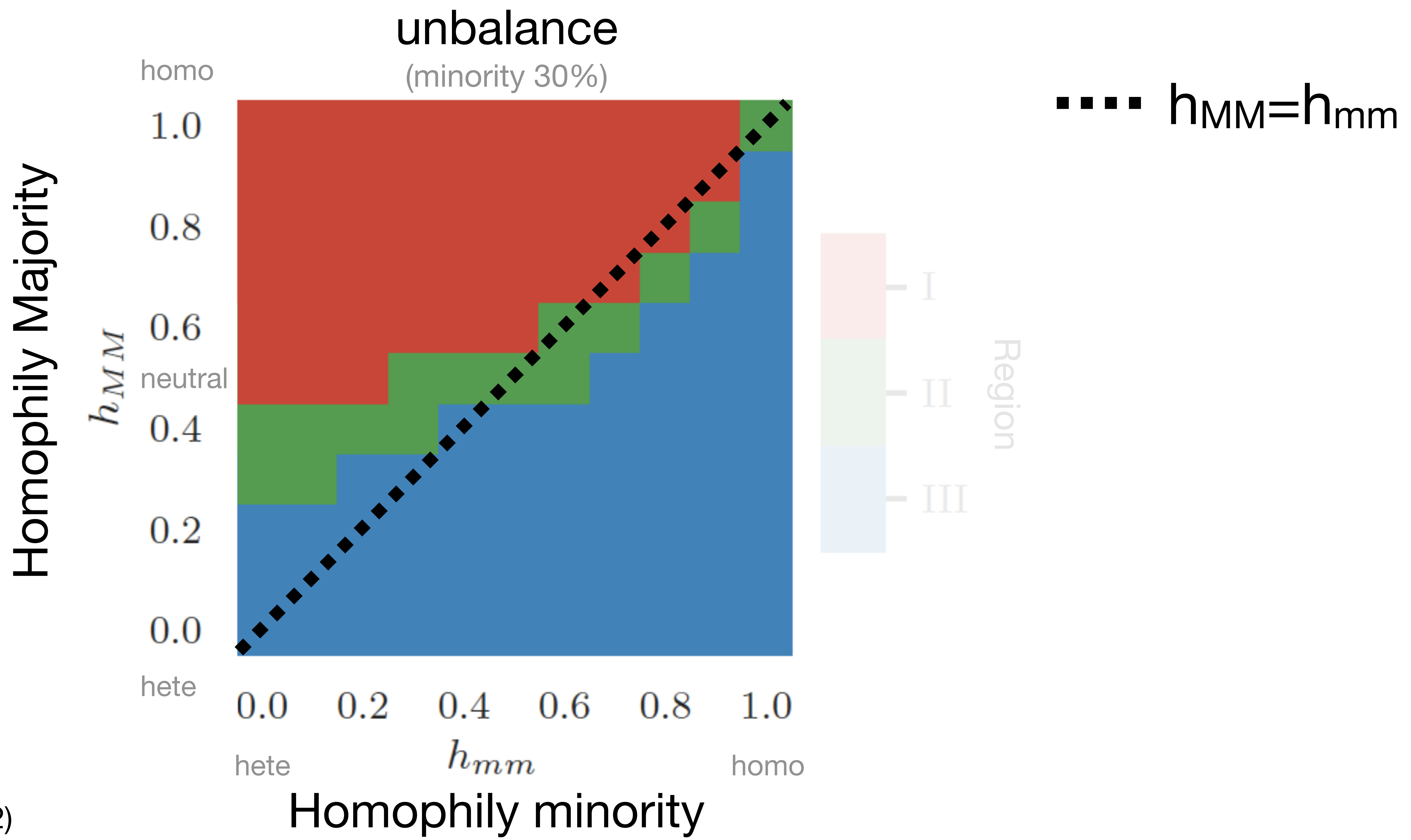
Over-representation
More than 30% in top-k

Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

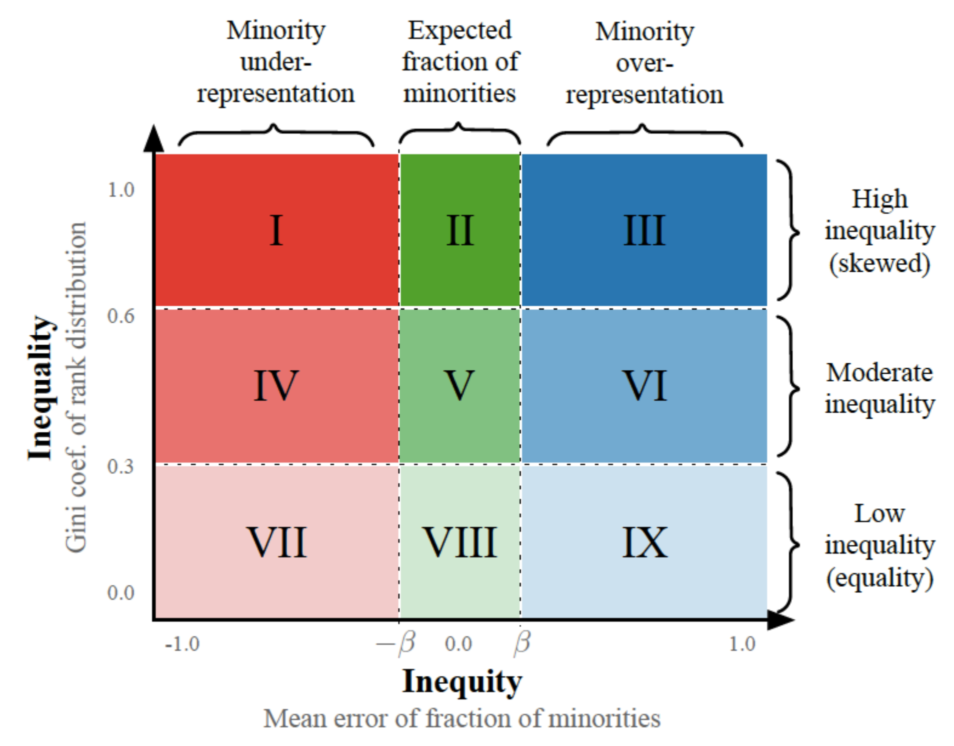




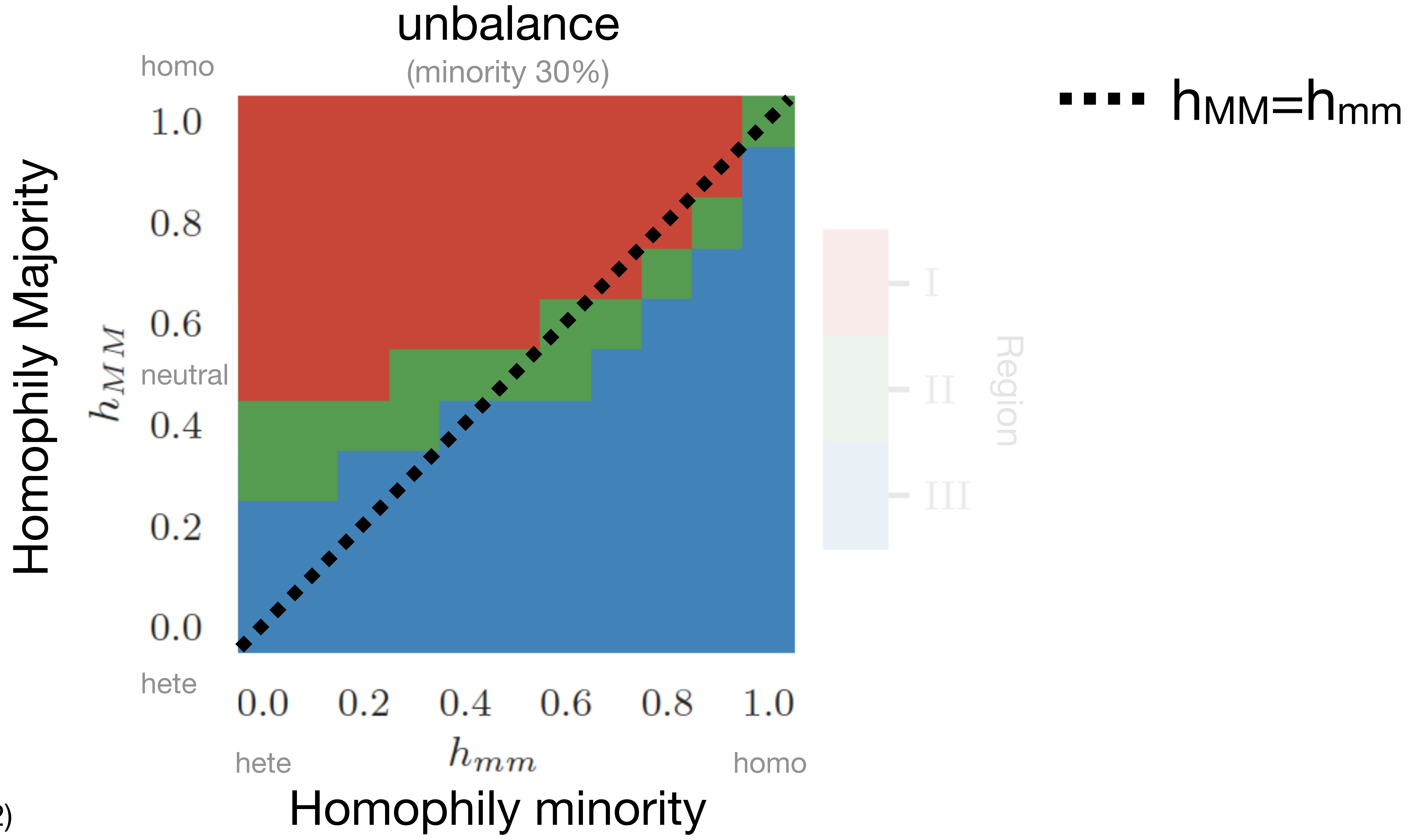
Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes



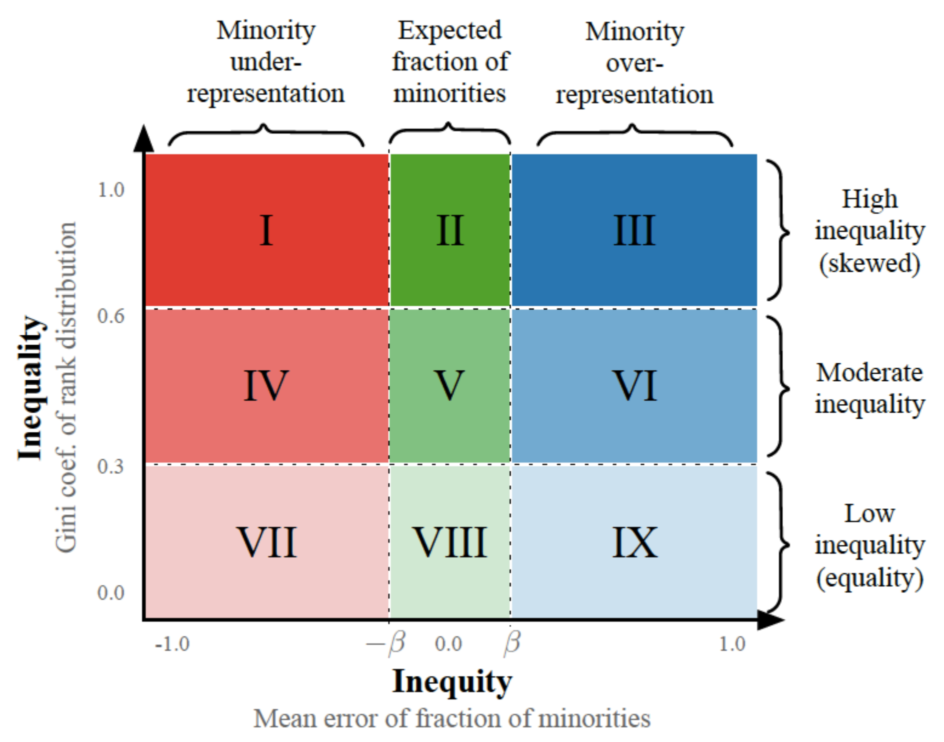
In 70-30 unbalanced networks, minorities are …



Espín-Noboa et al. Nature Scientific Reports (2022)

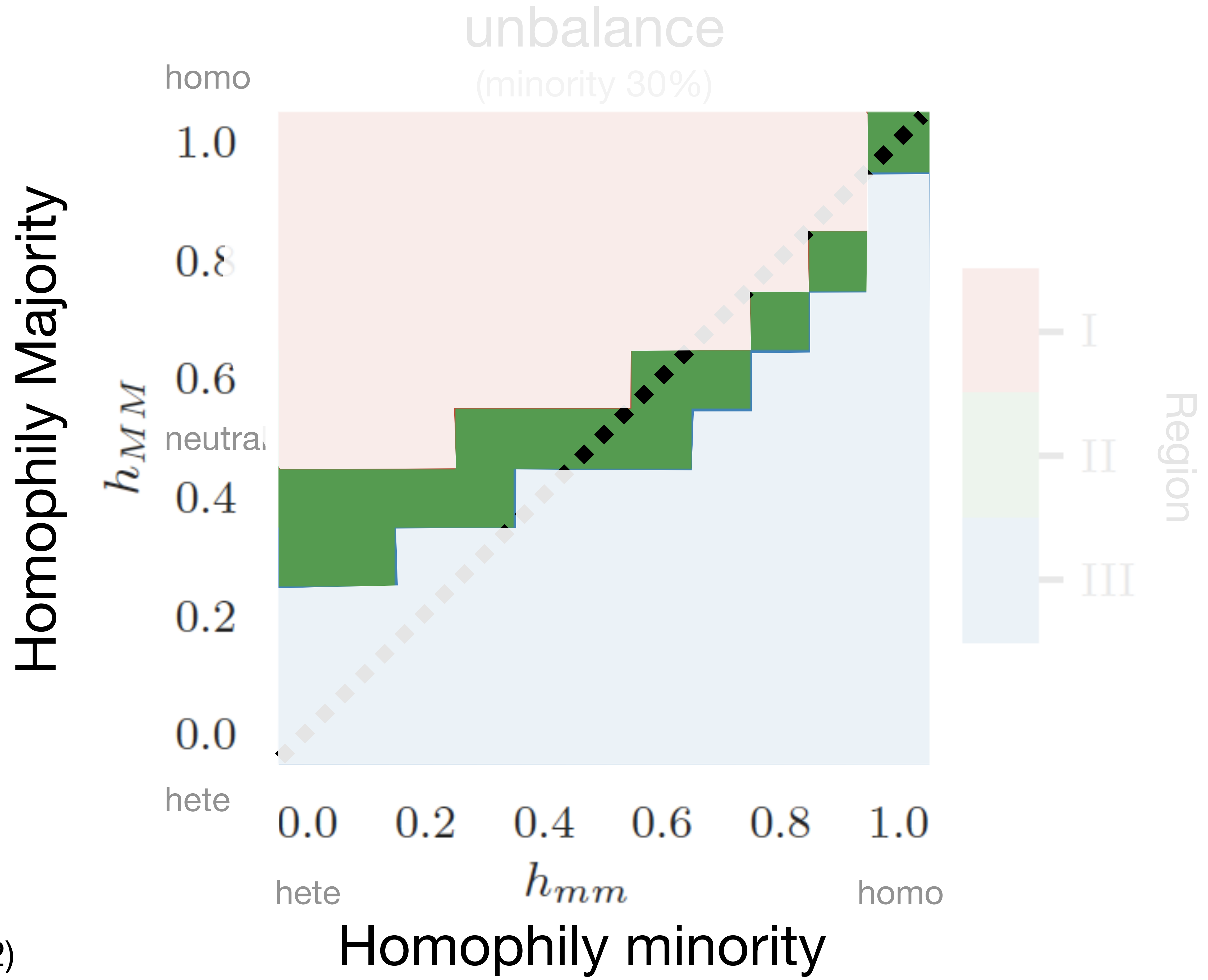
Disparity in PageRank

as a function of homophily and fraction of minority nodes



In 70-30 unbalanced networks, minorities are …

1 Well represented only in a few cases:
- $h_{MM}<0.5$ and $h_{mm}<h_{MM}$
- $h_{MM}>0.5$ and $h_{mm}=h_{MM}+1$

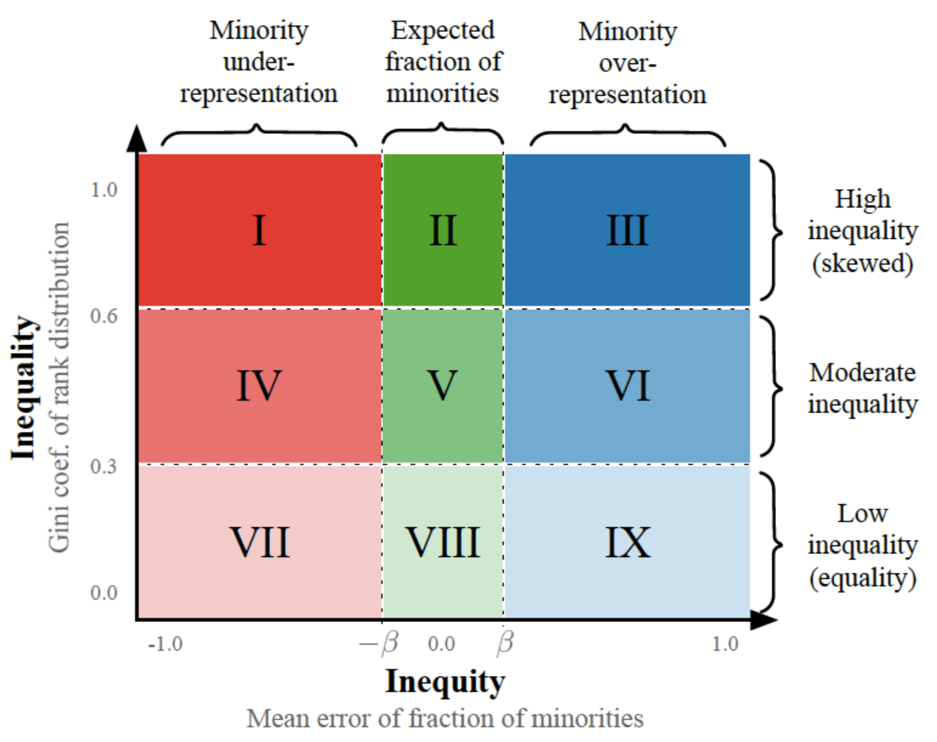


$h_{MM}=h_{mm}$

Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

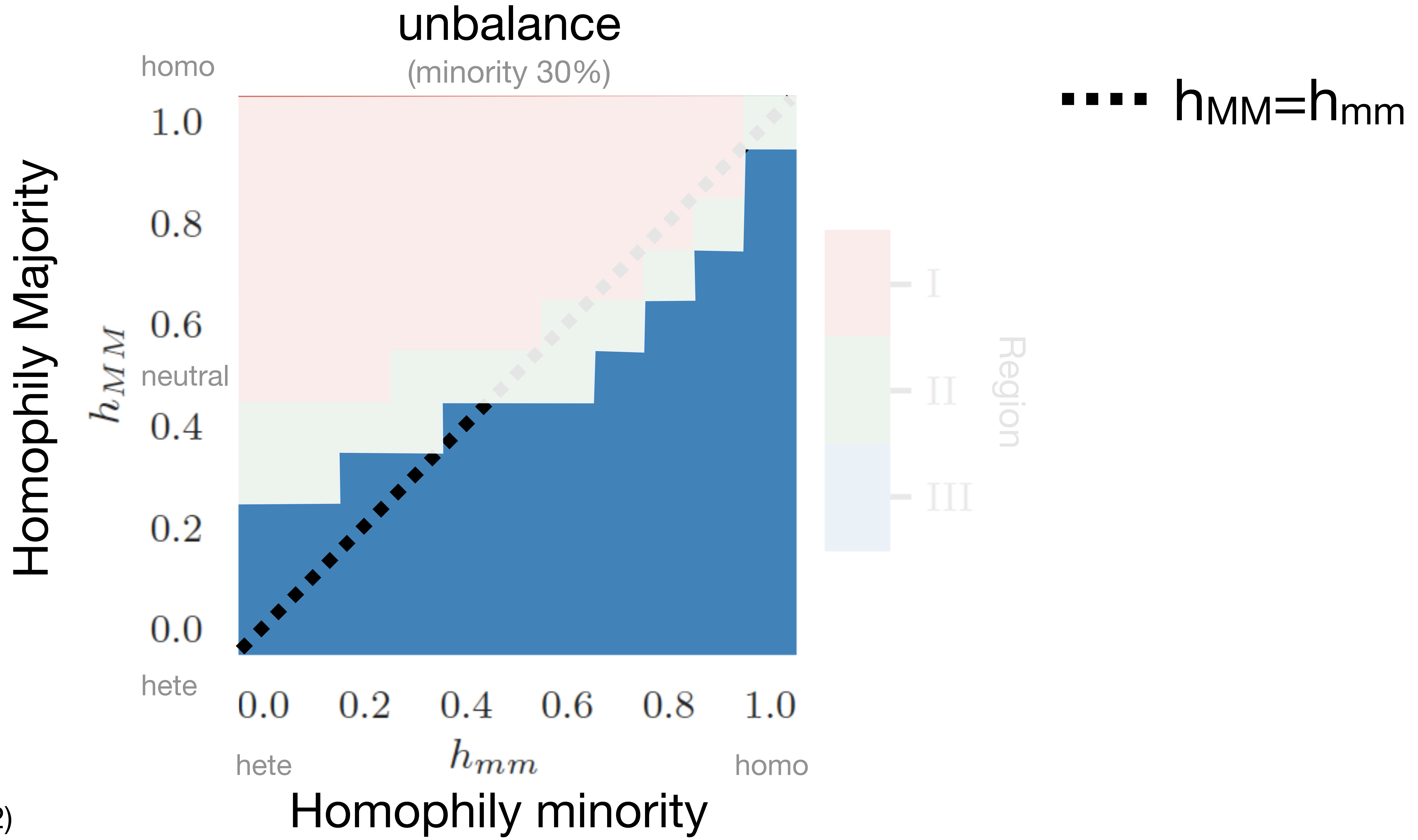
as a function of homophily and fraction of minority nodes



In 70-30 unbalanced networks, minorities are …

1 **Well represented only in a few cases:**
-h_{MM}<0.5 and h_{mm}<h_{MM}
-h_{MM}>0.5 and h_{mm}=h_{MM}+1

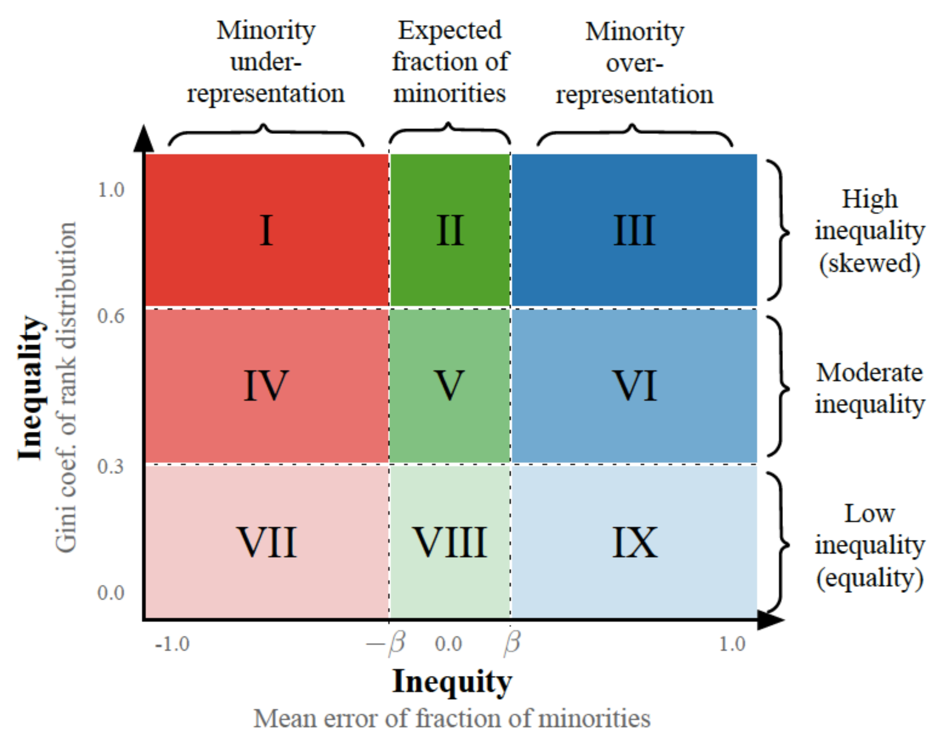
2 **Over-represented in most cases:**
-h_{MM}<h_{mm}
-h_{mm}<=h_{MM} and h_{MM}<0.3



Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

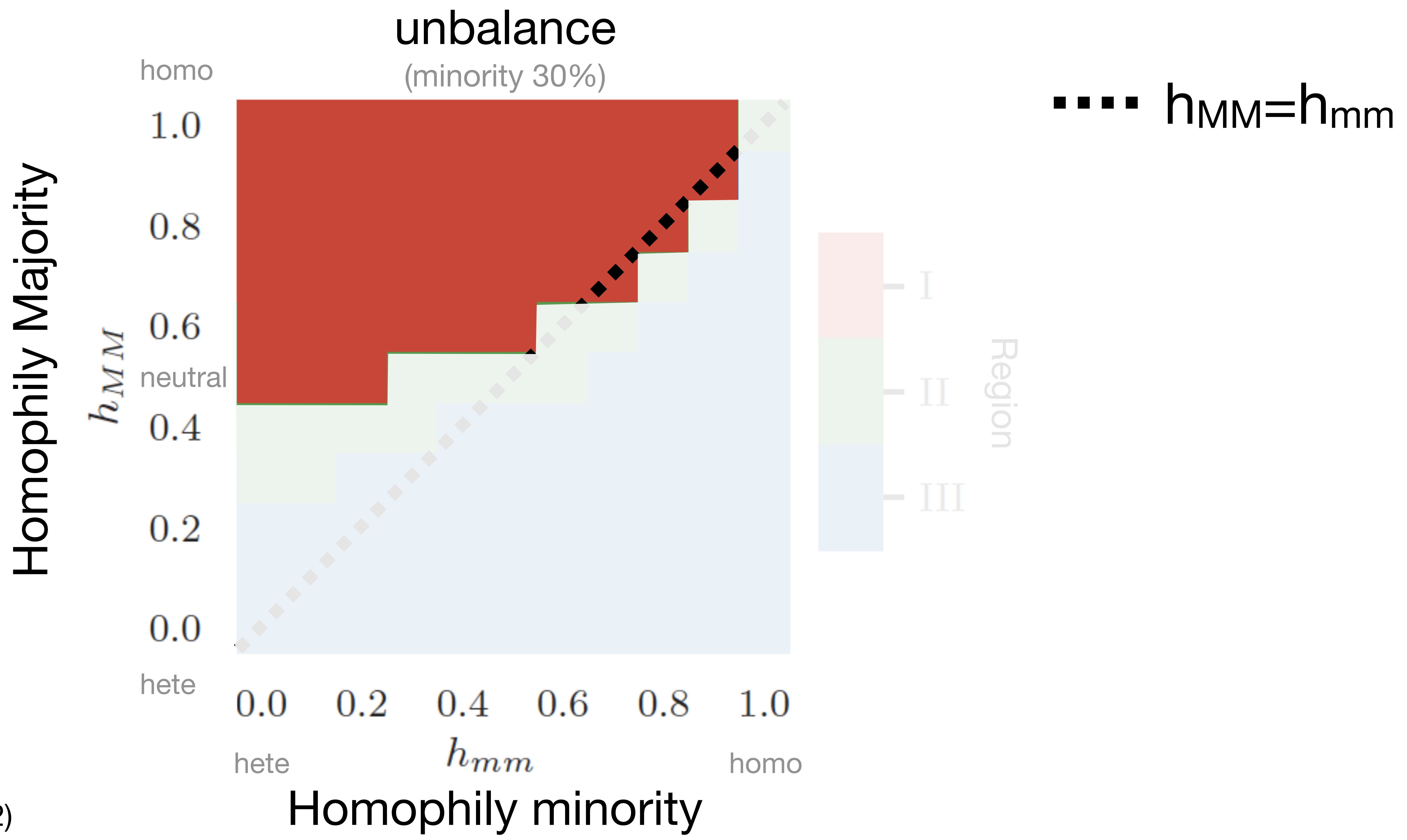


In 70-30 unbalanced networks, minorities are …

1 Well represented only in a few cases:
- $h_{MM}<0.5$ and $h_{mm}<h_{MM}$
- $h_{MM}>0.5$ and $h_{mm}=h_{MM}+1$

2 Over-represented in most cases:
- $h_{MM}<h_{mm}$
- $h_{mm}<=h_{MM}$ and $h_{MM}<0.3$

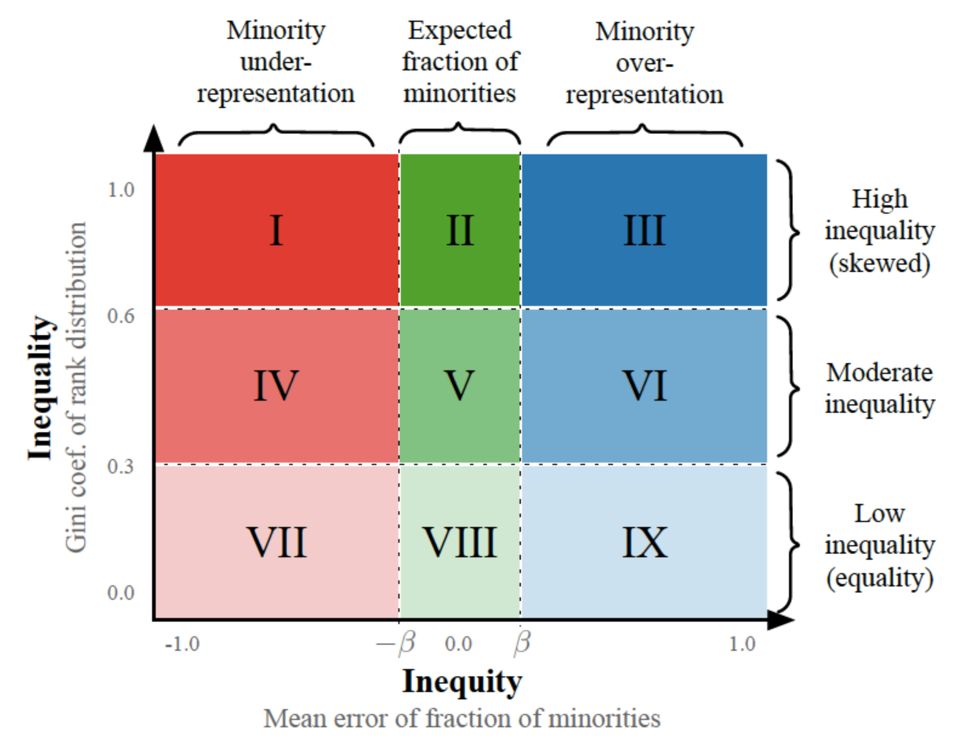
3 Under-represented in some cases:
- $h_{MM}>h_{mm}$



Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

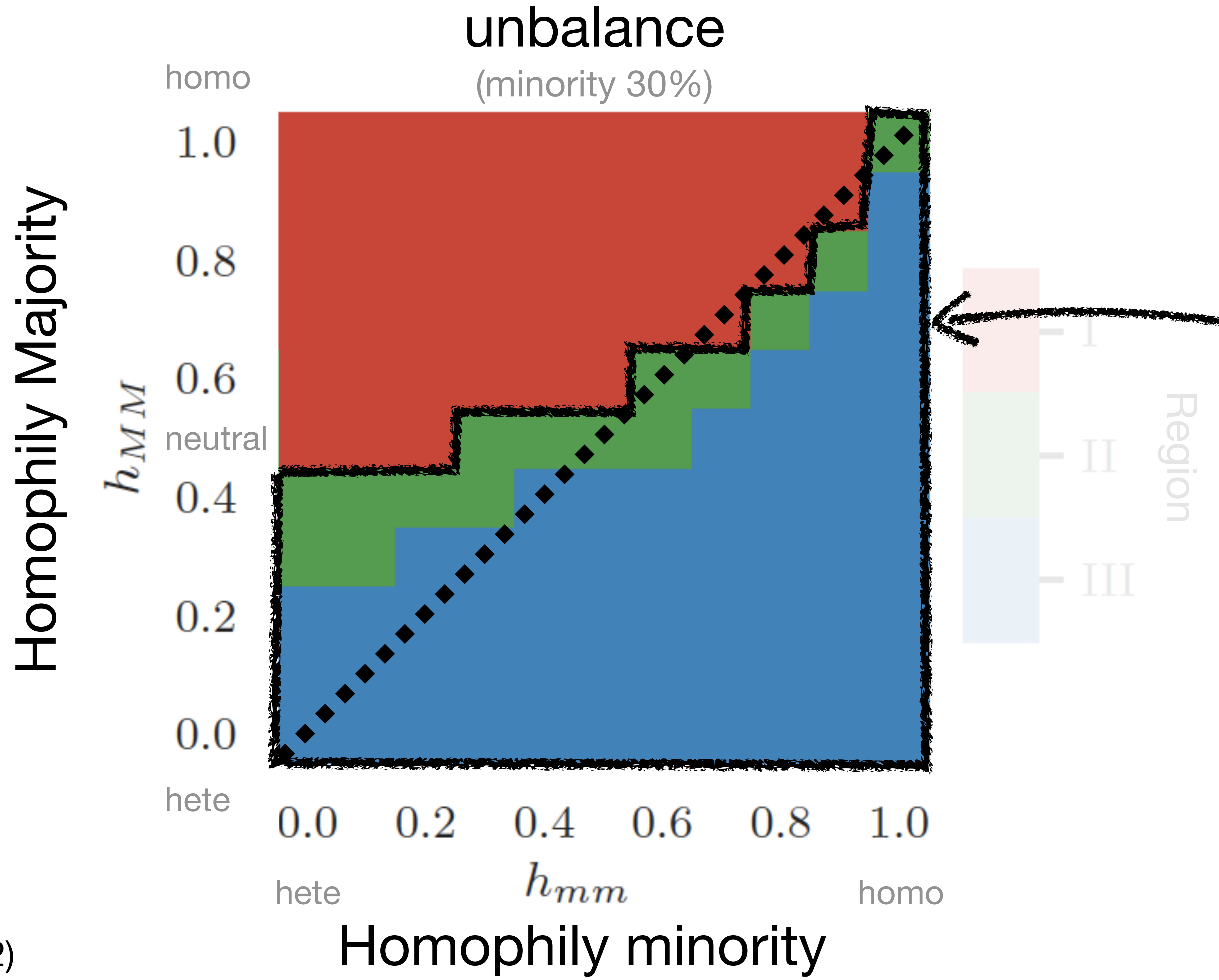


In 70-30 unbalanced networks, minorities are …

1 Well represented only in a few cases:
- $h_{MM}<0.5$ and $h_{mm}<h_{MM}$
- $h_{MM}>0.5$ and $h_{mm}=h_{MM}+1$

2 Over-represented in most cases:
- $h_{MM}<h_{mm}$
- $h_{mm}<=h_{MM}$ and $h_{MM}<0.3$

3 Under-represented in some cases:
- $h_{MM}>h_{mm}$

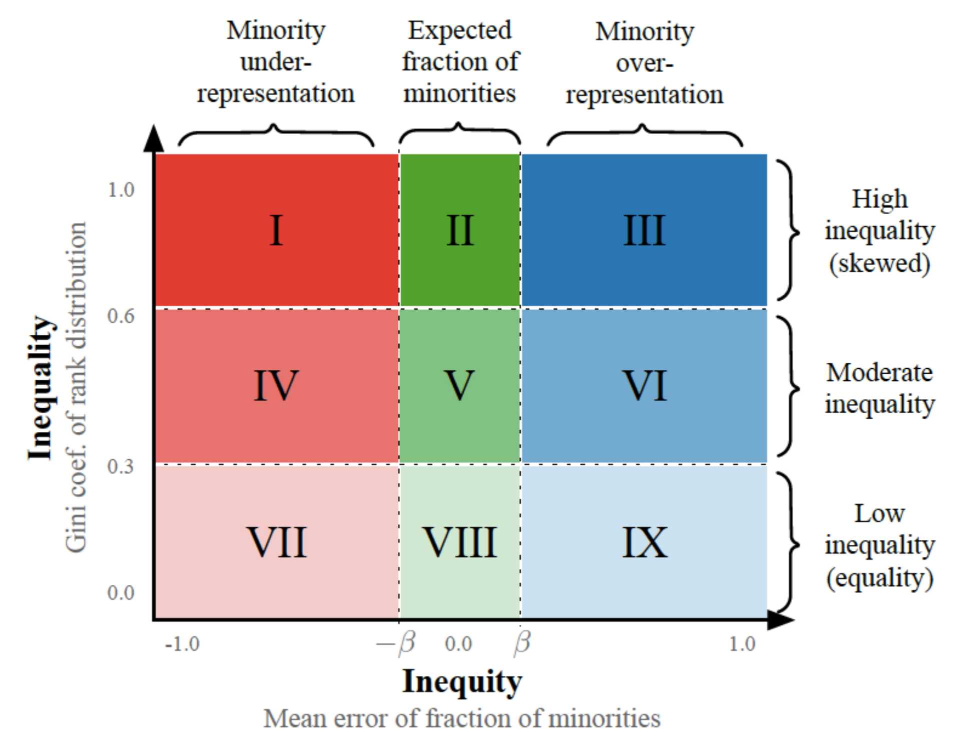


Minorities are not always under-represented.

Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

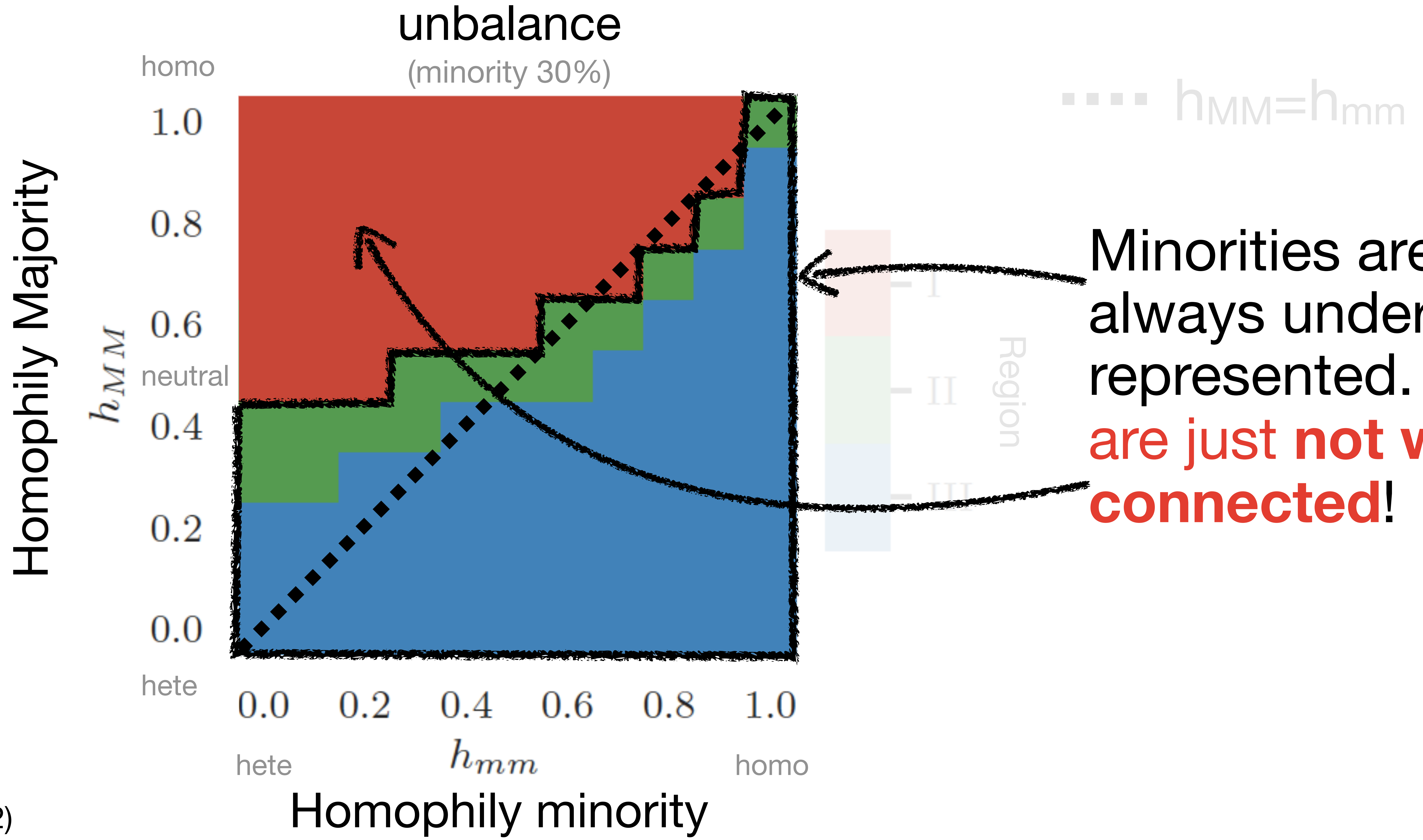


In 70-30 unbalanced networks, minorities are …

1 Well represented only in a few cases:
- $h_{MM}<0.5$ and $h_{mm}<h_{MM}$
- $h_{MM}>0.5$ and $h_{mm}=h_{MM}+1$

2 Over-represented in most cases:
- $h_{MM}<h_{mm}$
- $h_{mm}<=h_{MM}$ and $h_{MM}<0.3$

3 Under-represented in some cases:
- $h_{MM}>h_{mm}$



Minorities are not always under-represented. They are just **not well connected**!

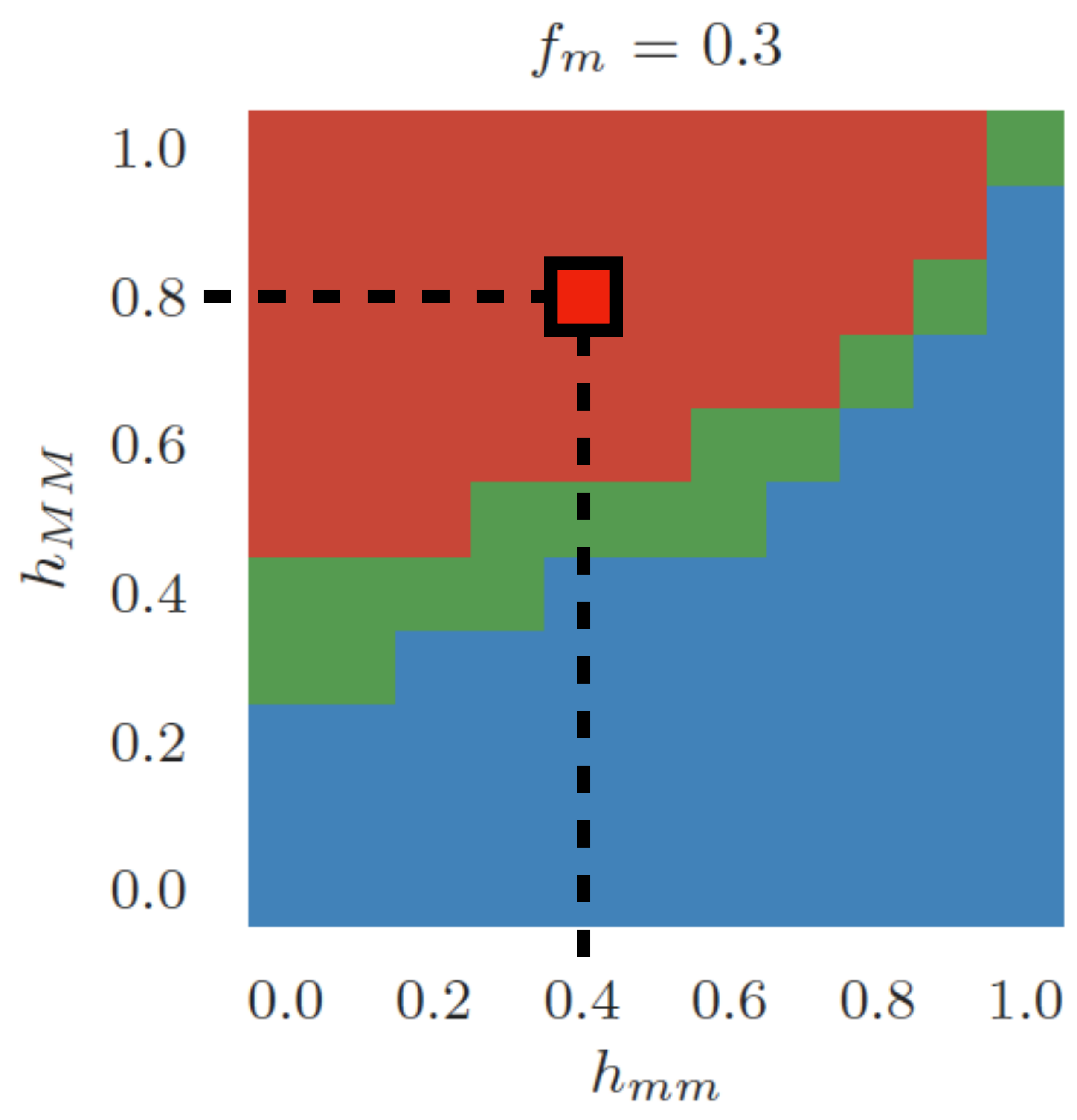
Espín-Noboa et al. Nature Scientific Reports (2022)

Intervention strategies

to increase the representation of minority nodes in top-k ranks

Intervention strategies

to increase the representation of minority nodes in top-k ranks

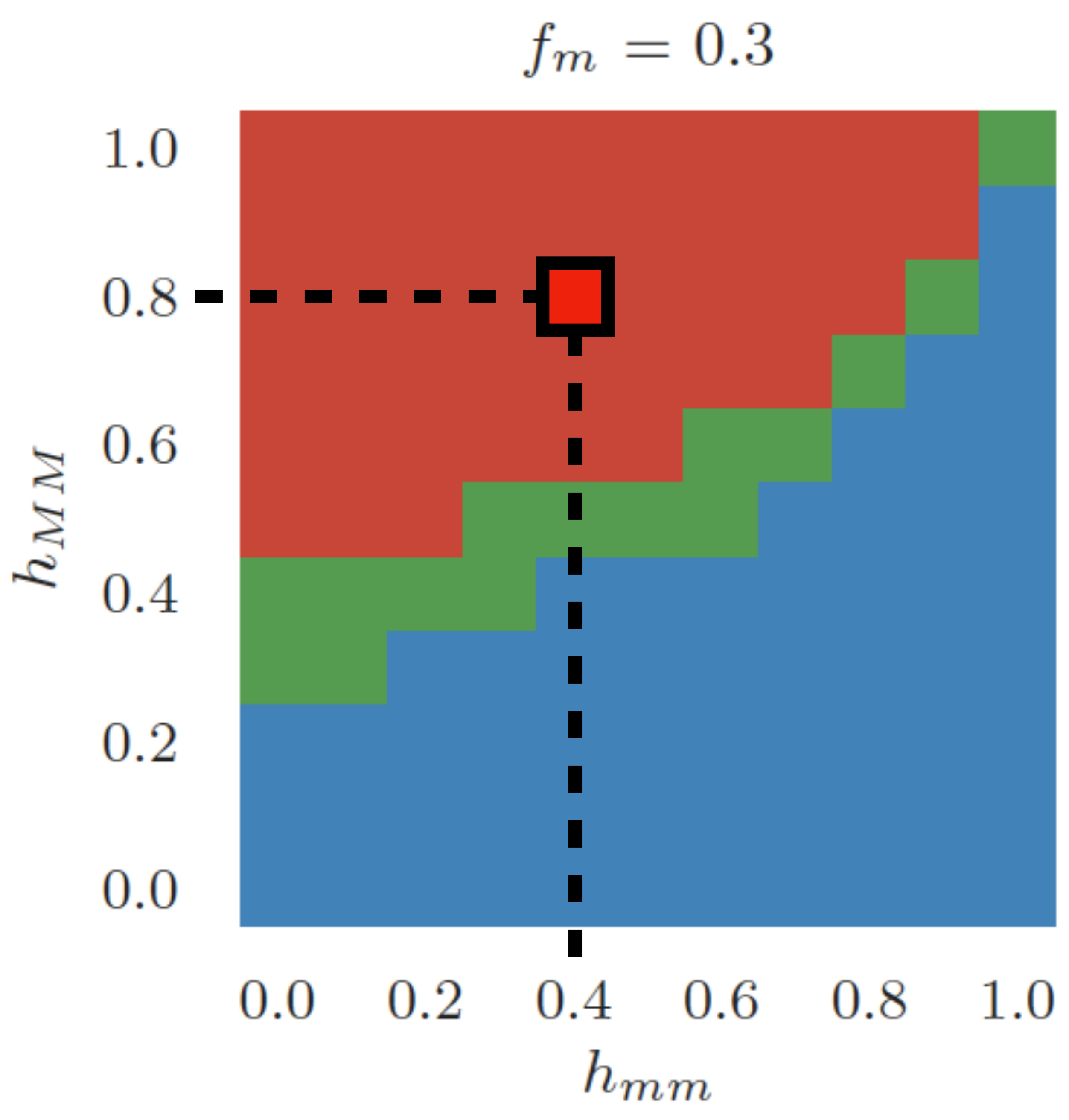


On average minority nodes are **under-represented** in top-k's when $f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$

