Diagnosing Algorithmic Inequality in Social Networks

Ana-Andreea Stoica,

Max Planck Institute for Intelligent Systems, Tübingen, Germany

May, 2023

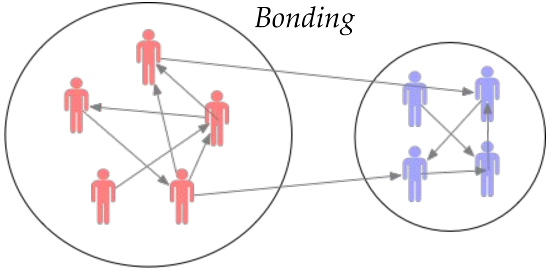
Networks and inequality: empirical studies

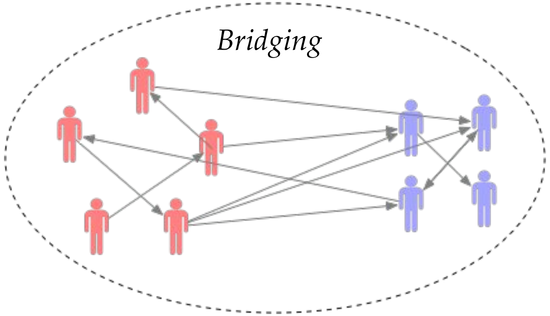
Empirical studies set the grounds for models

Structures & patterns that we see in real networks:

- Bigger cities ⇒ higher average degree & communication activity volume [Schlapher et al, 2014]
- Probability of friends-of-friends edges independent of city size [Schlapher et al, 2014]
- Decreased communication to and from a certain area ↔ poverty [Smith-Clarke et al, 2014]

Social network utility: social capital [Putnam, 2000]





Bonding and bridging communities in empirical studies

- [Gündoğdu et al, 2019] finds that poverty correlates to ‘bridging’ communities and wealth to ‘bonding’ communities
- Network of 378 mobile cell towers in Côte d’Ivoire
  - edges weighted by amount of communication of users in the cell towers
  - agreggate by area (commune)
- ‘Bonding’ (closed) or ‘bridging’ (open) measures:

Table 6. Mean values of the degree centrality, betweenness centrality, effective size, efficiency, and local clustering coefficient in the communication network. These measures were calculated for each of the ten communes of Abidjan.

		Bridging measures			Bonding measures
Commune	**Degree**	**Between. centrality**	**Eff. size**	**Efficiency**	**Local clust. coeff.**
Abobo	1200.200	1128.866	624.555	0.521	0.946
Adjame	1195.565	1123.036	609.005	0.509	0.946
Attecoube	1202.538	1147.459	619.986	0.516	0.945
Cocody	1141.789	1054.691	560.872	0.476	0.948
Koumassi	1188.118	1125.058	654.206	0.551	0.947
Marcory	1025.103	841.657	573.949	0.552	0.958
Plateau	1001.588	789.809	427.779	0.394	0.961
Port-Bouet	1067.750	896.124	607.666	0.568	0.956
Treichville	1113.850	950.254	612.798	0.550	0.954
Yopougon	1175.697	1143.794	599.460	0.502	0.946

Bonding and bridging communities in empirical studies

- [Gündoğdu et al, 2019] finds that poverty correlates to 'bridging' communities and wealth to 'bonding' communities
- Network of 378 mobile cell towers in Côte d'Ivoire
  - edges weighted by amount of communication of users in the cell towers
  - agreggate by area (commune)
- 'Bonding' (closed) or 'bridging' (open) measures:

Table 6. Mean values of the degree centrality, betweenness centrality, effective size, efficiency, and local clustering coefficient in the communication network. These measures were calculated for each of the ten communes of Abidjan.