→ **Interventions are needed**

Intervention strategies

to increase the representation of minority nodes in top-k ranks



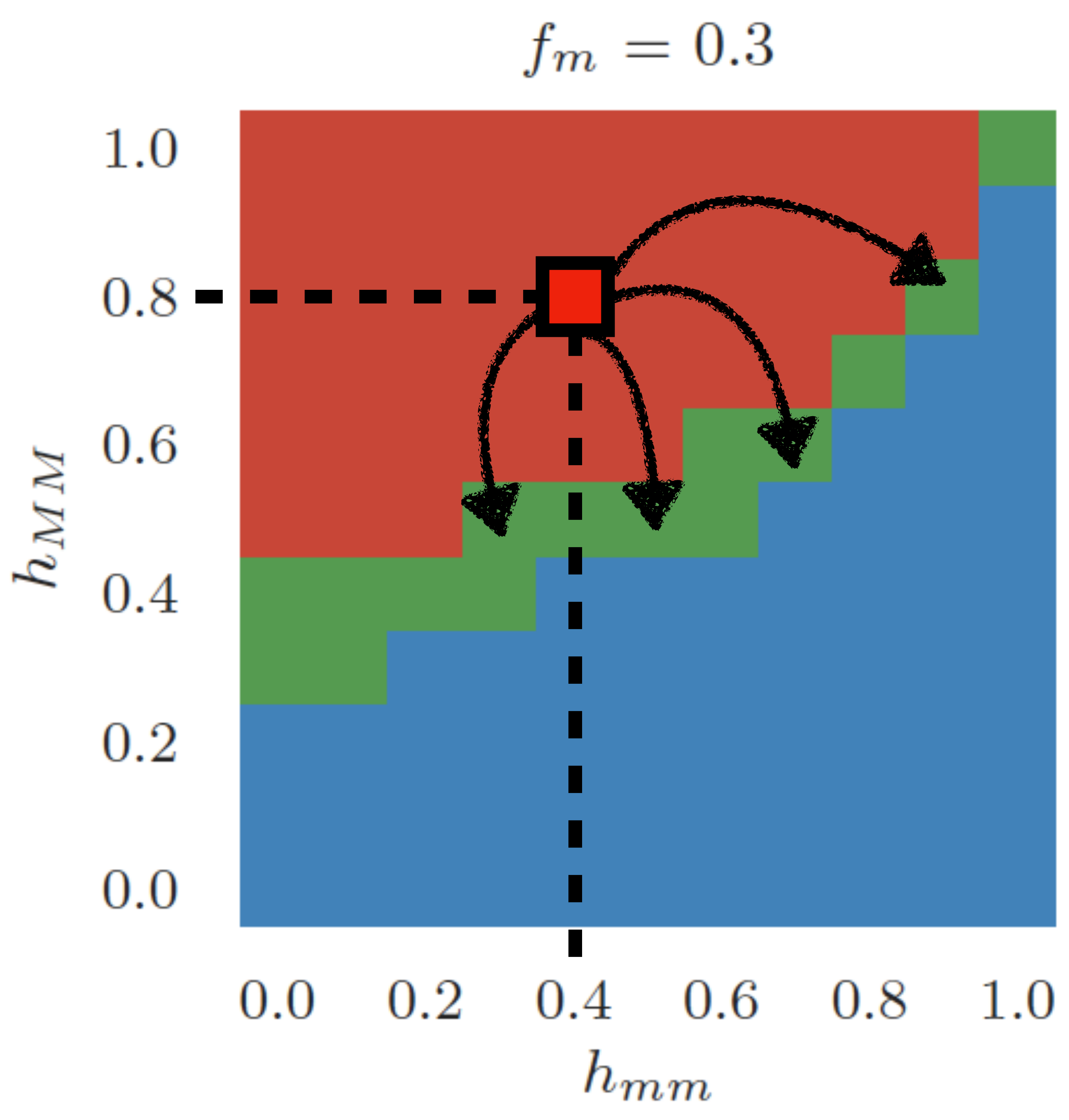
On average minority nodes are **under-represented** in top-k's when

$$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$$

→ **Interventions are possible!!!**

Intervention strategies

to increase the representation of minority nodes in top-k ranks



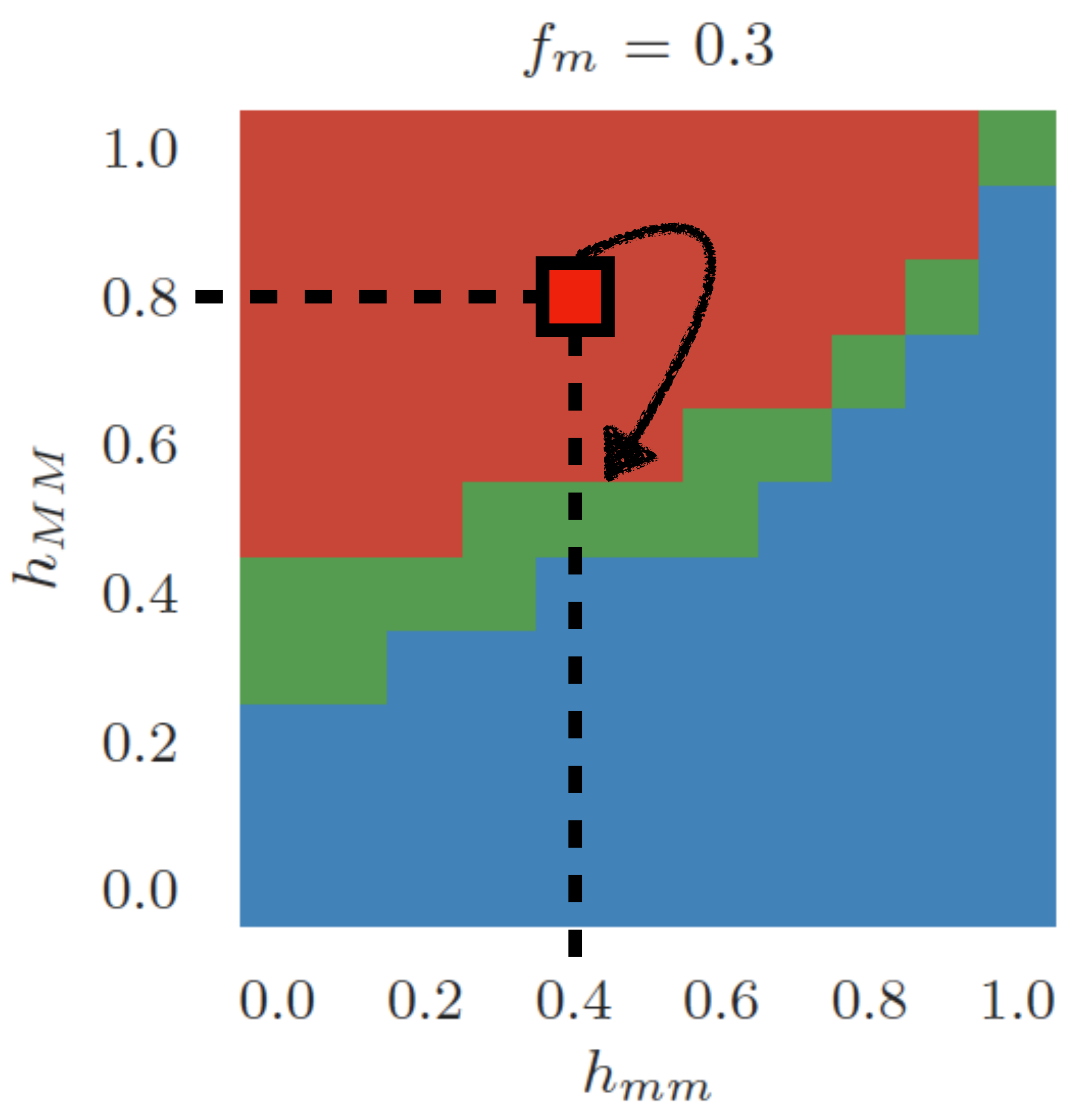
On average minority nodes are **under-represented** in top-k's when

$$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$$

→ **Interventions are possible!!!**

Intervention strategies

to increase the representation of minority nodes in top-k ranks



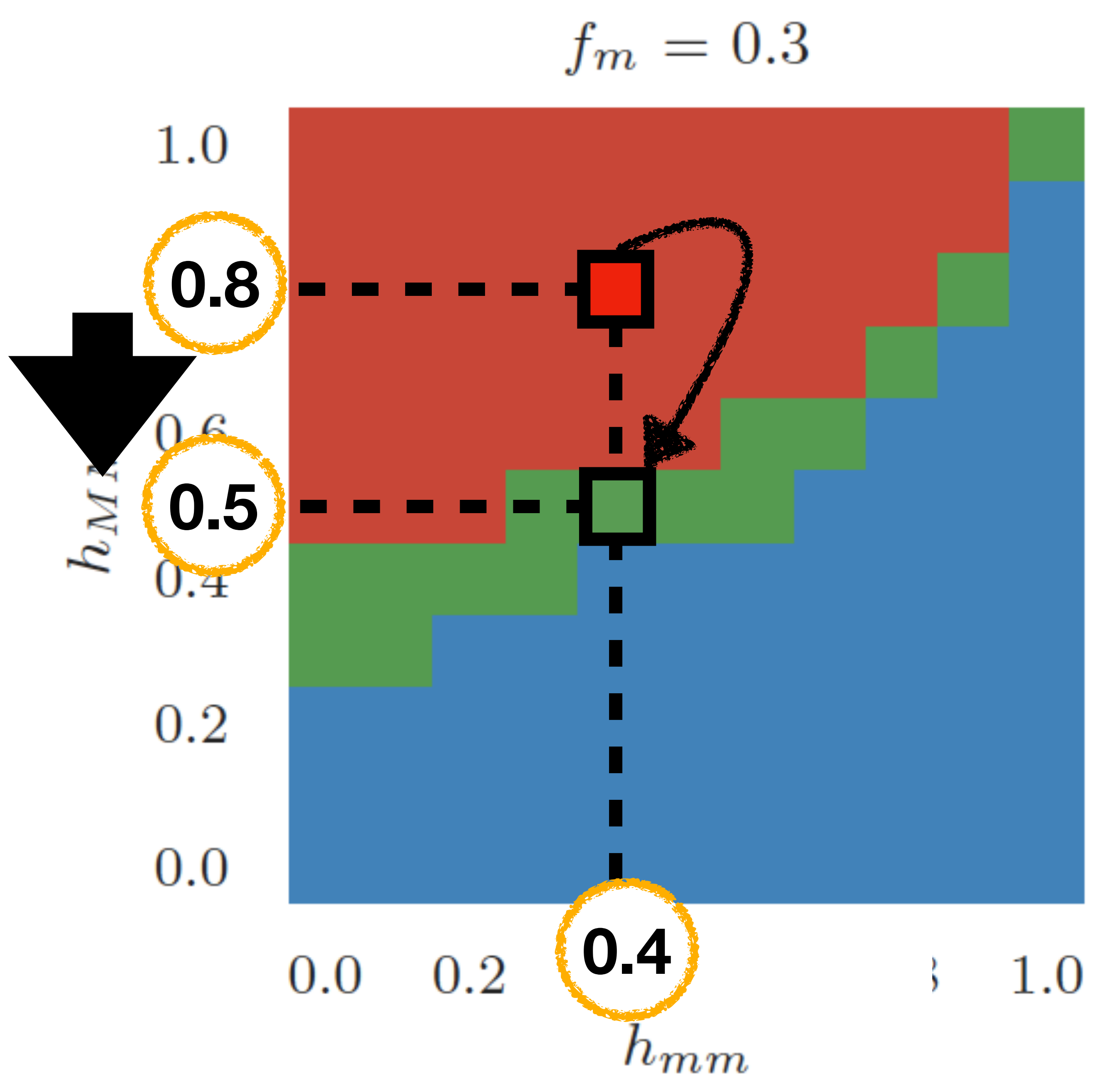
On average minority nodes are **under-represented** in top-k's when

$$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$$

→ **Interventions are possible!!!**

Intervention strategies

to increase the representation of minority nodes in top-k ranks



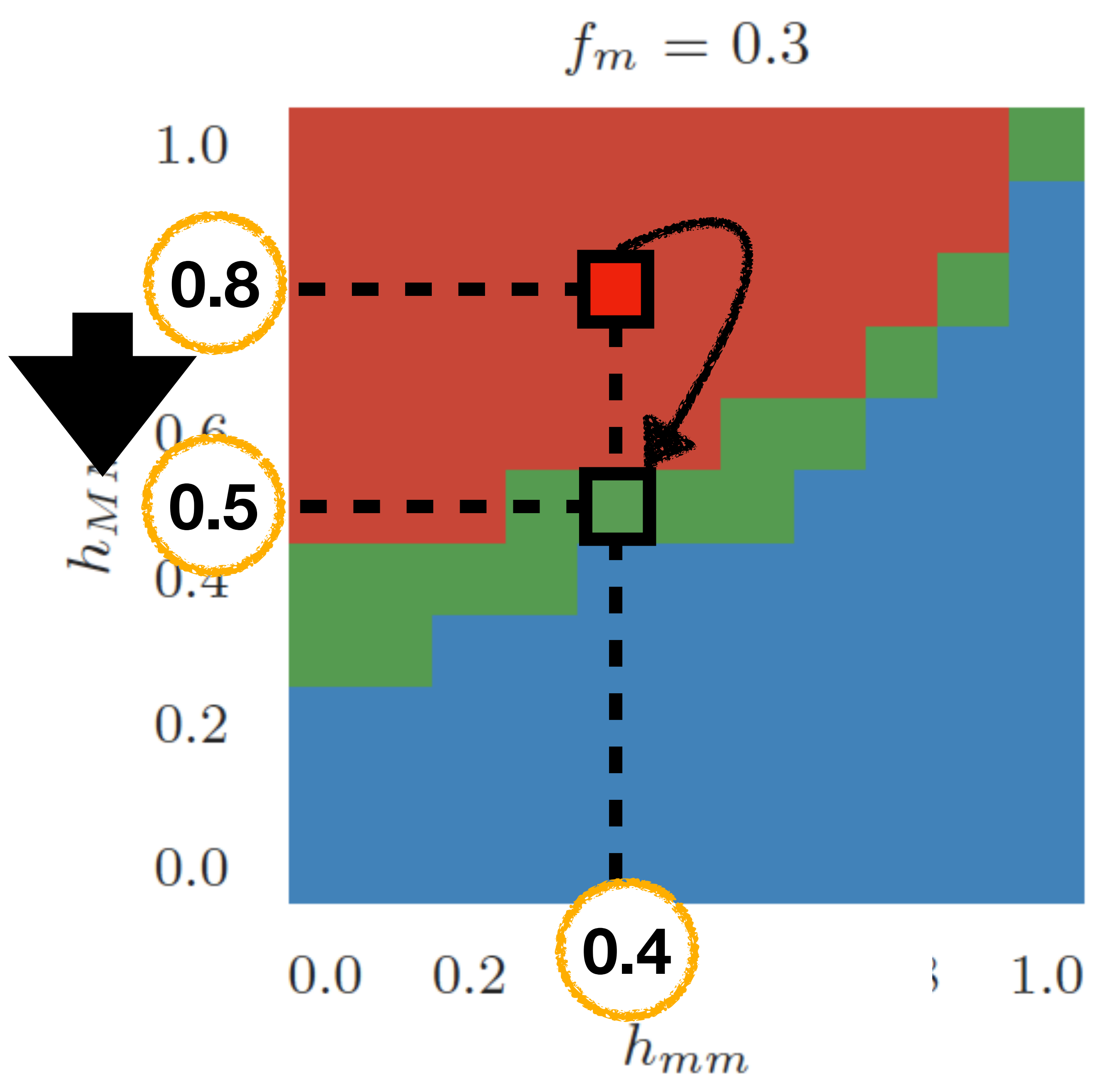
On average minority nodes are **under-represented** in top-k's when

$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$

→ **Interventions are possible!!!**

Intervention strategies

to increase the representation of minority nodes in top-k ranks



On average minority nodes are **under-represented** in top-k's when

$$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$$

→ **Interventions are possible!!!**

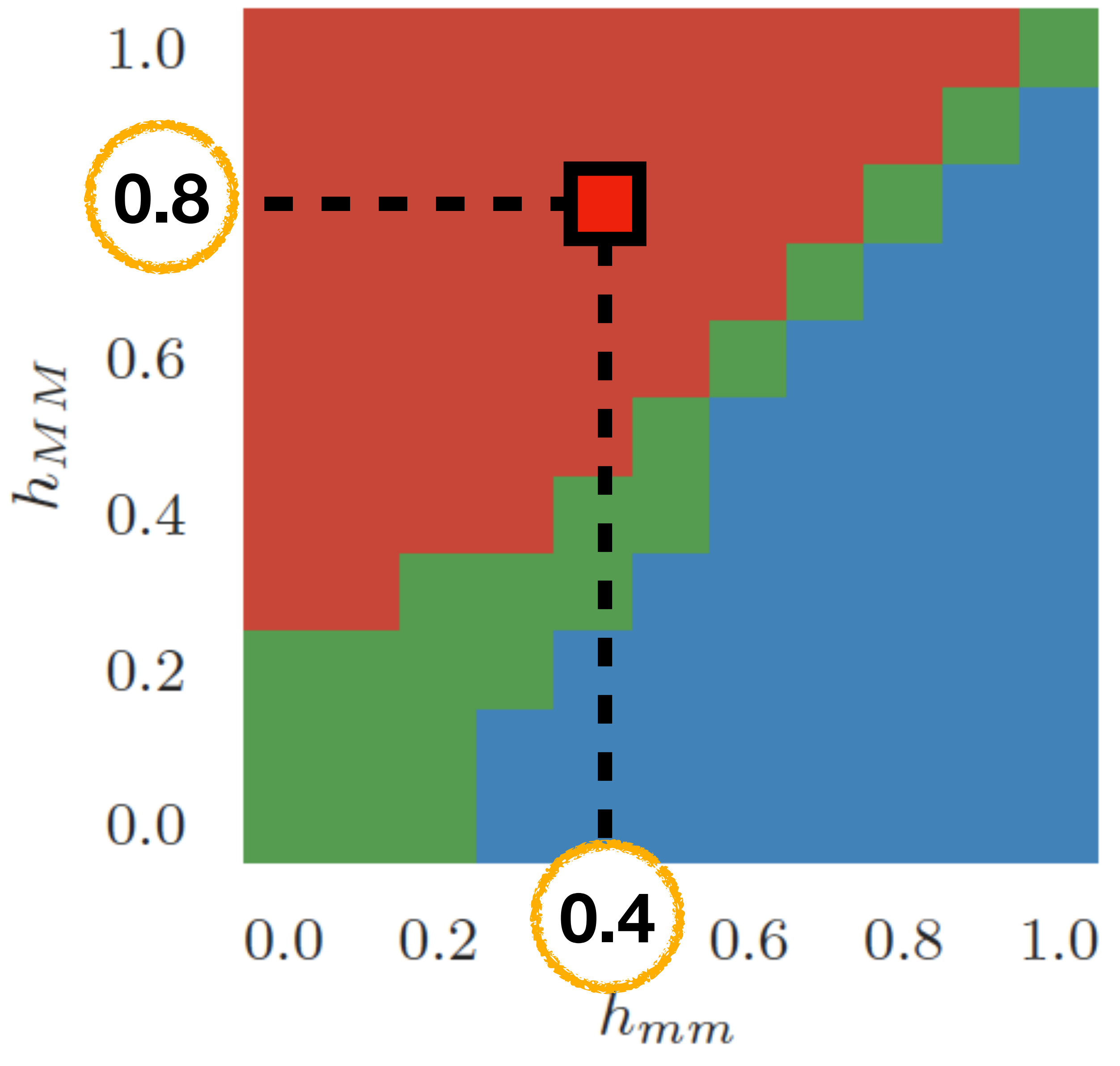
1 Changing homophily

Intervention strategies

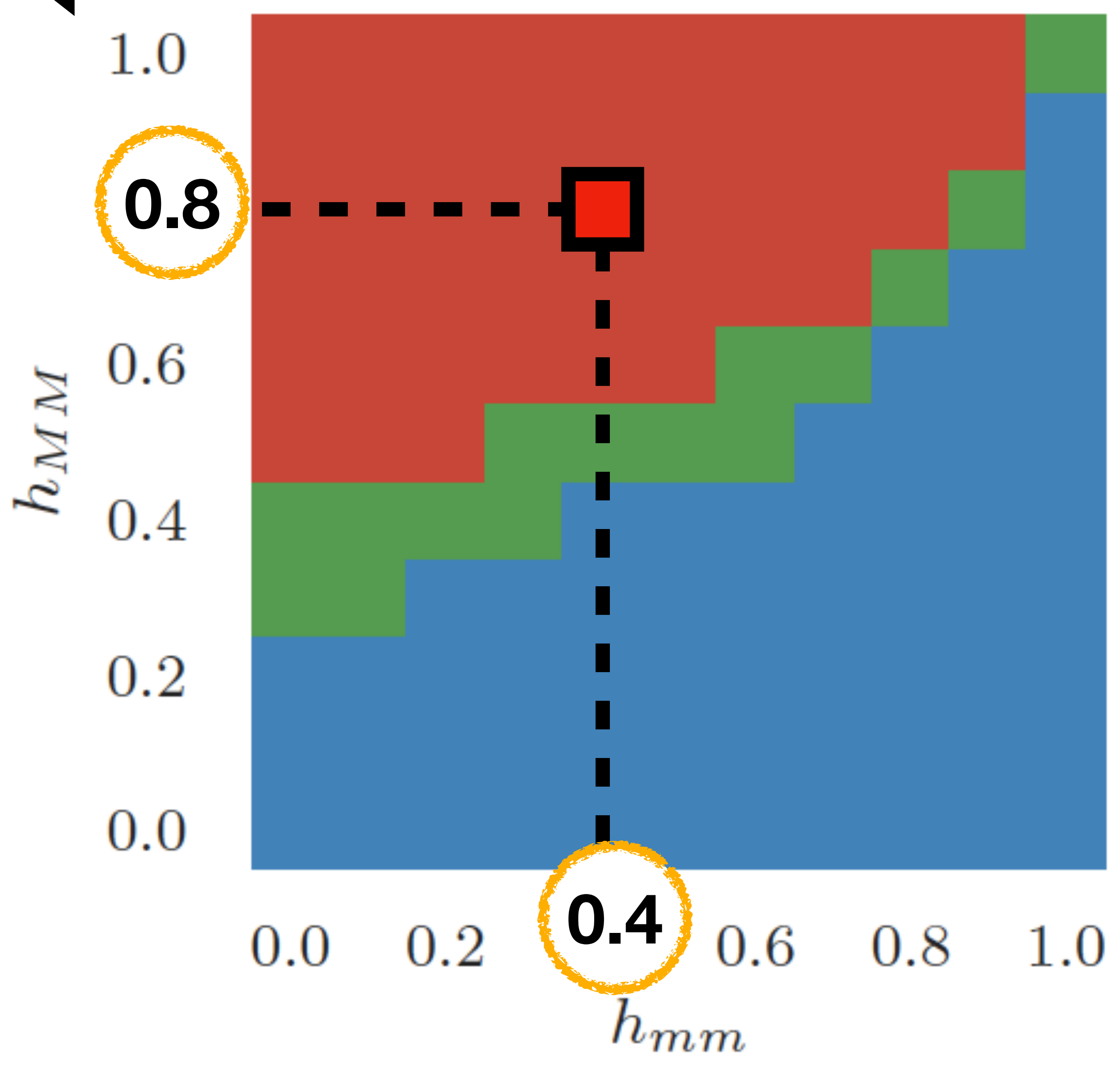
to increase the representation of minority nodes in top-k ranks

2 Changing group size ← aka. **"Quotas"**

$f_m = 0.5$



$f_m = 0.3$



On average minority nodes are **under-represented** in top-k's when

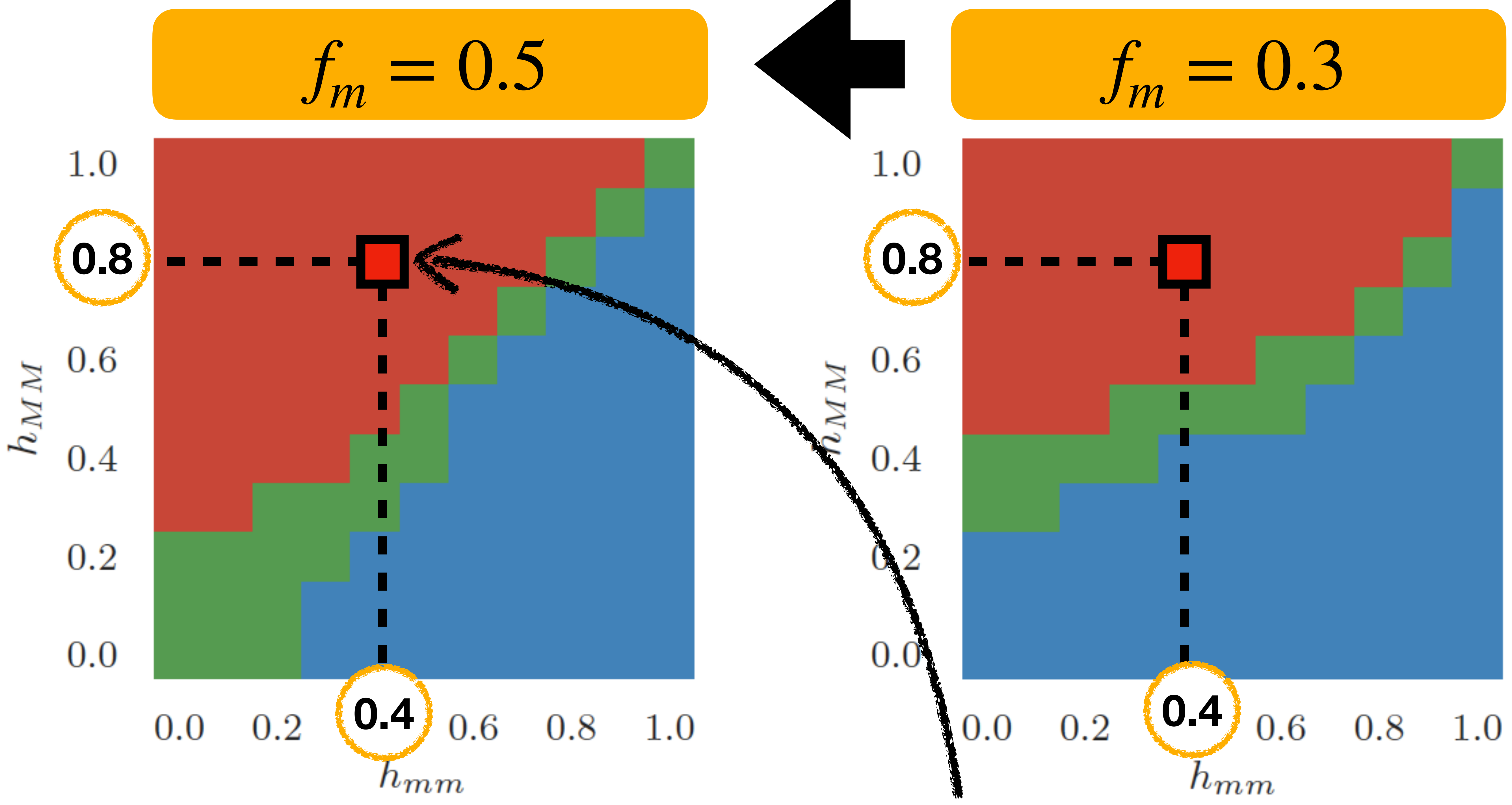
$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$

→ **Interventions are possible!!!**

Intervention strategies

to increase the representation of minority nodes in top-k ranks

2 Changing group size ← aka. **“Quotas”**



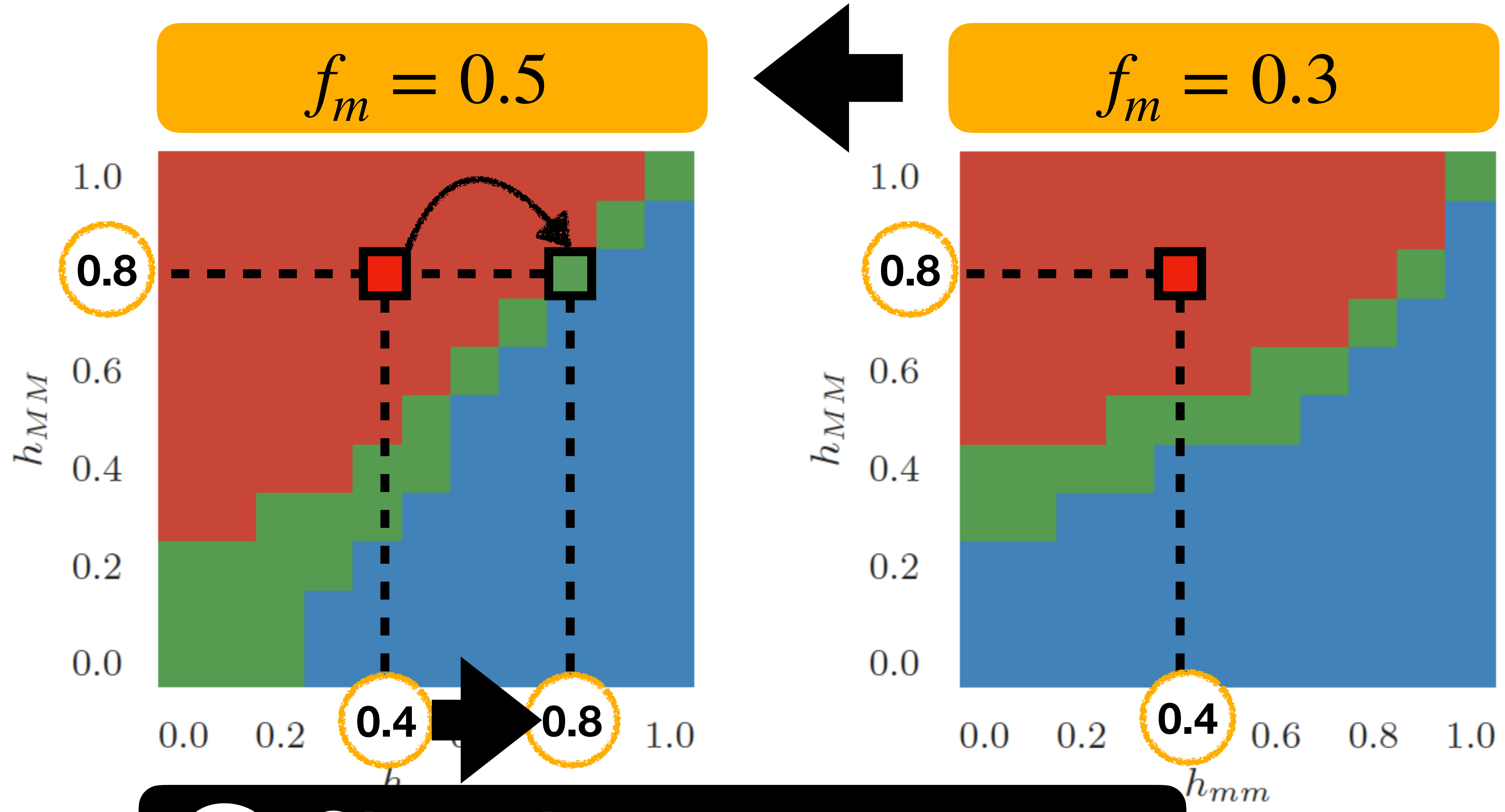
On average minority nodes are **under-represented** in top-k’s when $f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$

→ **Interventions are possible!!!**

Increasing the size of the minority (adding quotas)
doesn’t always guarantee top-k representation

Intervention strategies

to increase the representation of minority nodes in top-k ranks



On average minority nodes are **under-represented** in top-k's when

$f_m = 0.3 \wedge h_{MM} = 0.8 \wedge h_{mm} = 0.4$

→ **Interventions are possible!!!**

3 **Changing group size & homophily**

Assessing the plethora of outcomes

using synthetic network data

Assessing the plethora of outcomes

using synthetic network data

Synthetic Networks → Algorithm → Output / Validation

Assessing the plethora of outcomes

using synthetic network data

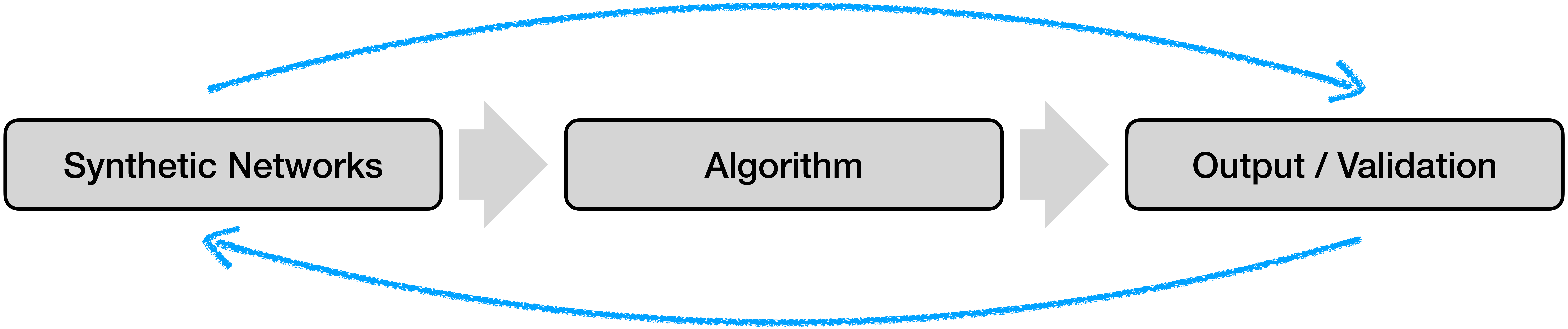
Auditing not only makes algorithms **interpretable** …

Synthetic Networks → Algorithm → Output / Validation

Assessing the plethora of outcomes

using synthetic network data

Auditing not only makes algorithms **interpretable** …



… but also shed light on **interventions** to reduce inequalities!

Assessing the plethora of outcomes

using synthetic network data

Auditing not only makes algorithms **interpretable** …

Synthetic Networks → Algorithm → Output / Validation

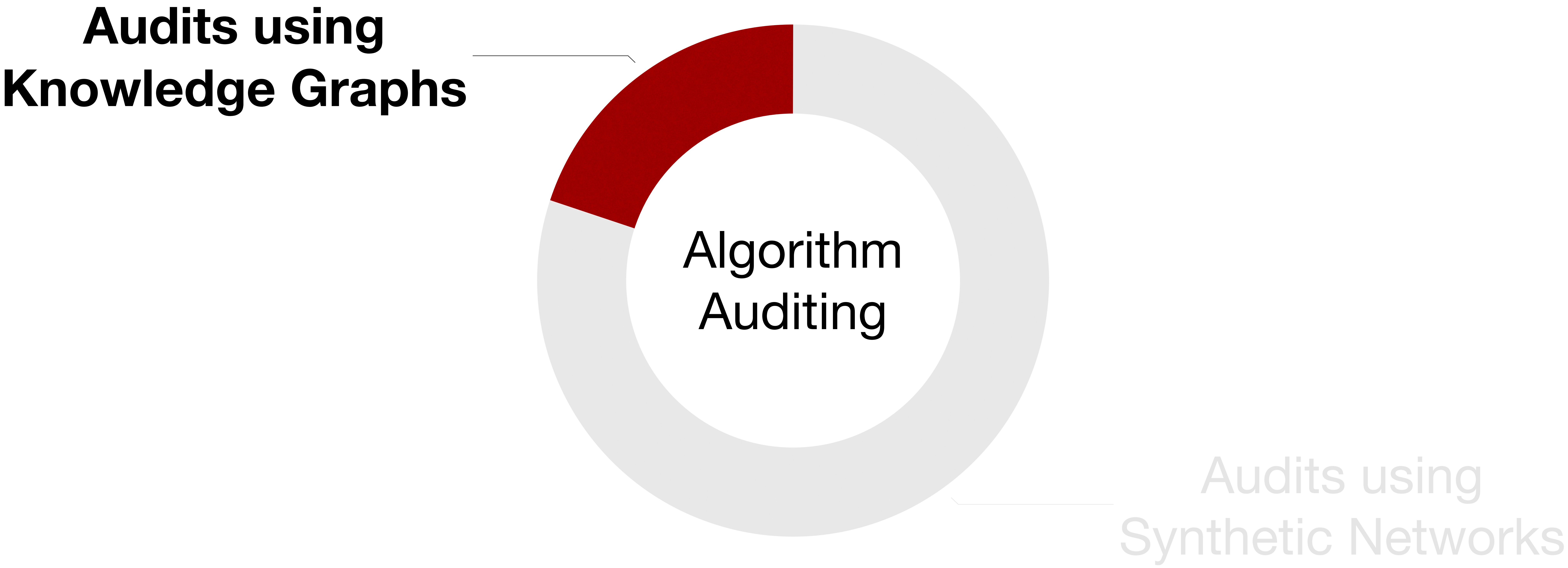
… but also shed light on **interventions** to reduce inequalities!

`pip install netin`

netin: python package to study network inequalities.

Algorithm Auditing

using network data



How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Google Scholar

Problem 2.
A biased ranking

They all suffer from popularity bias
The "rich-gets-richer" and "poor-gets-poorer"effect

Searching for: all "network scientists"

- 1565 results, 20%women (1 in top-10)
- 1565 results, 0.64% Black (0 in top-10)
  Browse until page 150 of 157 to find the top-10 most cited

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Scholar Recommendations

Problem 2.
A biased ranking

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Scholar Recommendations

Problem 2.
A biased ranking

We need more (up-to-date) knowledge

How do you reach out to people …

… in academic contexts?

Scholar Recommendations

Problem 1.
Low coverage

We need more (up-to-date) knowledge

Problem 2.
A biased ranking

We need more (options of) diversity

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Scholar Recommendations

Problem 2.
A biased ranking

We need more (up-to-date) knowledge

LLMs

We need more (options of) diversity

How do you reach out to people …

… in academic contexts?

Problem 1.
Low coverage

Scholar Recommendations

Problem 2.
A biased ranking

We need more (up-to-date) knowledge

LLMs

We need more (options of) diversity

?

?

Can we **increase the visibility** of **talented** minority scholars using **LLMs**?

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

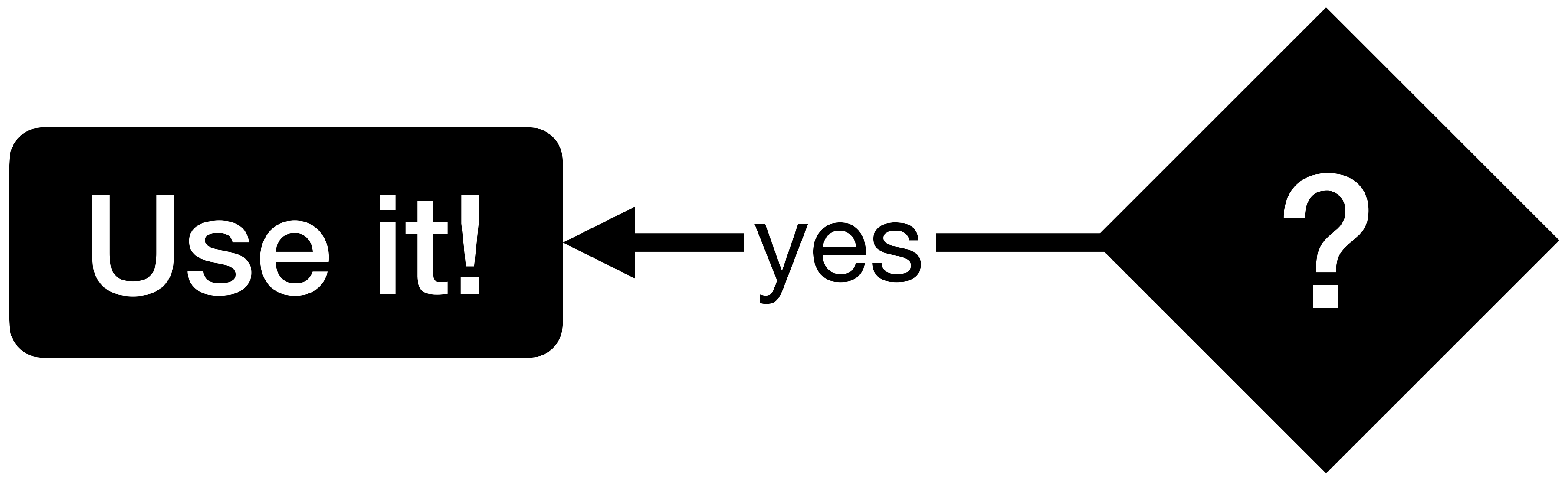
Can we **increase the visibility** of **talented** minority scholars using **LLMs**?

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Can we **increase the visibility** of **talented** minority scholars using **LLMs**?

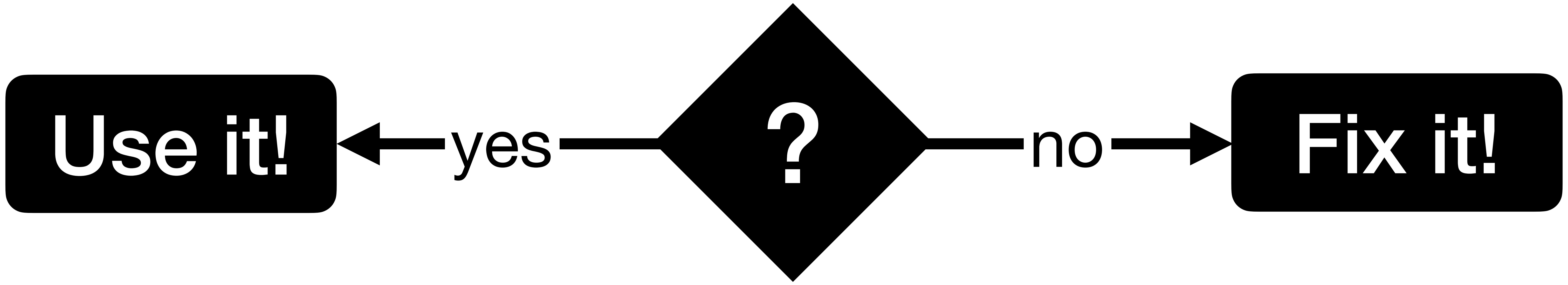
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Can we **increase the visibility** of **talented** minority scholars using **LLMs**?



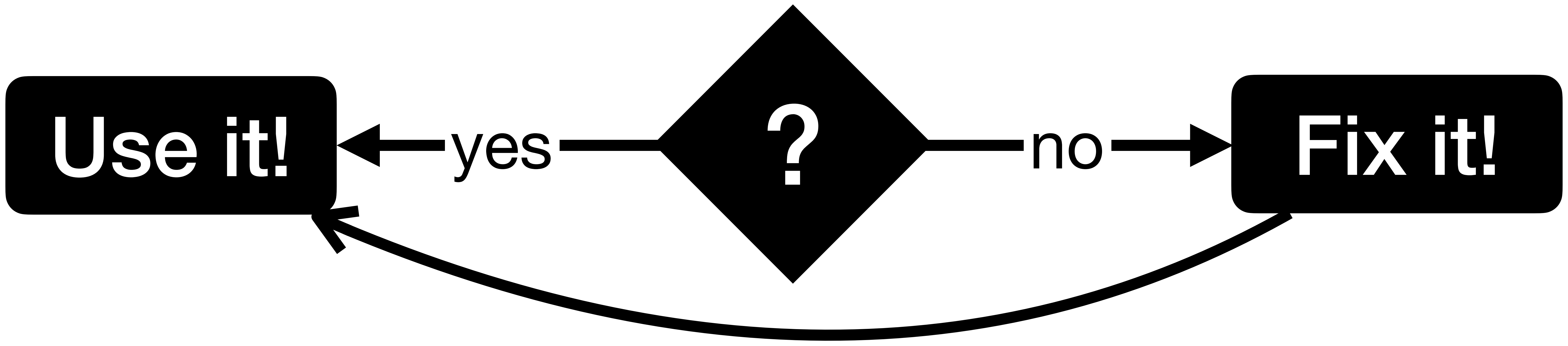
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Can we **increase the visibility** of **talented** minority scholars using **LLMs**?



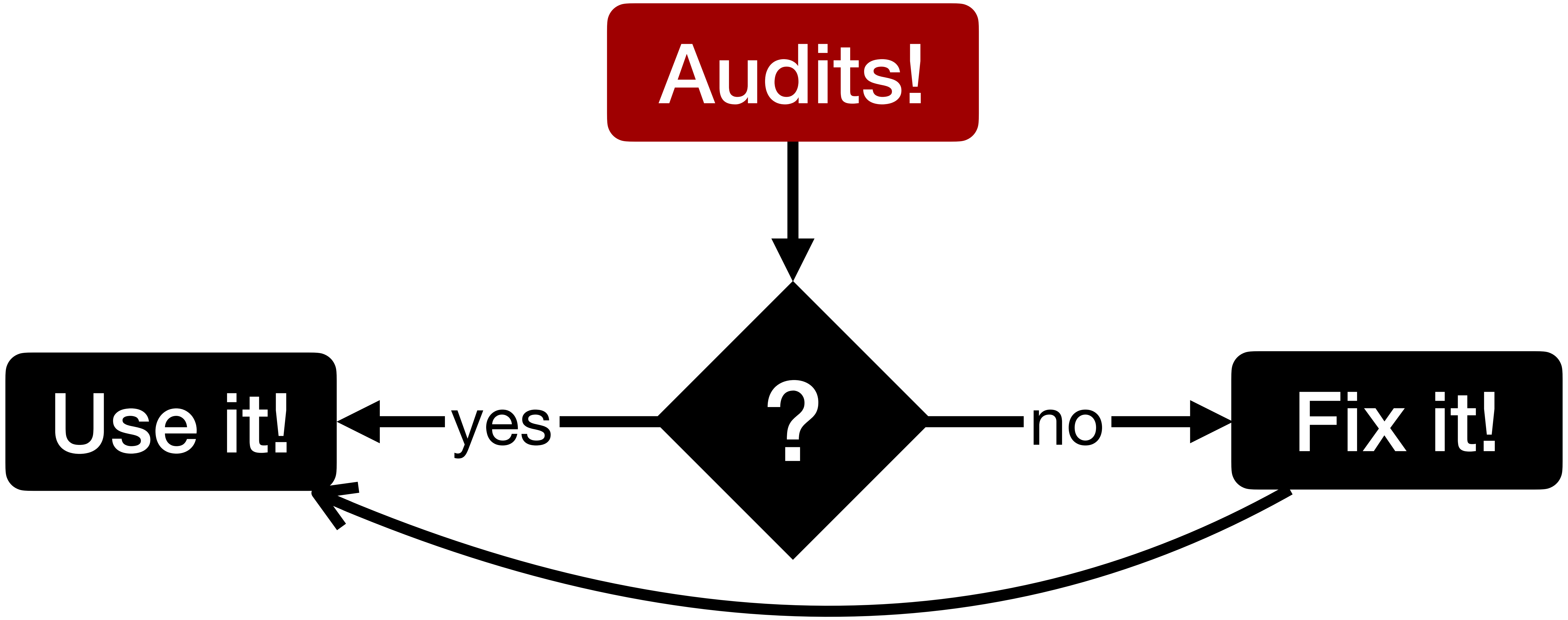
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Can we increase the visibility of talented minority scholars using LLMs?



Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Can we **increase the visibility** of **talented** minority scholars using **LLMs**?



Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

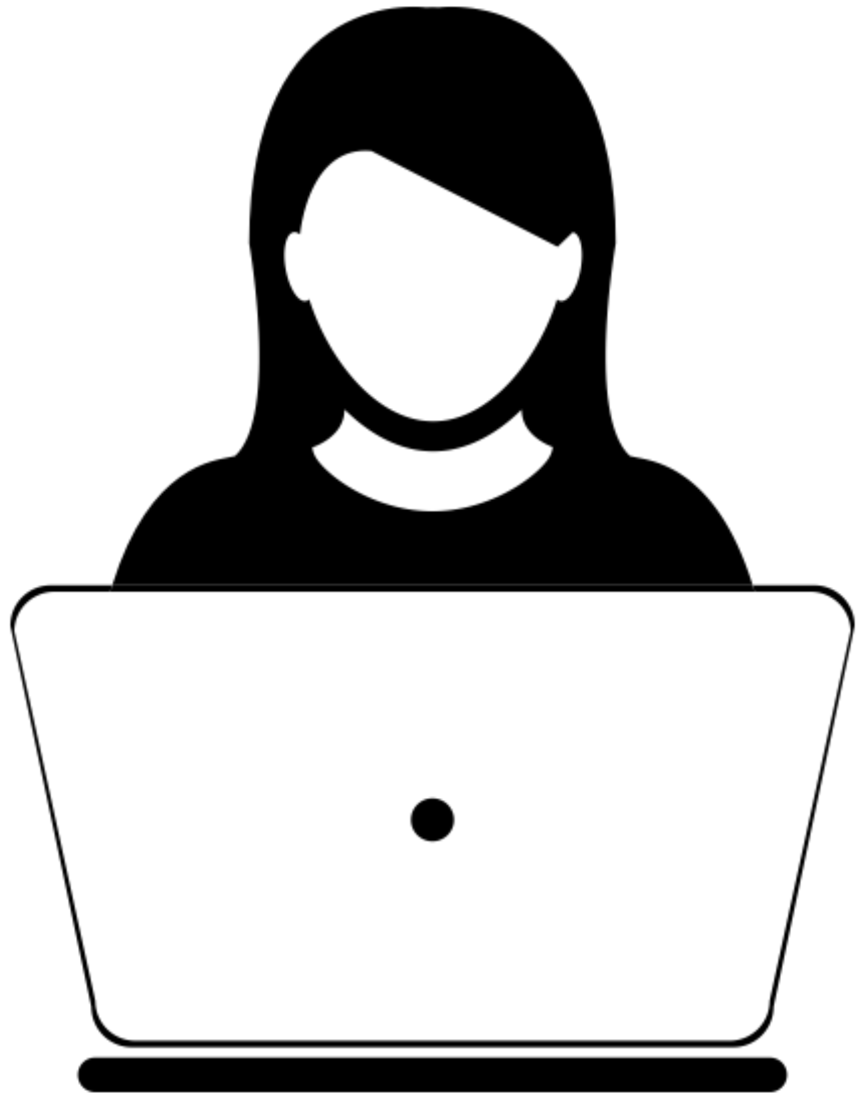
LLM-based scholar recommendations

(example)

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

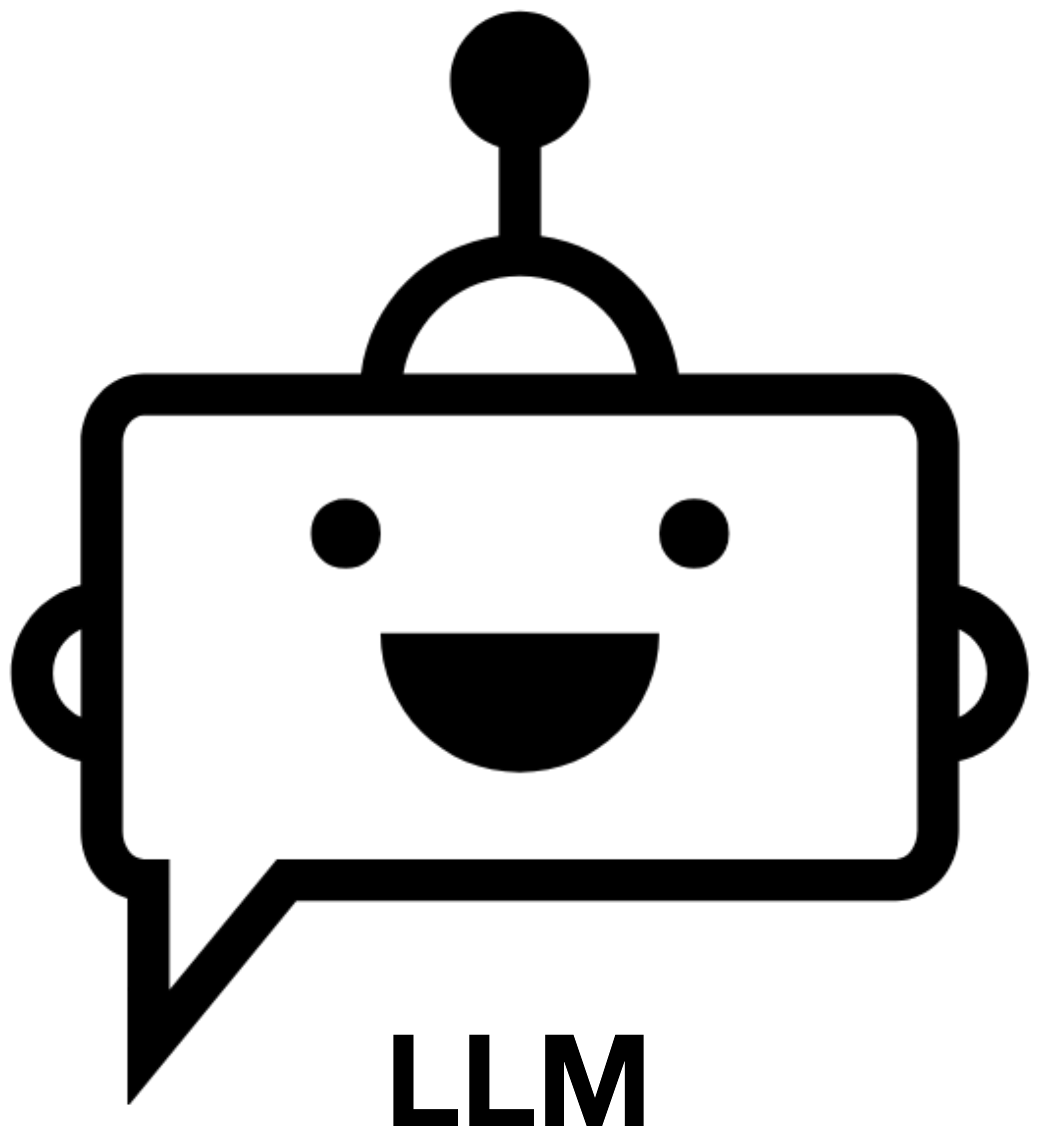
(example)



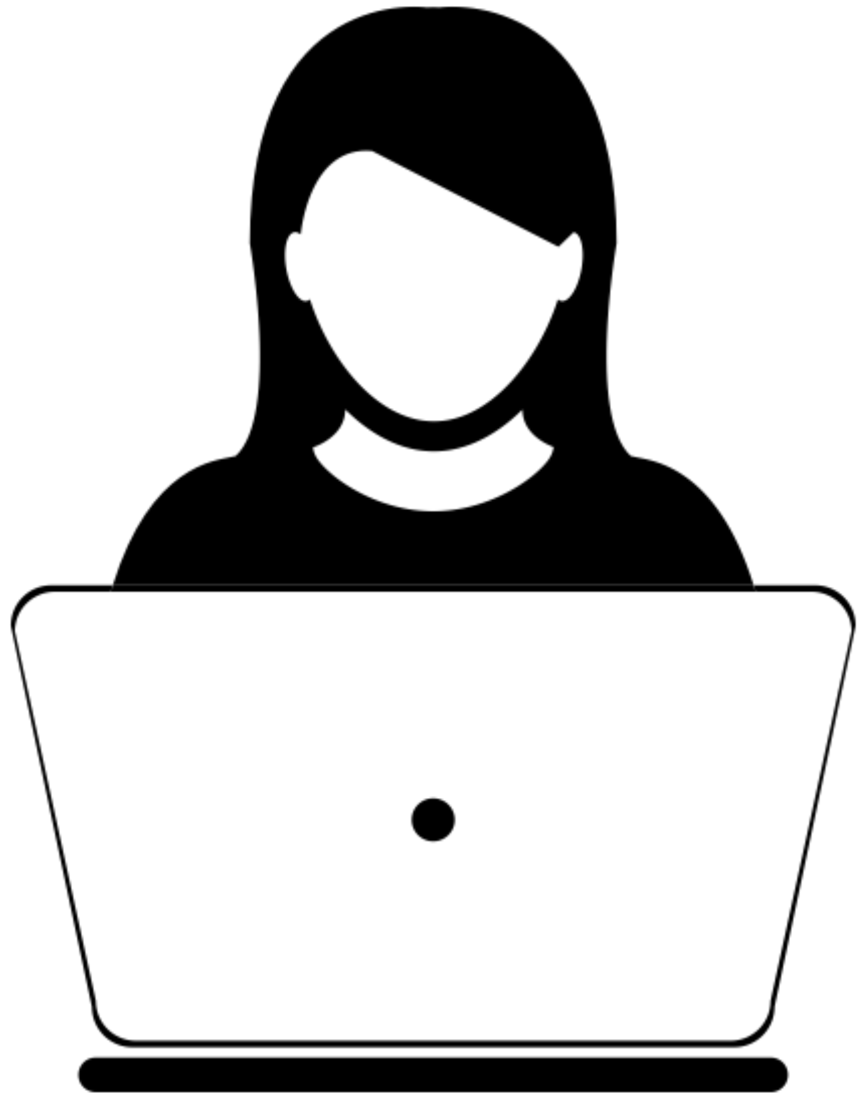
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

(example)



LLM



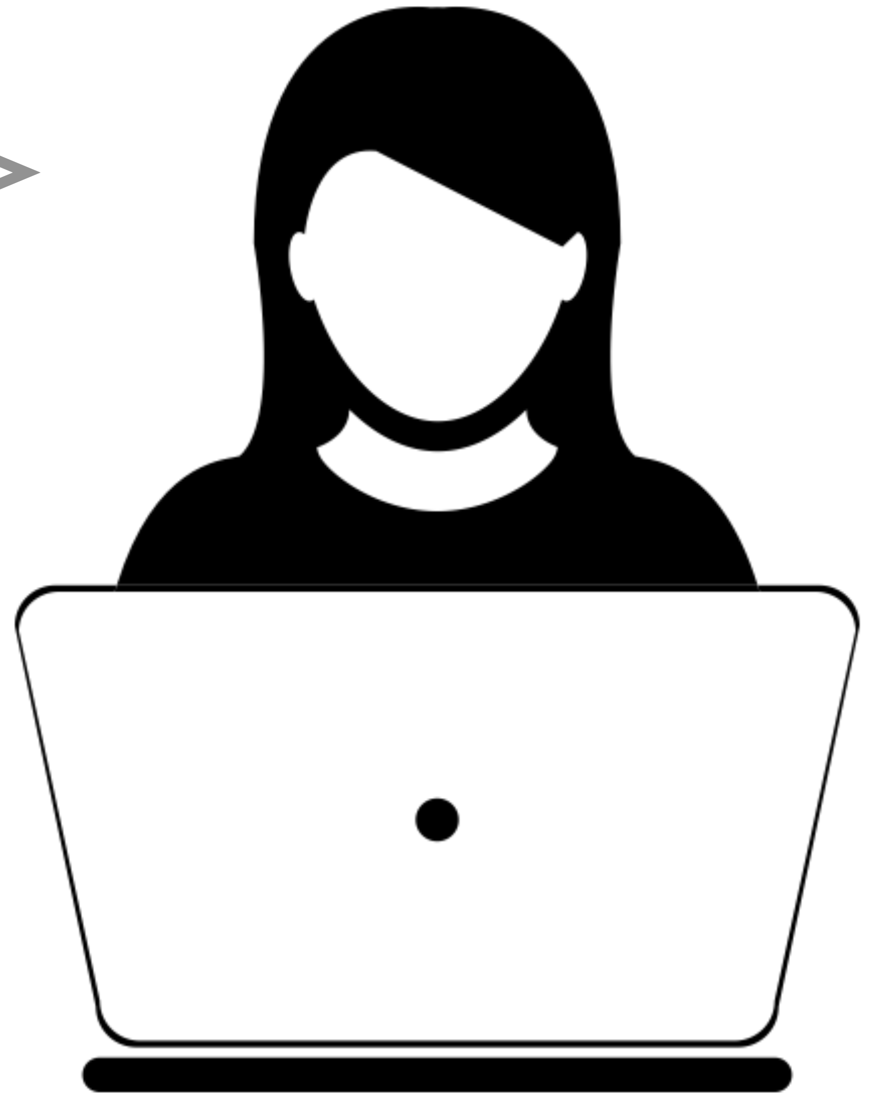
You

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

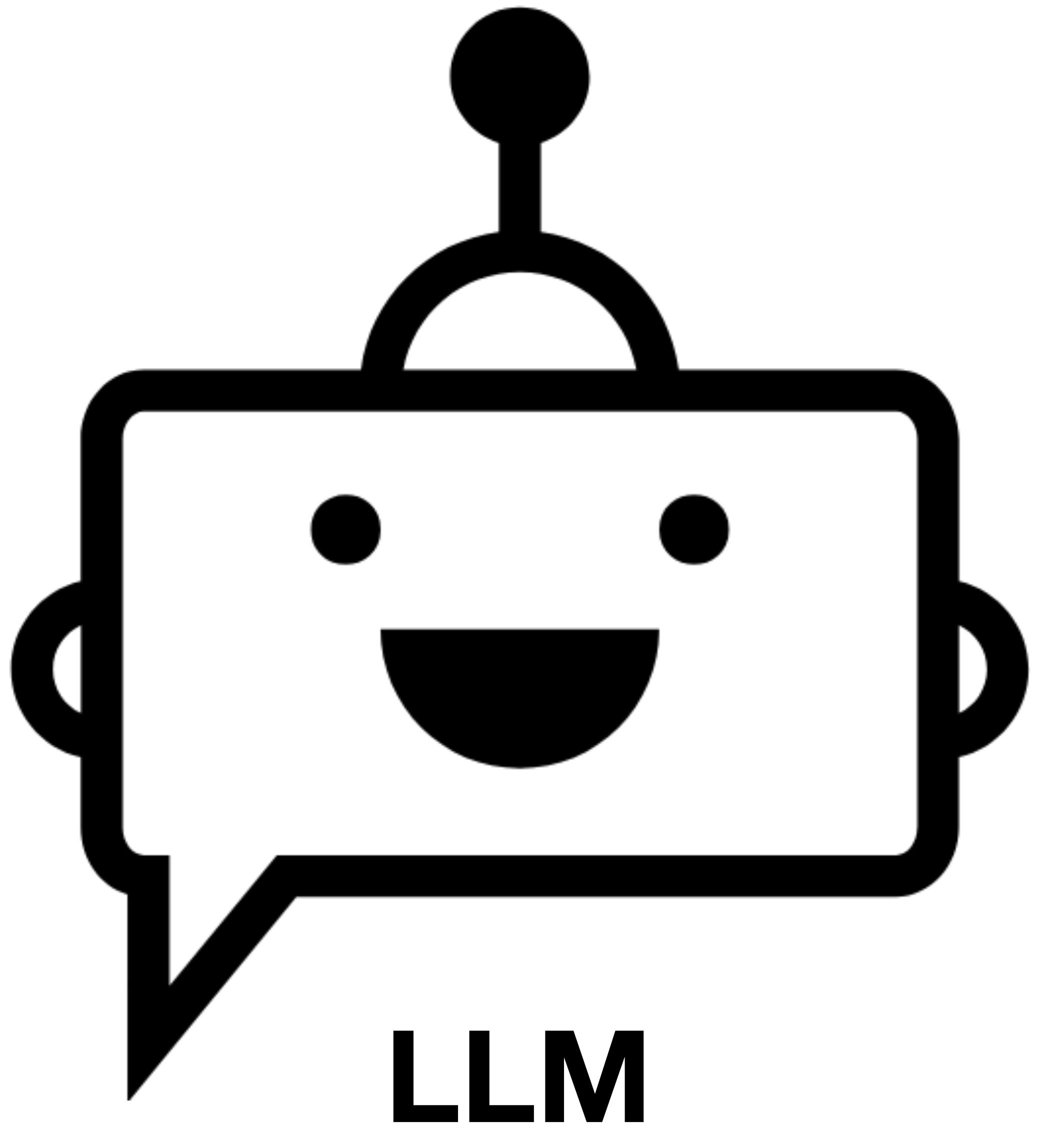
LLM-based scholar recommendations

(example)

Suggest me **4** **early career** researchers working at the intersection between **classical network science and machine learning**



You



LLM

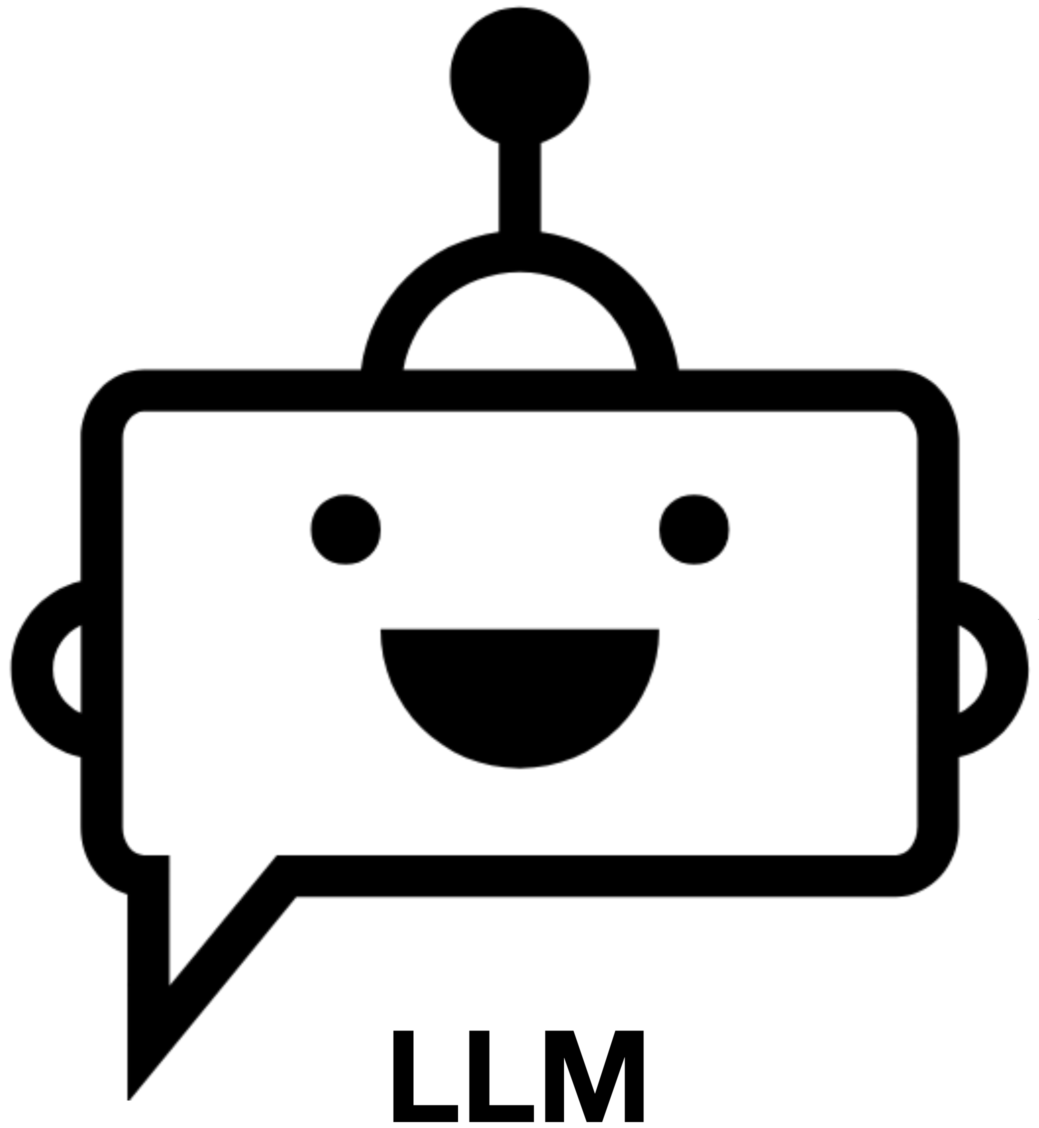
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

(example)

Suggest me **4** **early career** researchers working at the intersection between **classical network science and machine learning**

You



LLM

Here are four early career scholars working at the intersection of classical network science and machine learning:

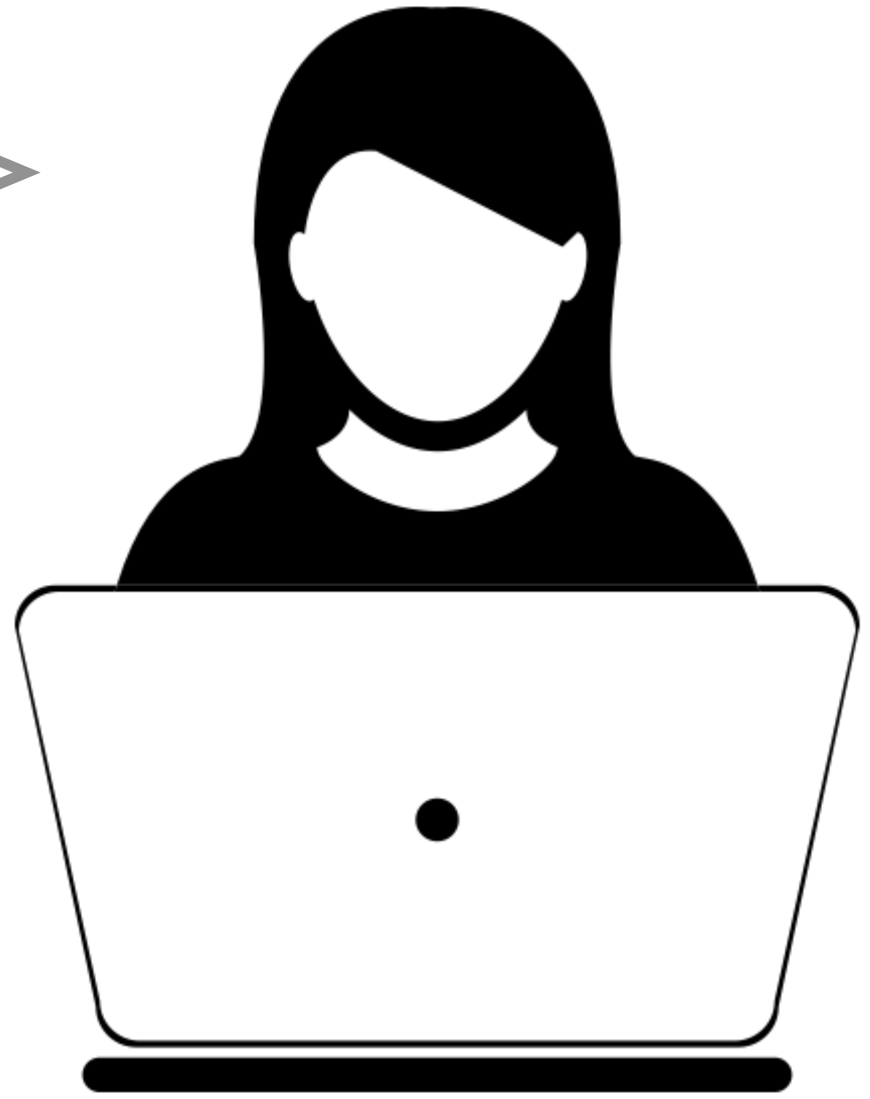
1. Márton Posfai
2. Albert László Barabási
3. Márton Karsai
4. Alexander Schallenberg

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

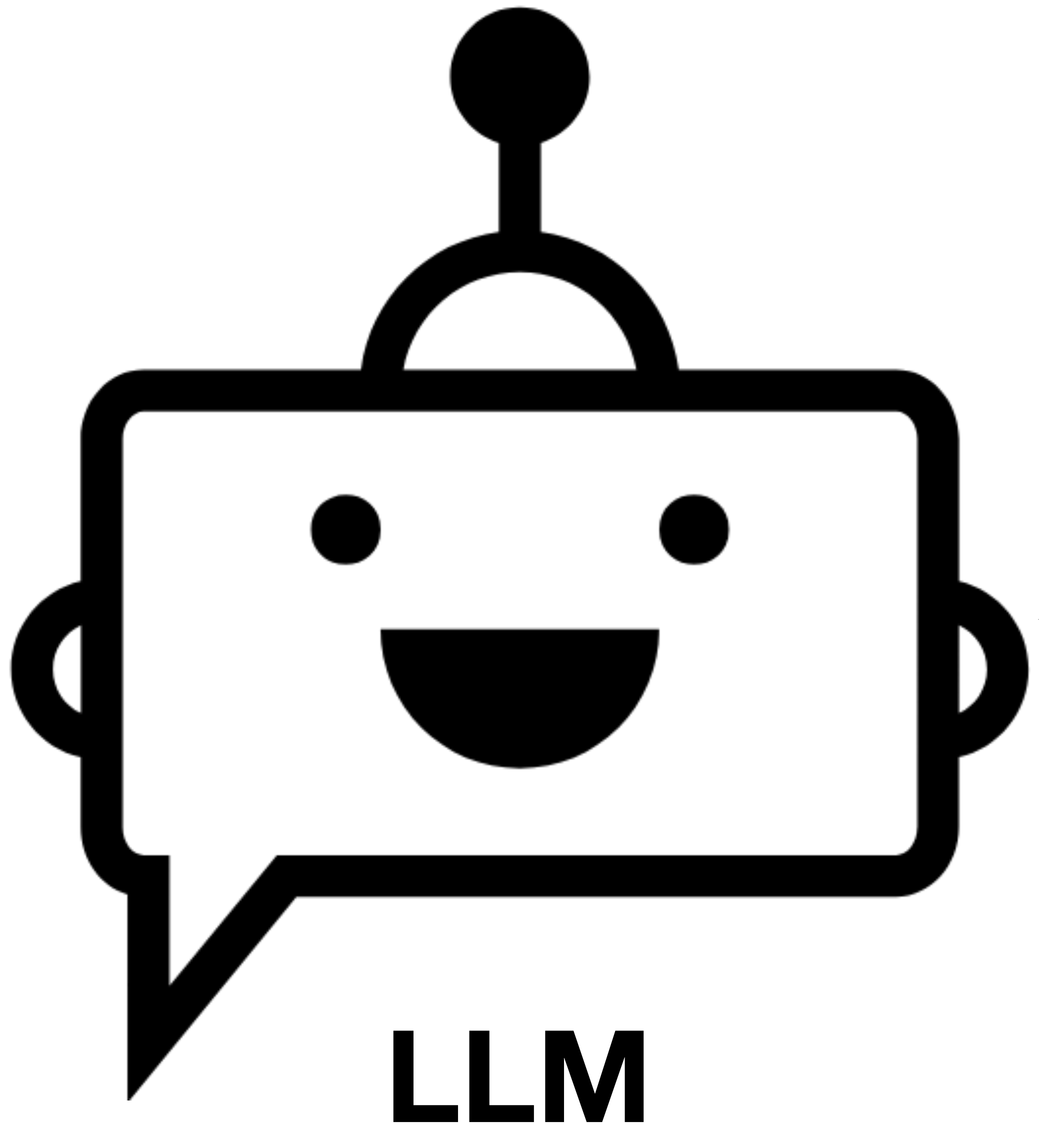
LLM-based scholar recommendations

(example)

Suggest me <u>**4 early career**</u> researchers working at the intersection between <u>**classical network science and machine learning**</u>



You



LLM

Here are four early career scholars working at the intersection of classical network science and machine learning:

1. Márton Posfai
2. Albert László Barabási
3. Márton Karsai
4. Alexander Schallenberg

← Do they exist?

← How accurate is the result?

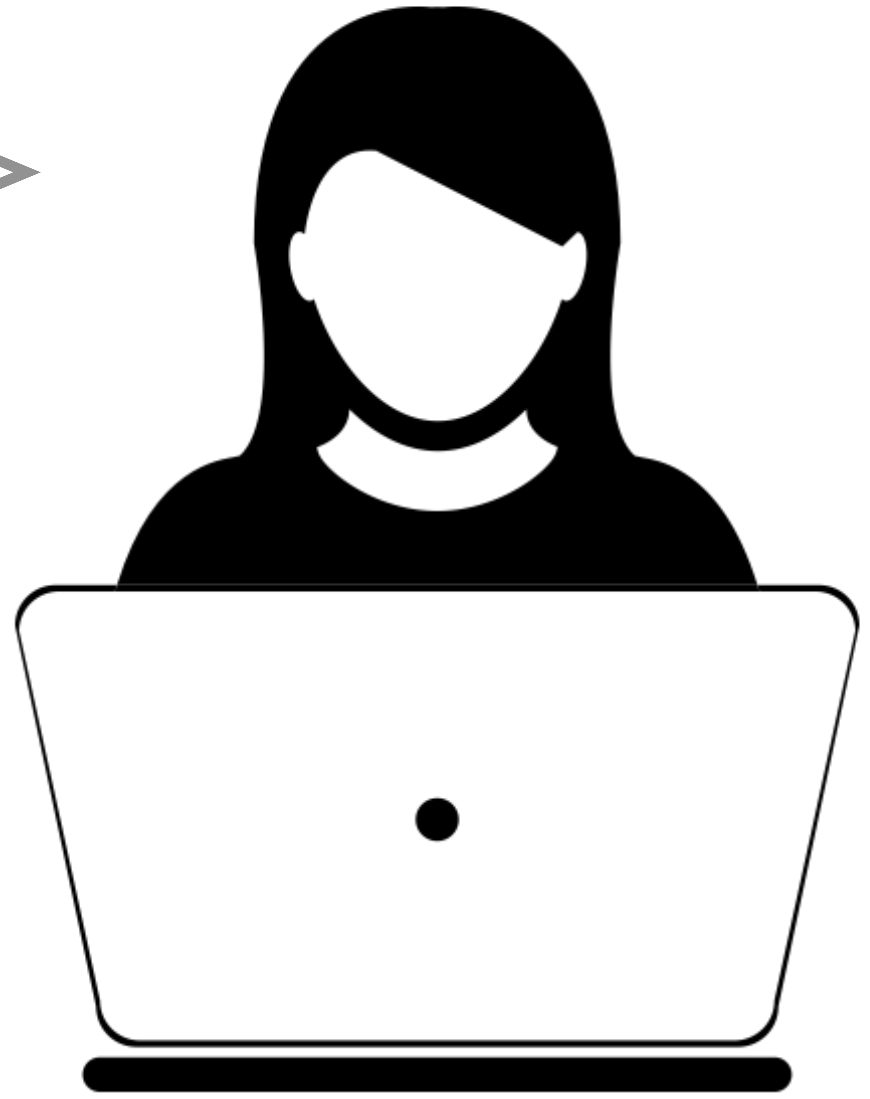
← How diverse is the result?

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

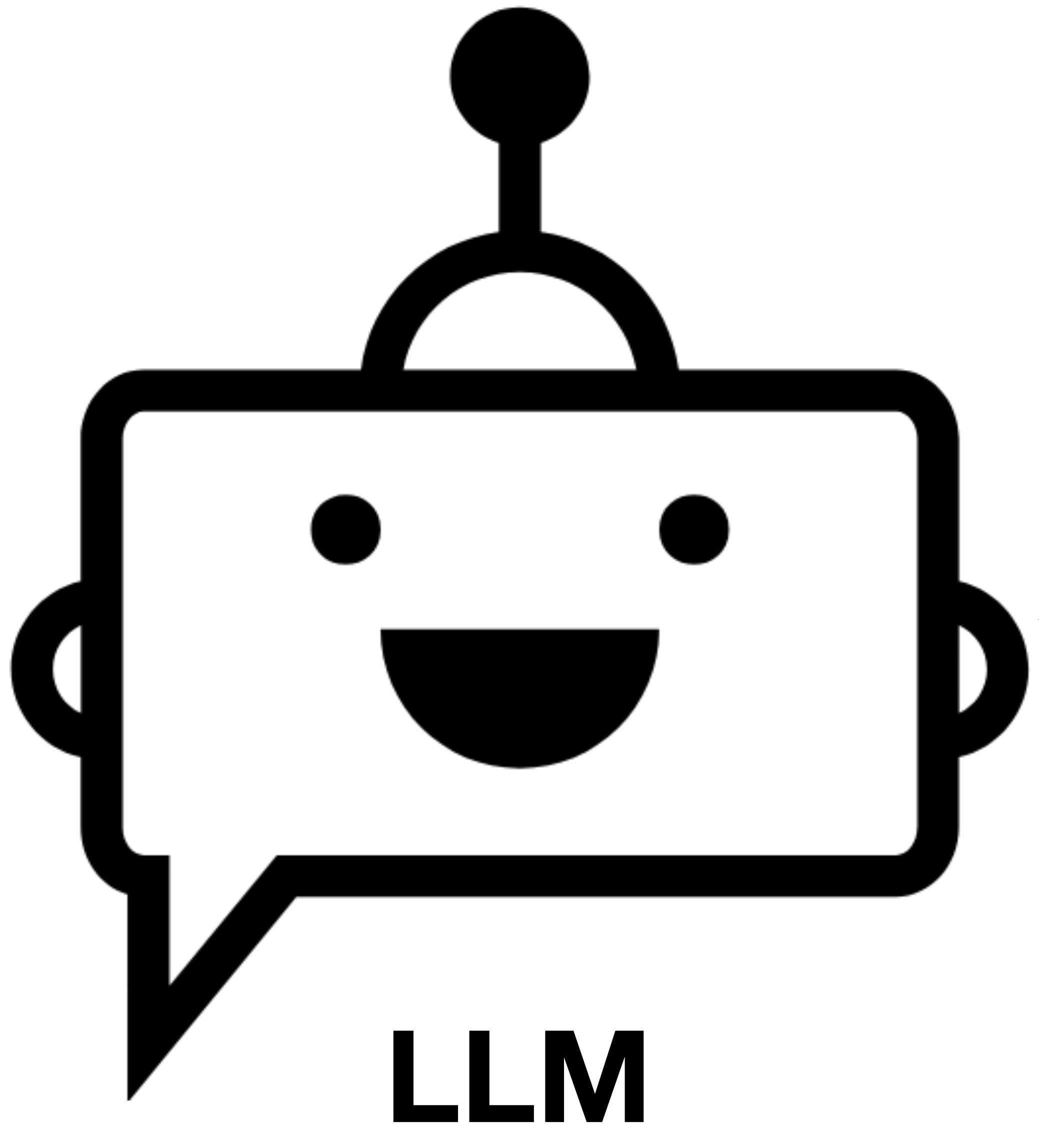
LLM-based scholar recommendations

(audits)

Suggest me **<u>4</u> <u>early career</u>** researchers working at the intersection between **<u>classical network science and machine learning</u>**



You



LLM

Here are four early career scholars working at the intersection of classical network science and machine learning:

1. Márton Posfai
2. Albert László Barabási
3. Márton Karsai
4. Alexander Schallenberg

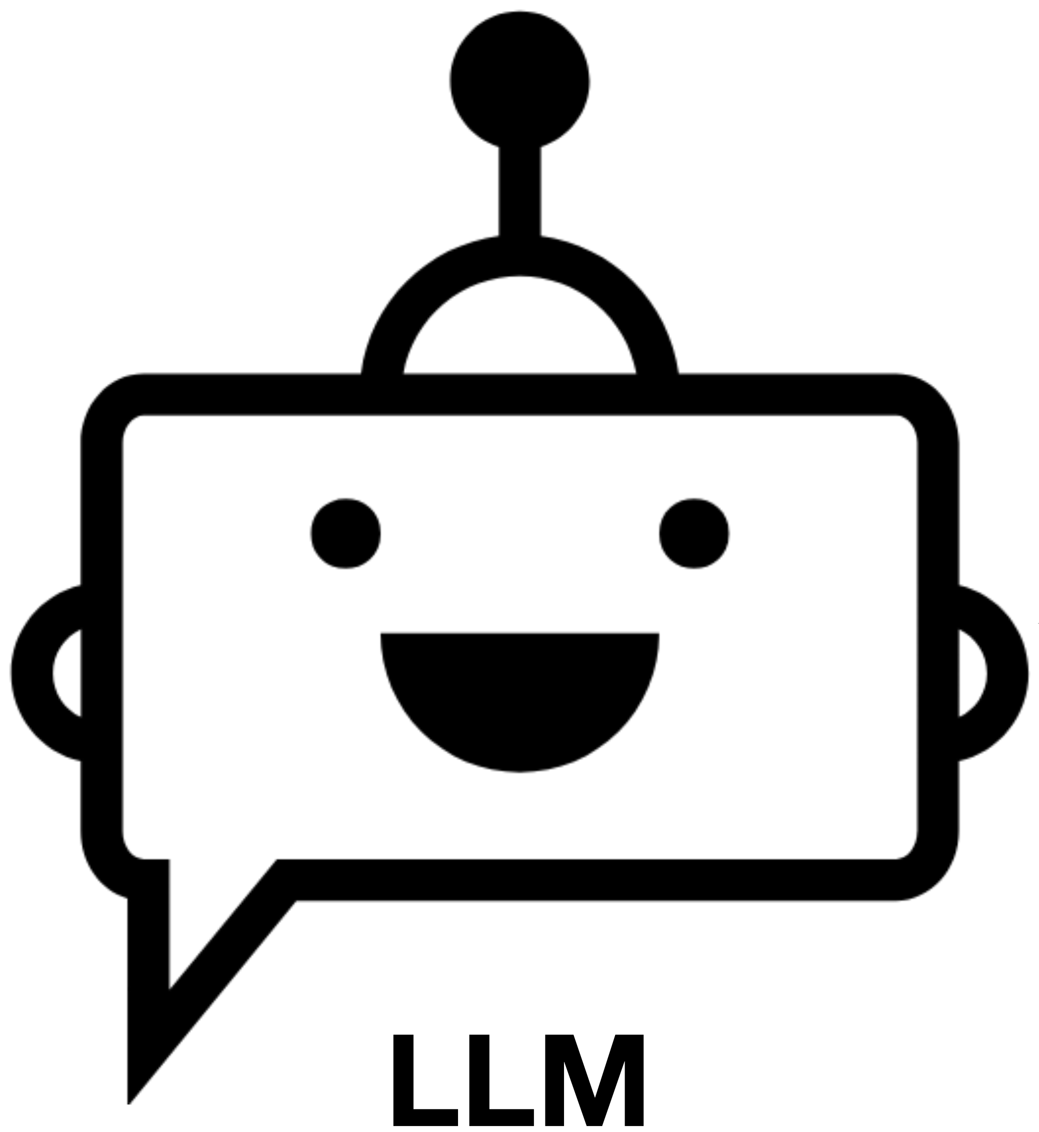
Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

(audits)

Suggest me **4** **early career** researchers working at the intersection between **classical network science and machine learning**

You



LLM

Here are four early career scholars working at the intersection of classical network science and machine learning:

1. Márton Posfai ← Exists, but is not an early career
2. Albert László Barabási ← Exists, but is not an early career
3. Márton Karsai ← Exists, but is not an early career
4. Alexander Schallenberg ← Does not exist in academia

The audit

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

(audits)

Suggest me 4 early career researchers working at the intersection between classical network science and machine learning

... rs working ... twork
science and machine learning:

1. Márton Posfai
2. Albert László Barabási
3. Márton Karsai
4. Alexander Schallenberg

eg. The LLM is biased towards men, and Hungarians

← Exists, but is not an early career
← Exists, but is not an early career
← Exists, but is not an early career
← Does not exist in academia

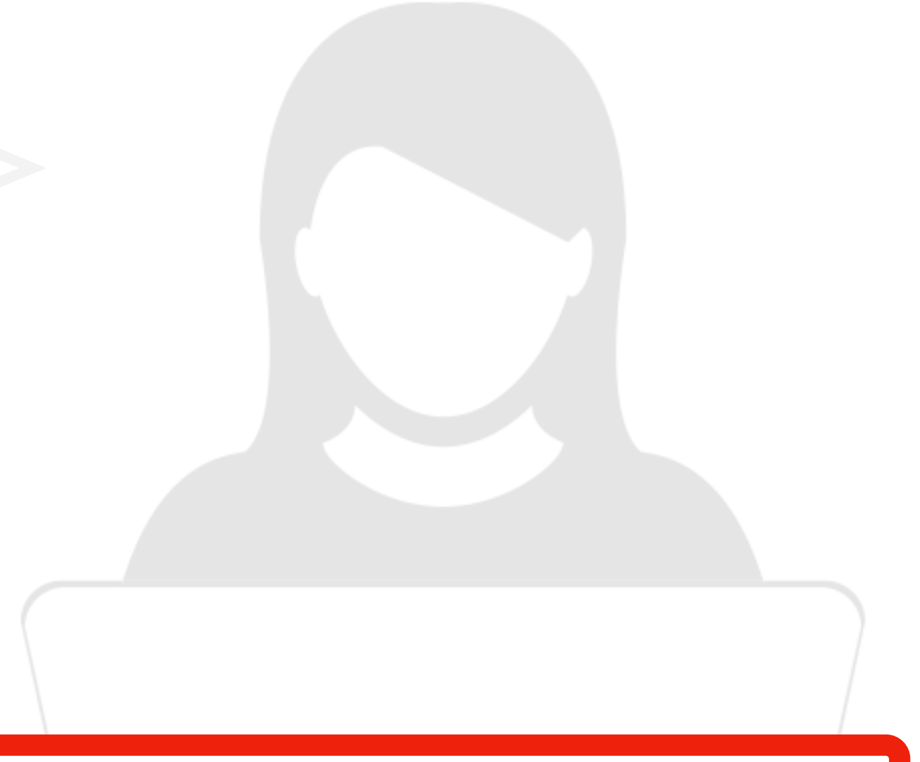
The audit

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

LLM-based scholar recommendations

(audits)

Suggest me 4 early career researchers working at the intersection between classical network science and machine learning



You

LLM

science and machine learning:

1. Márton Posfai
2. Albert László Barabási
3. Márton Karsai
4. Alexander Schallenberg

eg. The LLM is biased towards men, and Hungarians

eg. The LLM under-represents Asian scholars

← Exists, but is not an early career
← Exists, but is not an early career
← Exists, but is not an early career
← Does not exist in academia

The audit

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

To what extent do LLMs represent minorities in scholar recommendations?

RQ1

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

To what extent do LLMs represent minorities in scholar recommendations?

RQ1

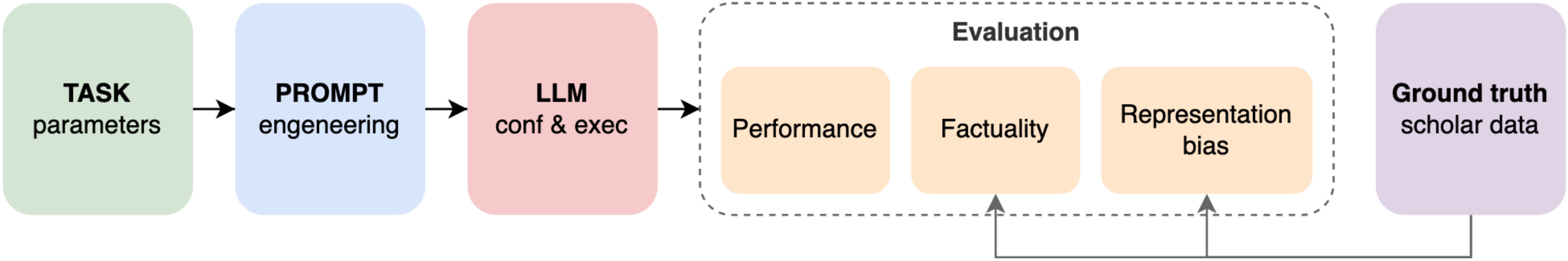
Background:

Lahoti, P., et al. Improving diversity of demographic representation in large language models via collective-critiques and self-voting. EMNLP 2023.

Barolo et al. 2025 (AISci workshop)
Barolo et al. 2025 (under-review)

Auditing LLMs

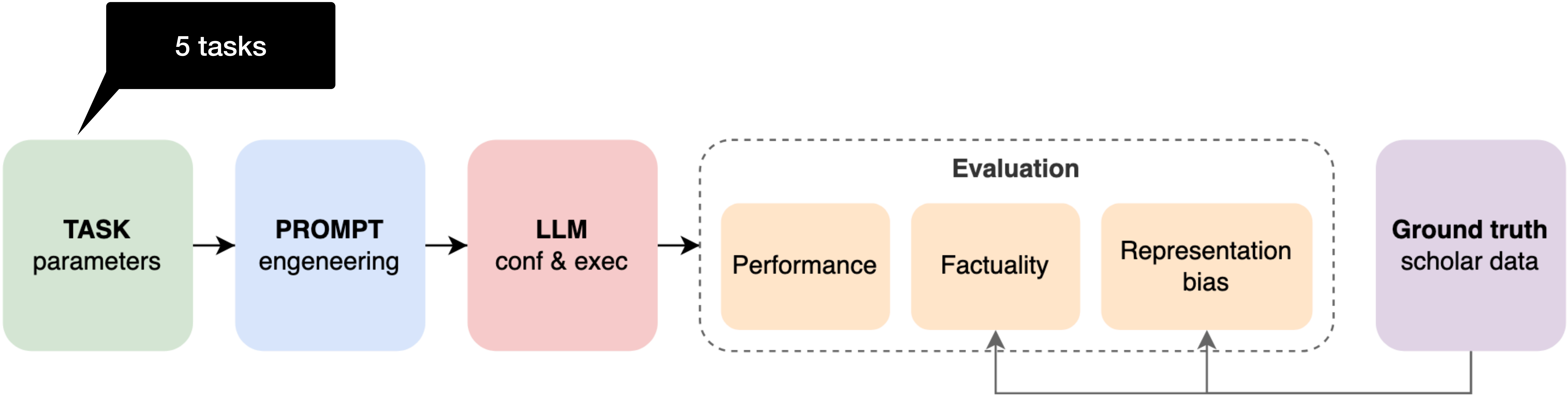
RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

TASK
parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Tasks

Rank by **expertise**

Who are the **top 5** experts in **Physics**?

Who are the **top 100** experts in **Physics**?

TASK parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Tasks

Rank by **expertise**

Search by **epoch**

Who were the most influential **physicists** of the **1950s**?

Who were the most influential **physicists** of the **2000s**?

TASK parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Tasks

Rank by expertise

Search by epoch

Search by Field

Suggest me influential scientists working on **Physics Education Research**

Suggest me influential scientists working on **Condensed Matter and Materials Physics**

TASK parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Tasks

Rank by expertise

Search by epoch

Search by Field

Search by seniority

Suggest me **early-career physicists**

Suggest me **senior physicists**

TASK parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Tasks

Rank by expertise

Search by epoch

Search by Field

Search by seniority

Search by statistical twin

Who is the **statistical twin** of **Albert Laszlo Barabási**?

Who is the **statistical twin** of **Lisette Espín-Noboa**?

Who is the **statistical twin** of **Sheldon Cooper**?

TASK parameters

Barolo et al. 2025 (under-review)

Auditing LLMs: Tasks

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

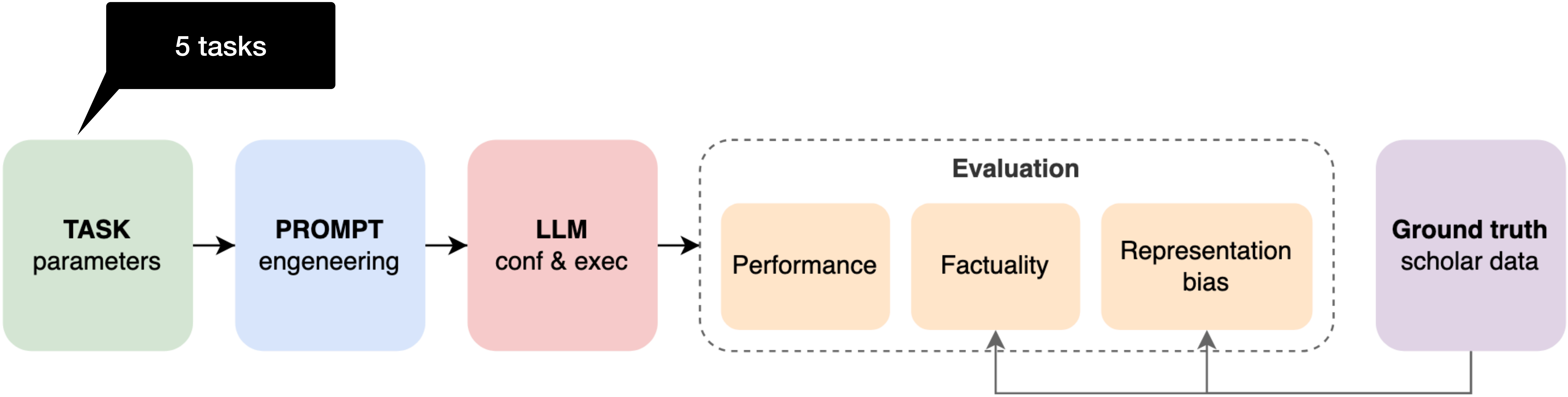
	Rank by **expertise**	Search by **epoch**	Search by **Field**	Search by **seniority**	Search by **statistical twin**
Tasks	Rank by **expertise**	Search by **epoch**	Search by **Field**	Search by **seniority**	Search by **statistical twin**
Parameters	• Top-5 • Top-100	• 1950s • 2000s	• PER • CM&MP	• Early career • Senior	• Famous • Random • Politician • TV character • Fictitious

TASK
parameters

Barolo et al. 2025 (under-review)

Auditing LLMs

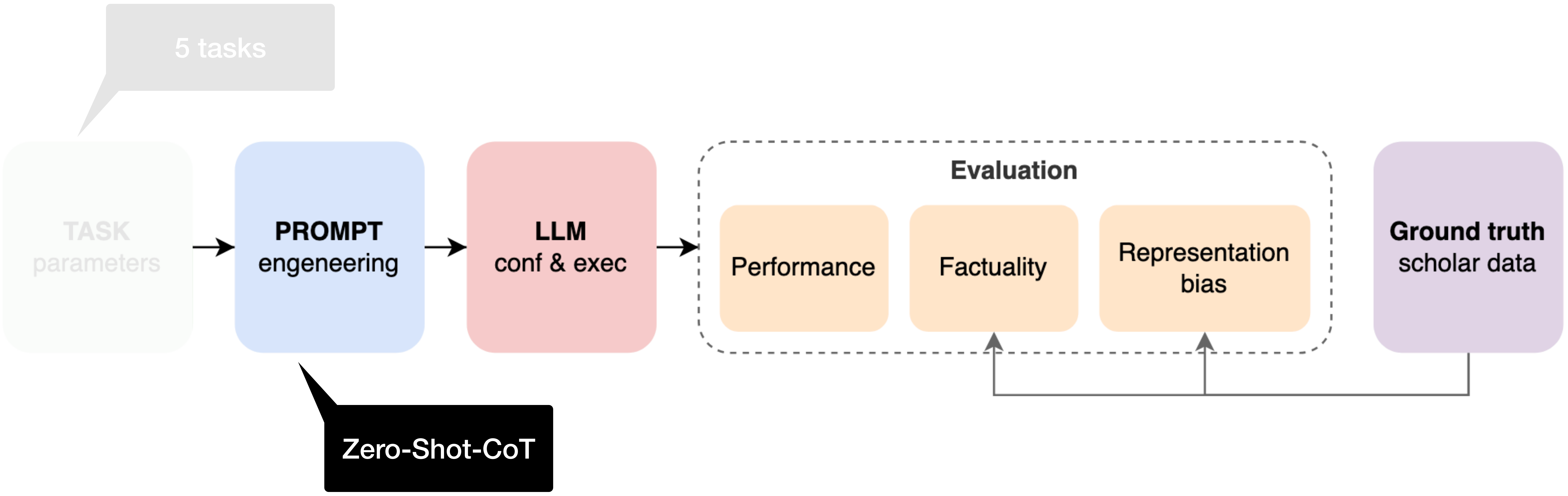
RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs

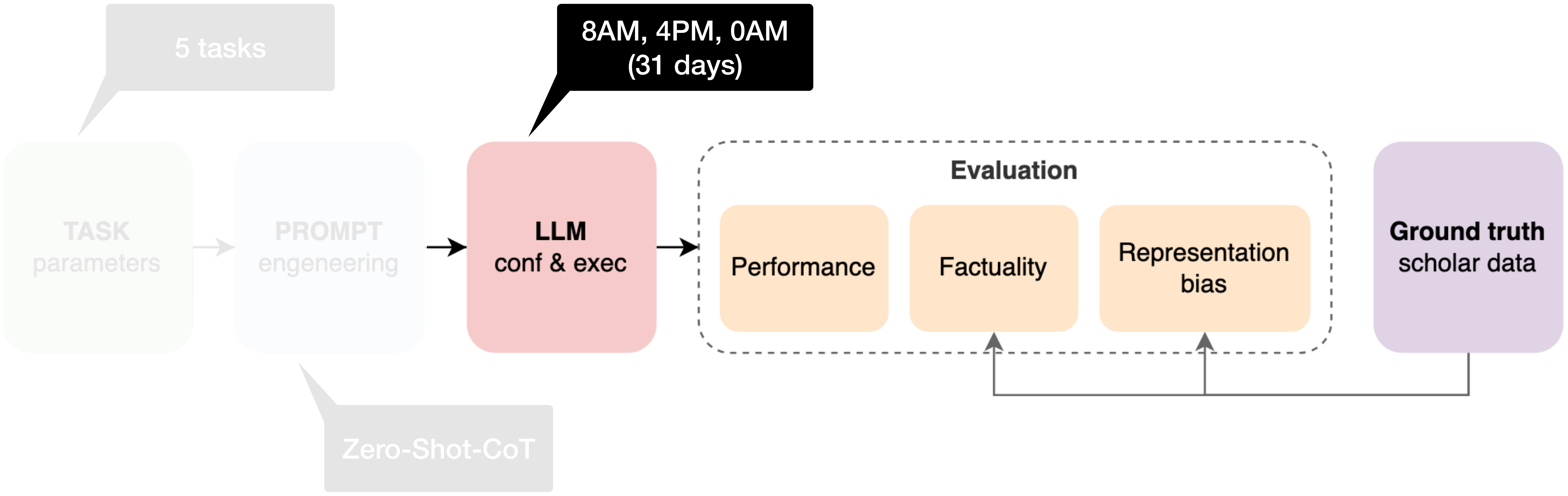
RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs

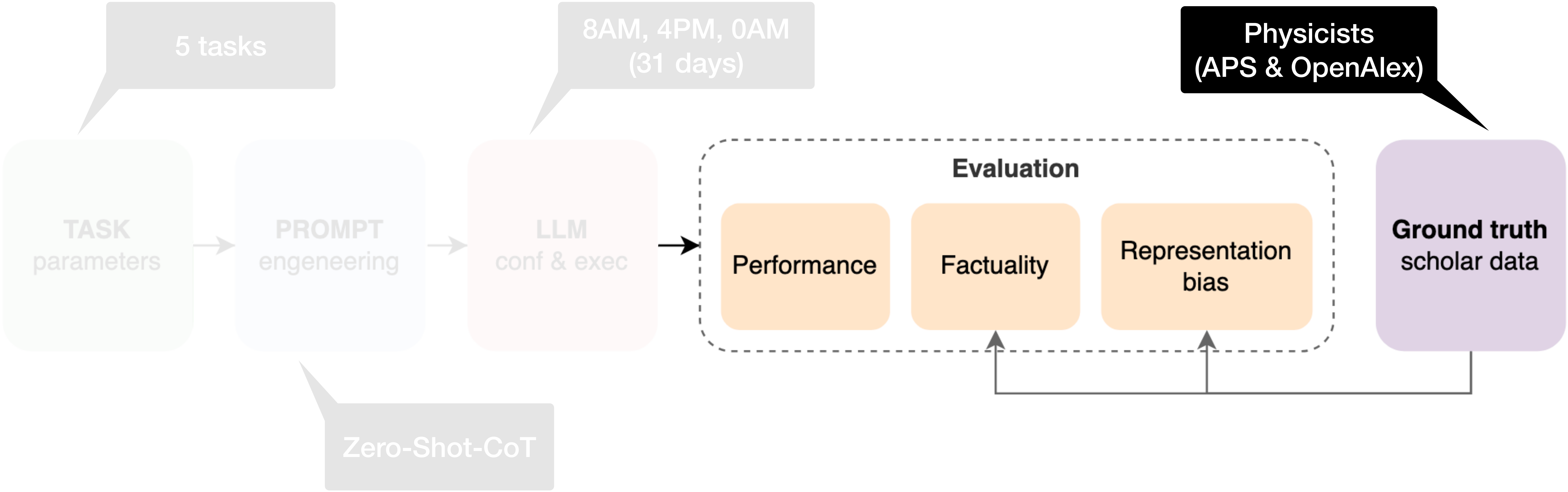
RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

Ground truth
scholar data

Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

AMERICAN PHYSICAL SOCIETY

Ground truth
scholar data

OpenAlex

Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?

APS
AMERICAN PHYSICAL SOCIETY

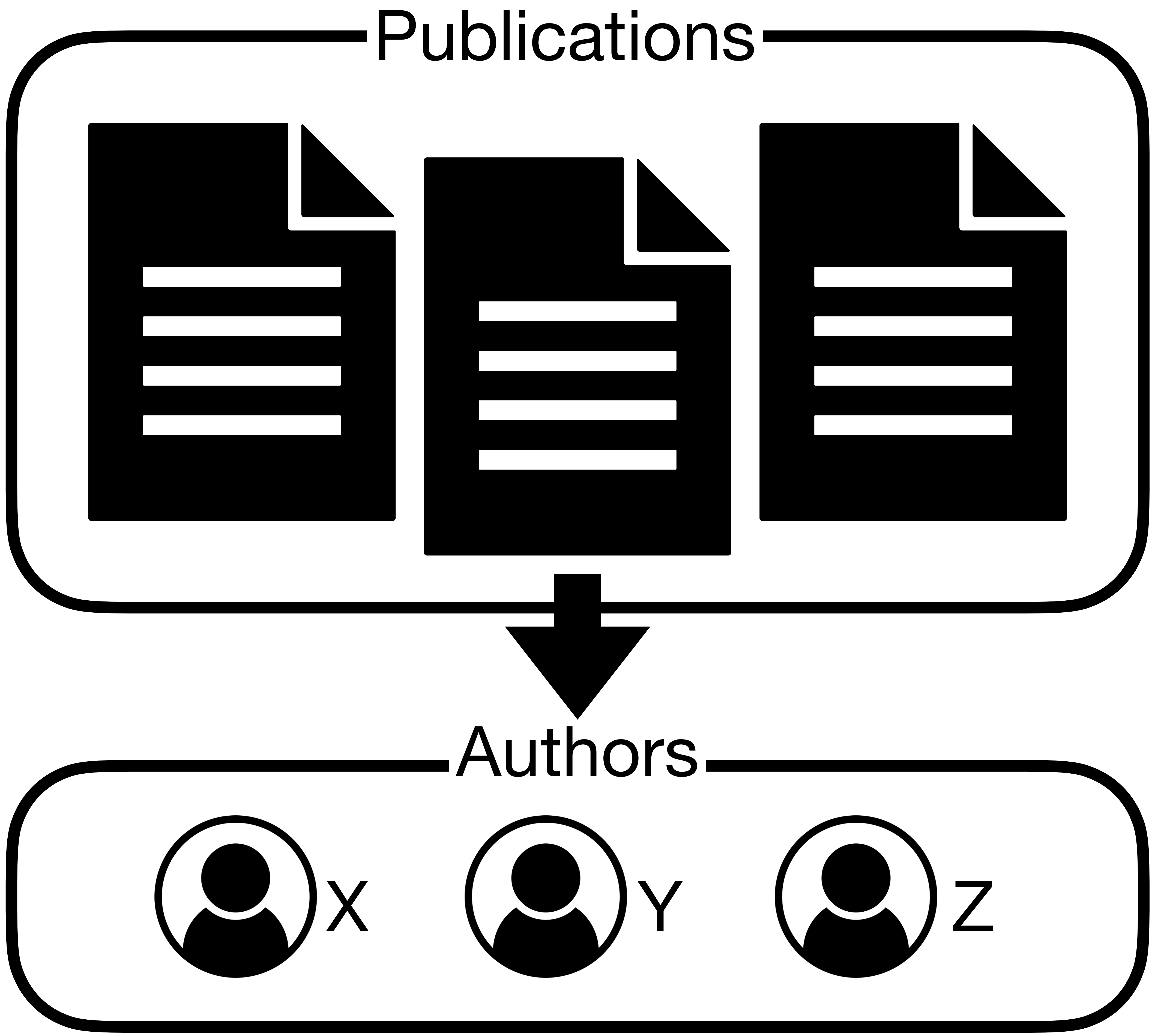
Ground truth scholar data

OpenAlex

Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



AMERICAN PHYSICAL SOCIETY

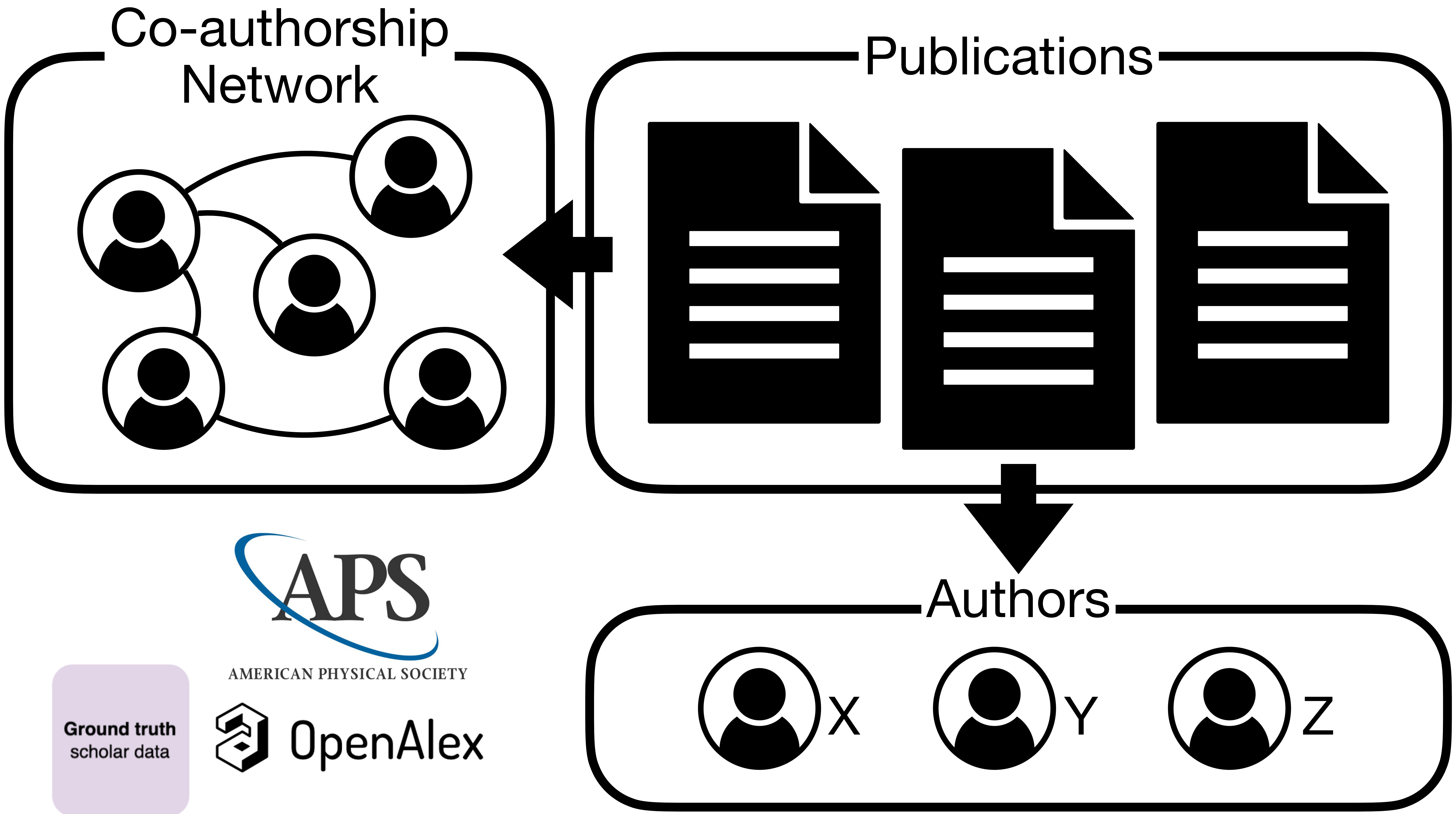
Ground truth
scholar data

OpenAlex

Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Ground truth scholar data

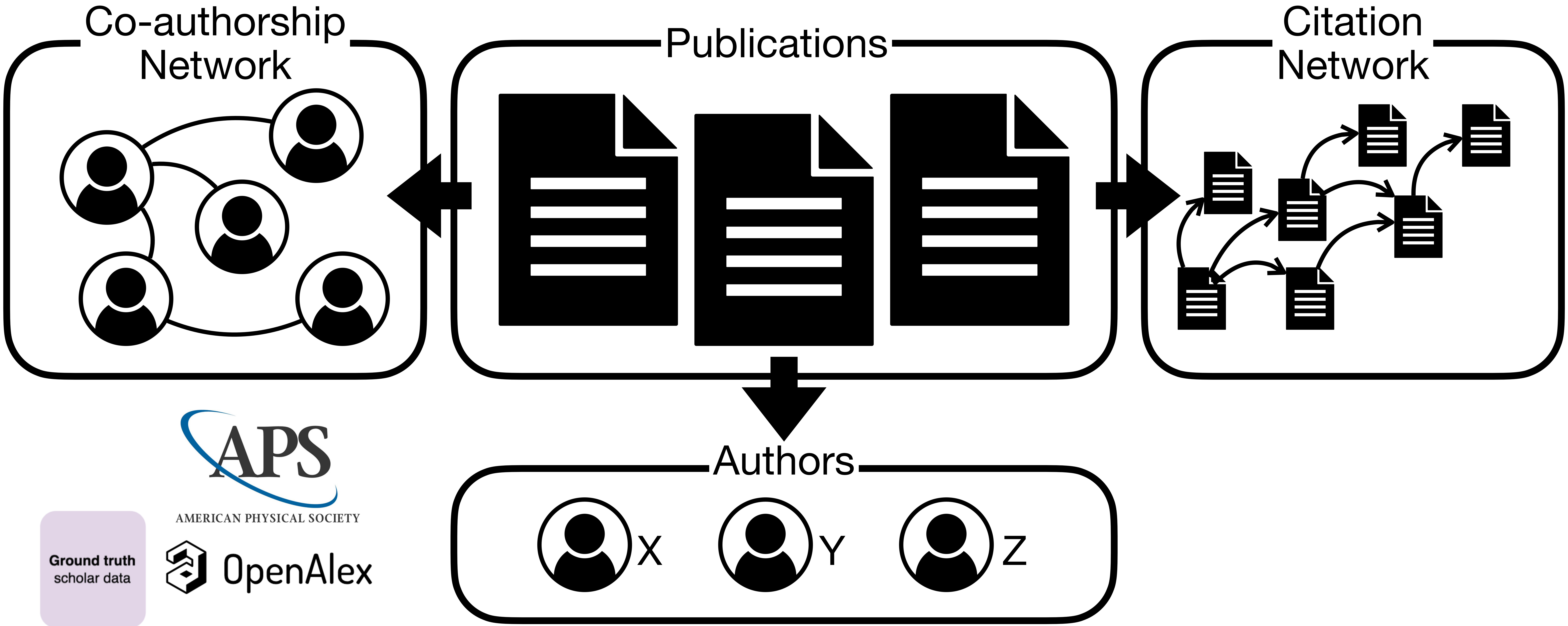
AMERICAN PHYSICAL SOCIETY

OpenAlex

Barolo et al. 2025 (under-review)

Auditing LLMs: Ground-truth data

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Ground truth scholar data

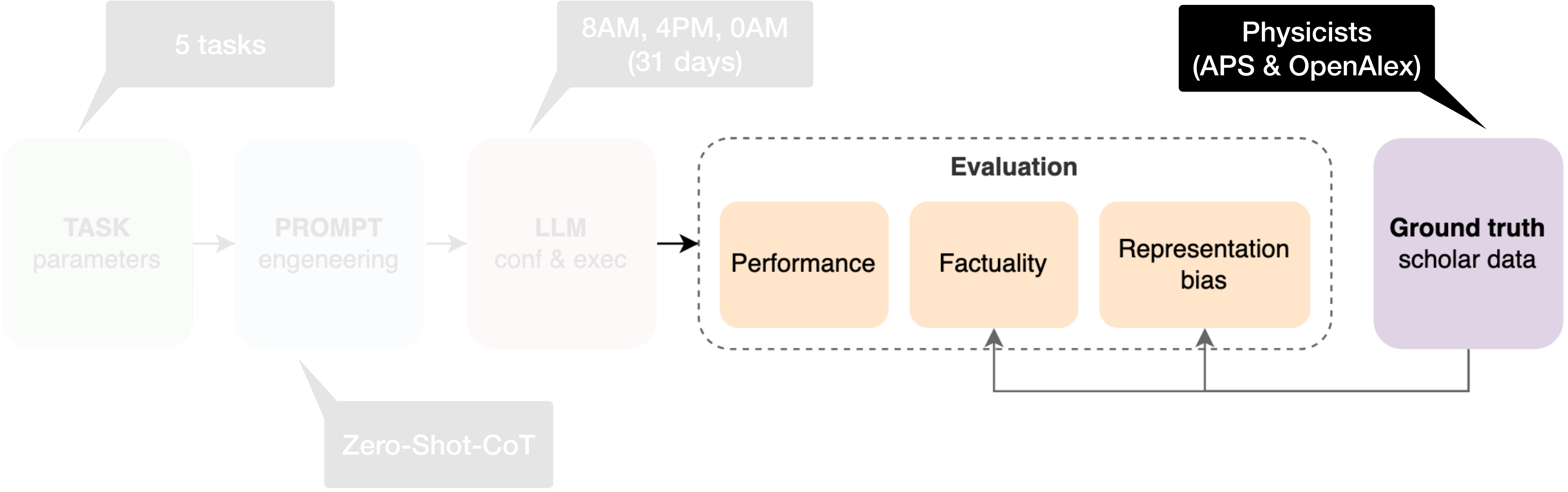
AMERICAN PHYSICAL SOCIETY

OpenAlex

Barolo et al. 2025 (under-review)

Auditing LLMs

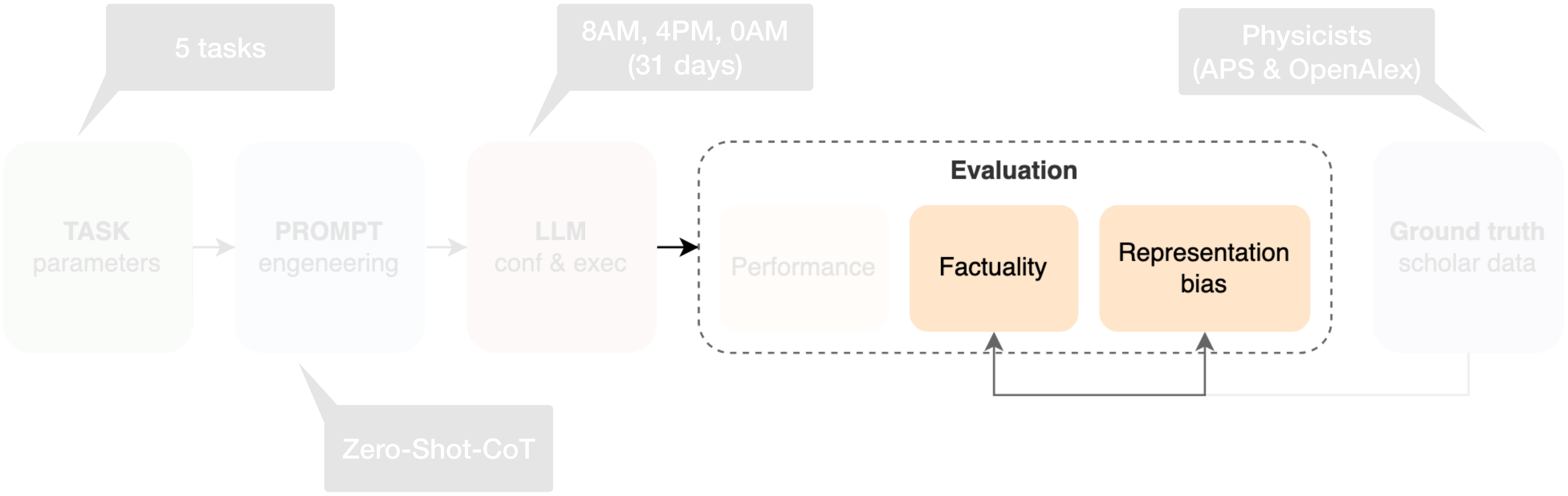
RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

Auditing LLMs

RQ1: To what extent do LLMs represent minorities in scholar recommendations?



Barolo et al. 2025 (under-review)

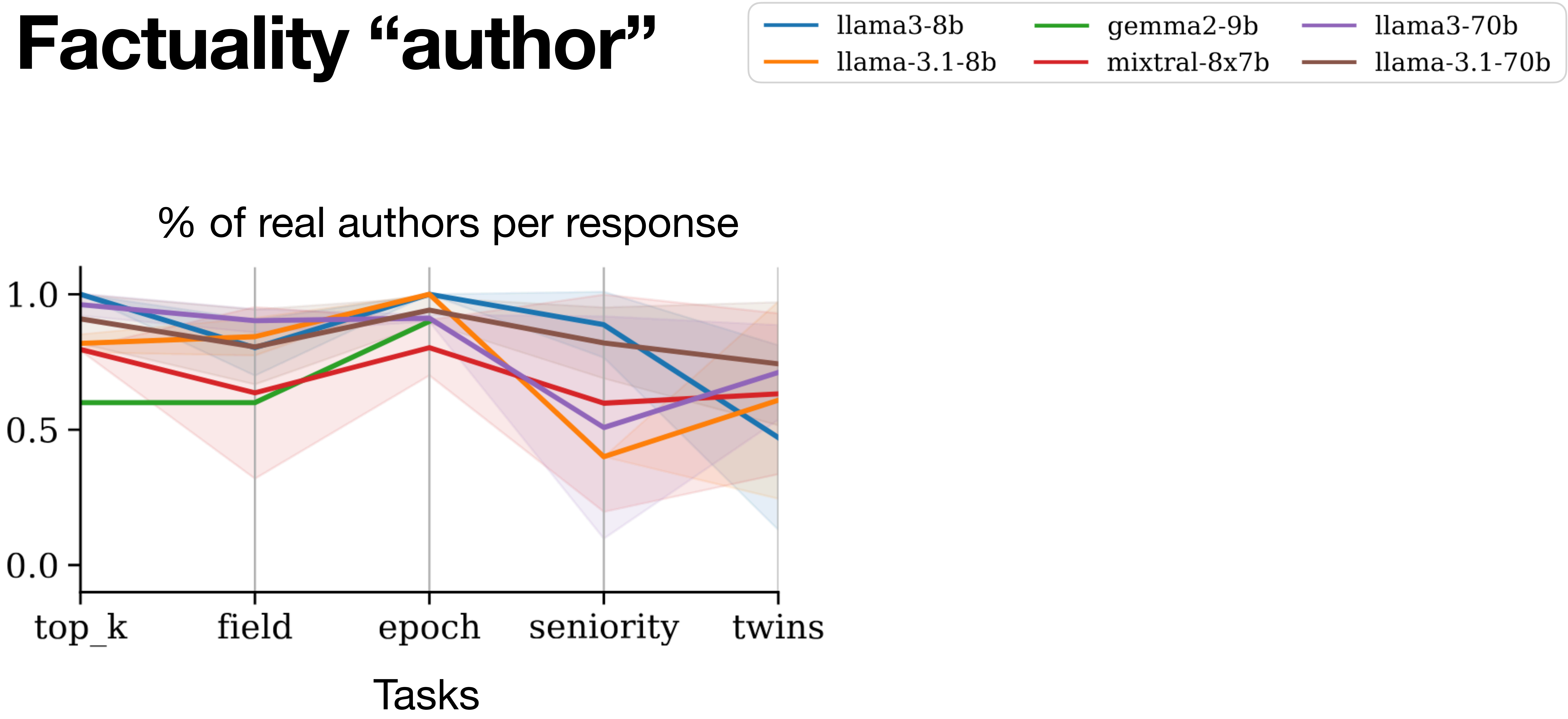
Factuality

LLM Audits

Factuality "author"

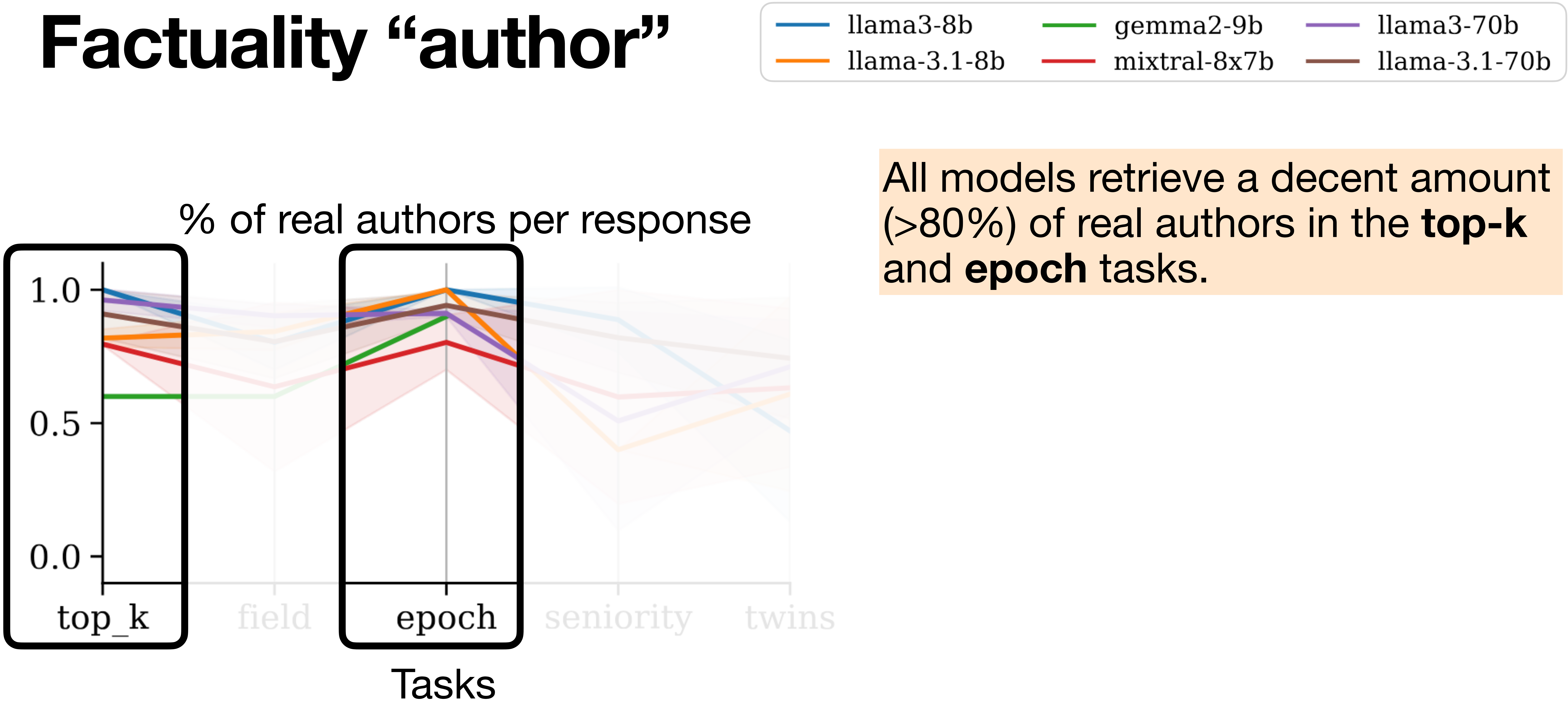
Barolo et al. 2025 (under-review)

Factuality "author"



Barolo et al. 2025 (under-review)

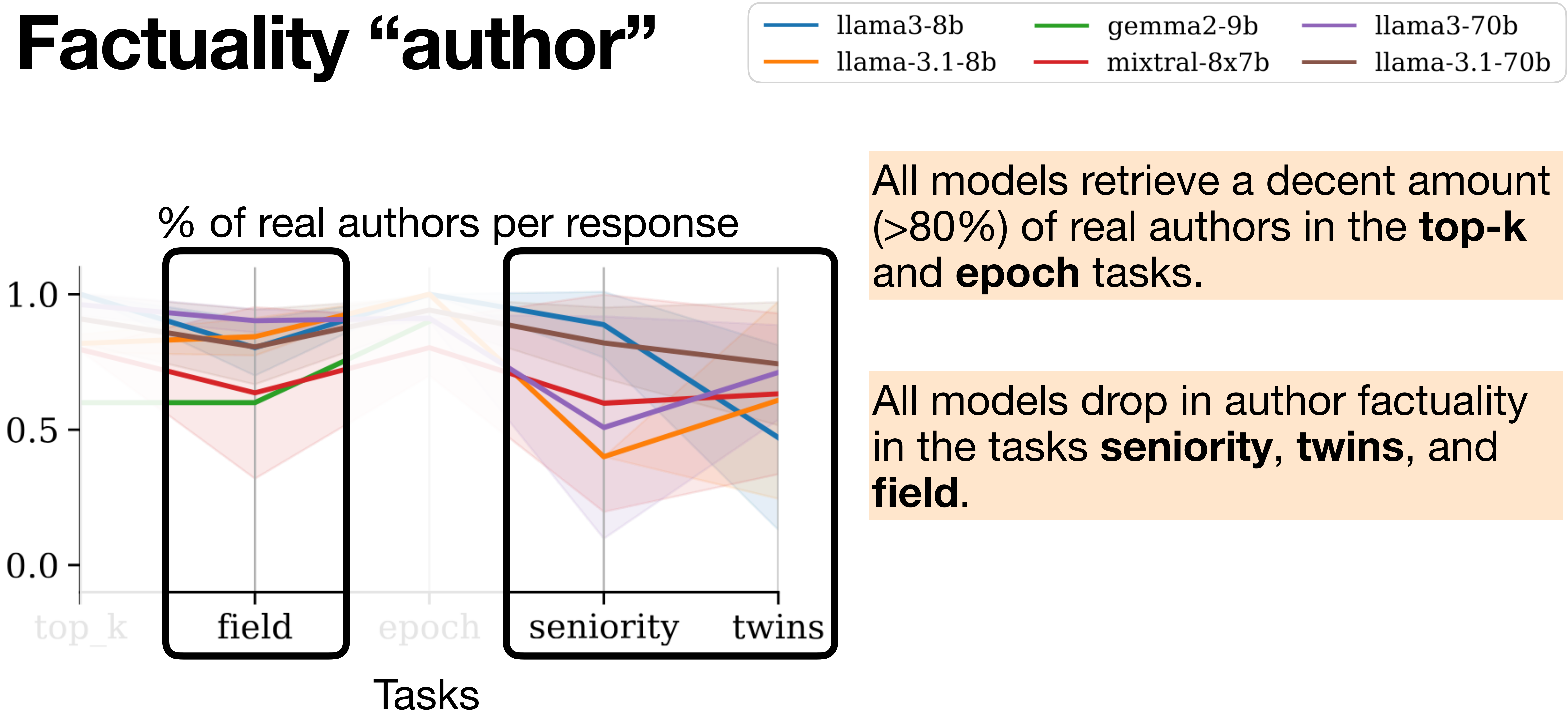
Factuality "author"



All models retrieve a decent amount (>80%) of real authors in the **top-k** and **epoch** tasks.

Barolo et al. 2025 (under-review)

Factuality "author"

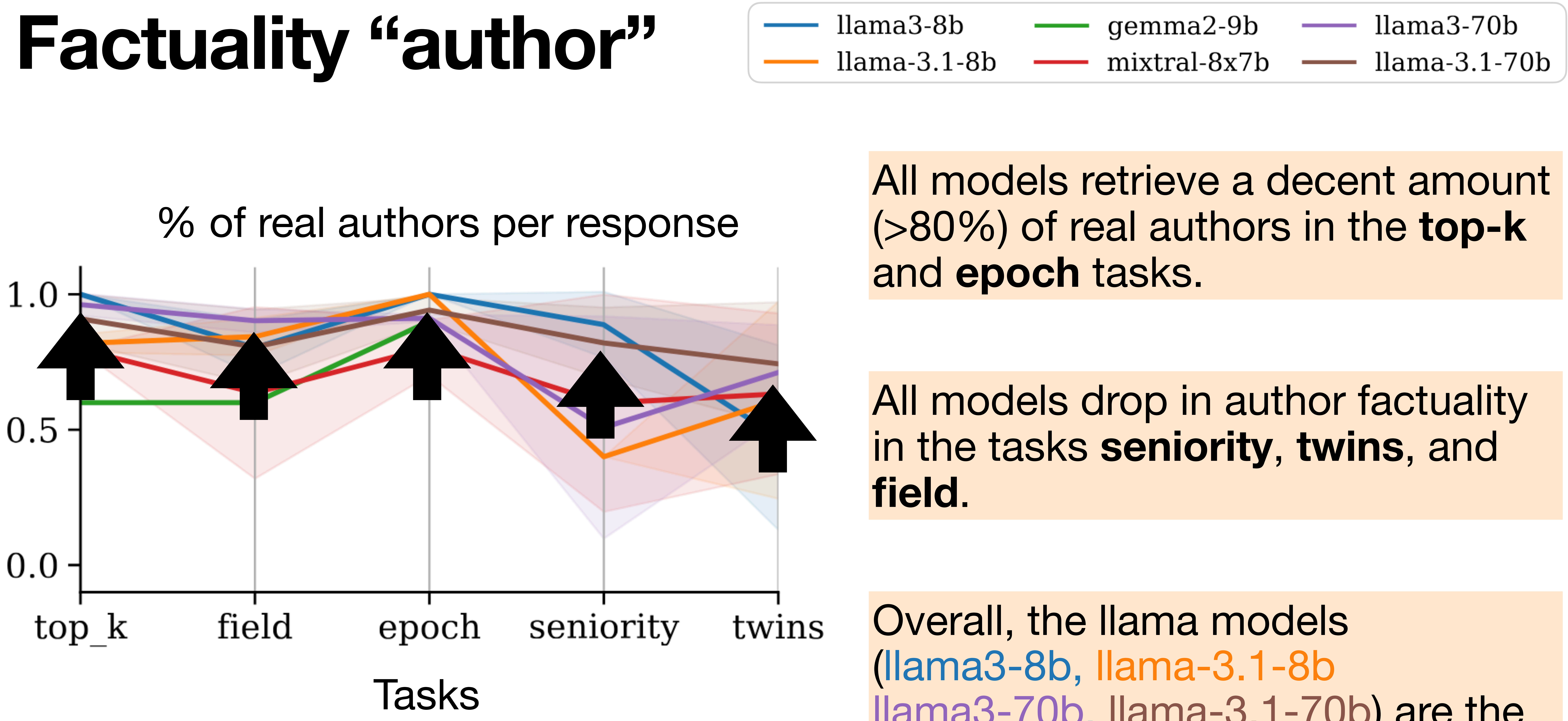


All models retrieve a decent amount (>80%) of real authors in the **top-k** and **epoch** tasks.

All models drop in author factuality in the tasks **seniority**, **twins**, and **field**.

Barolo et al. 2025 (under-review)

Factuality "author"



All models retrieve a decent amount (>80%) of real authors in the **top-k** and **epoch** tasks.

All models drop in author factuality in the tasks **seniority**, **twins**, and **field**.

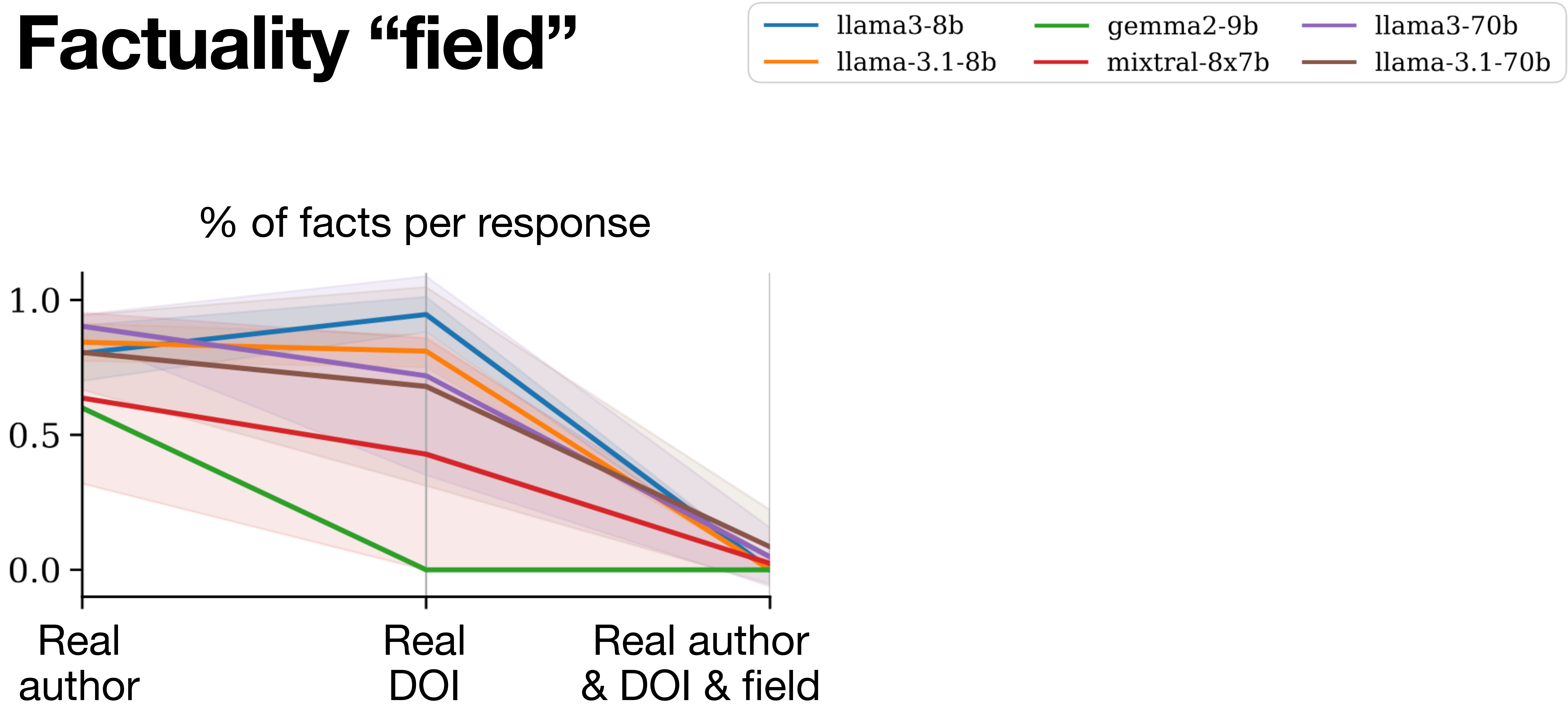
Overall, the llama models (llama3-8b, llama-3.1-8b llama3-70b, llama-3.1-70b) are the most factual.

Barolo et al. 2025 (under-review)

Factuality “field”

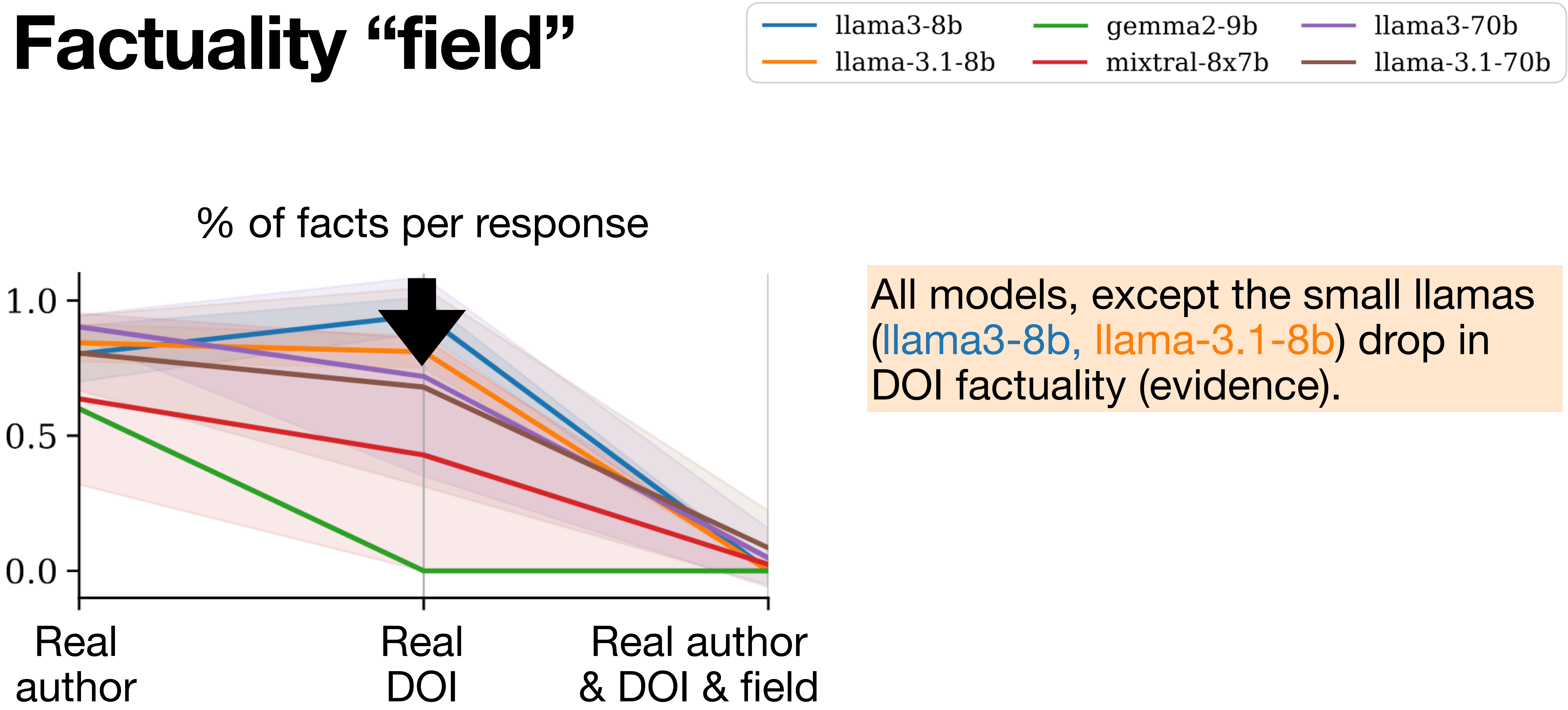
Barolo et al. 2025 (under-review)

Factuality "field"



Barolo et al. 2025 (under-review)

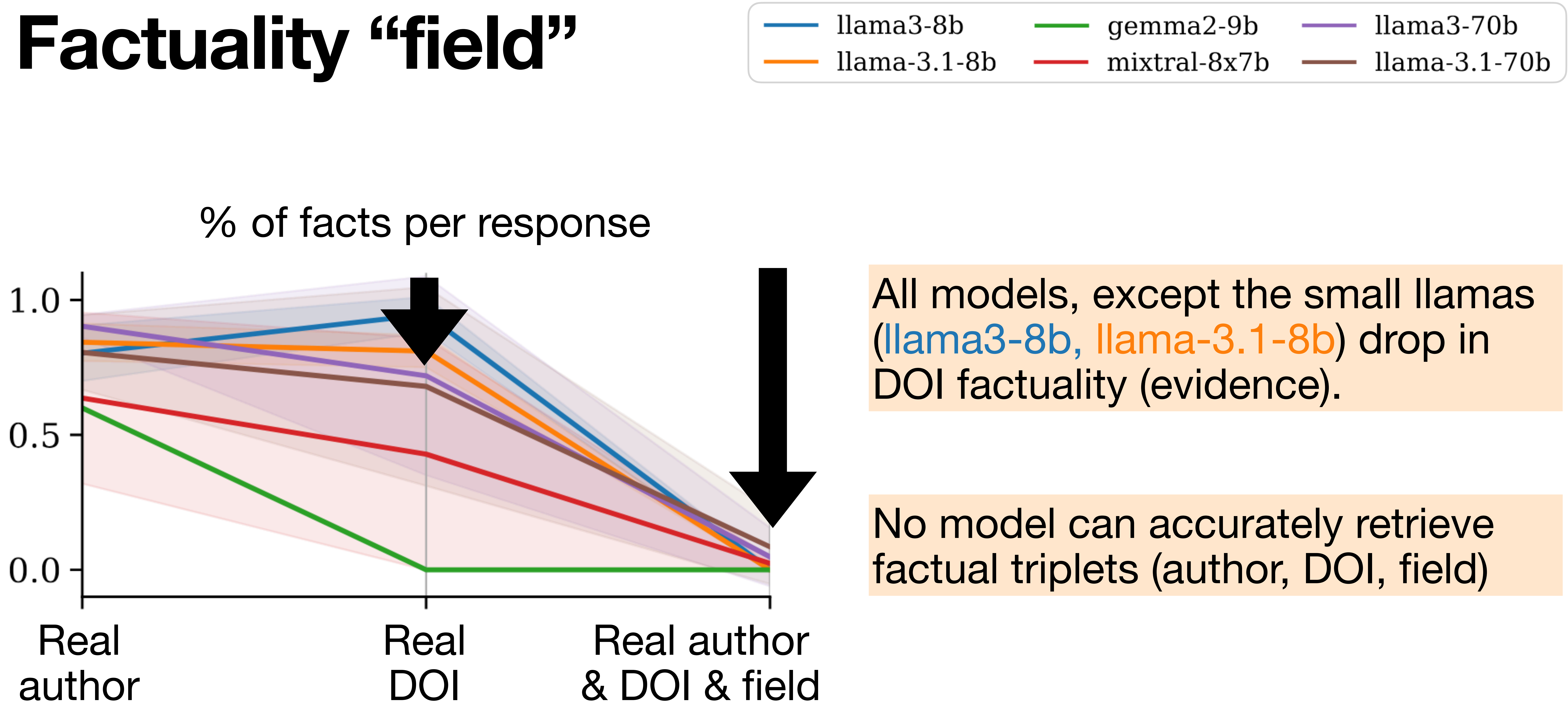
Factuality “field”



All models, except the small llamas (llama3-8b, llama-3.1-8b) drop in DOI factuality (evidence).

Barolo et al. 2025 (under-review)

Factuality “field”



All models, except the small llamas (llama3-8b, llama-3.1-8b) drop in DOI factuality (evidence).