		Bridging measures			Bonding measures
Commune	**Degree**	**Between. centrality**	**Eff. size**	**Efficiency**	**Local clust. coeff.**
Abobo	1200.200	1128.866	624.555	0.521	0.946
Adjame	1195.565	1123.036	609.005	0.509	0.946
Attecoube	1202.538	1147.459	619.986	0.516	0.945
Cocody	1141.789	1054.691	560.872	0.476	0.948
Koumassi	1188.118	1125.058	654.206	0.551	0.947
Marcory	1025.103	841.657	573.949	0.552	0.958
Plateau	1001.588	789.809	427.779	0.394	0.961
Port-Bouet	1067.750	896.124	607.666	0.568	0.956
Treichville	1113.850	950.254	612.798	0.550	0.954
Yopougon	1175.697	1143.794	599.460	0.502	0.946

poorer

Bonding and bridging communities in empirical studies

- [Gündoğdu et al, 2019] finds that poverty correlates to 'bridging' communities and wealth to 'bonding' communities
- Network of 378 mobile cell towers in Côte d'Ivoire
  - edges weighted by amount of communication of users in the cell towers
  - agreggate by area (commune)
- 'Bonding' (closed) or 'bridging' (open) measures:

Table 6. Mean values of the degree centrality, betweenness centrality, effective size, efficiency, and local clustering coefficient in the communication network. These measures were calculated for each of the ten communes of Abidjan.

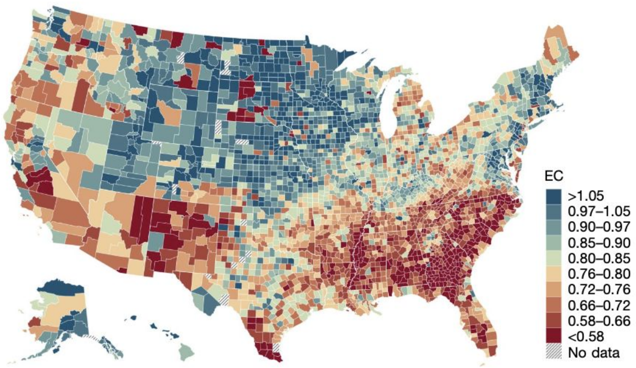
		Bridging measures			Bonding measures
Commune	**Degree**	**Between. centrality**	**Eff. size**	**Efficiency**	**Local clust. coeff.**
Abobo	1200.200	1128.866	624.555	0.521	0.946
Adjame	1195.565	1123.036	609.005	0.509	0.946
Attecoube	1202.538	1147.459	619.986	0.516	0.945
Cocody	1141.789	1054.691	560.872	0.476	0.948
Koumassi	1188.118	1125.058	654.206	0.551	0.947
Marcory	1025.103	841.657	573.949	0.552	0.958
Plateau	1001.588	789.809	427.779	0.394	0.961
Port-Bouet	1067.750	896.124	607.666	0.568	0.956
Treichville	1113.850	950.254	612.798	0.550	0.954
Yopougon	1175.697	1143.794	599.460	0.502	0.946

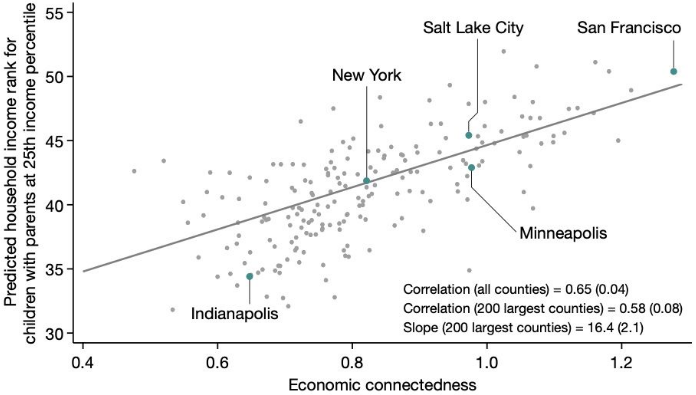
richer

Homophily and inequality

[Chetty et al, 2022]:

“the share of high socio-economic status friends among individuals with low socio-economic status is among the strongest predictors of upward income mobility identified to date”

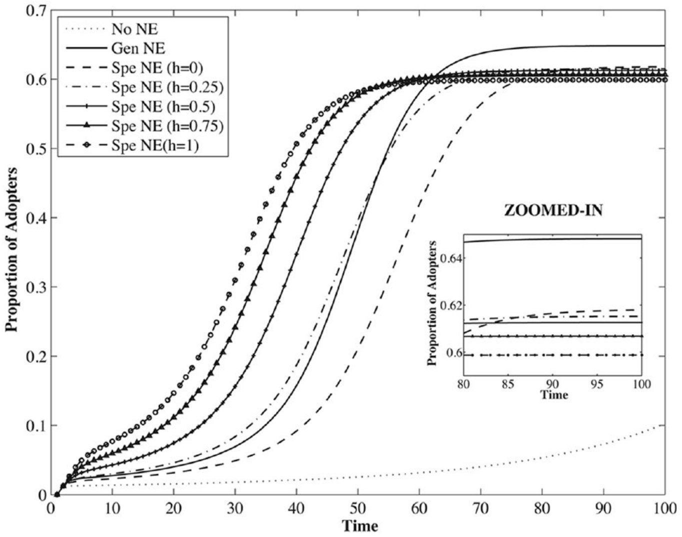




Homophily and inequality

[DiMaggio & Garip, 2011] shows that homophily (bonding) leads to increasing inter-group inequality r.e. Internet adoption in the US:

- 2,257 African-American and white respondents to the 2002 General Social Survey (GSS), which included items on network size, race, education, and income
    - Create networks with individual features, vary homophily
- Simulate diffusion through threshold model + a fixed initial price of Internet



⇒ Homophily decreases adoption with time

Homophily and inequality

[DiMaggio & Garip, 2011] shows that homophily (bonding) leads to increasing inter-group inequality r.e. Internet adoption in the US:

- 2,257 African-American and white respondents to the 2002 General Social Survey (GSS), which included items on network size, race, education, and income
    - Create networks with individual features, vary homophily
- Simulate diffusion through threshold model + a fixed initial price of Internet

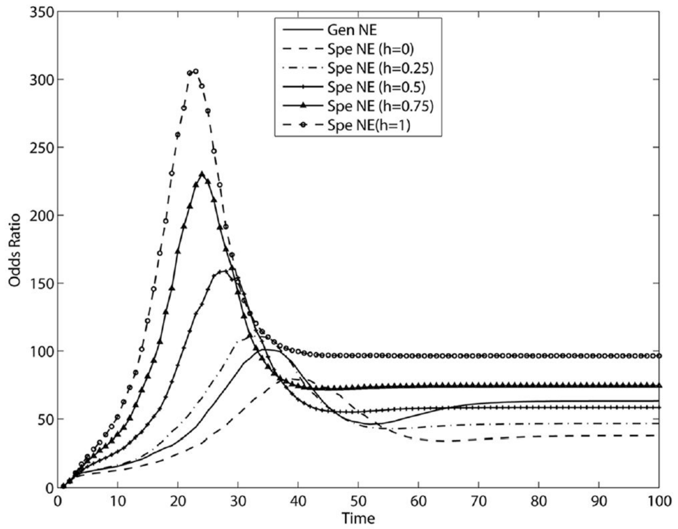


FIG. 4.—Odds ratios of diffusion rates for highest- as compared to lowest-income classes in six conditions of externalities and homophily.

⇒ Homophily increases inter-group inequality in adoption

Homophily and inequality

[DiMaggio & Garip, 2011] shows that homophily (bonding) leads to increasing inter-group inequality r.e. Internet adoption in the US:

- 2,257 African-American and white respondents to the 2002 General Social Survey (GSS), which included items on network size, race, education, and income
    - Create networks with individual features, vary homophily
- Simulate diffusion through threshold model + a fixed initial price of Internet

TABLE 2
LINEAR REGRESSION OF ADOPTION LEVELS ON EXPERIMENTAL CONDITIONS

		RACE		INCOME		EDUCATION	
	ALL	Whites	Blacks	High	Low	BA	Less than High School
No network externalities	−.516**	−.536**	−.399**	−.685**	−.238**	−.611**	−.351**
General network externalities	.030**	.028**	.043**	.032**	.017**	.023**	.030**
Homophily = .25	−.003**	−.001	−.012**	.009**	−.014**	.005**	−.011**
Homophily = .5	−.005**	−.002**	−.024**	.017**	−.028**	.010**	−.024**
Homophily = .75	−.011**	−.006**	−.040**	.024**	−.046**	.012**	−.043**
Homophily = 1	−.019**	−.012**	−.061**	.029**	−.067**	.015**	−.068**
Intercept	.618**	.647**	.454**	.925**	.249**	.788**	.392**
R^2	.99	.99	.97	.99	.96	.99	.96

NOTE.—All independent variables are binary. Both dependent and independent variables are measured on the final period of simulations (t = 100). Reference: homophily = 0; N = 7,000.
* $P < .05$.
** $P < .01$.