No model can accurately retrieve factual triplets (author, DOI, field)

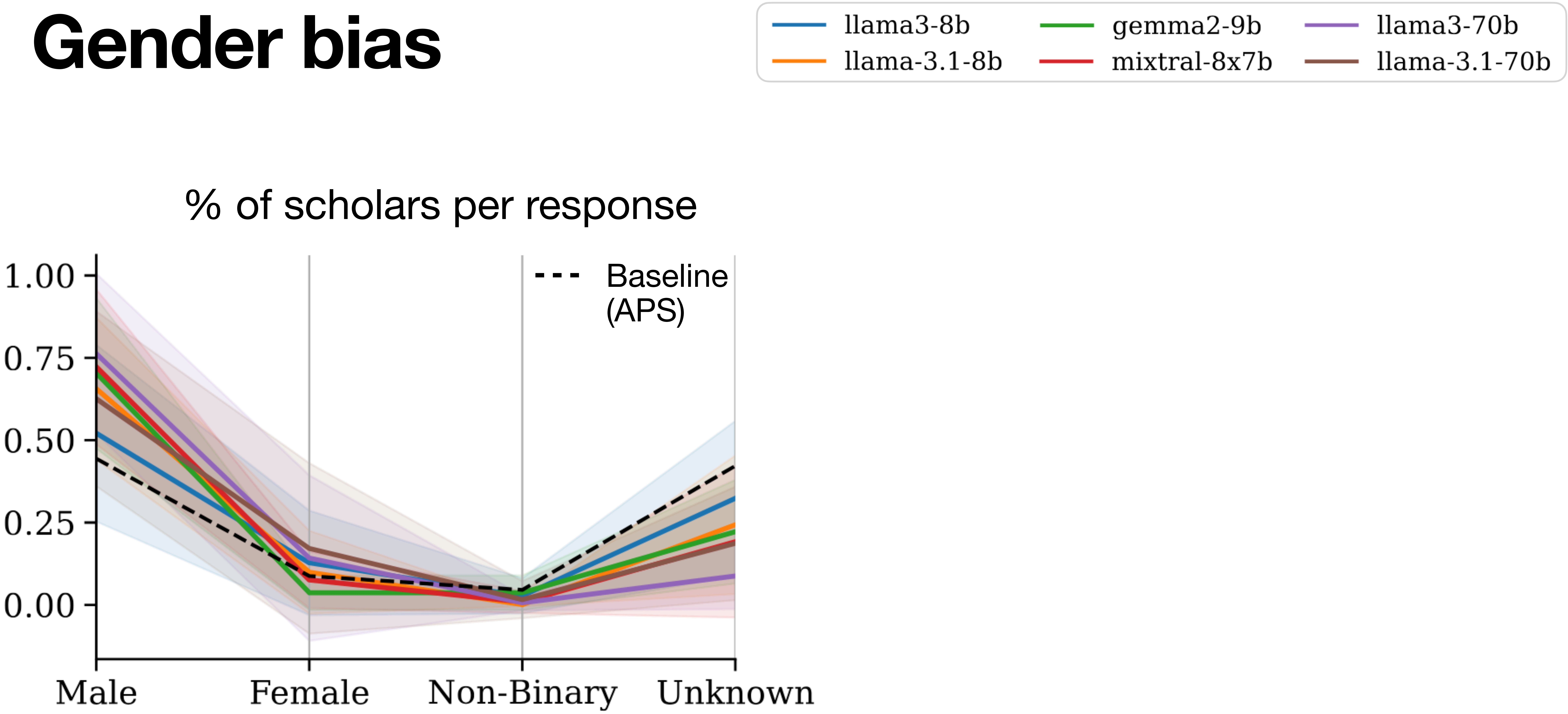
Representation bias

LLM Audits

Gender bias

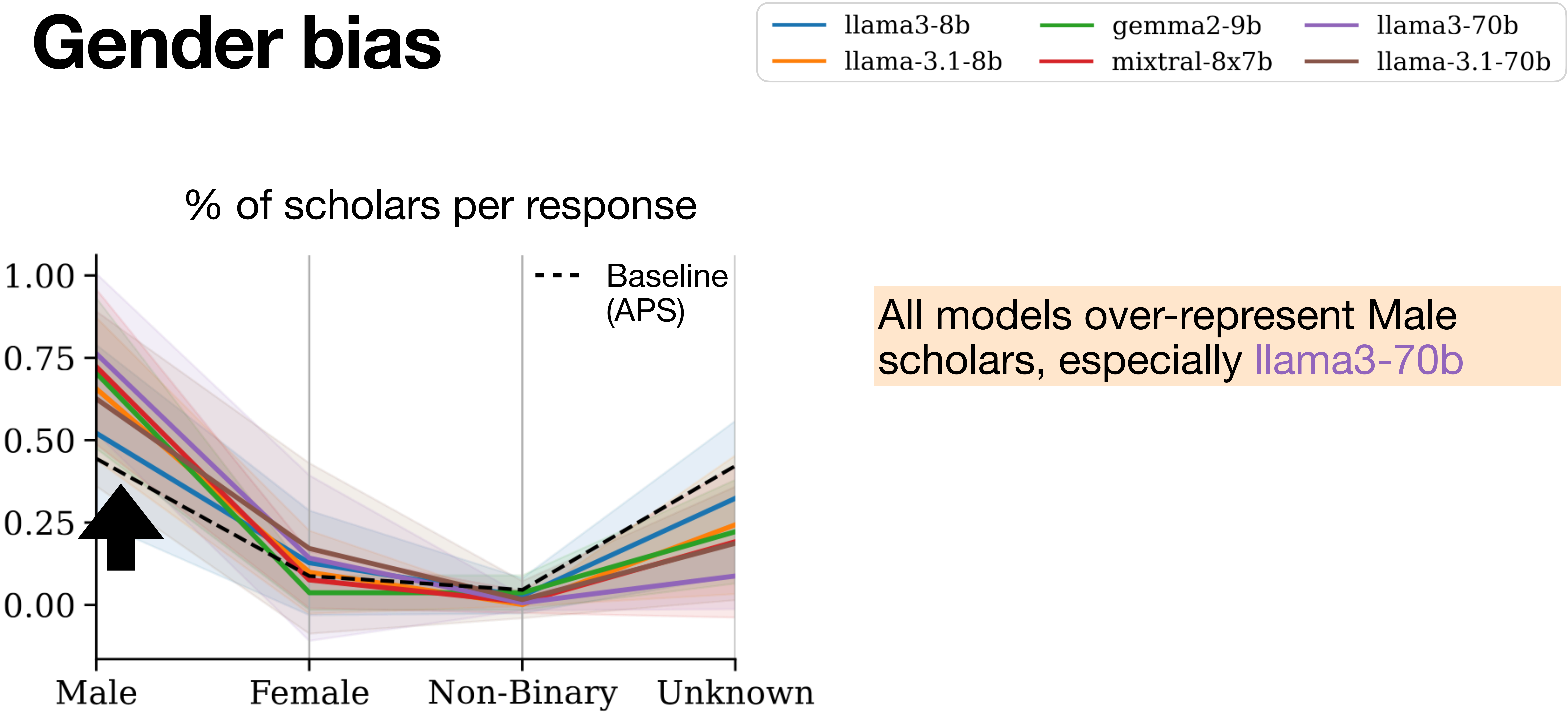
Barolo et al. 2025 (under-review)

Gender bias



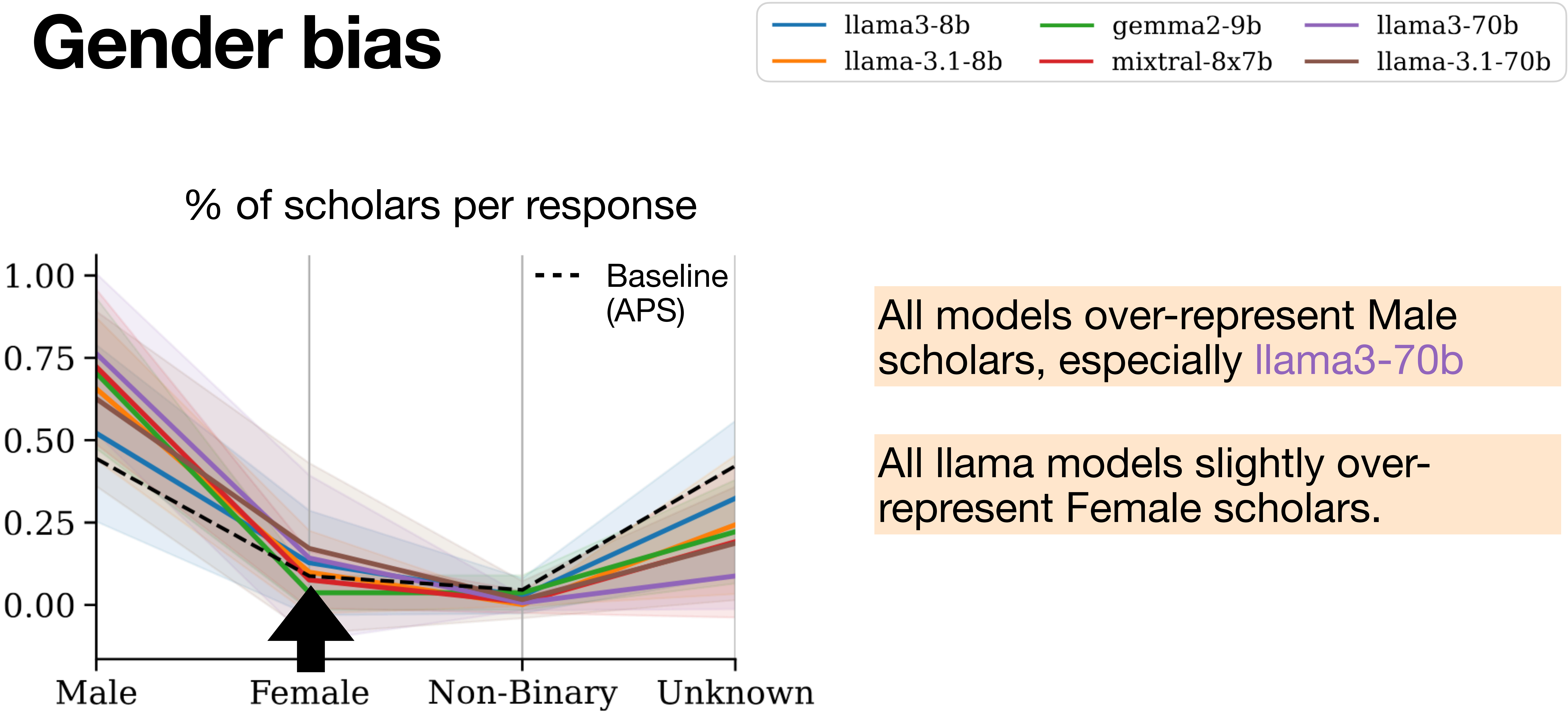
Barolo et al. 2025 (under-review)

Gender bias



All models over-represent Male scholars, especially llama3-70b

Gender bias

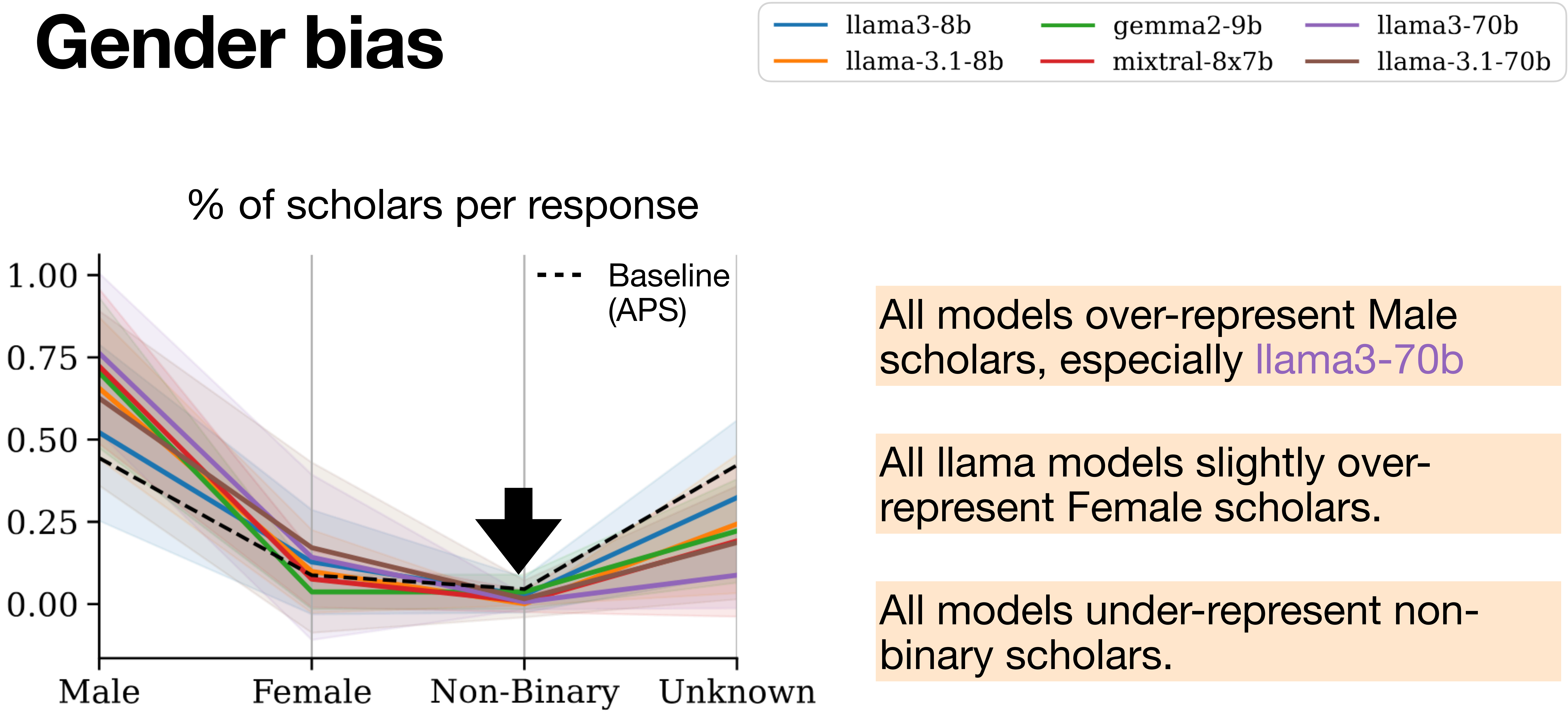


All models over-represent Male scholars, especially llama3-70b

All llama models slightly over-represent Female scholars.

Barolo et al. 2025 (under-review)

Gender bias



All models over-represent Male scholars, especially llama3-70b

All llama models slightly over-represent Female scholars.

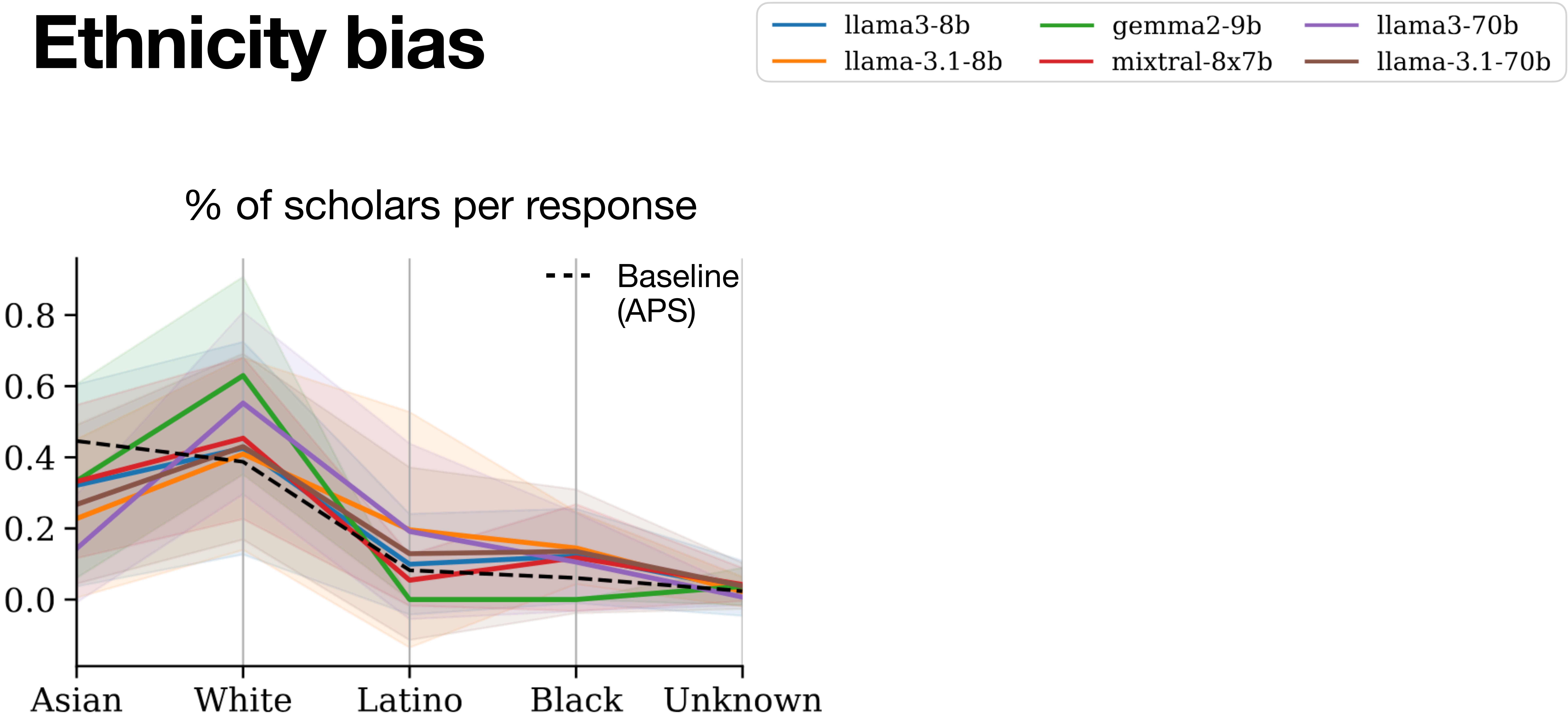
All models under-represent non-binary scholars.

Barolo et al. 2025 (under-review)

Ethnicity bias

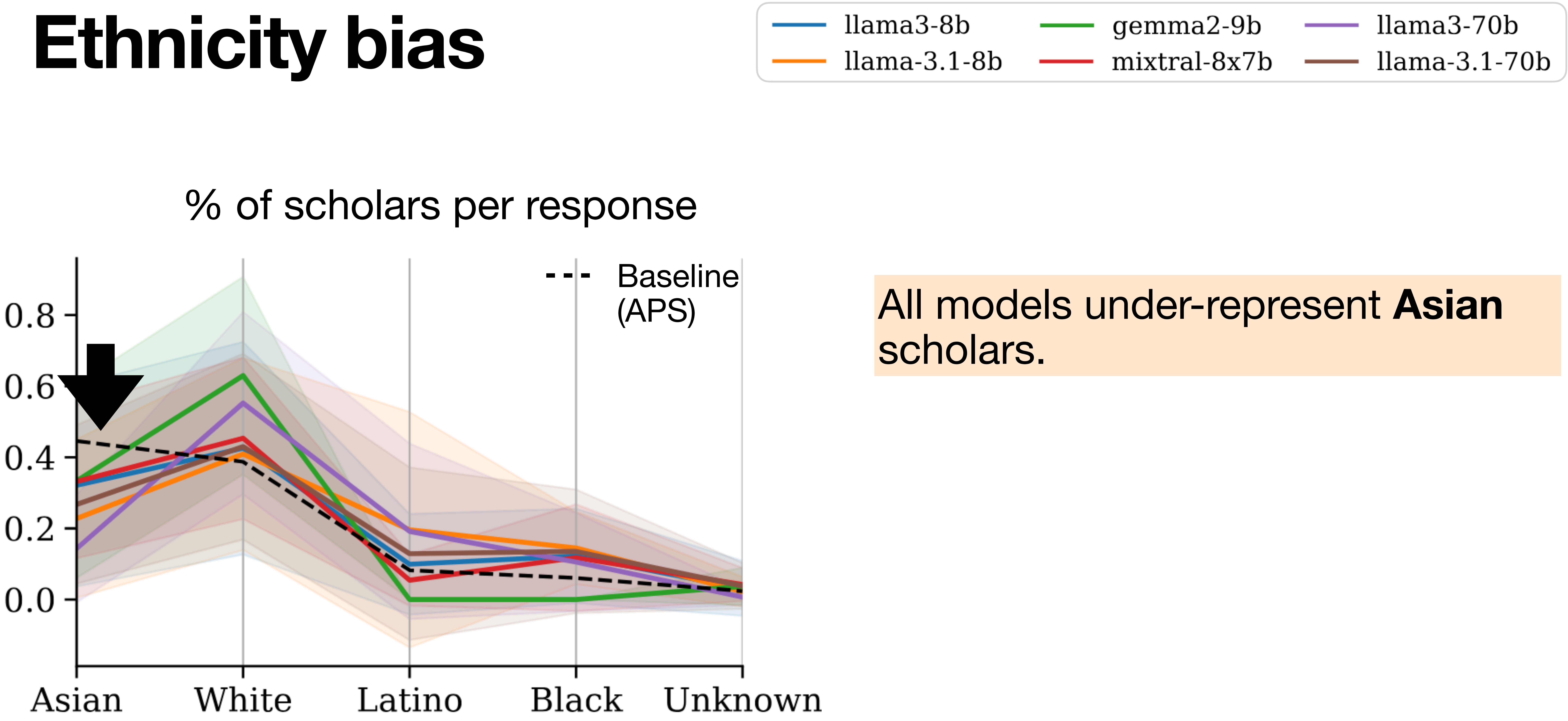
Barolo et al. 2025 (under-review)

Ethnicity bias



Barolo et al. 2025 (under-review)

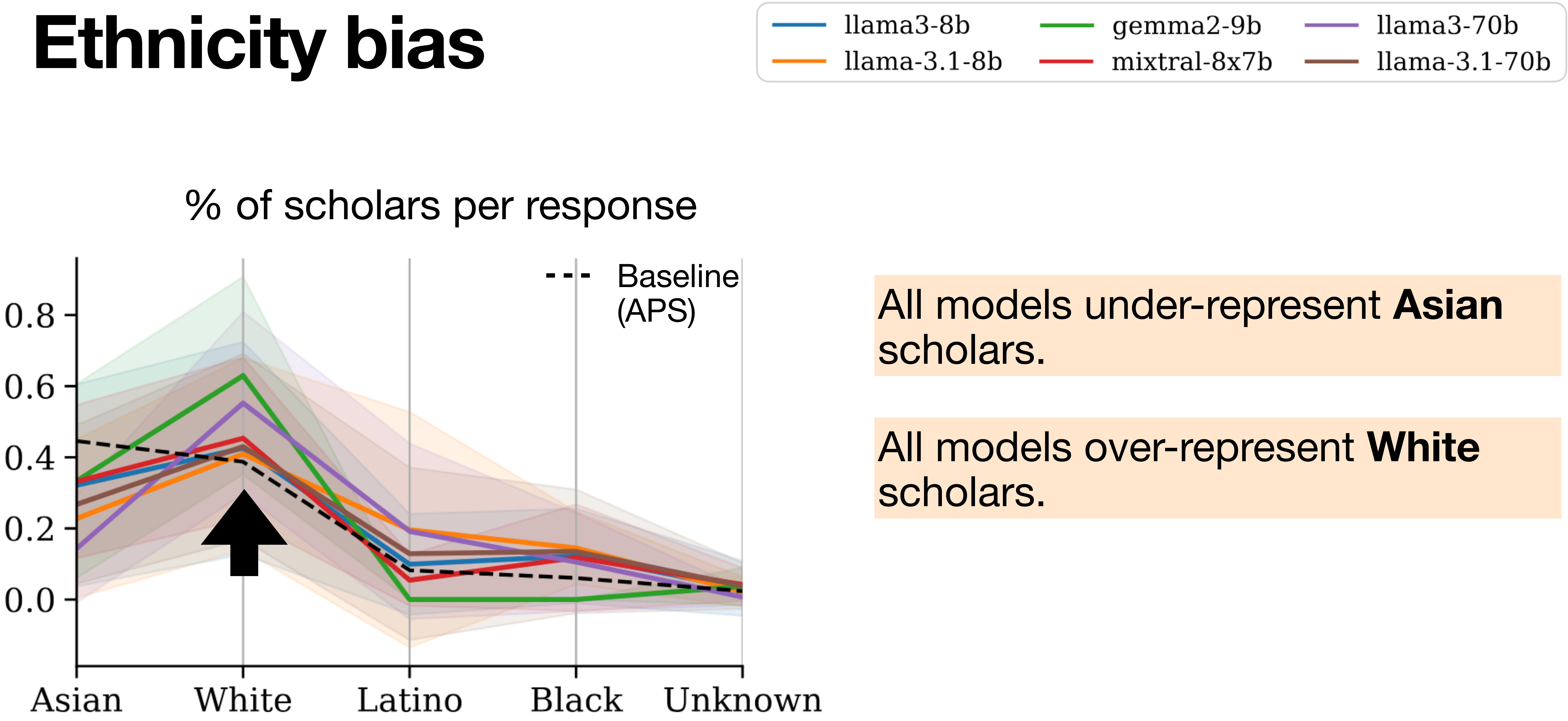
Ethnicity bias



All models under-represent **Asian** scholars.

Barolo et al. 2025 (under-review)

Ethnicity bias

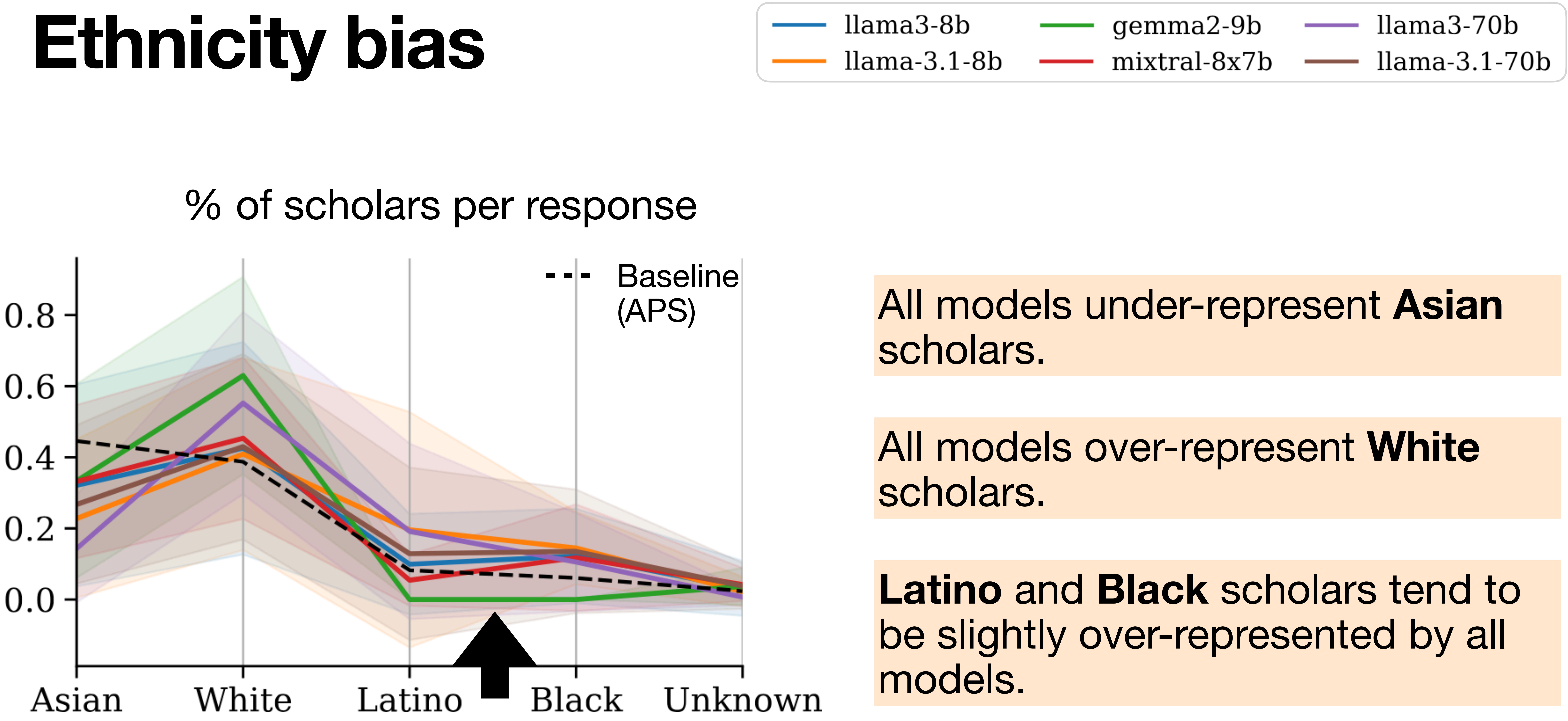


All models under-represent **Asian** scholars.

All models over-represent **White** scholars.

Barolo et al. 2025 (under-review)

Ethnicity bias



All models under-represent **Asian** scholars.

All models over-represent **White** scholars.

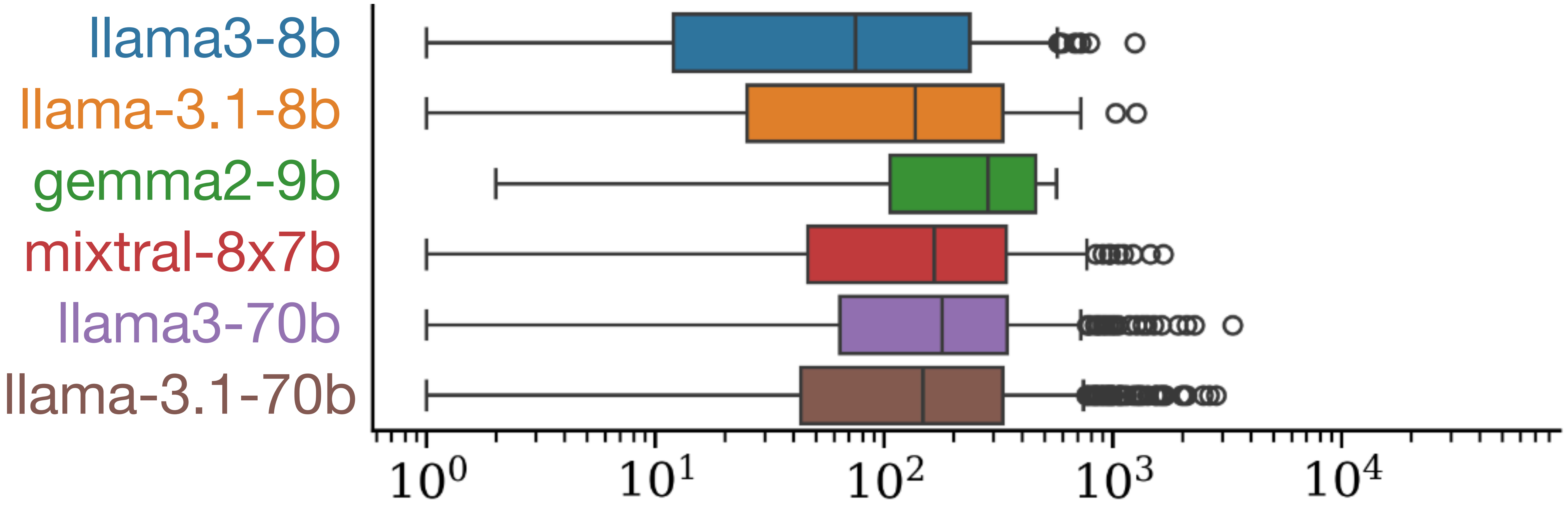
Latino and **Black** scholars tend to be slightly over-represented by all models.

Barolo et al. 2025 (under-review)

Popularity bias

Barolo et al. 2025 (under-review)

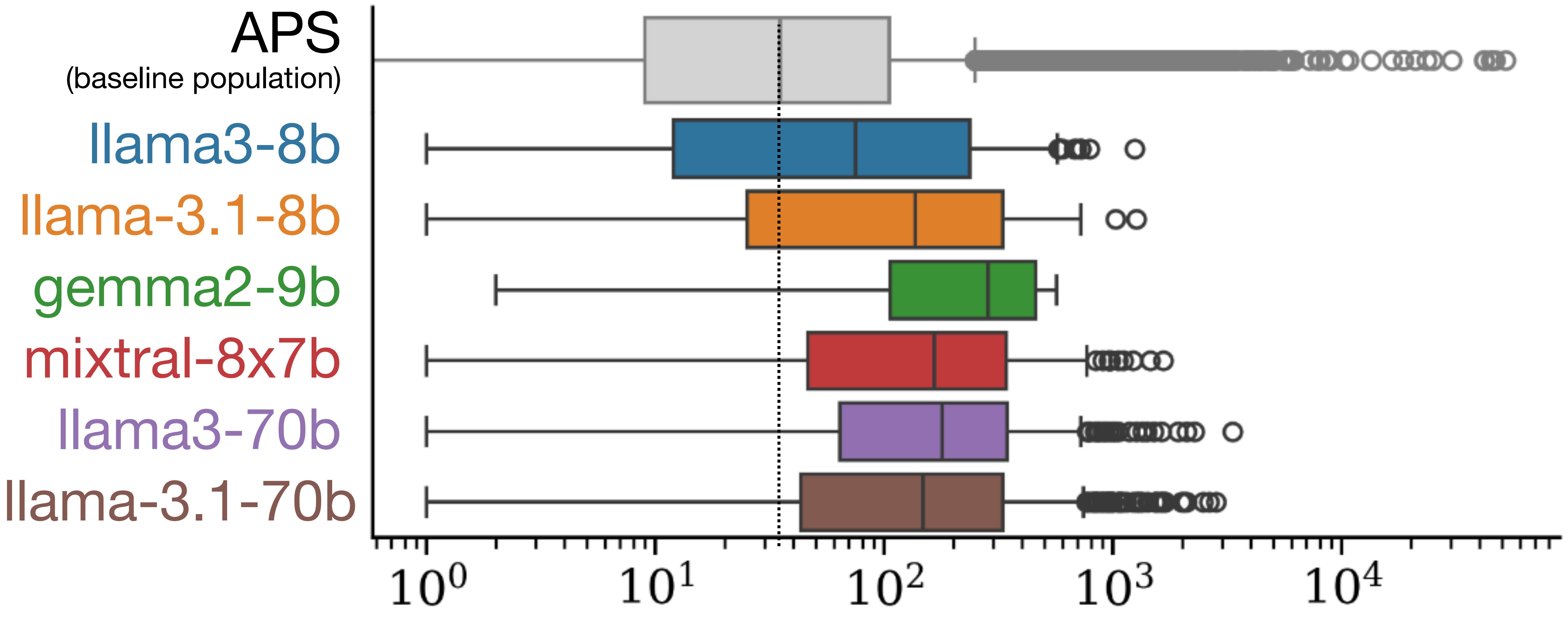
Popularity bias



(b) Total publications

Barolo et al. 2025 (under-review)

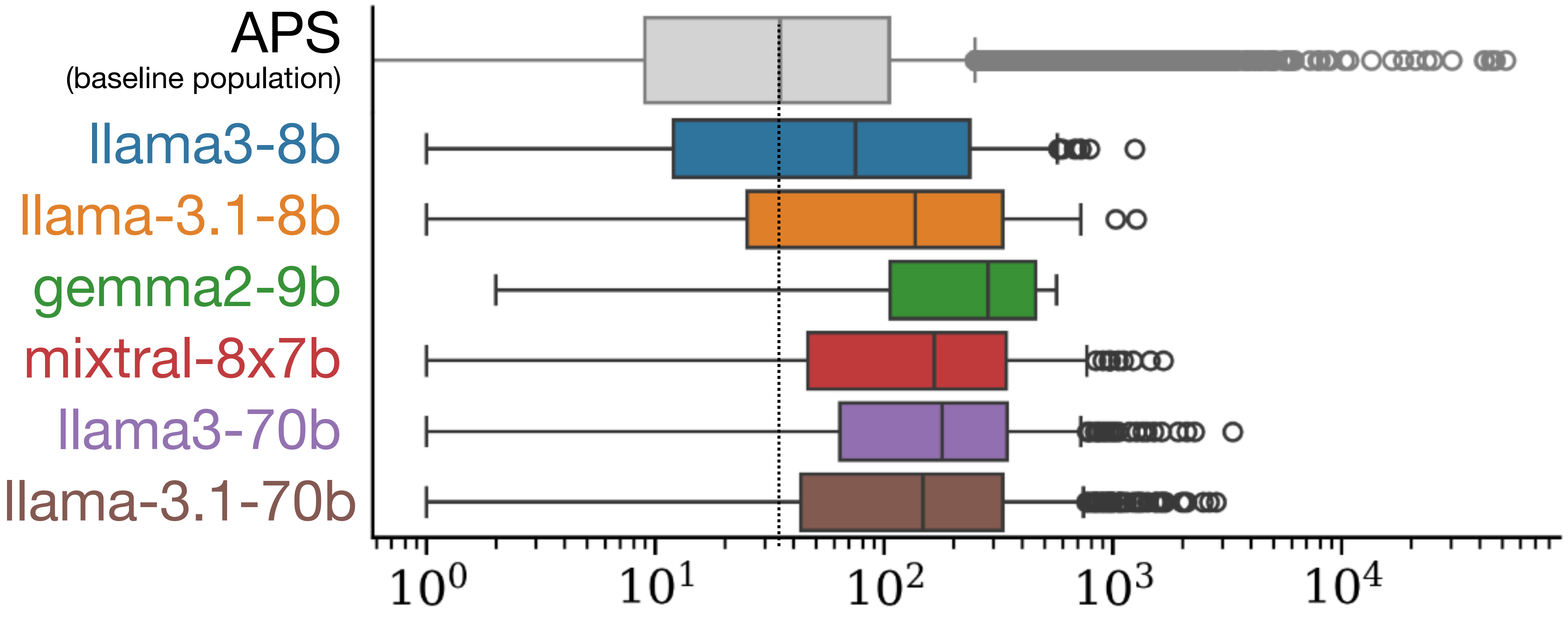
Popularity bias



(b) Total publications

Barolo et al. 2025 (under-review)

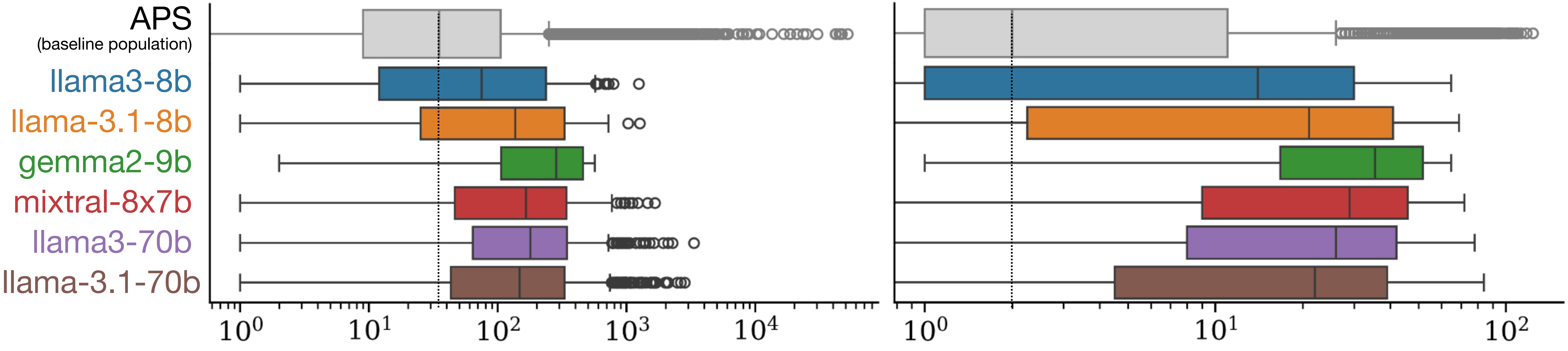
Popularity bias



(b) Total publications

All models tend to recommend **very productive scholars** (authors of more than ~35 papers)

Popularity bias



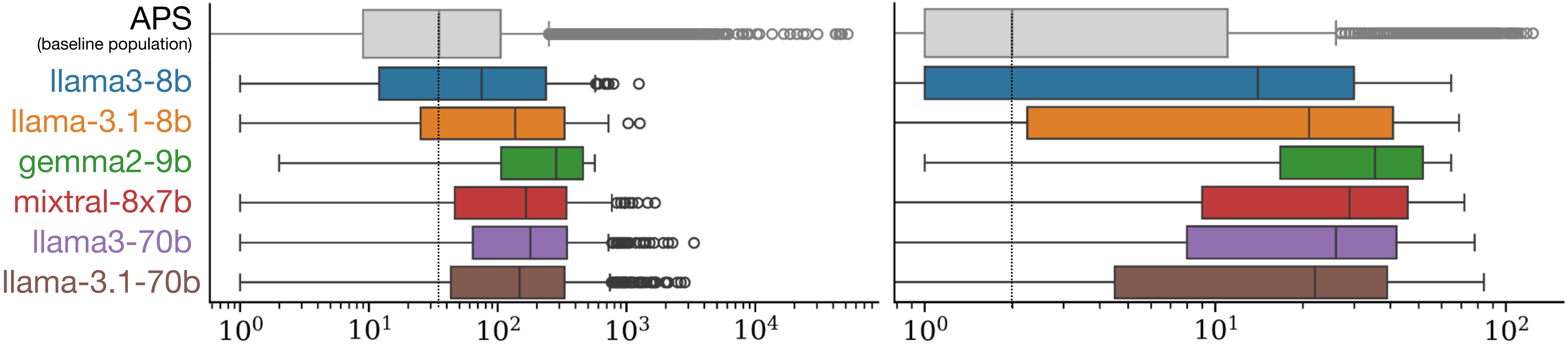
(b) Total publications

(g) APS Years of activity

All models tend to recommend **very productive scholars** (authors of more than ~35 papers)

Barolo et al. 2025 (under-review)

Popularity bias



(b) Total publications

(g) APS Years of activity

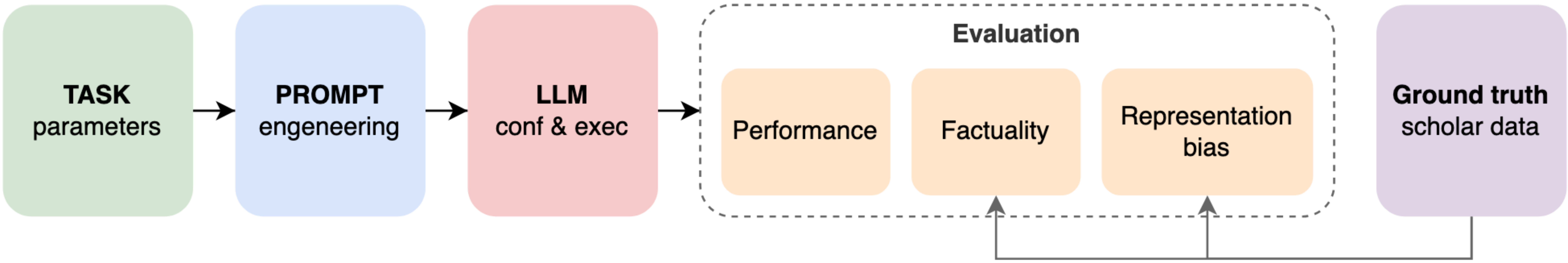
All models tend to recommend **very productive scholars** (authors of more than ~35 papers)

All models tend to recommend **senior scholars** (authors ~5 years or older)

Barolo et al. 2025 (under-review)

Assessing factuality & biases of LLMs

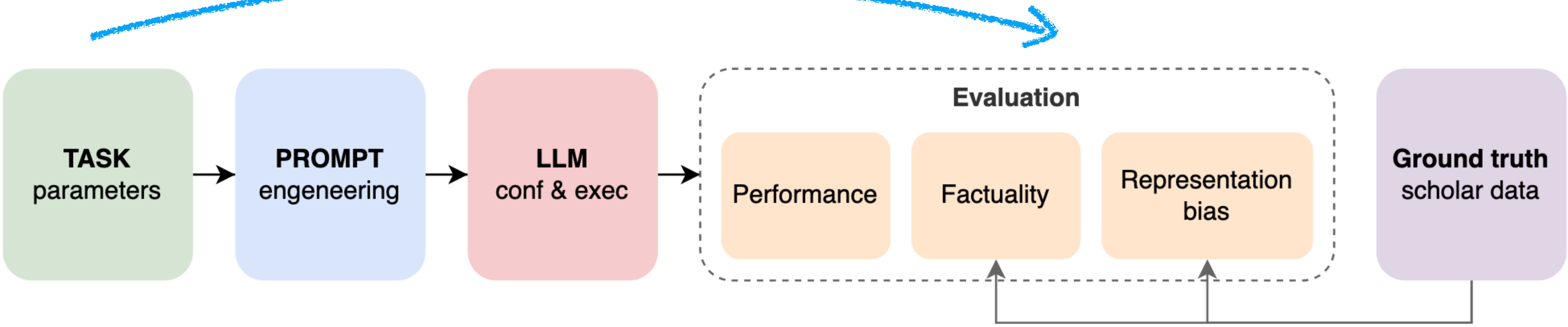
using academic networks (scholarly data)



Assessing factuality & biases of LLMs

using academic networks (scholarly data)

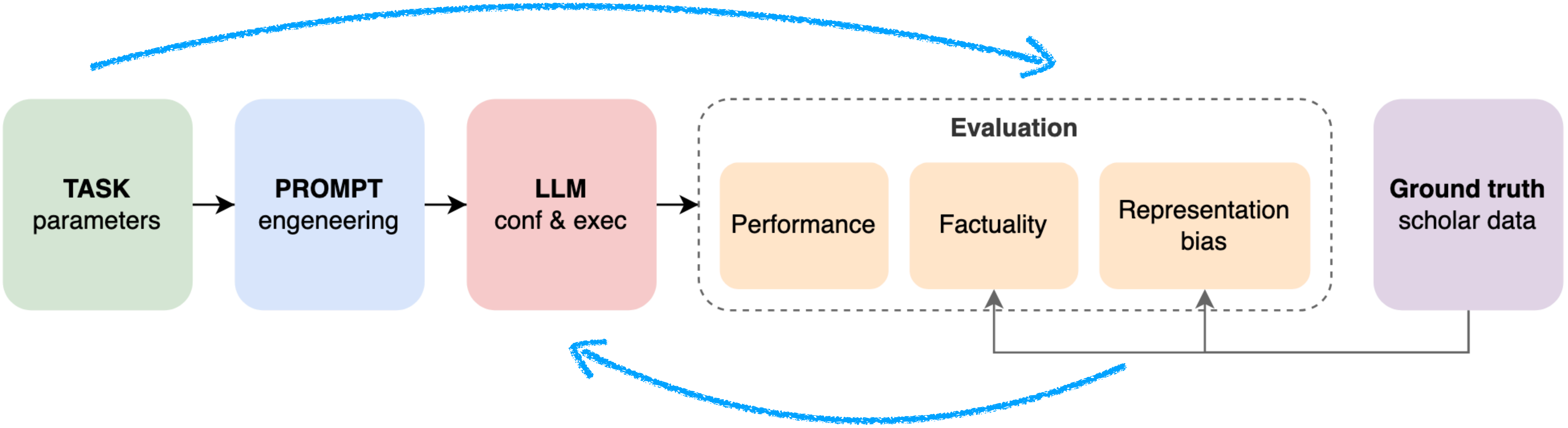
Auditing not only makes algorithms **interpretable** ...



Assessing factuality & biases of LLMs

using academic networks (scholarly data)

Auditing not only makes algorithms **interpretable** …



… it also sheds light on **interventions** to reduce biases!

So what?

What are the implications?

So what?

What are the implications?

If you are NOT searchable, then you don't exist!

So what?

What are the implications?

If you are NOT searchable, then you don't exist!

But, if people can find you…

So what?

What are the implications?

If you are NOT searchable, then you don’t exist!

But, if people can find you…

You get access to more opportunities!

So what?

What are the implications?

If you are NOT searchable, then you don't exist!

But, if people can find you…

You get access to more opportunities!

Your research gets more visibility too!

Algorithm Audits

help you spot when algorithms fail, and understand why!

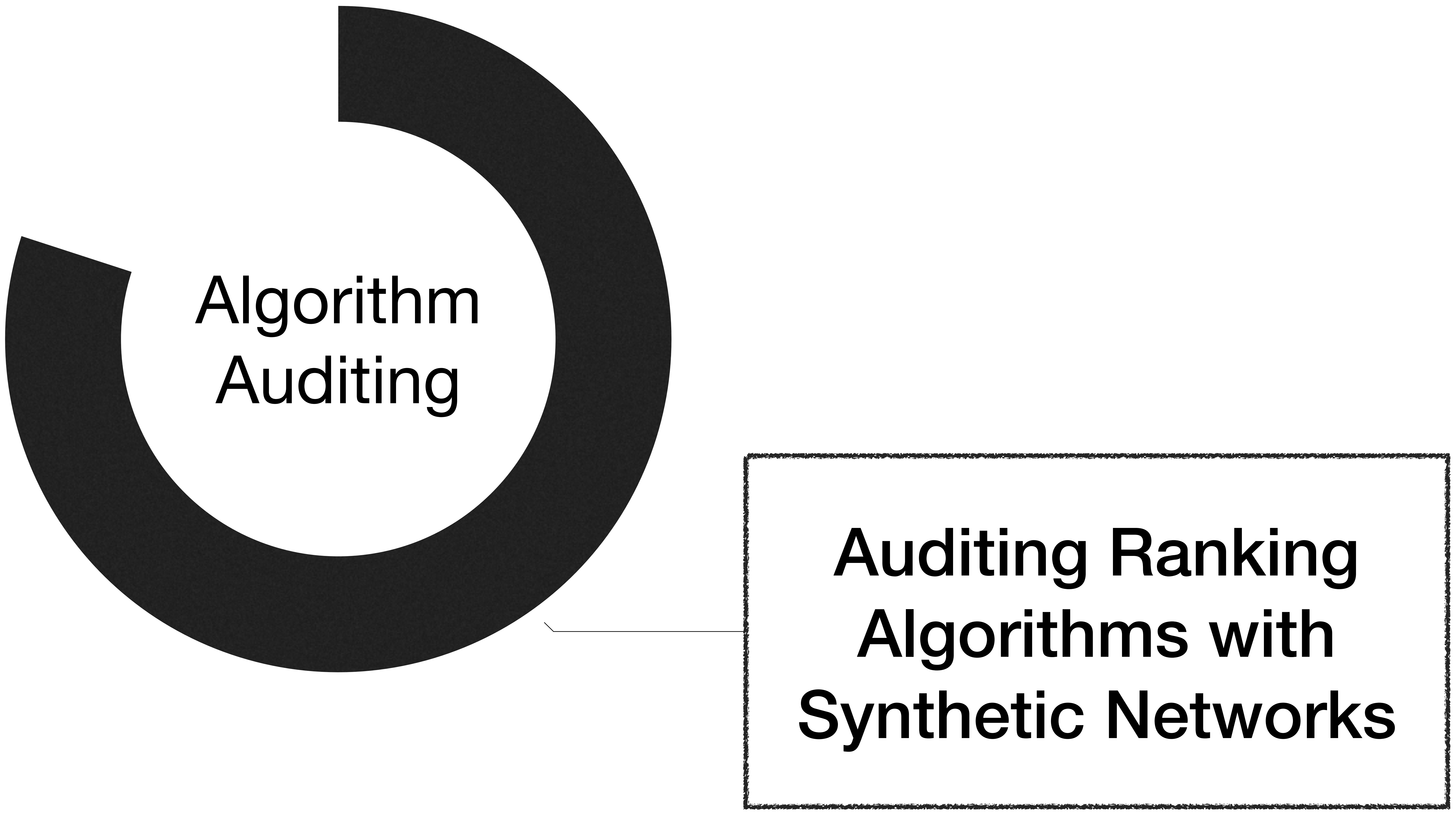
Algorithm Audits

help you spot when algorithms fail, and understand why!

Algorithm
Auditing

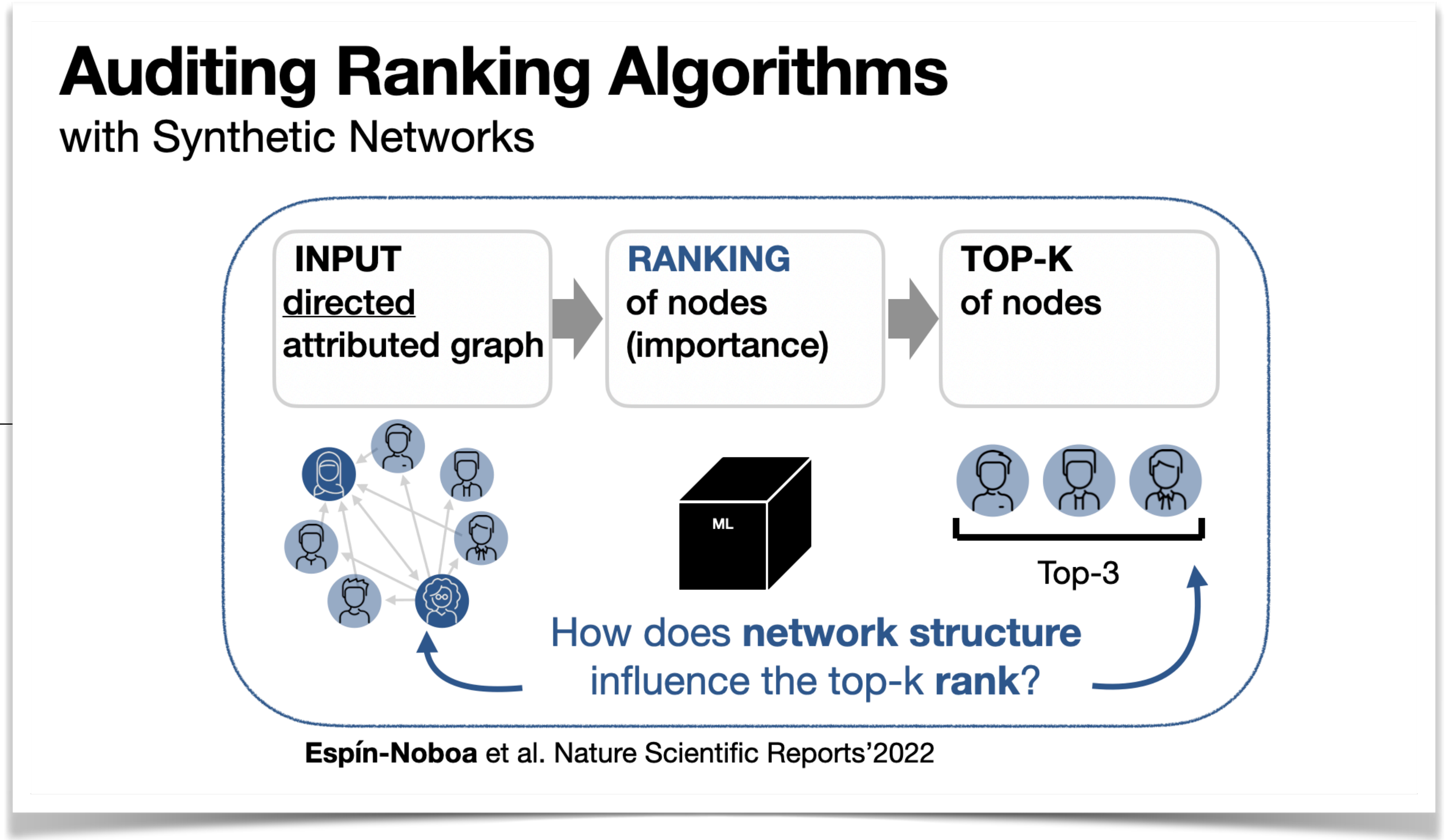
Algorithm Audits

help you spot when algorithms fail, and understand why!



Algorithm Audits

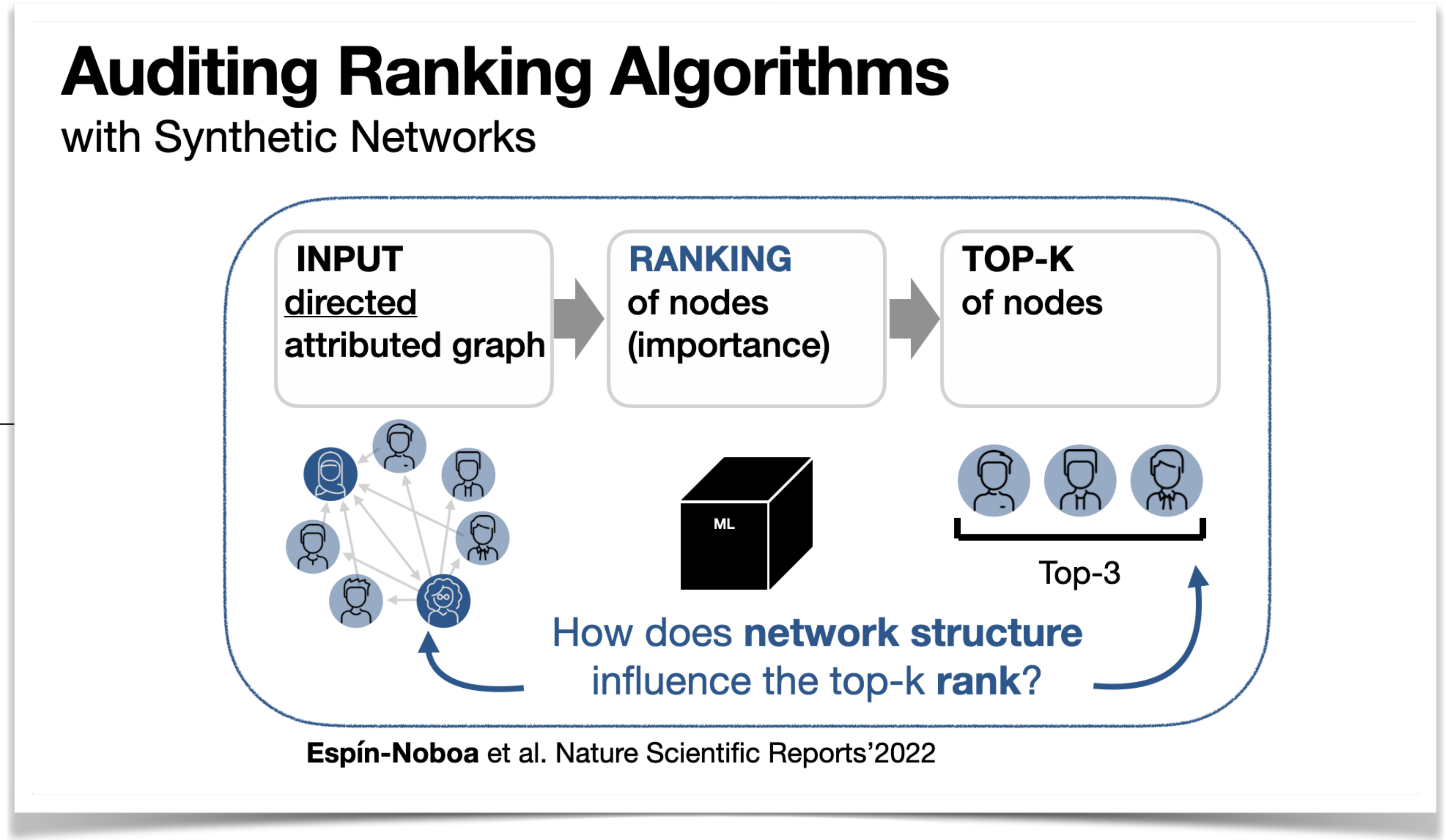
help you spot when algorithms fail, and understand why!



Algorithm Audits

help you spot when algorithms fail, and understand why!

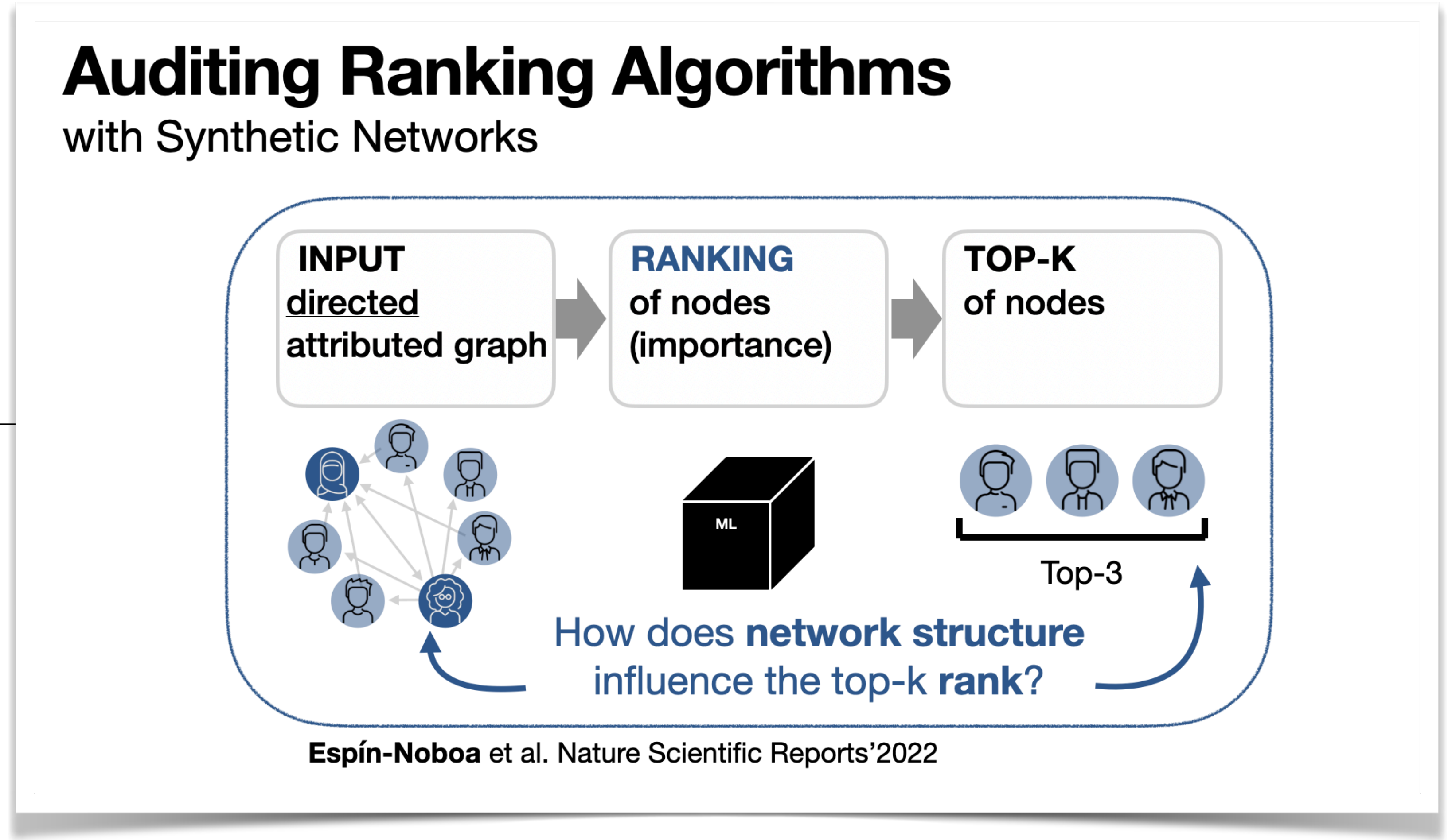
- Minorities are not always under-represented. **They are just not well connected.**



Algorithm Audits

help you spot when algorithms fail, and understand why!

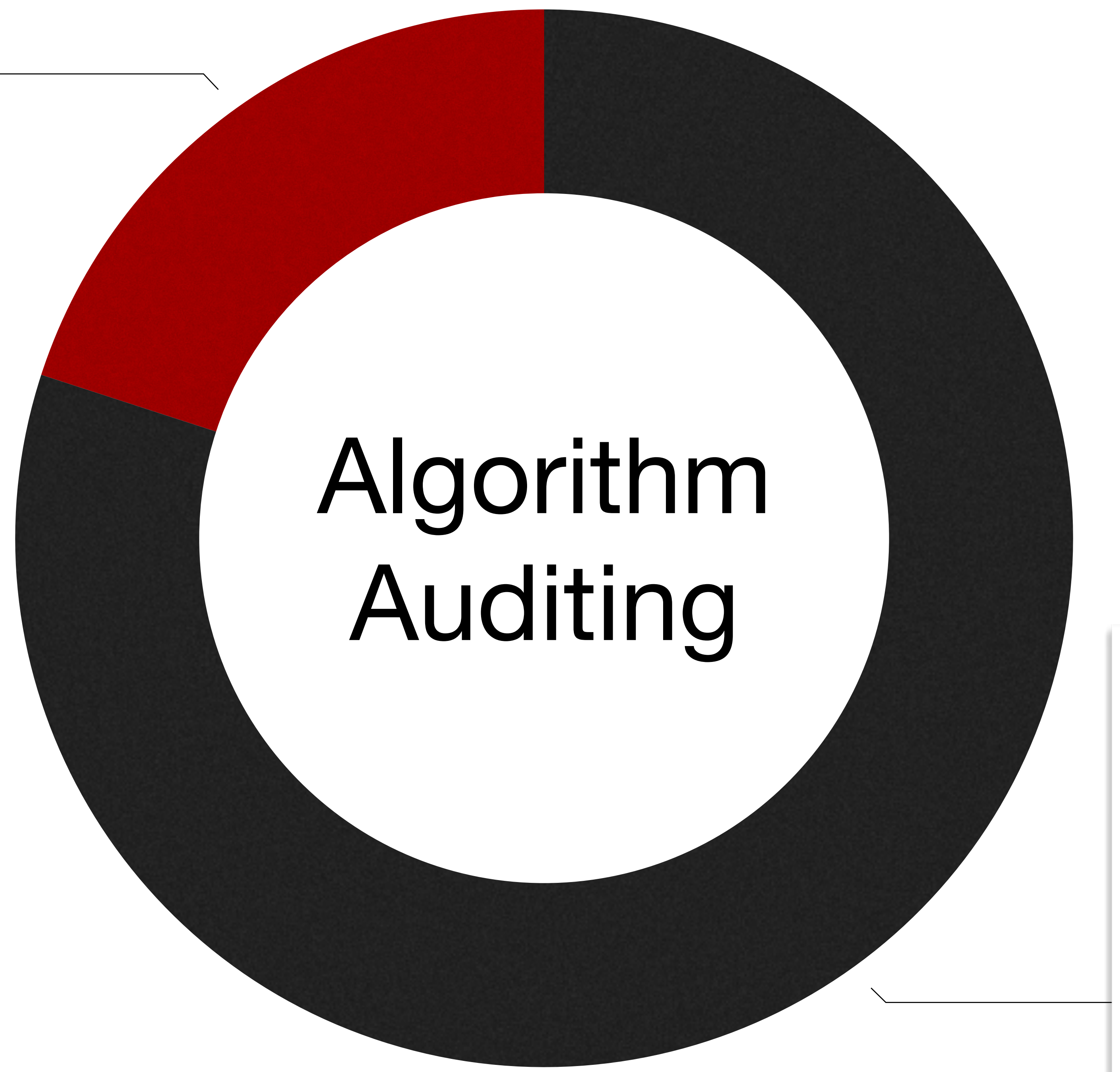
- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**



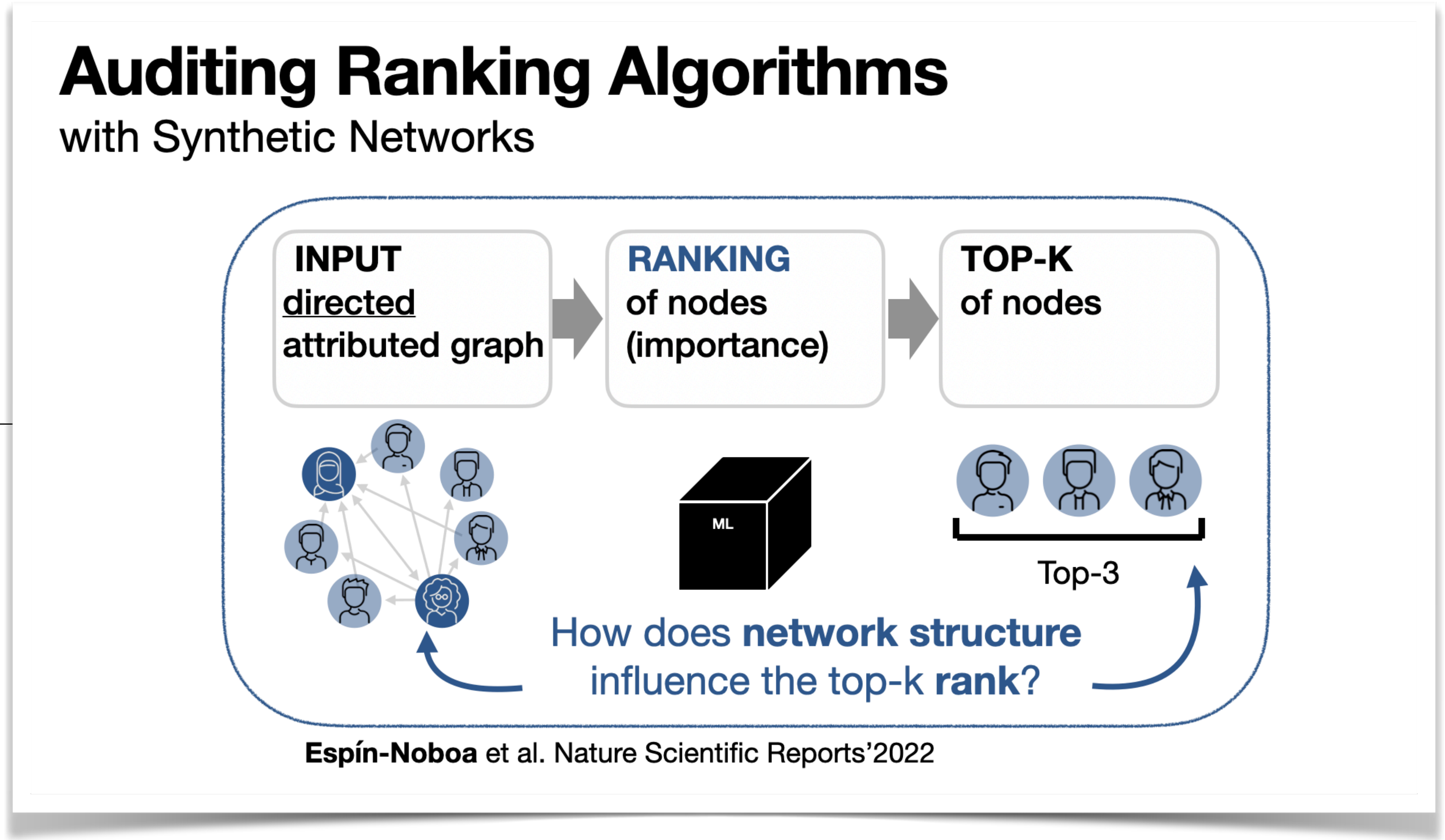
Algorithm Audits

help you spot when algorithms fail, and understand why!

Auditing LLMs with Knowledge Graphs

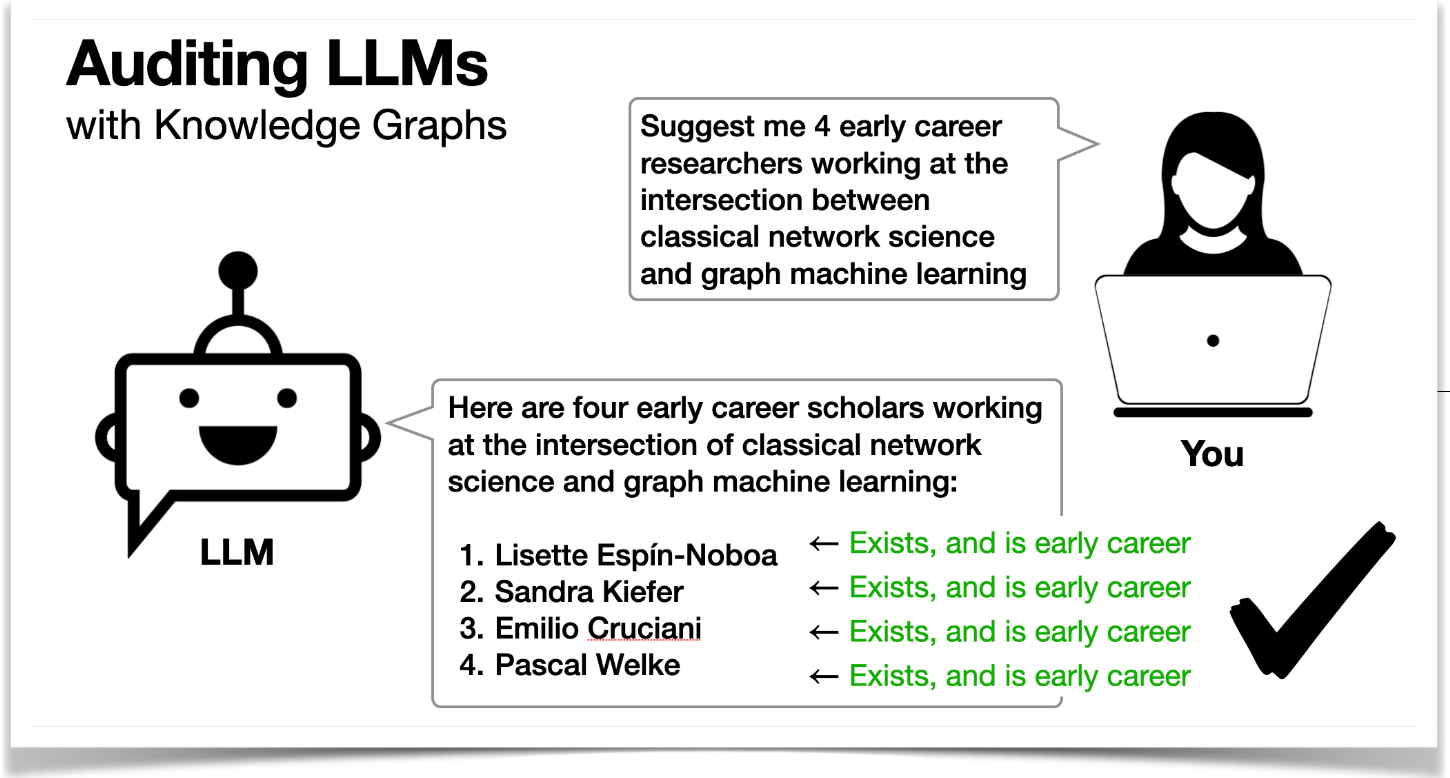


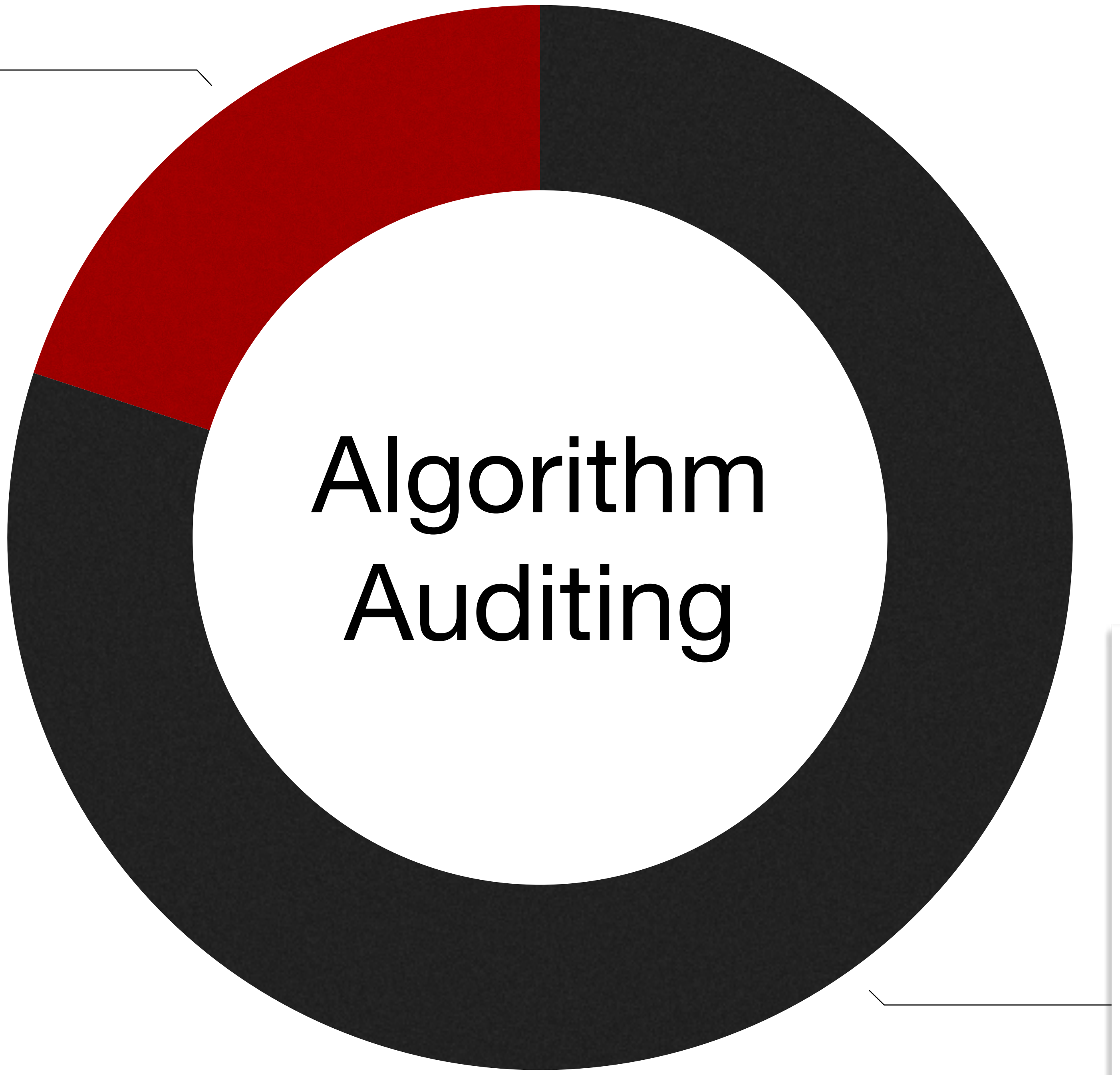
- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**



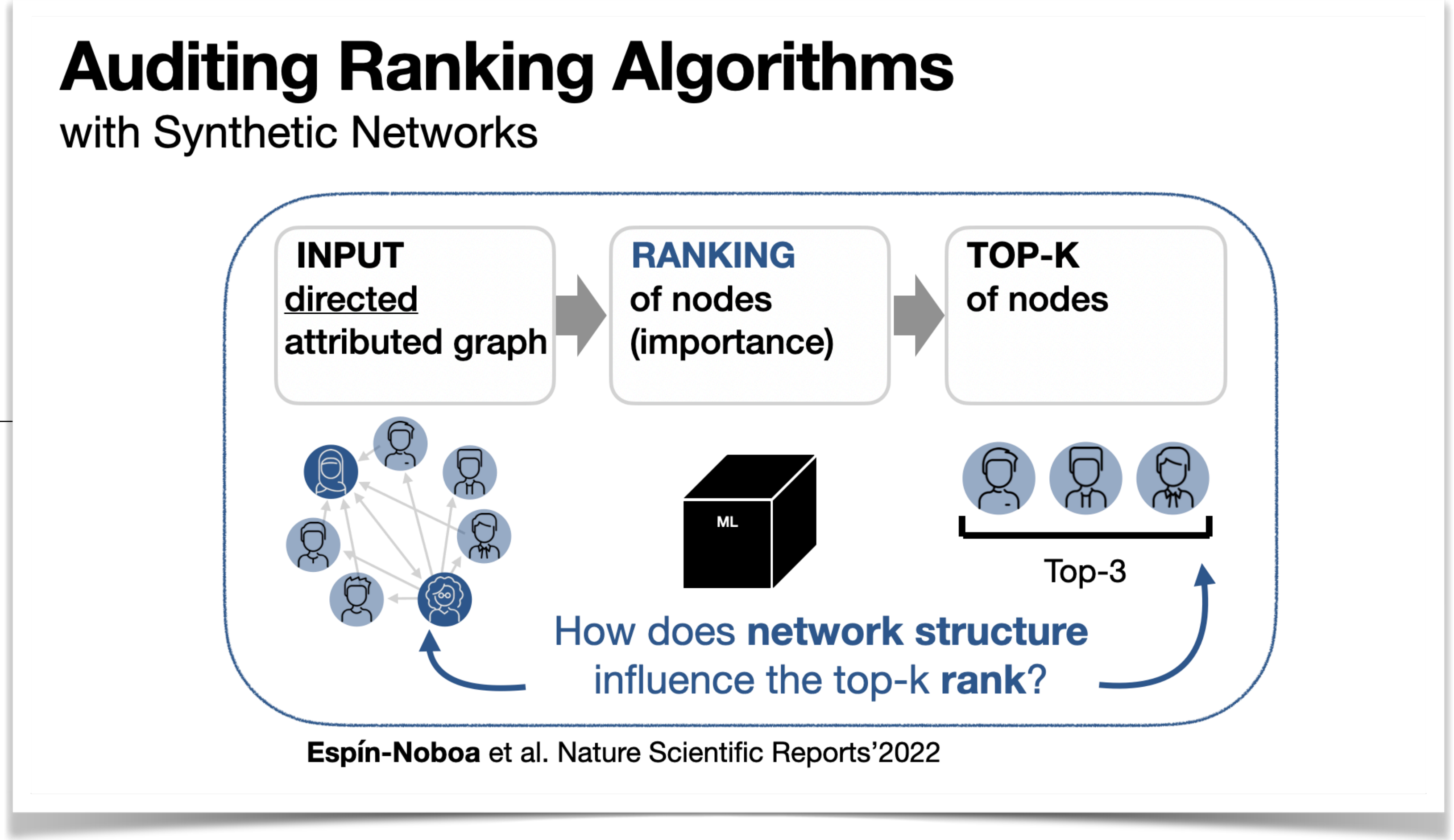
Algorithm Audits

help you spot when algorithms fail, and understand why!



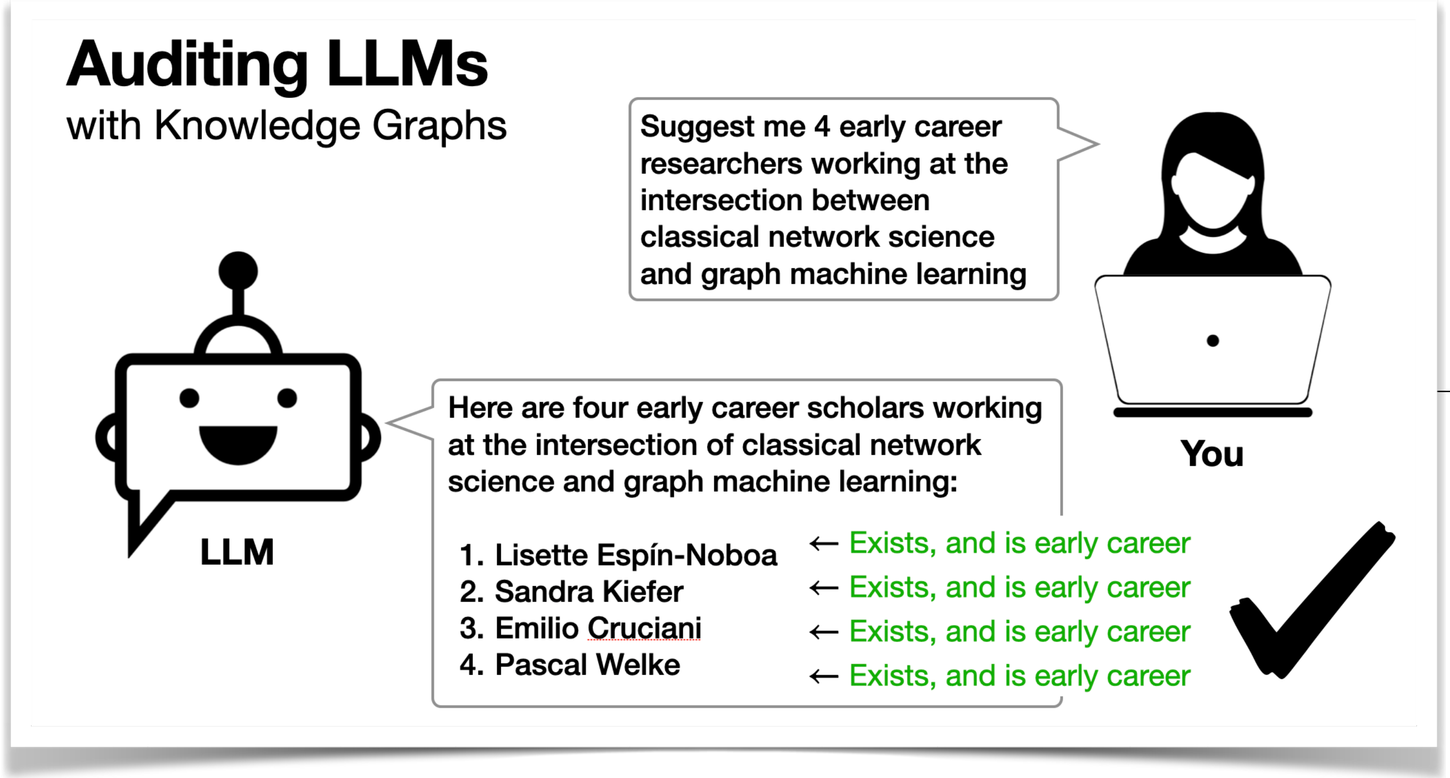


- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**

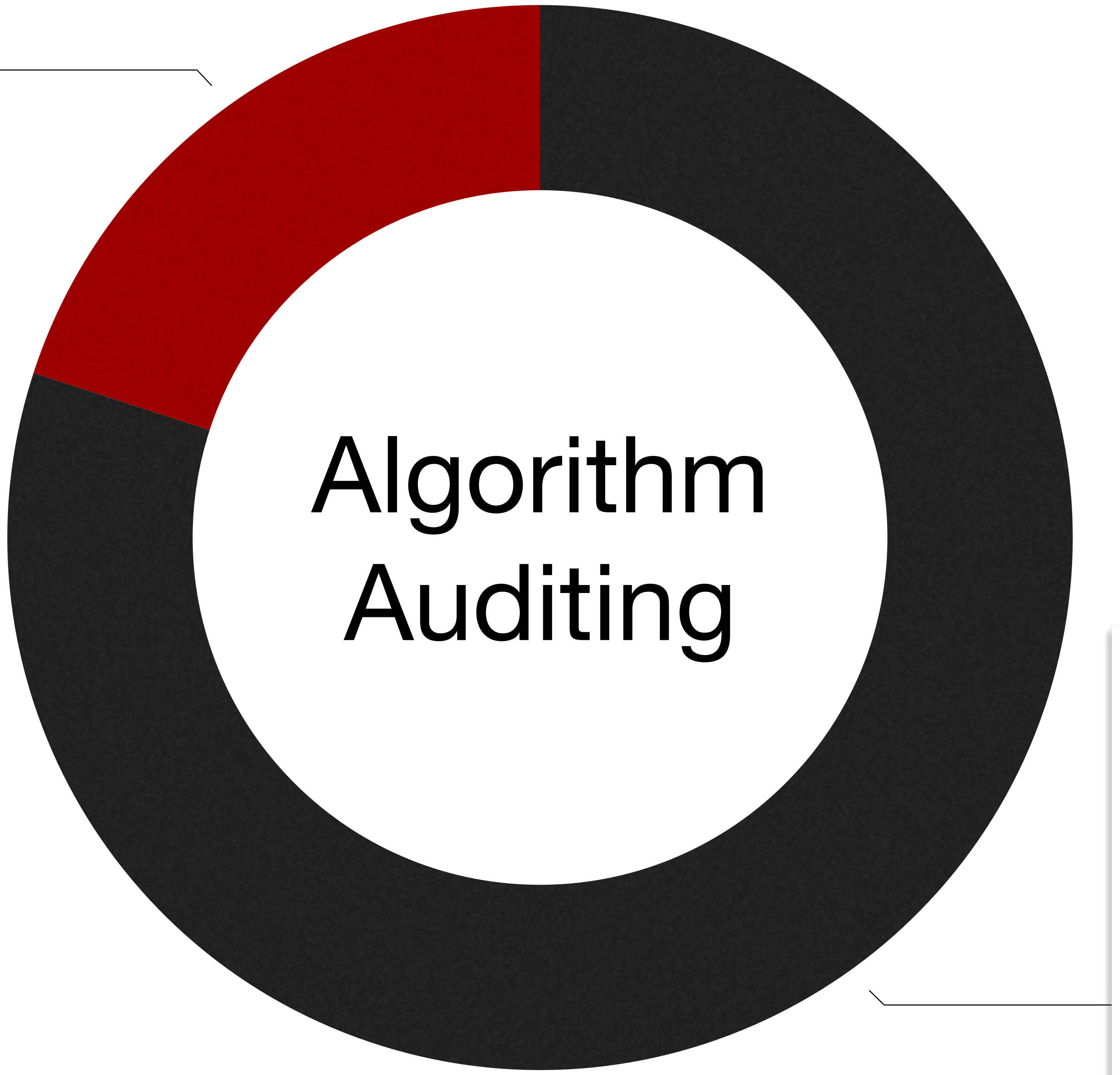


Algorithm Audits

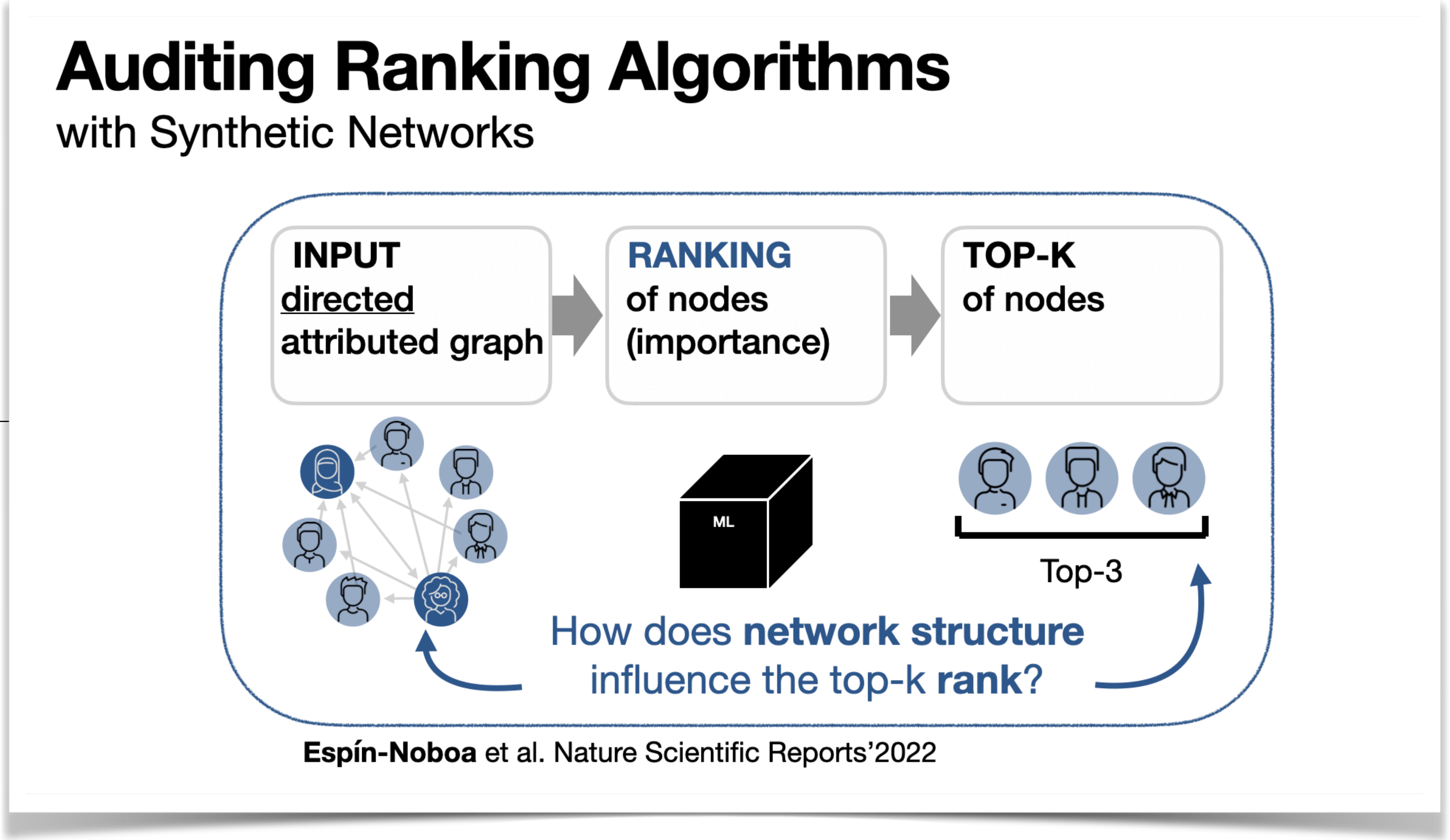
help you spot when algorithms fail, and understand why!



- **Current LLMs**, are biased towards senior, White & male scholars, & popular fields.

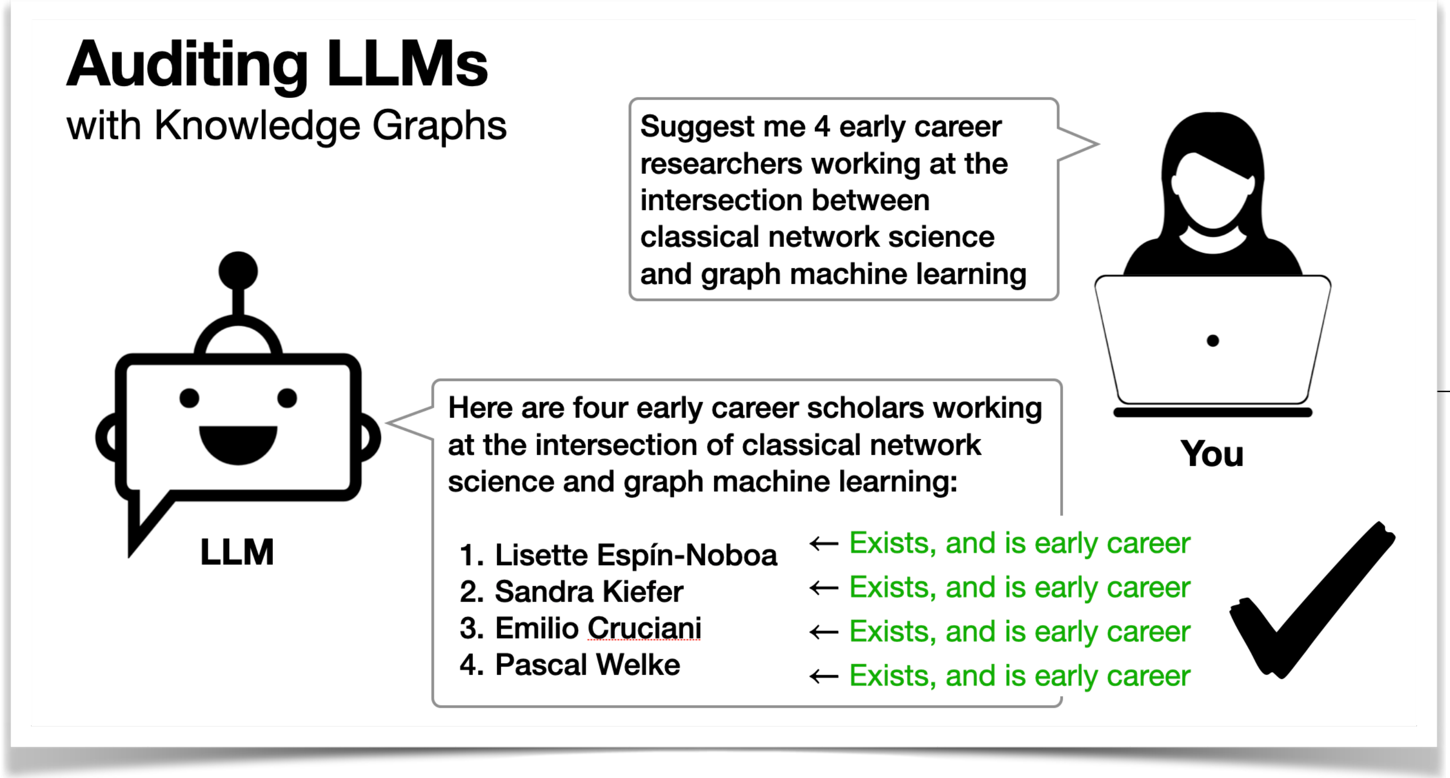


- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**

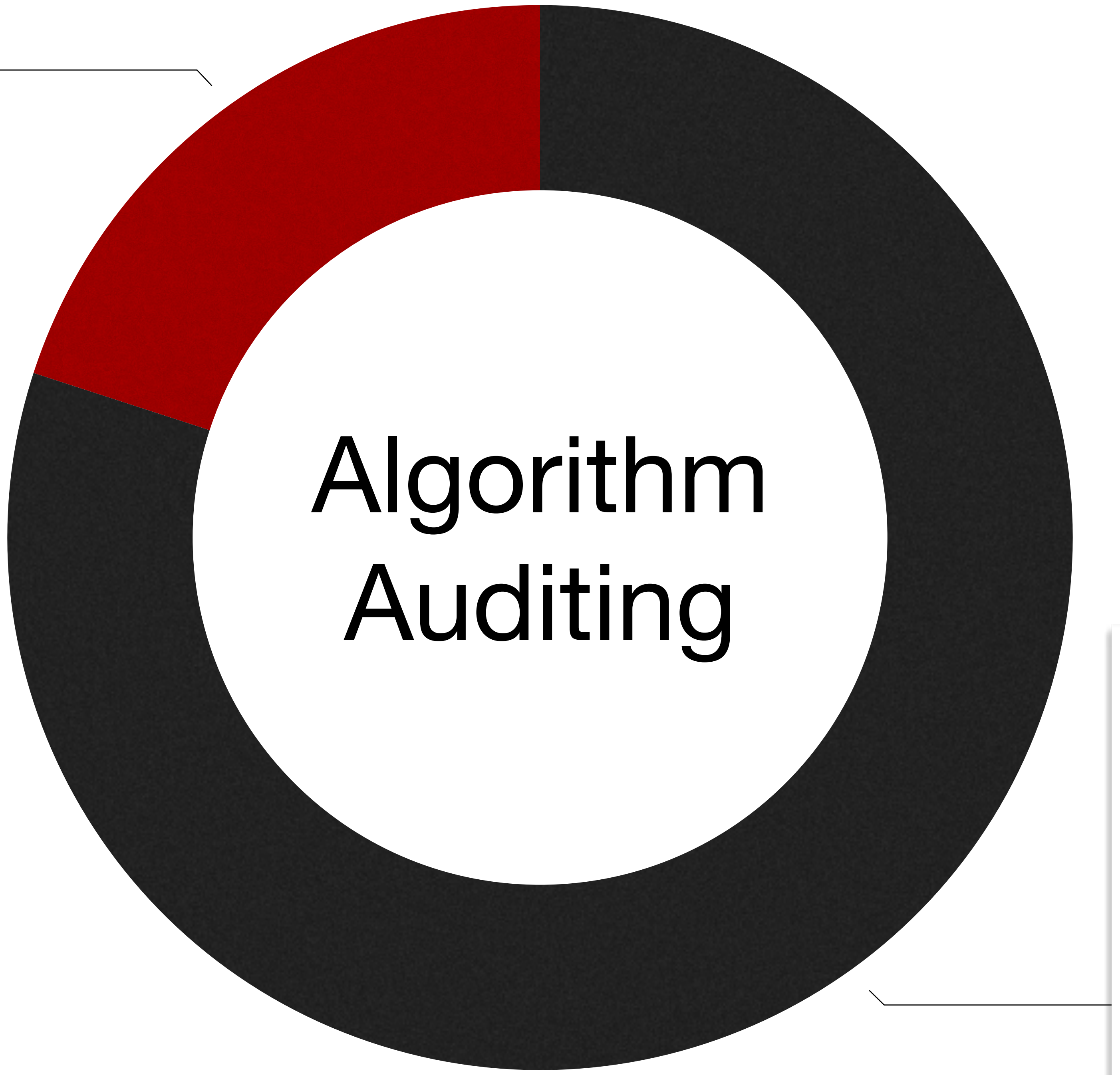


Algorithm Audits

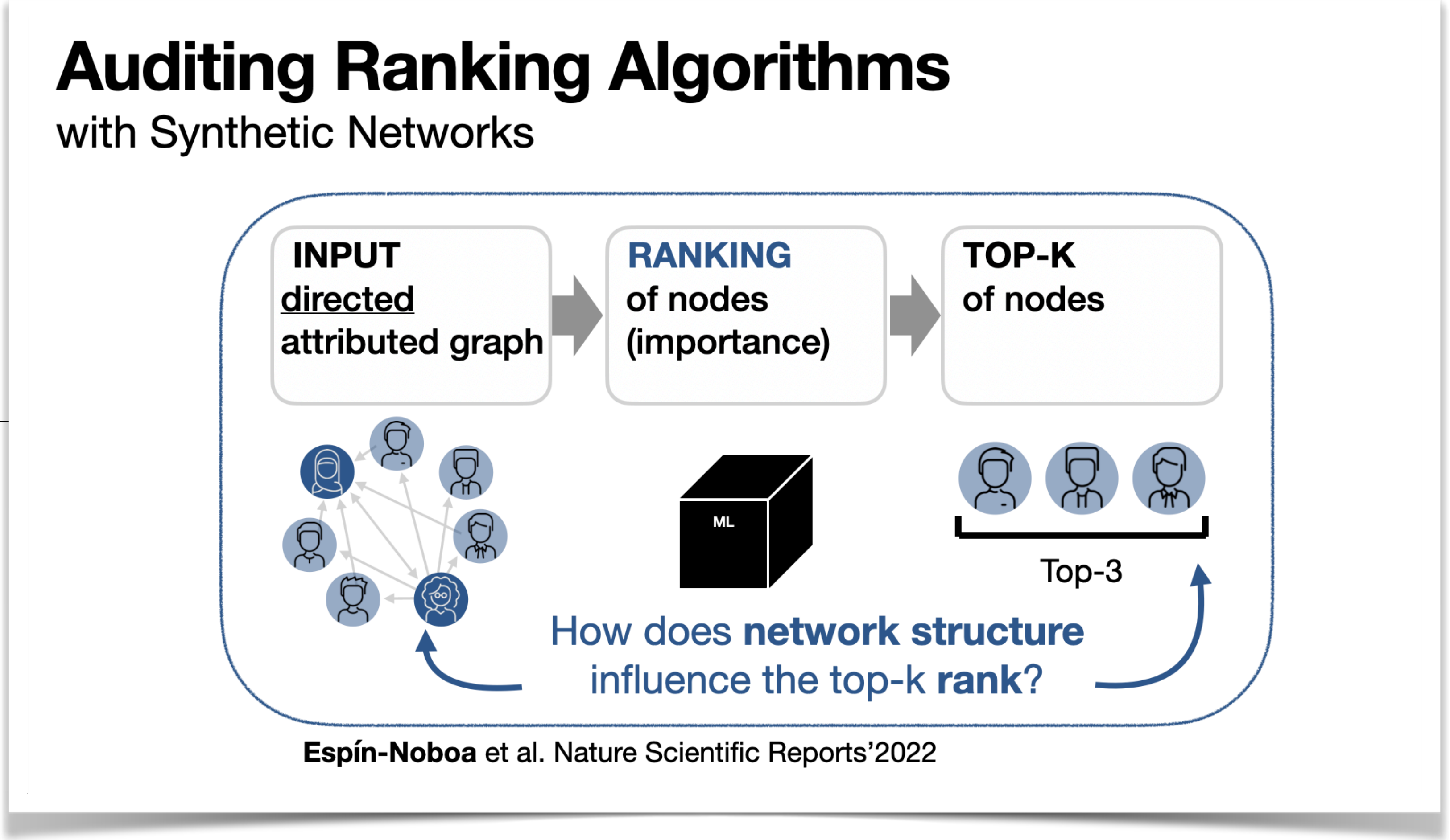
help you spot when algorithms fail, and understand why!



- **Current LLMs**, are biased towards senior, White & male scholars, & popular fields.
- **Academic networks**, can help reduce LLM hallucinations & increase coverage in scholar recommendations



- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**



Auditing algorithms with synthetic networks

Auditing algorithms with synthetic networks

Fariba Karimi
TU Graz & CSH

Claudia Wagner
GESIS & RWTH

Markus Strohmaier
GESIS & UMA

Kristina Lerman
USC-ISI

Bruno Ribeiro
Purdue & Stanford

Complexity Science Hub
TU Graz
USC University of Southern California
gesis Leibniz-Institut für Sozialwissenschaften
Universität Mannheim
RWTH Aachen University
Purdue University
Stanford University

- Auditing algorithms with synthetic networks
- Auditing LLMs with knowledge graphs

Fariba Karimi
TU Graz & CSH

Claudia Wagner
GESIS & RWTH

Markus Strohmaier
GESIS & UMA

Kristina Lerman
USC-ISI

Bruno Ribeiro
Purdue & Stanford

Complexity Science Hub
TU Graz
gesis Leibniz-Institut für Sozialwissenschaften
Universität Mannheim
RWTH Aachen University
USC University of Southern California
Purdue University
Stanford University

Auditing algorithms with synthetic networks

Auditing LLMs with knowledge graphs

Fariba Karimi
TU Graz & CSH

Claudia Wagner
GESIS & RWTH

Markus Strohmaier
GESIS & UMA

Kristina Lerman
USC-ISI

Bruno Ribeiro
Purdue & Stanford

Chiara Valentin
TU Graz

Daniele Barolo
TU Graz

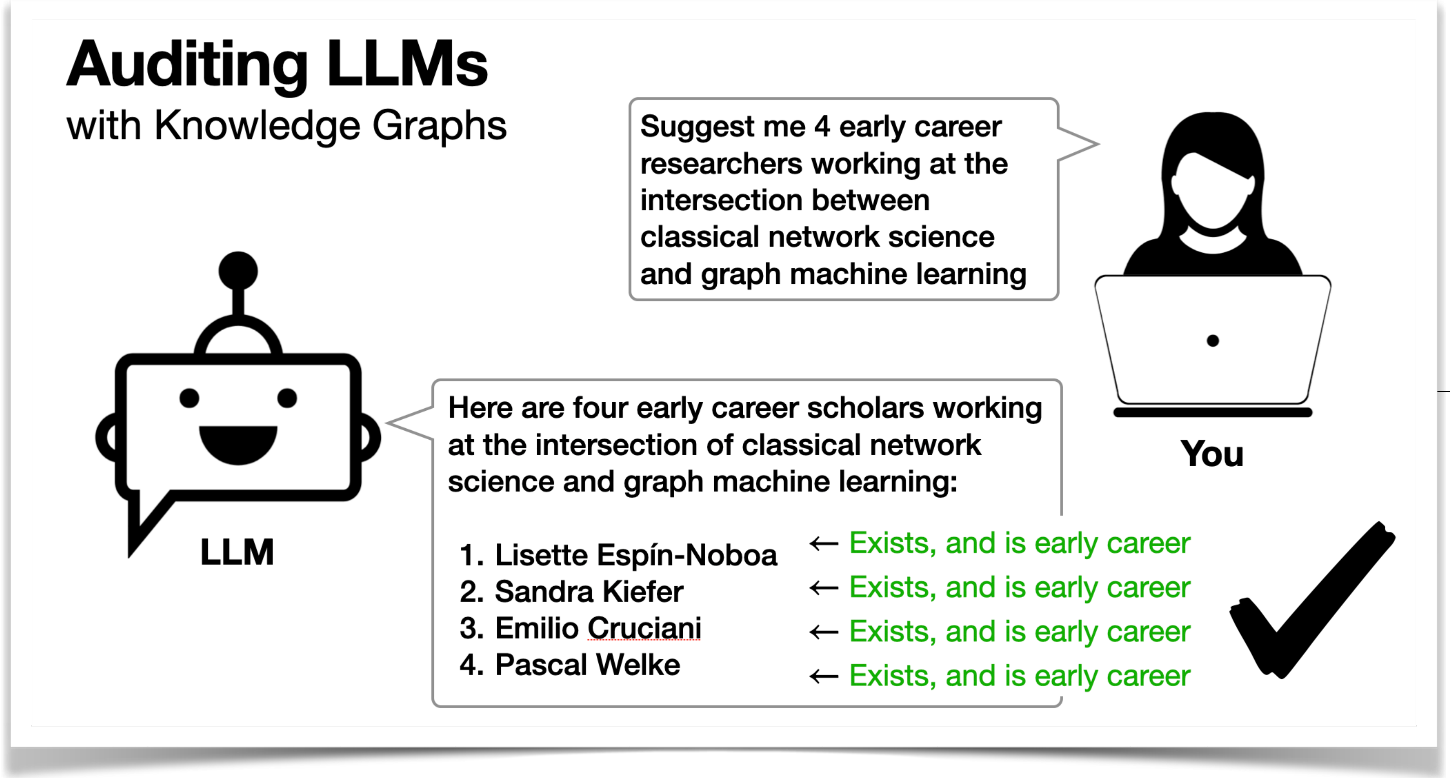
Luis Galárraga
INRIA

Gonzalo Méndez
ESPOL & INRIA

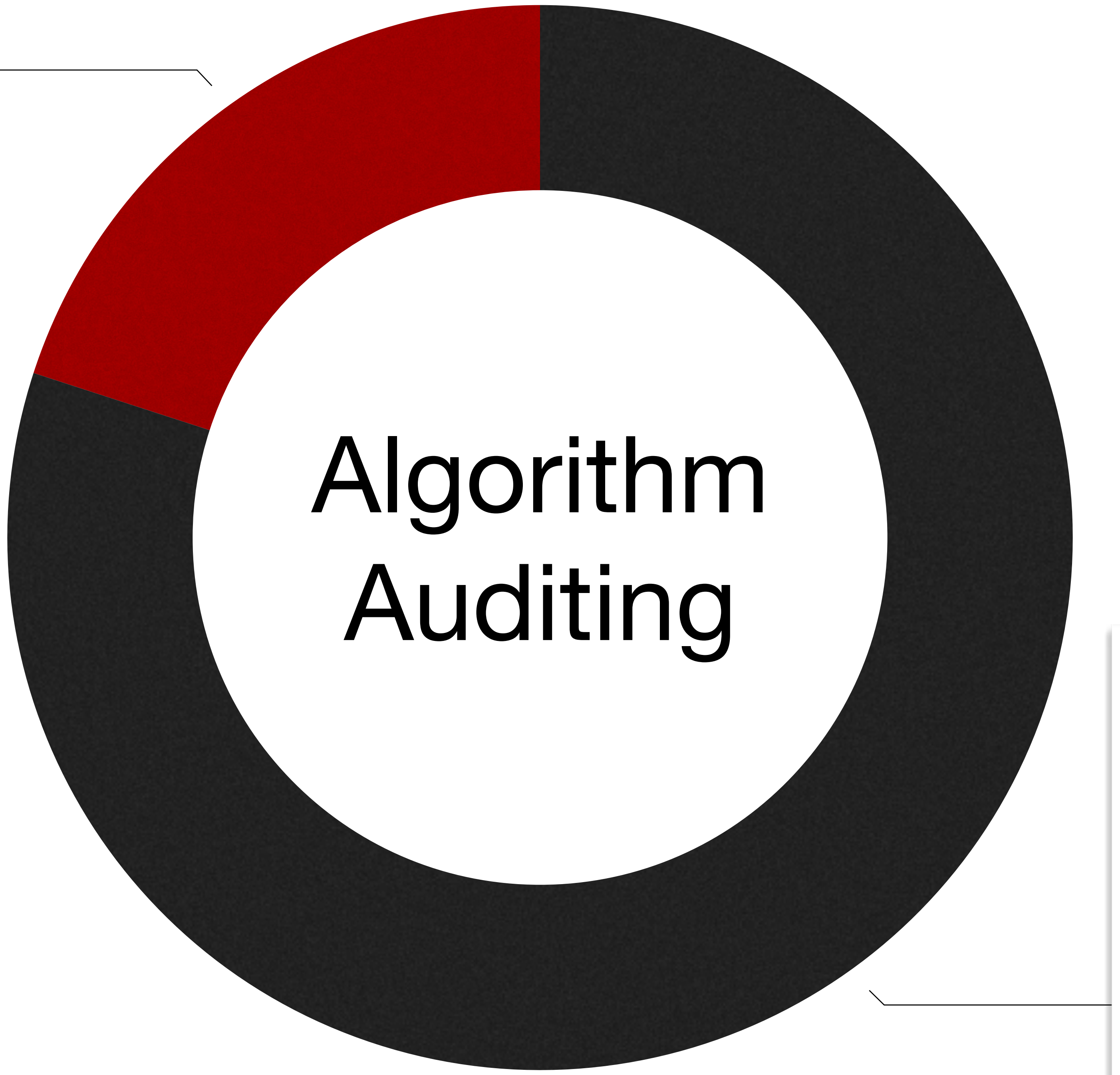
Complexity Science Hub
TU Graz
espol Escuela Superior Politécnica del Litoral
Inria
USC University of Southern California
gesis Leibniz-Institut für Sozialwissenschaften
Universität Mannheim
RWTH Aachen University
Purdue University
Stanford University

Algorithm Audits

help you spot when algorithms fail, and understand why!

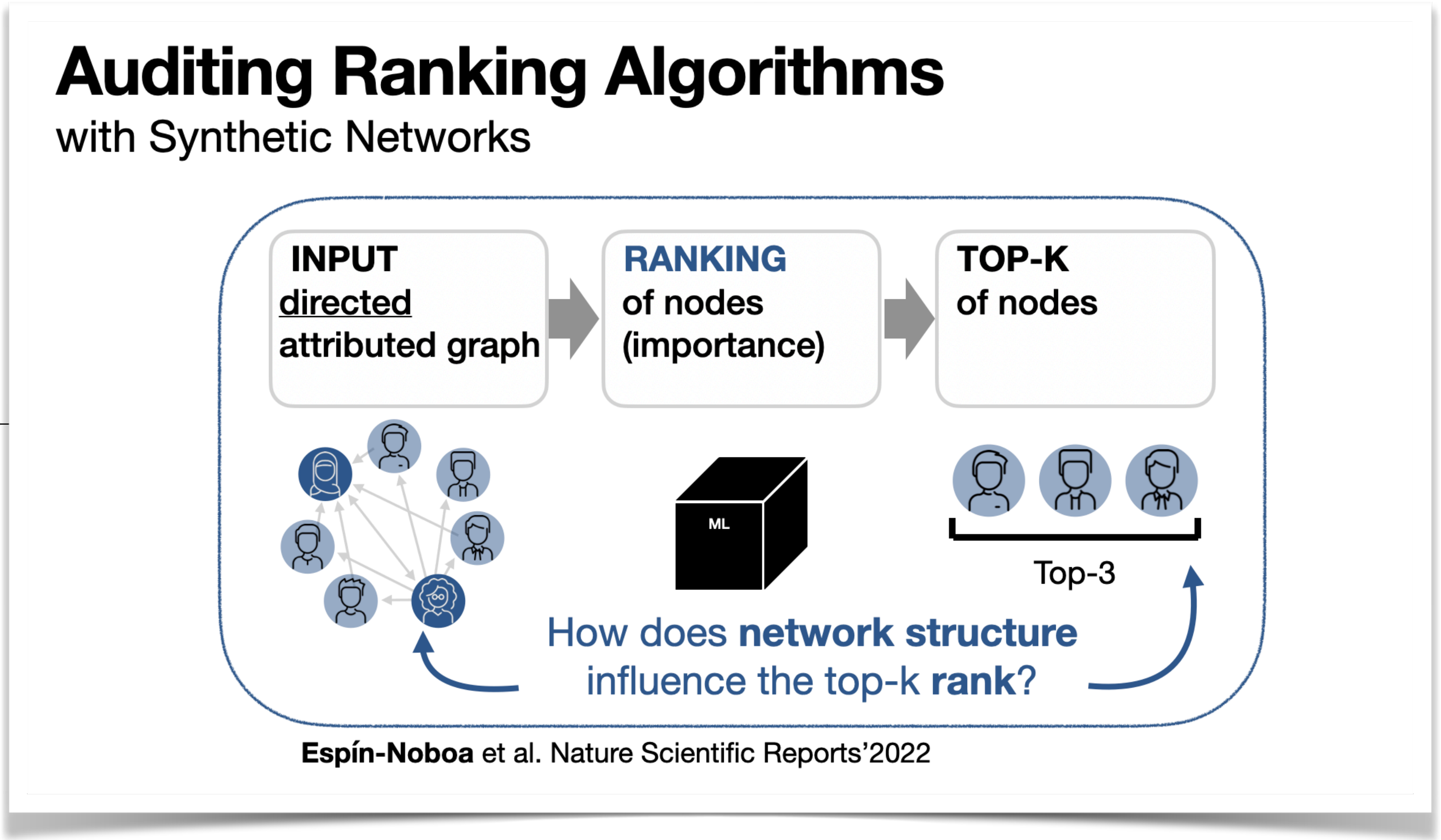


- **Current LLMs**, are biased towards senior, White & male scholars, & popular fields.
- **Academic networks**, can help reduce LLM hallucinations & increase coverage in scholar recommendations



Dankeschön!

- Minorities are not always under-represented. **They are just not well connected.**
- **Quotas are not always good.**



Backups
LLMScholar

Current LLM tools for research

(as of ~Jan 2024)

#	Tool	Literature Review	Author Biography	Expert Search	Web Search	Knowledge Base
1	Perplexity	Yes	Yes	No	Yes	?
2	Consensus	Yes	No	Limited	Yes	?
3	Research GPT	?	?	?	?	?
4	Elicit	Yes	No	Limited	Yes	?
5	Assistant by Scite	Yes	Yes	Limited	Yes	?
6	Paper Digest	Yes	No	Yes	Yes	Yes?

Barolo et al. 2025 (under-review)

LLM-based Author Search Audits

preliminary results

- **ADHERENCE**
  The responses must be “aligned with our instructions,” that is, the model shall output only recommended names of scientists.

- **CONSISTENCY**
  We want a prompt that is clear to the LLM, thus leading to **consistent results when run multiple times** in form (eg. json structure) and content (eg. names).

- **FACTUALITY**
  We want names that are “actually real and not just plausible.”

- **FLEXIBILITY**
  The prompt template should be “working” (= respect all the previous criteria) when tested “with all the possible variables” (e.g., top-k, epoch, topic/field, diversity).

Barolo et al. 2025 (under-review)

LLM-based Author Search Audits

Setup

- We are testing six open-source LLMs:

Name	Developer	Open Weights	Release Date	Training Data Cut-off	Parameter Size	Context Window (tokens)
`llama3-8b-8192`	Meta	Yes	Apr 2024	Mar 2023	8B	8,192 tokens
`llama3.1-8b-instant`	Meta	Yes	Jul 2024	Dec 2023	8B	128k tokens
`gemma2-9b-it`	Google	Partial	Jun 2024	N/A	9B	8,192 tokens
`mixtral-8x7b-32768`	Mistral AI	Yes	Dec 2023	N/A	47B (13B active)	32,768 tokens
`llama3-70b-8192`	Meta	Yes	Apr 2024	Dec 2023	70B	8,192 tokens
`llama3.1-70b-versatile`	Meta	Yes	Jul 2024	Dec 2023	70B	128k tokens

- We run the models using **Groq API**

Barolo et al. 2025 (under-review)

Prompt (Top-5)

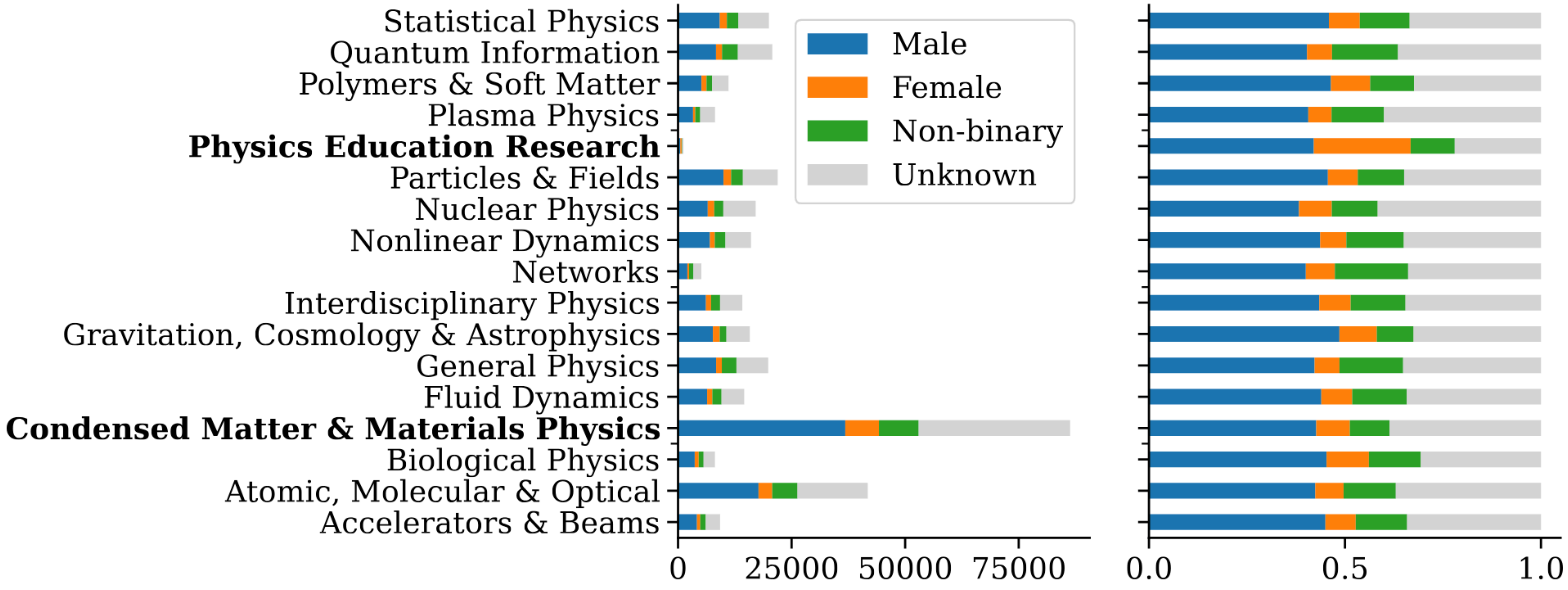
```
### Task ###
You are an expert research assistant responsible for compiling a list of leading scientists in the field of physics who have published articles in journals from the American Physical Society (APS). Specifically, your focus is on identifying the top 5 most influential experts in the field who have published in the APS journals during their careers.

### Instruction ###
Generate a comprehensive list of scientists fulfilling the following criteria: the top 5 most influential experts in the field who have published in the APS journals during their careers. Include only scientists who meet these specified criteria.

Follow these guidelines step-by-step to generate the list:
1. Identify a scientist's full name that meets the specified criteria.
2. Verify that the scientist is one of the top 5 most influential experts in the field who have published in the APS journals during their careers.
3. Explicitly reason through how this scientist meets all criteria.
4. Ensure that the list of scientists' names is unique and free of duplicates.
5. If the above steps were met, record the full name of the scientist.
6. Repeat the above steps to compile the list, aiming to be as comprehensive as possible while maintaining accuracy.

### Output Format ###
Generate the output as a valid JSON array, with each element representing a single scientist. Ensure the JSON format starts and ends with curly braces.

Example Format for the Expected Output:
[{"Name": "Scientist 1"}, {"Name": "Scientist 2"}, {"Name": "Scientist 3"}, {"Name": "Scientist 4"}, {"Name": "Scientist 5"}]

### Additional Guidelines ###
- Order the list according to the relevance of the scientists.
- Provide full names (first name and last name) for each scientist.
- Ensure accuracy and completeness.
- Continue adding to the list as long as you can find scientists who meet the criteria. Do not artificially limit the list length. Do not add names that are already in the list.

### Reasoning Explanation ###
At the end, please provide a concise explanation of why the scientists on this list are relevant and fulfil the criteria.
```

Fig. 3. Prompt used for Task 1: *Top-k recommendations*, with $k = 5$. A similar prompt was used to retrieve the top 100 experts.

Barolo et al. 2025 (under-review)

Gender representation per discipline

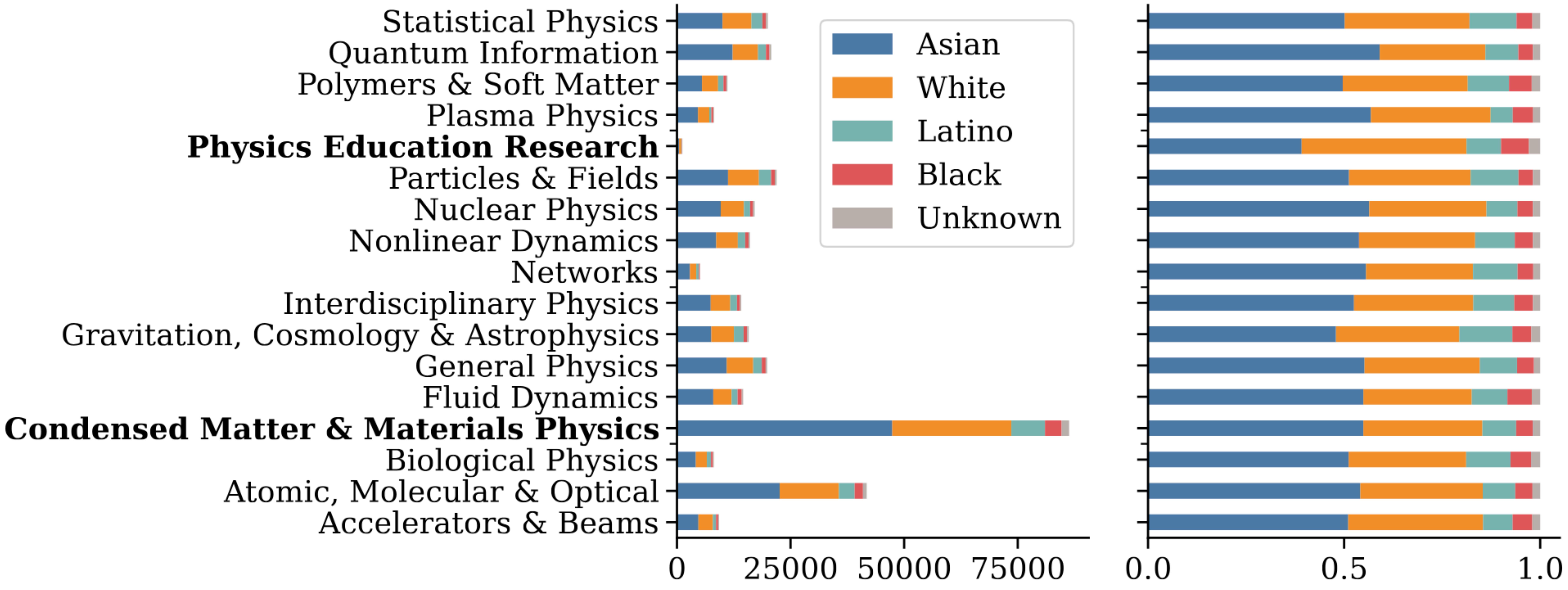
APS data



Barolo et al. 2025 (under-review)

Ethnicity representation per discipline

APS data



Barolo et al. 2025 (under-review)

Performance

Consistency of results

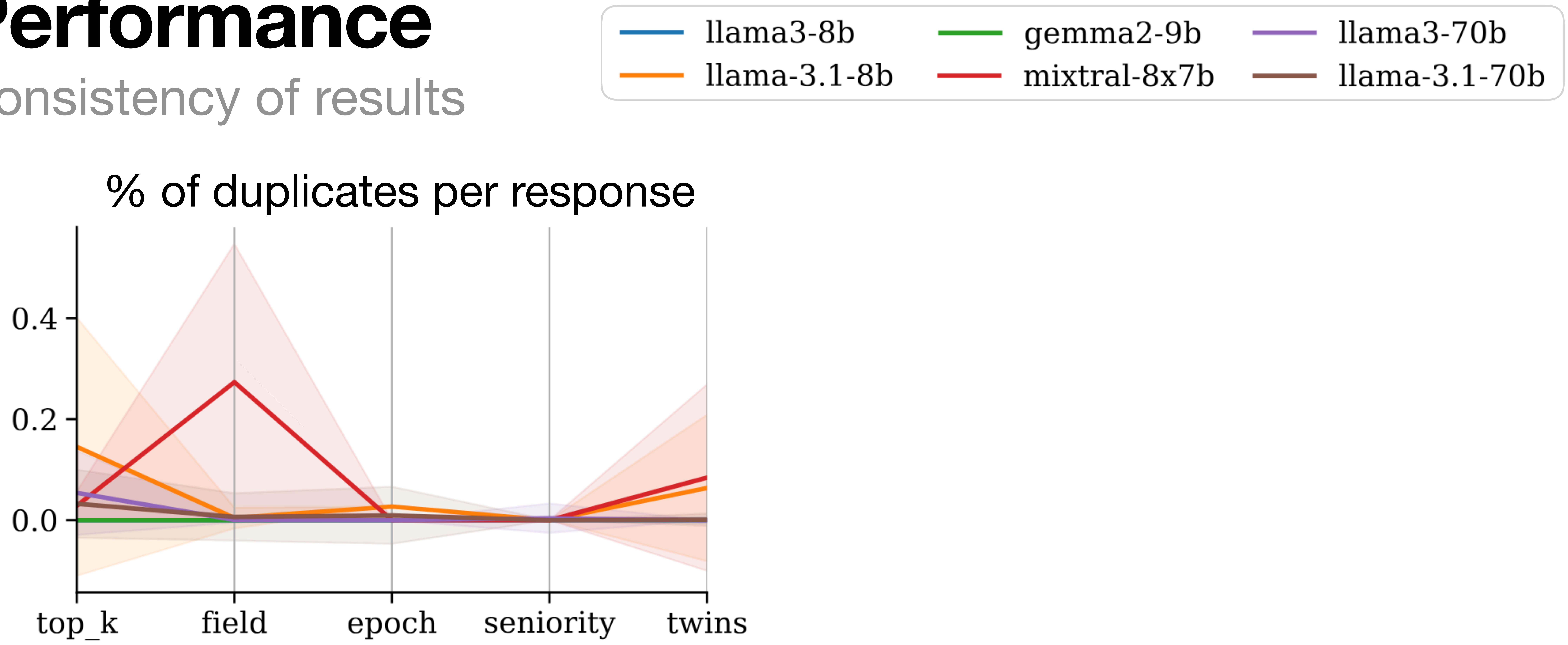
Performance

Consistency of results

Barolo et al. 2025 (under-review)

Performance

Consistency of results

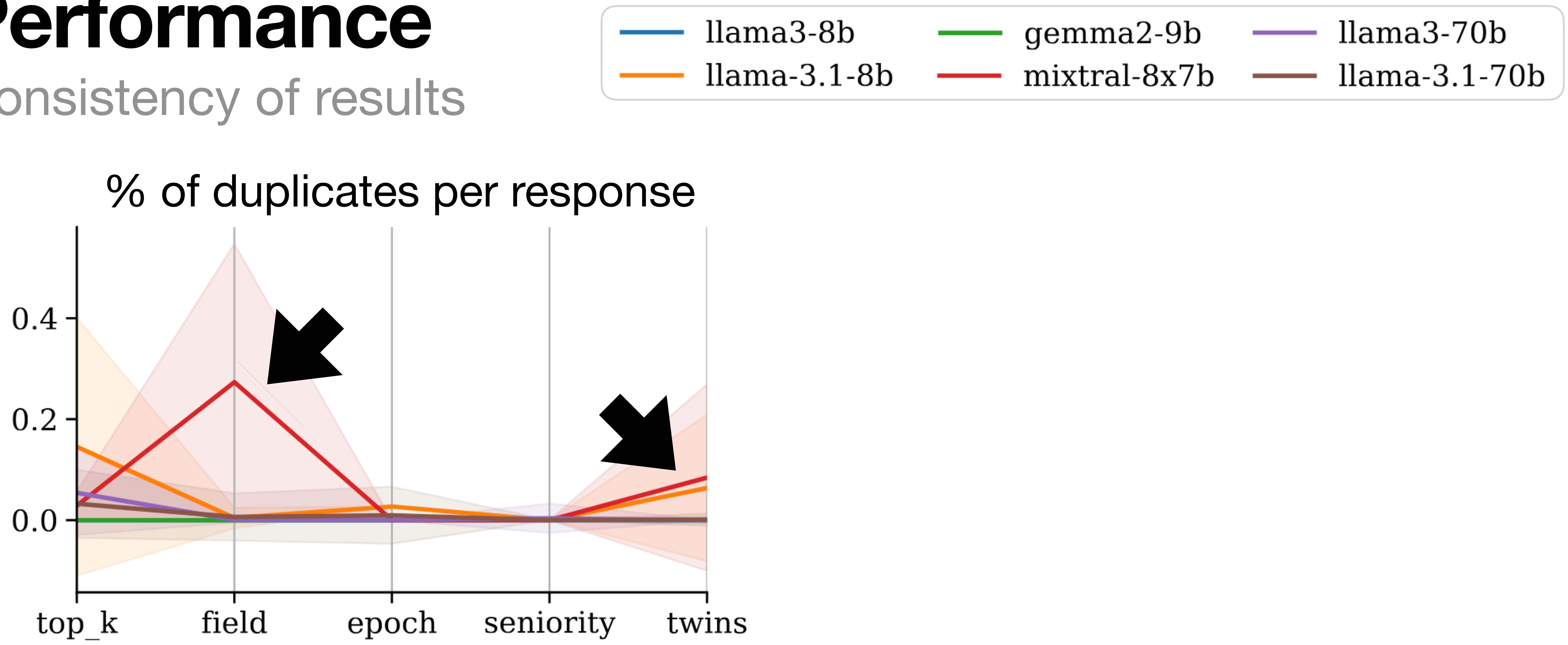


Barolo et al. 2025 (under-review)

Performance

Consistency of results

% of duplicates per response

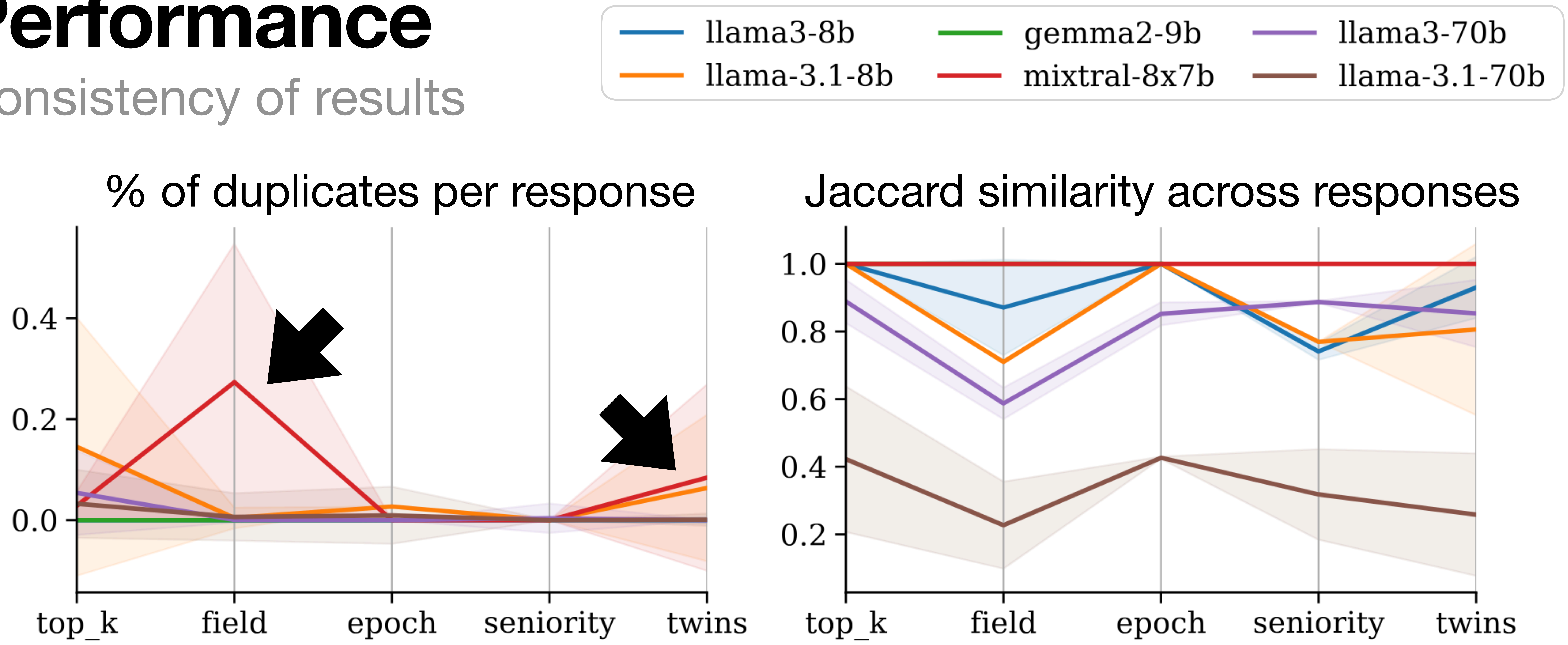


mixtral & llama-3.1-8b tend to repeat names in **the same response.**

Barolo et al. 2025 (under-review)

Performance

Consistency of results



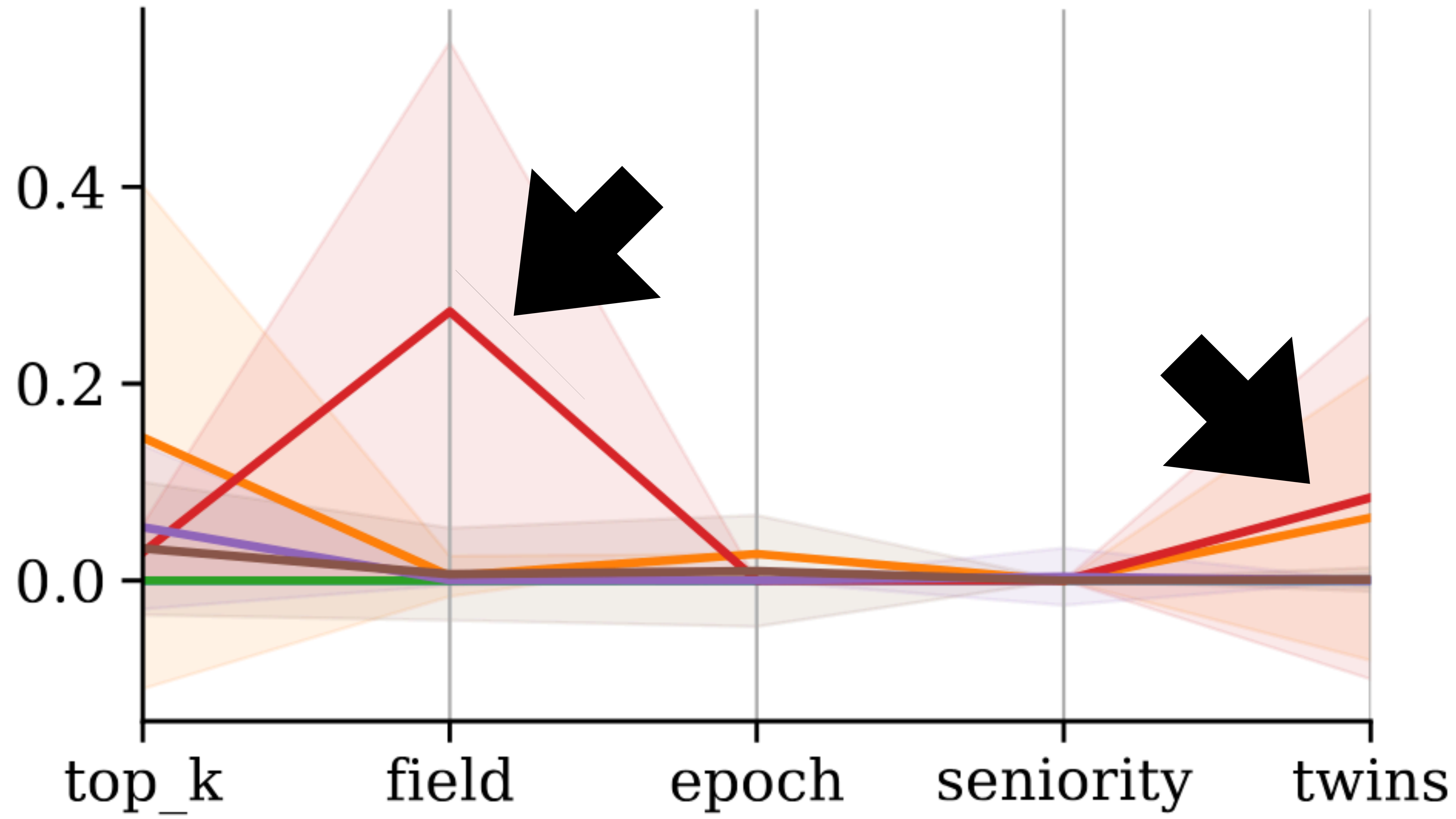
mixtral & llama-3.1-8b tend to repeat names in **the same response.**

Barolo et al. 2025 (under-review)

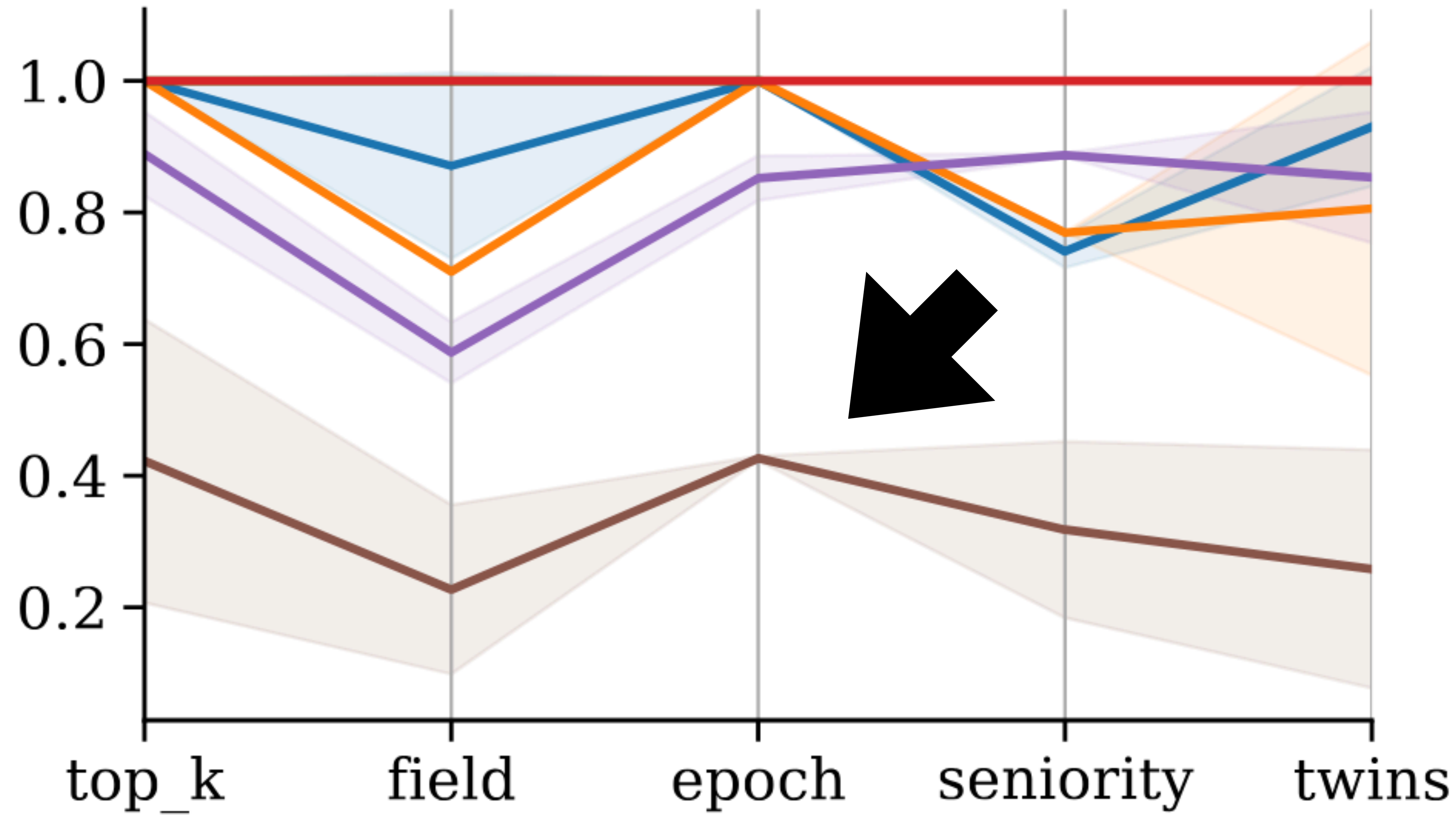
Performance

Consistency of results

% of duplicates per response



Jaccard similarity across responses

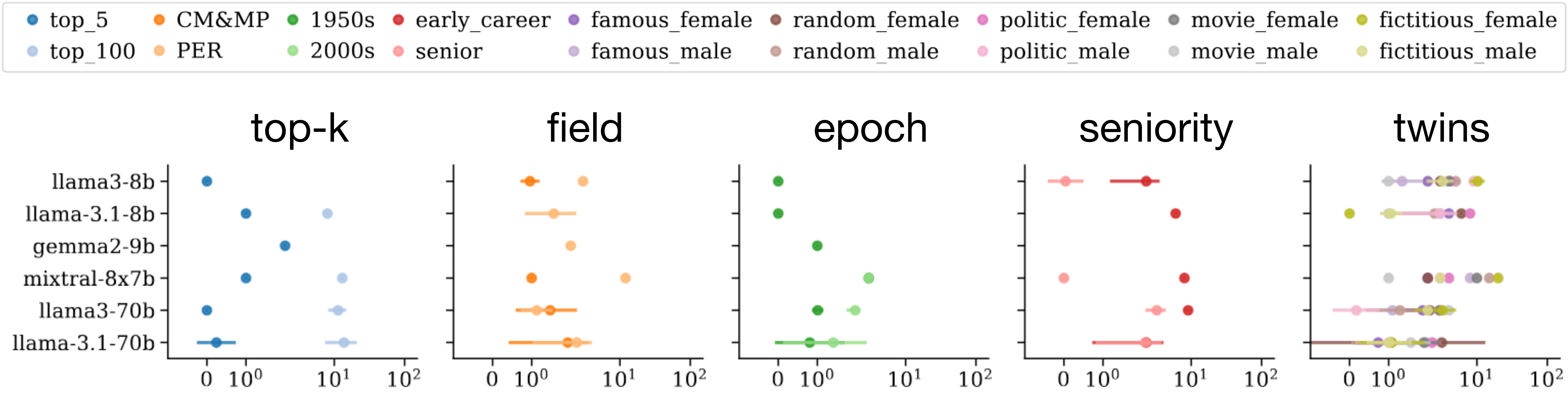


mixtral & llama-3.1-8b tend to repeat names in **the same response.**

llama-3.1-70b tends to be more creative by no repeating names **across responses.**

Barolo et al. 2025 (under-review)

Potential hallucinations



(c) Average number of potential hallucinations

Simpson diversity index [0 no diversity, 1 perfect diversity]

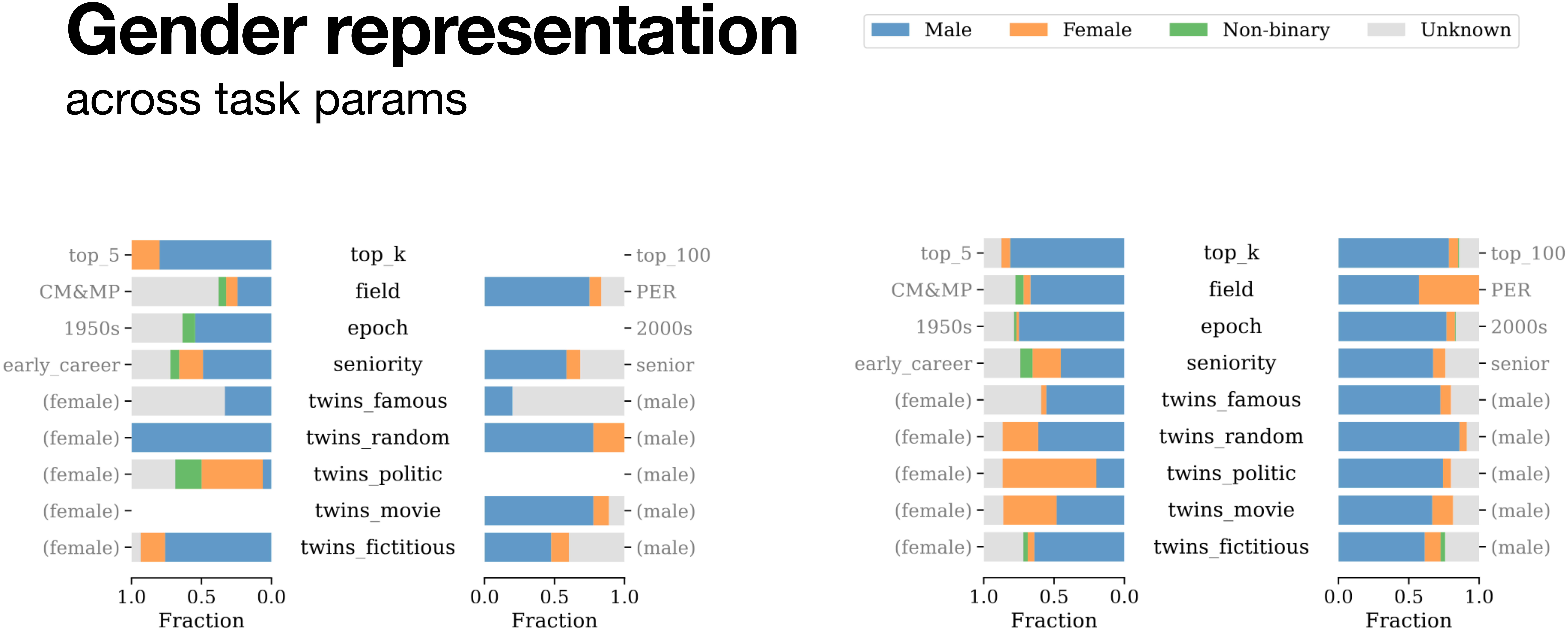
(a) Ethnicity

(b) Gender

Barolo et al. 2025 (under-review)

Gender representation

across task params



(a) llama3-8b

(f) llama3.1-70b

Barolo et al. 2025 (under-review)

Ethnicity representation

across task params

Barolo et al. 2025 (under-review)

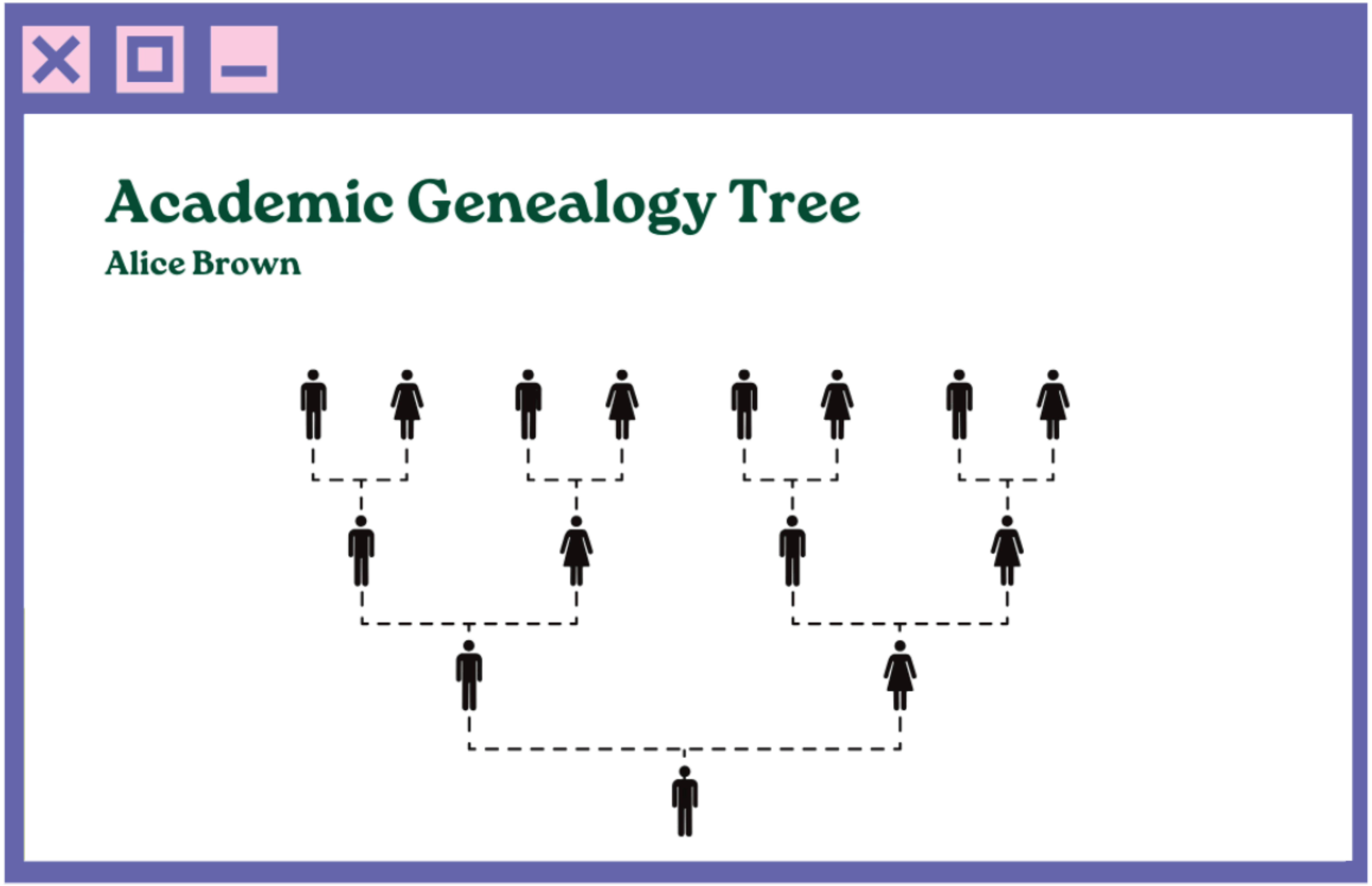
Academic roadmap

a mock-up

Espín-Noboa et al. (work-in-progress)

Academic genealogy tree

a mock-up

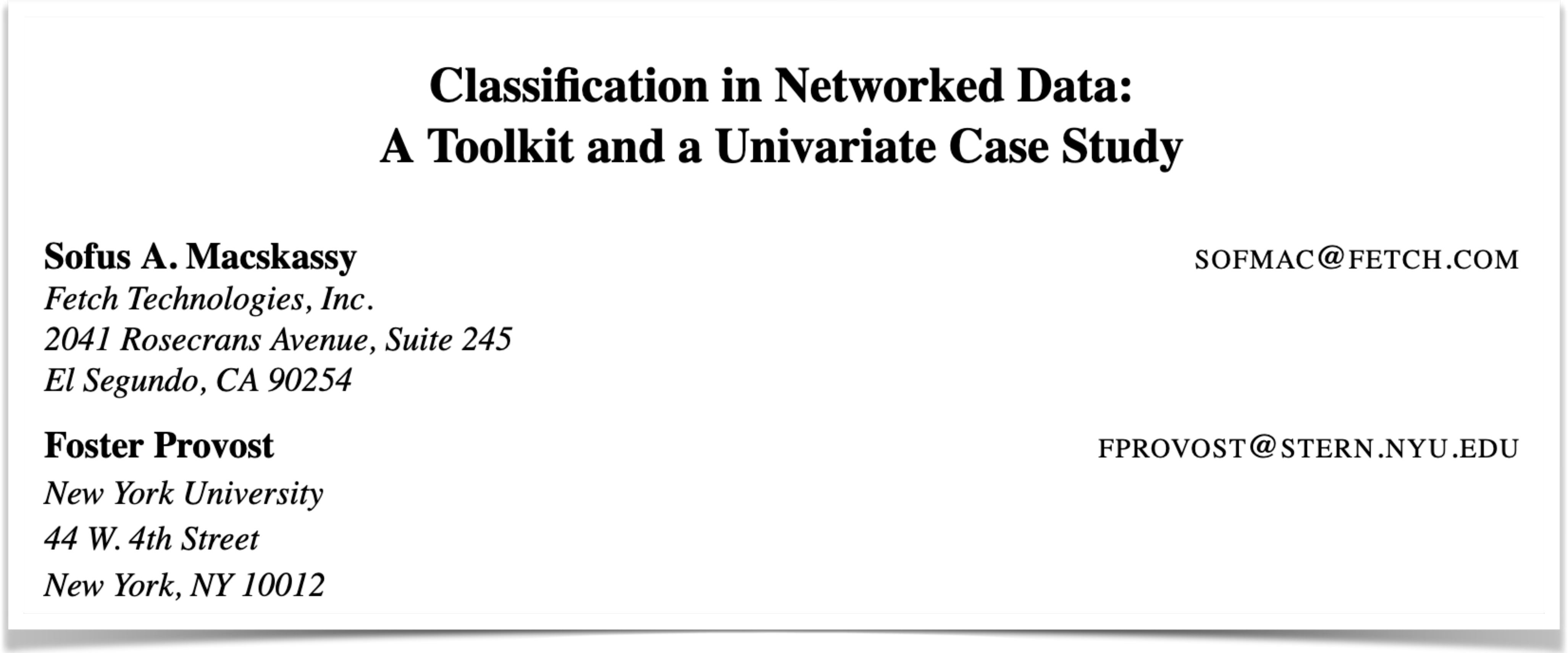


Espín-Noboa et al. (work-in-progress)

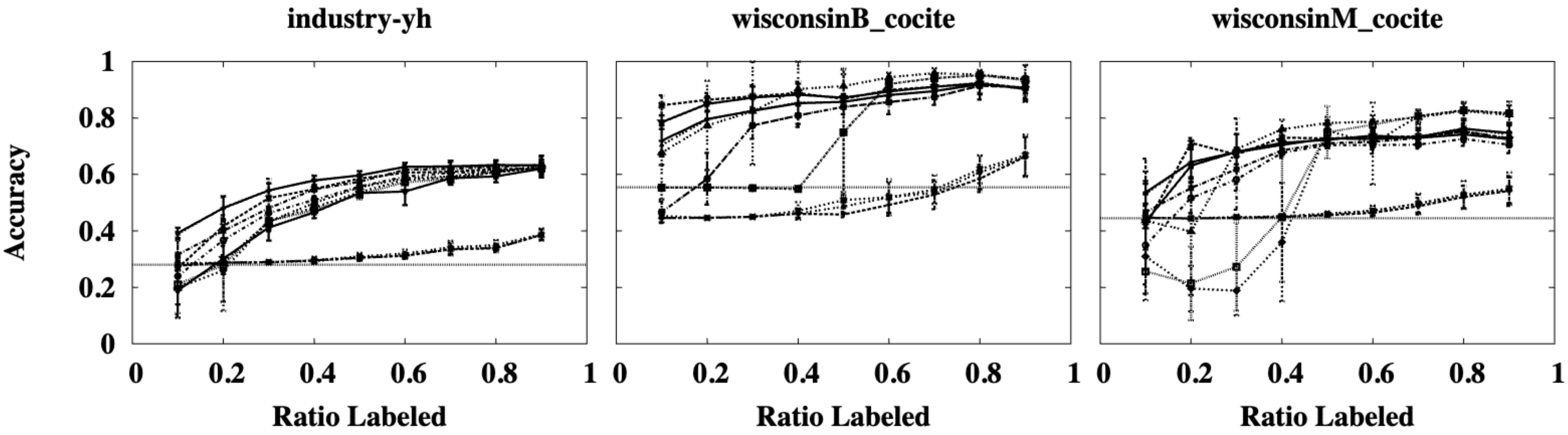
Backups
Synthetic Networks

The CS Approach

Real-network benchmarks



Overall classification accuracies



Example: Macskassy, S. A., & Provost, F. (2007). Classification in networked data: A toolkit and a univariate case study. *Journal of machine learning research*, *8*(5).

PageRank

by Google

PageRank of neighbor j

Following neighbors

$$PR(i) = (1 - \alpha) + \alpha \sum_{j \in N_i} \frac{PR(j)}{k_j^{out}}$$

Damping factor (teleportation)

Out-degree of neighbor j

Page, L., Brin, S., Motwani, R. & Winograd, T. The pagerank citation ranking: Bringing order to the web (Technical Report, Stanford InfoLab, 1999).

Gini Coefficient

by Italian Statistician Corrado Gini

$$Gini(X) = \frac{\sum_{i=1}^{\hat{n}} (2i - \hat{n} - 1)x_i}{\hat{n} \sum_{i=1}^{\hat{n}} x_i}$$

Sorted **Distribution**

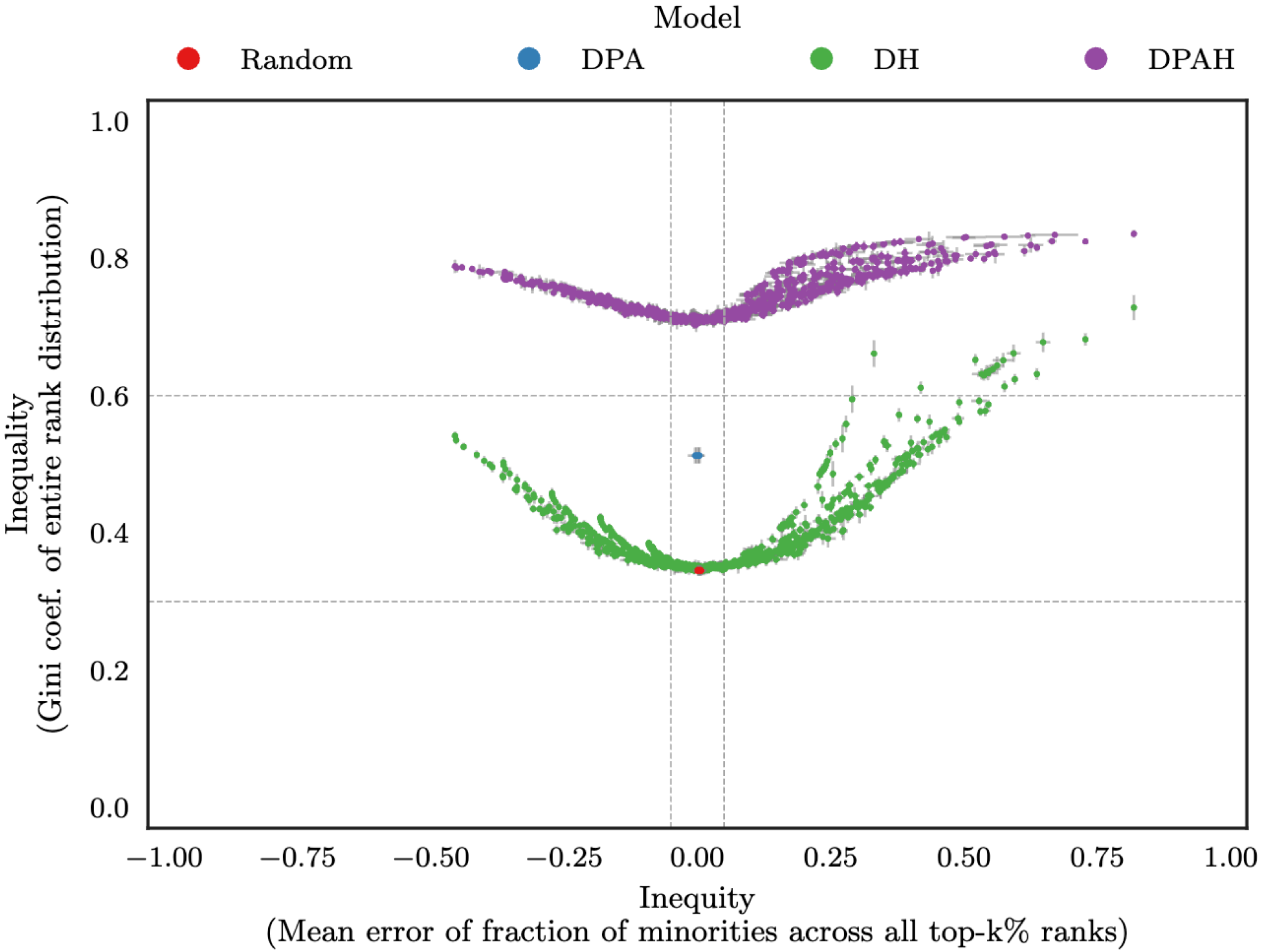
Size of X

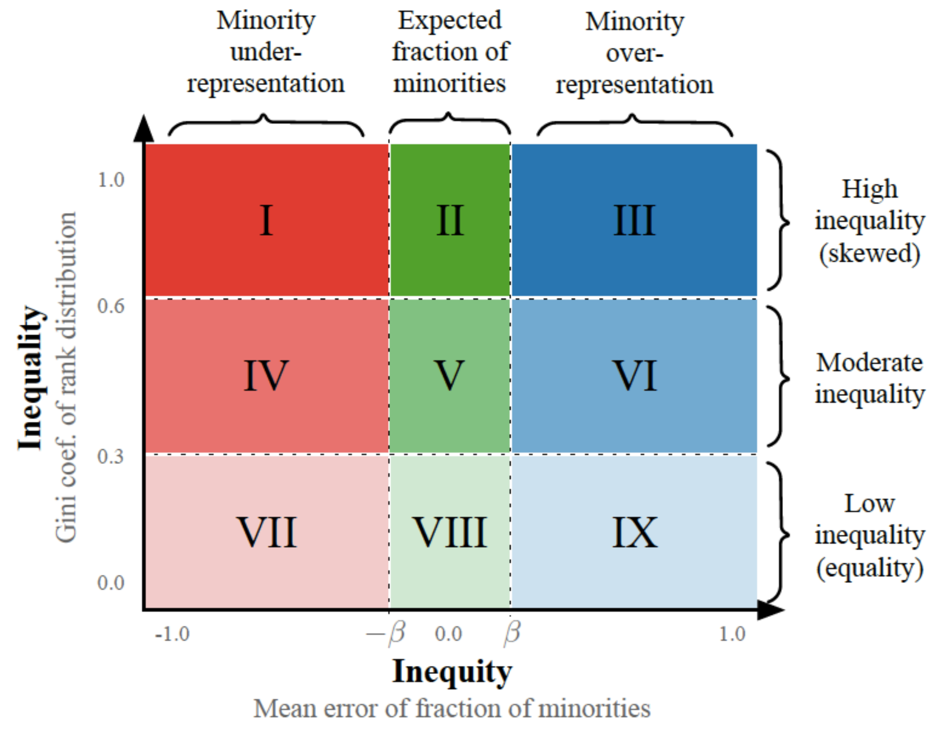
i-th element of X

Gini, C. (1921). Measurement of inequality of incomes. The economic journal, 31(121), 124-125.

Disparity in PageRank

as a function of homophily and preferential attachment

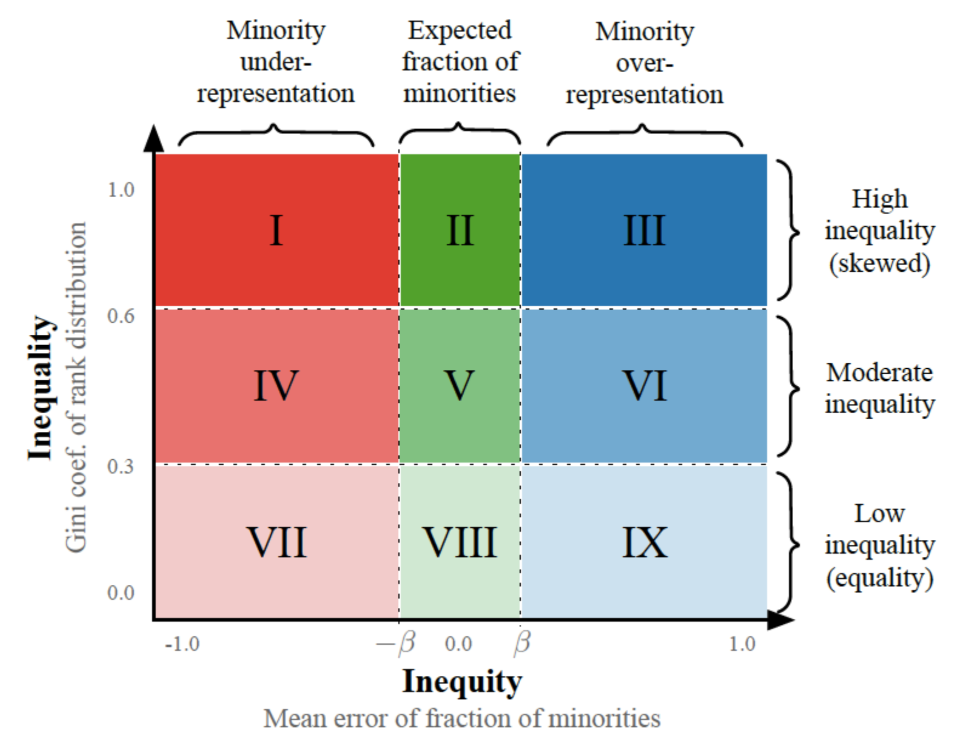




Espín-Noboa et al.
Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes



minority 10%

minority 20%

minority 30%

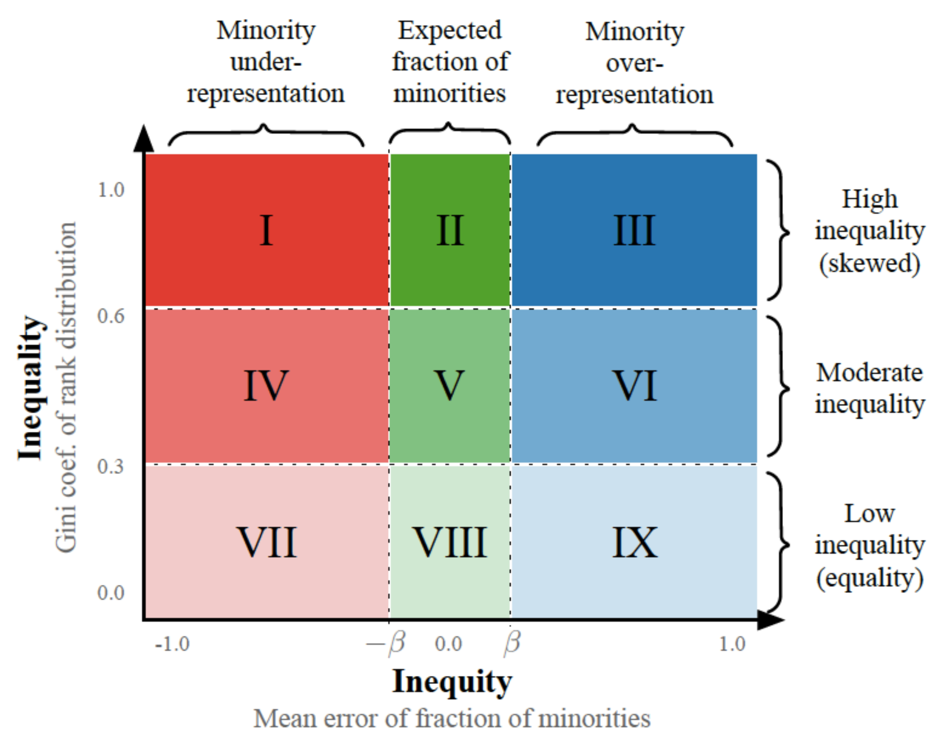
minority 40%

balanced 50%

Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

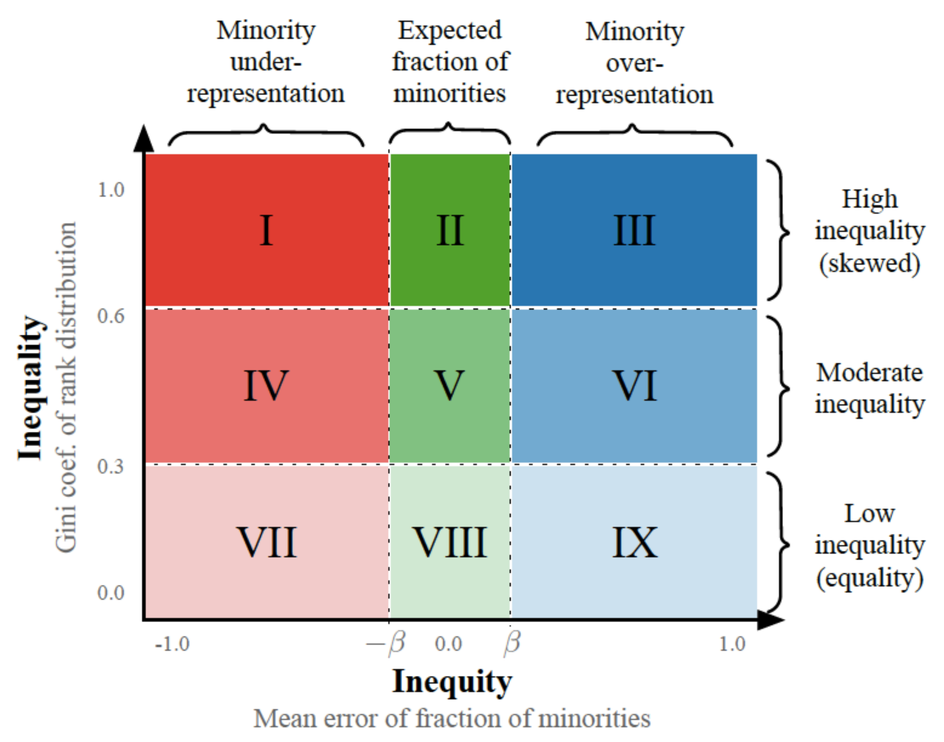
as a function of homophily and fraction of minority nodes



Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

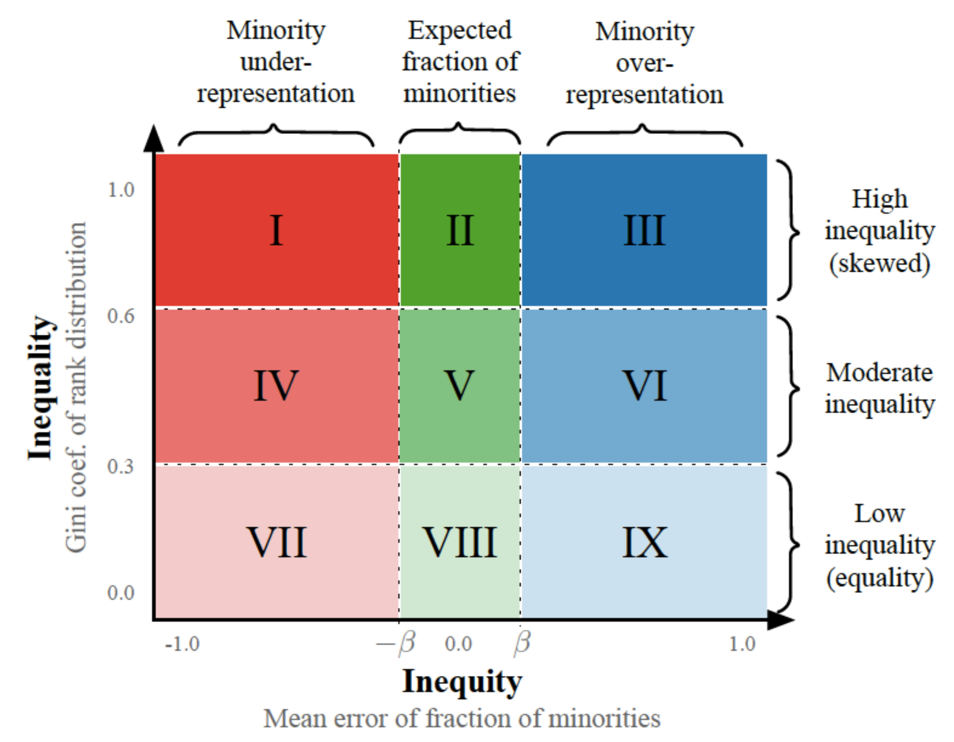
as a function of homophily and fraction of minority nodes

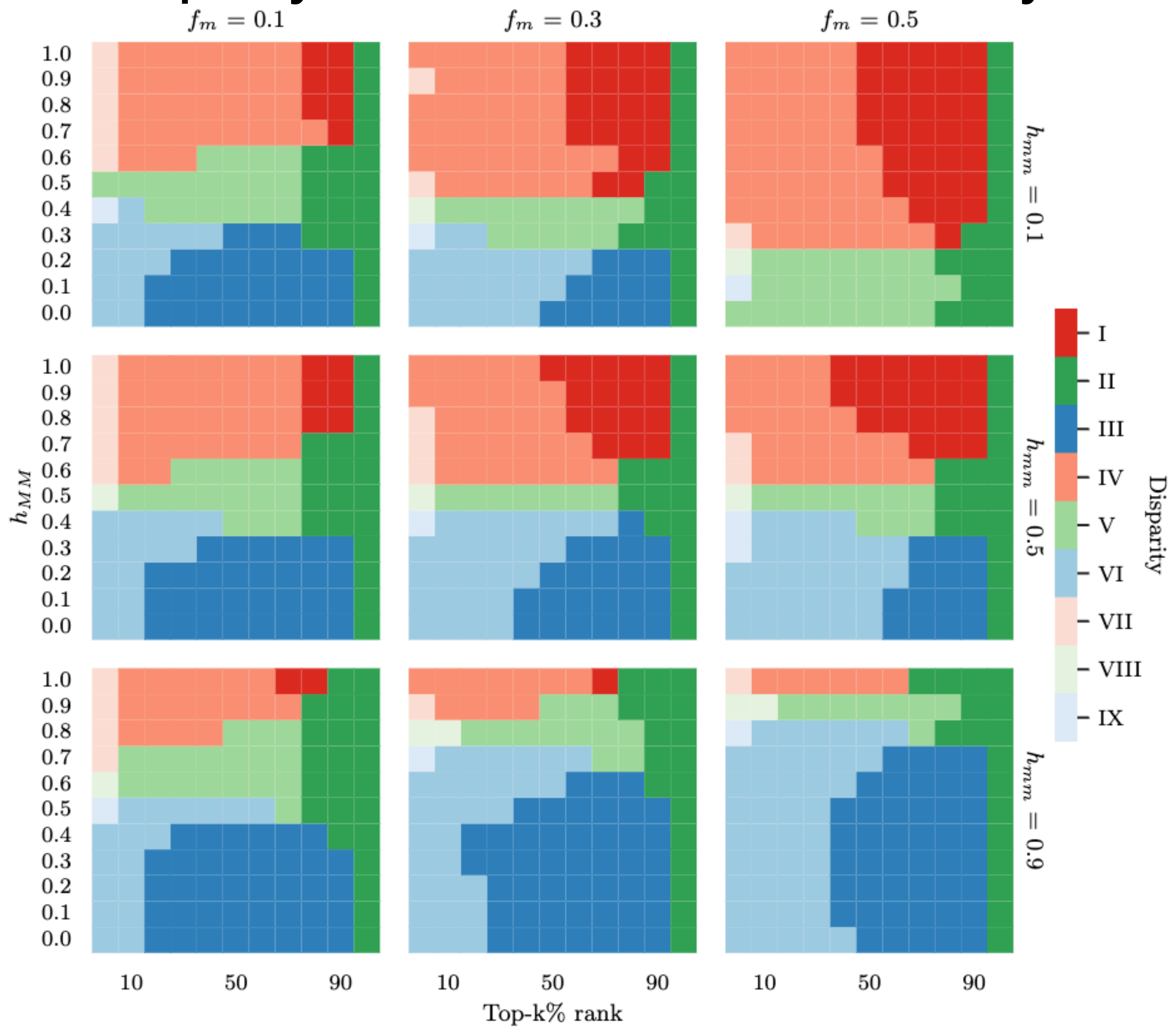


Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

as a function of homophily and fraction of minority nodes

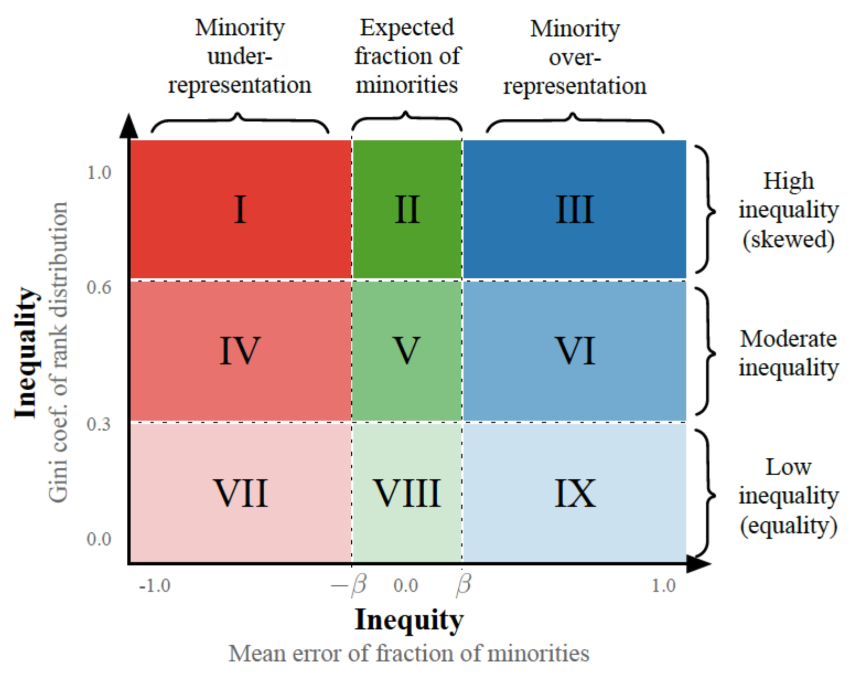


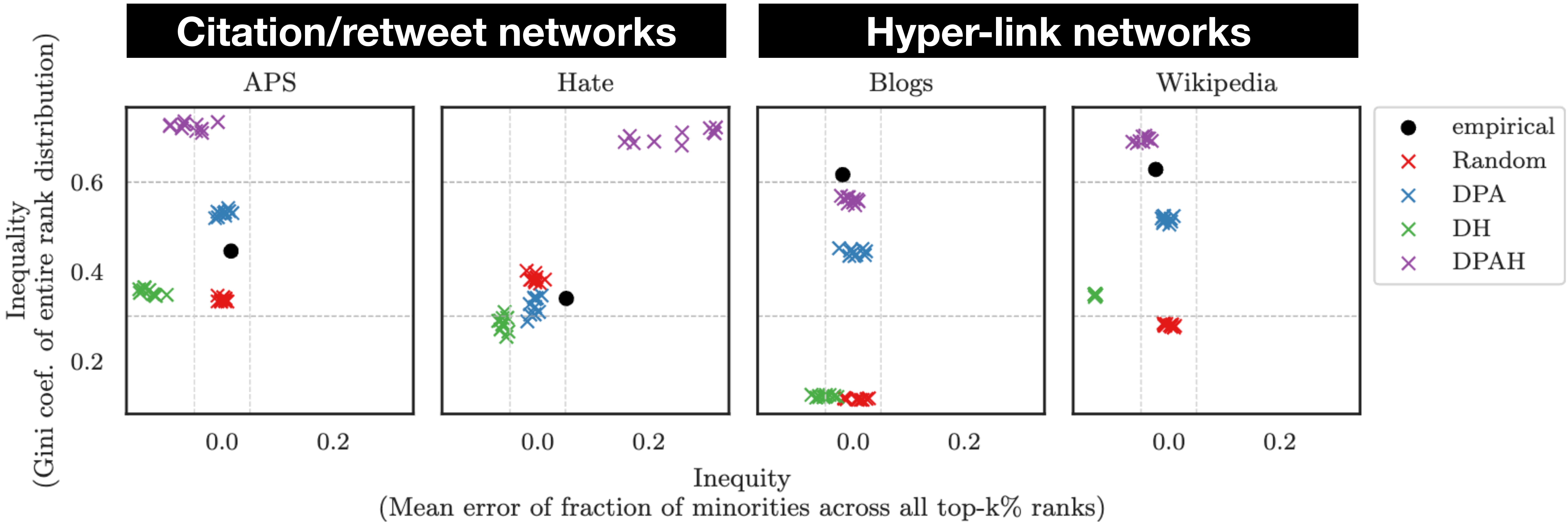


Espín-Noboa et al.
Nature Scientific Reports (2022)

Disparity in PageRank

on real networks

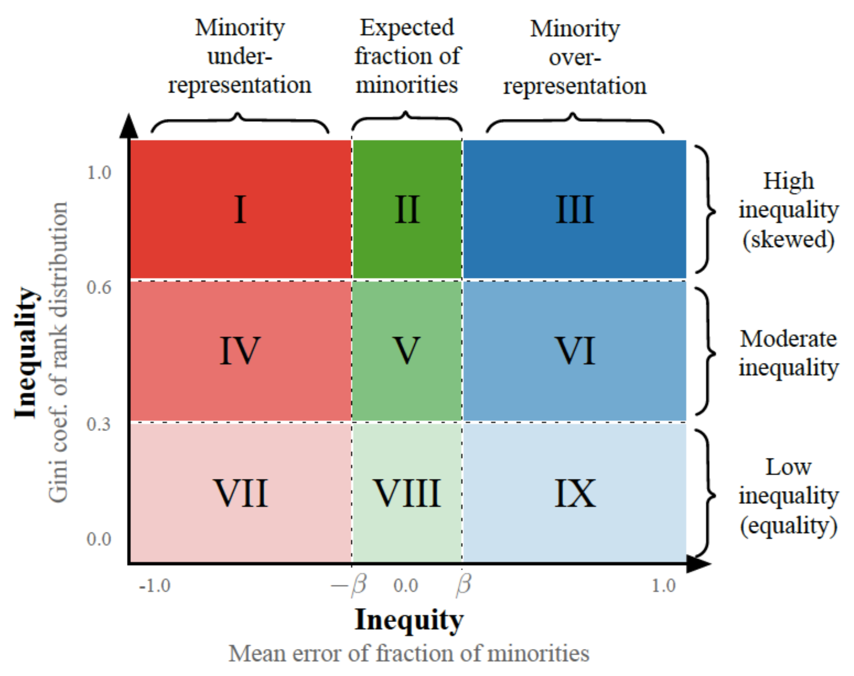


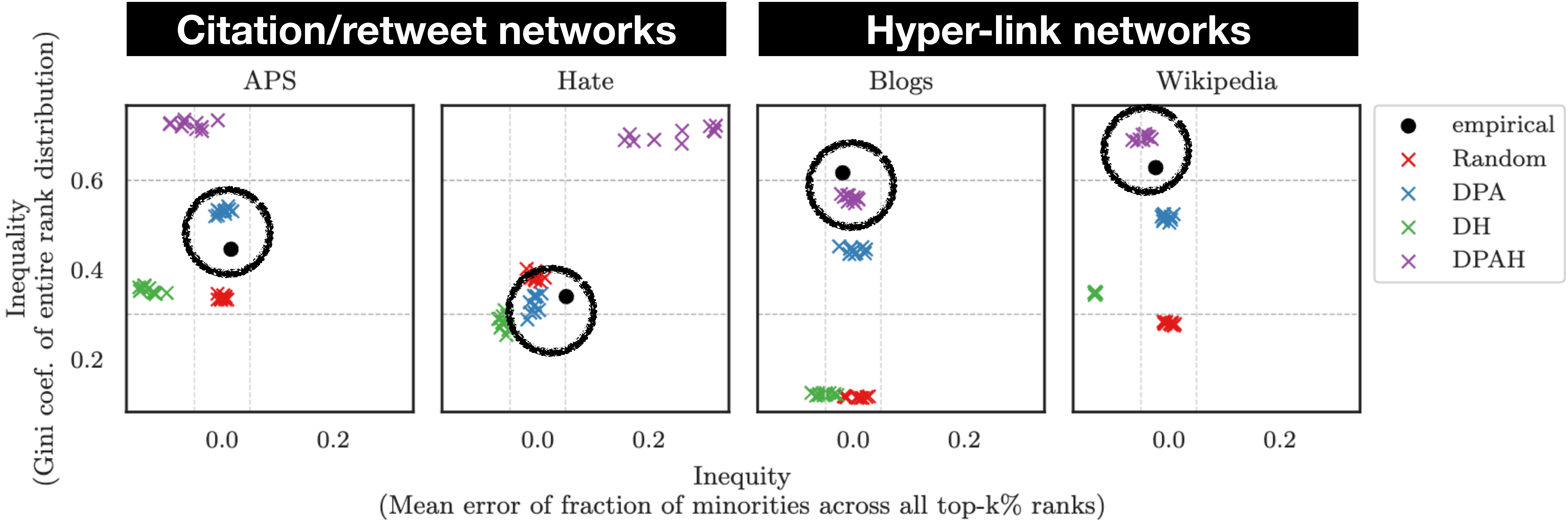


Espín-Noboa et al. Nature Scientific Reports (2022)

Disparity in PageRank

on real networks

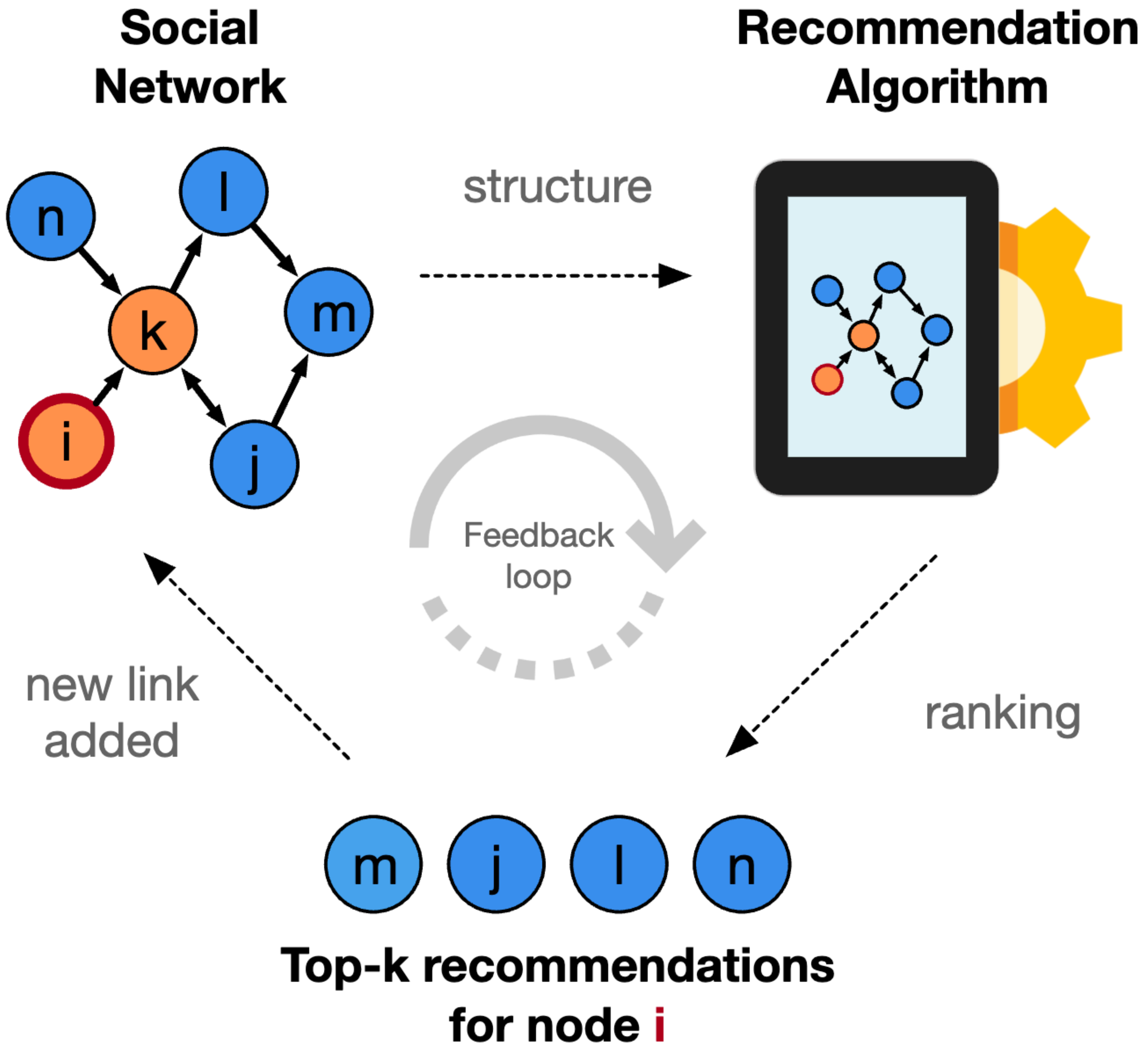




Espín-Noboa et al. Nature Scientific Reports (2022)

Link recommendations:

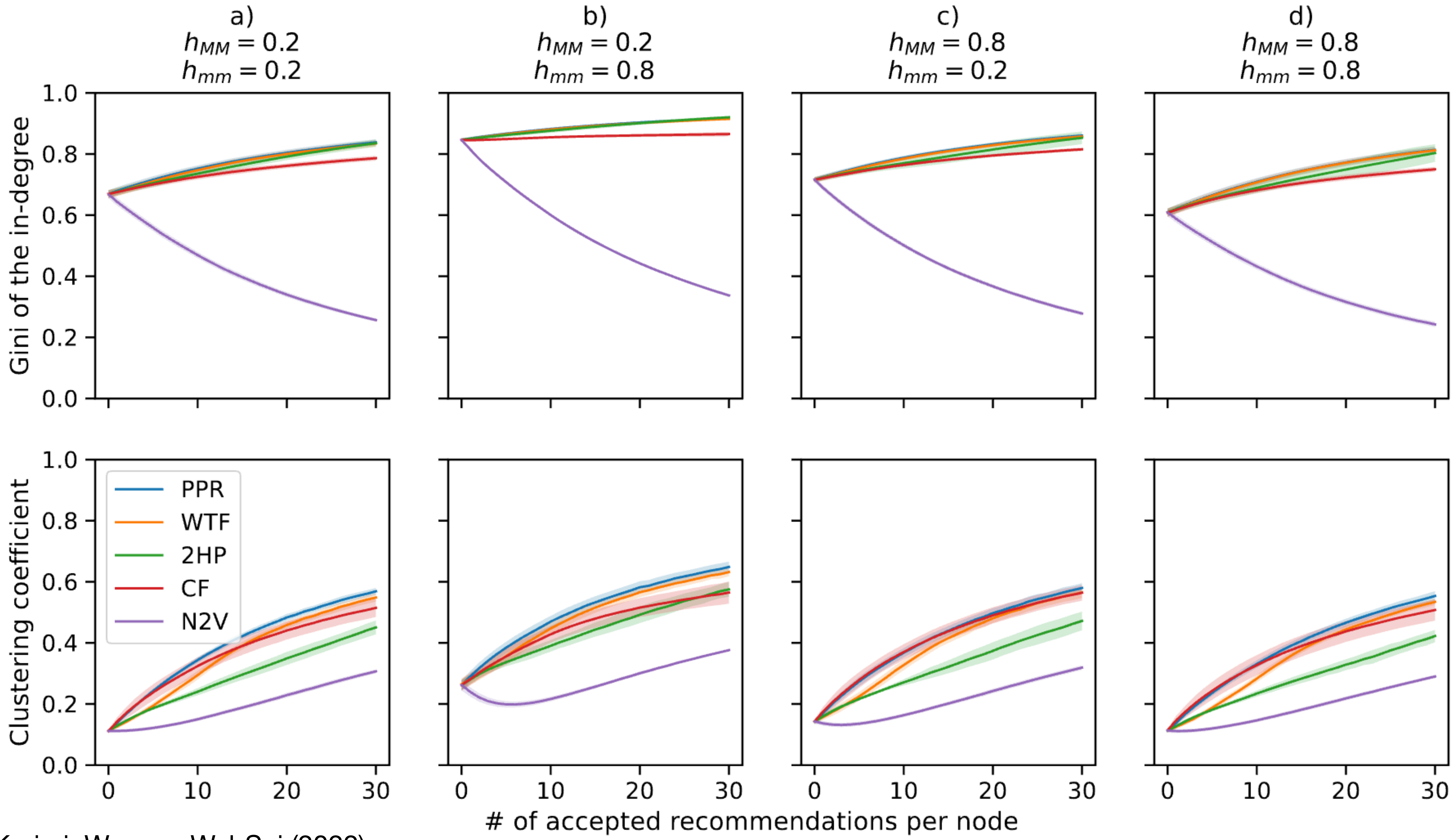
Their impact on network structure and minorities



Ferrara, **Espín-Noboa**, Karimi, Wagner. WebSci (2022)

Link recommendations:

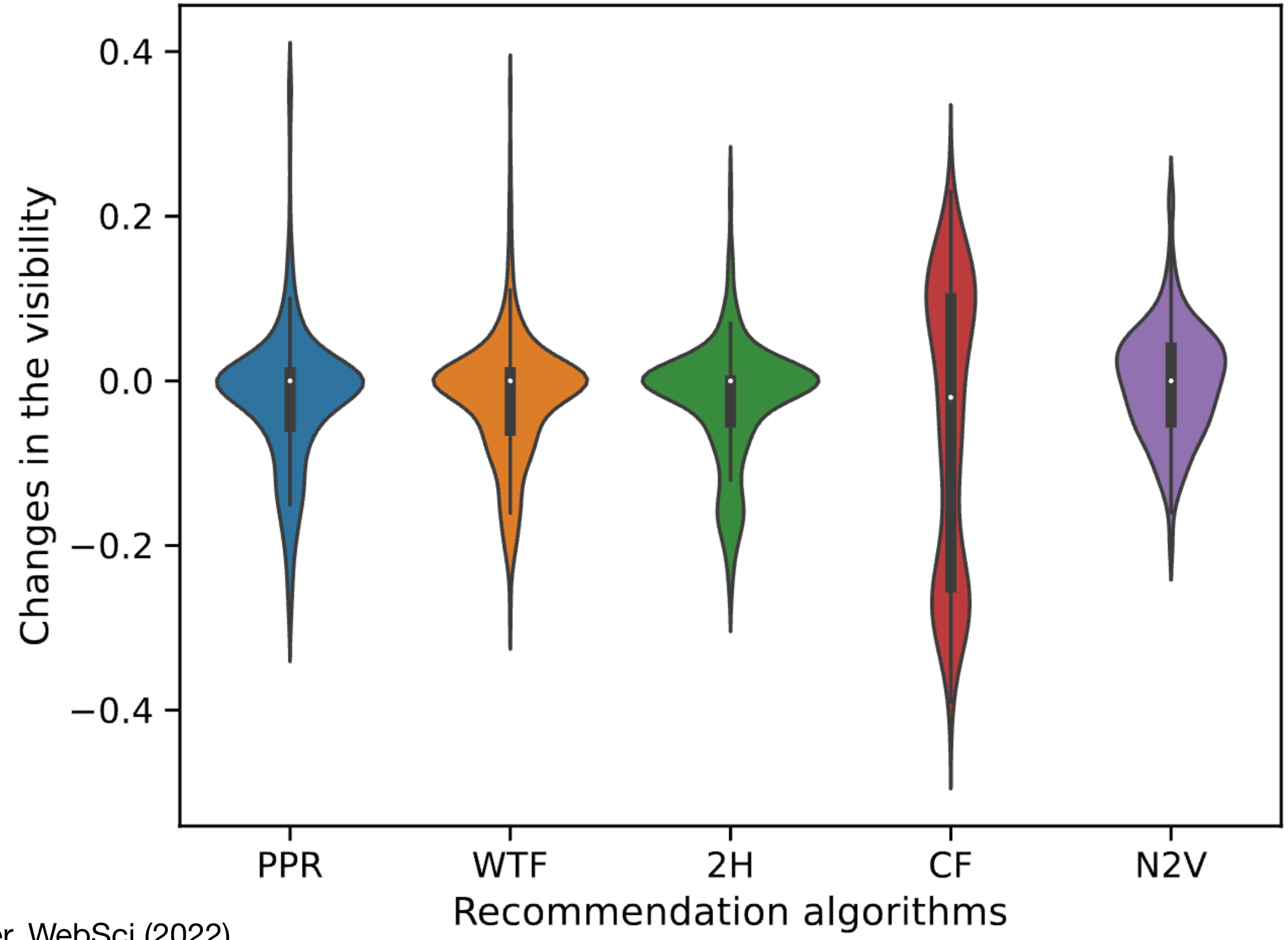
Their impact on network structure and minorities



Ferrara, **Espín-Noboa**, Karimi, Wagner. WebSci (2022)

Link recommendations:

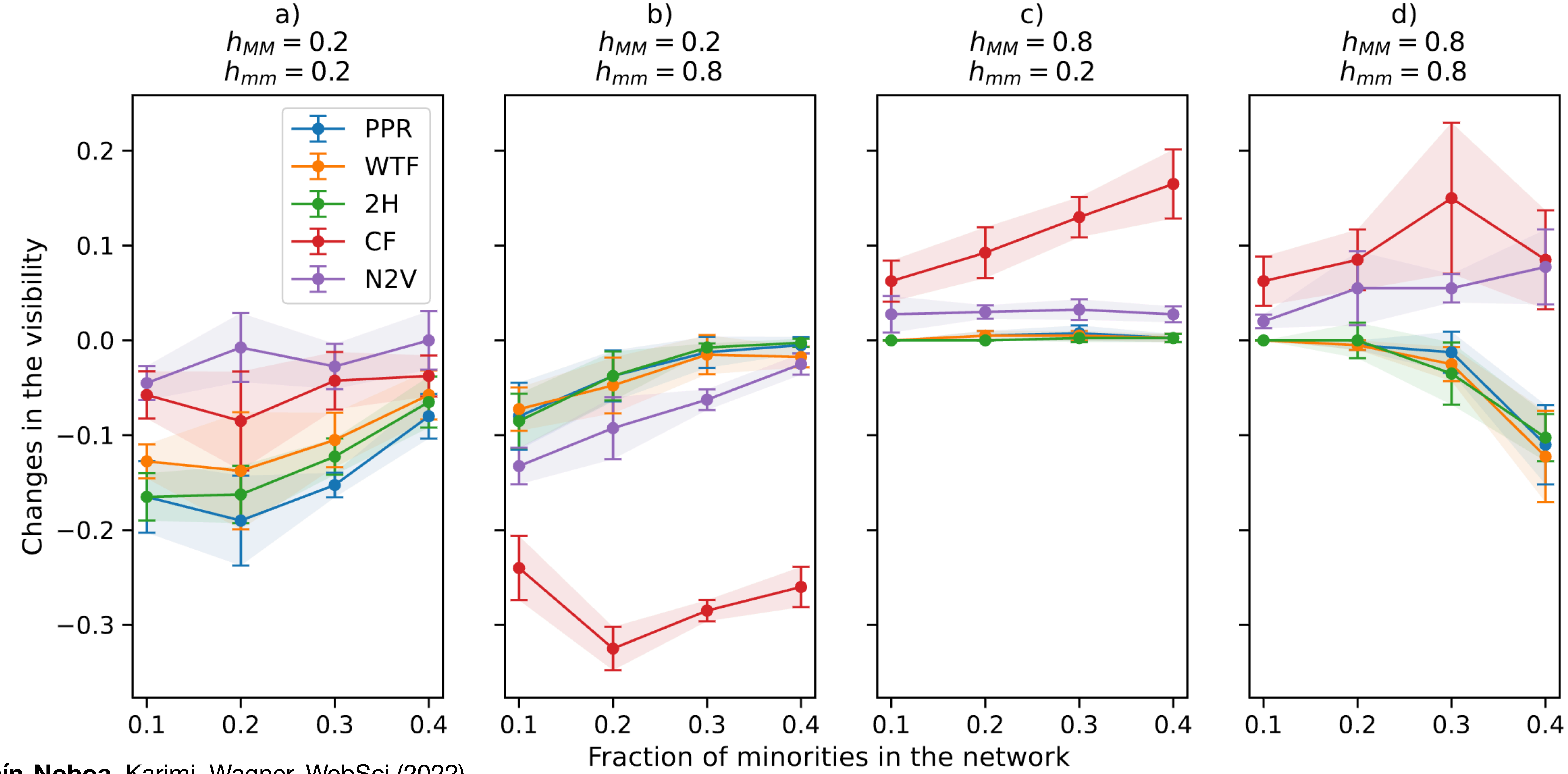
Their impact on network structure and minorities (overall)



Ferrara, **Espín-Noboa**, Karimi, Wagner. WebSci (2022)

Link recommendations:

Their impact on network structure and minorities (wrt group size)



Ferrara, **Espín-Noboa**, Karimi, Wagner. WebSci (2022)