⇒ Internet adoption increases among the most prosperous in the presence of homophily

Empirical studies on the Internet [Barabasi-Albert,1999]

Power law degree distribution in online networks: $P(k) \sim k^{\gamma}$

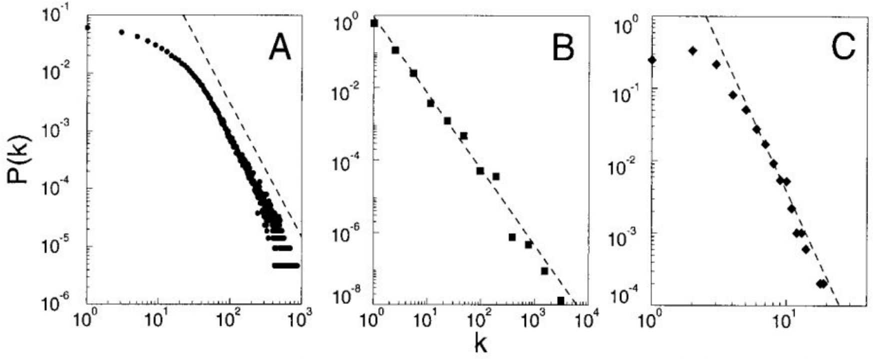


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Social capital

Resources, opportunities, …

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials & interference

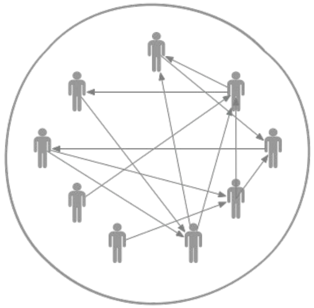
How do we use networks to design algorithms?

1. Diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials & interference

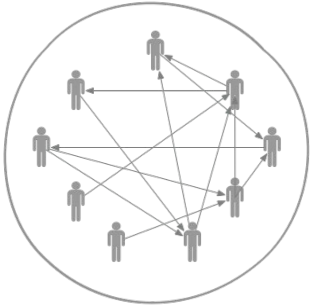
How do we use networks to design algorithms?

1. Diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials & interference
3.

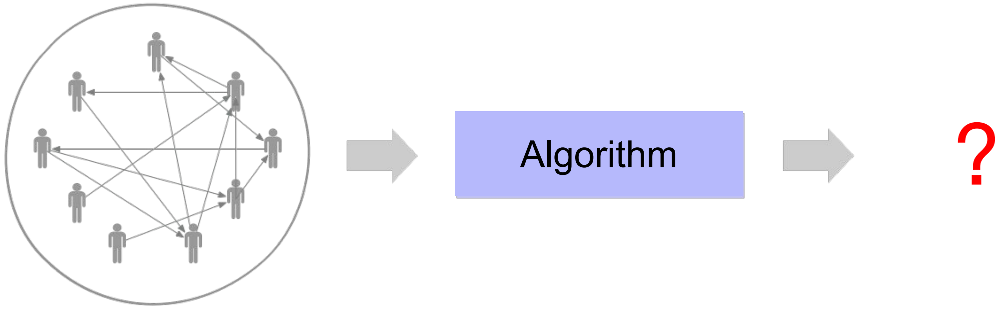
Diagnosing algorithmic bias



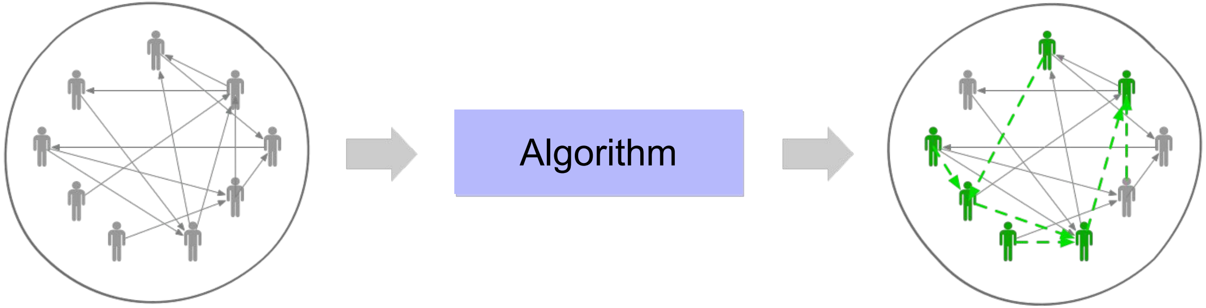
Diagnosing algorithmic bias



Diagnosing algorithmic bias

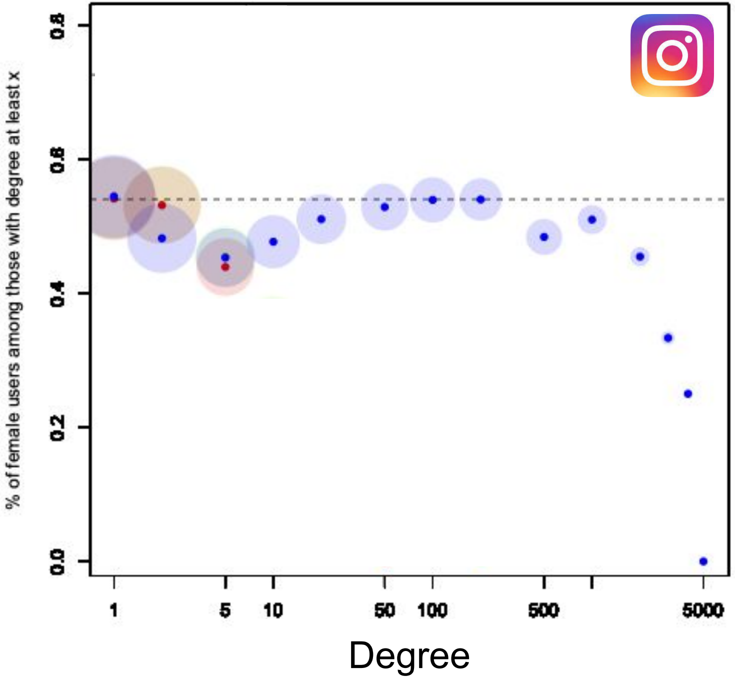


Diagnosing algorithmic bias



Distributional inequality in social capital

Instagram activity graph of likes and comments

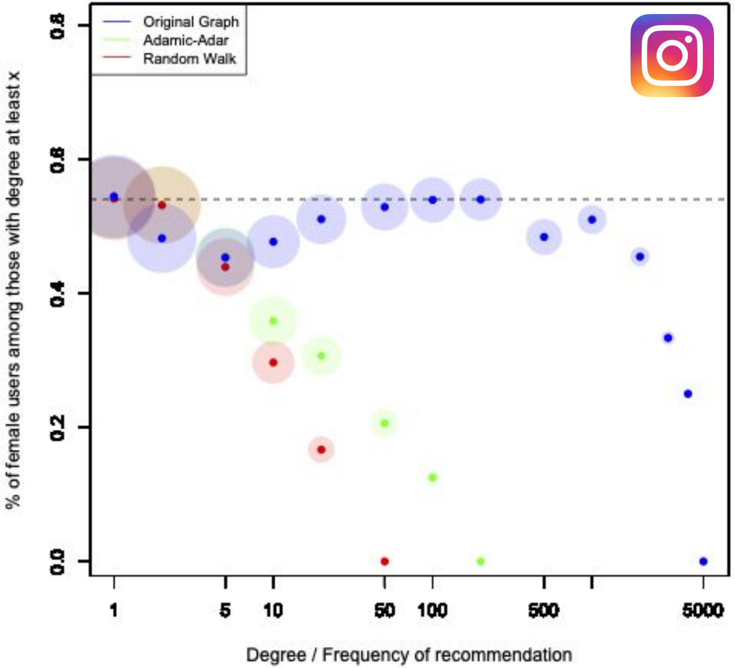


[Stoica et al, 2018]

- Groups: men (46%) and women (54%)
- Only **organic** connections
- Representation of women is *increasingly worse* for popular accounts

Distributional inequality in social recommendations

Instagram activity graph of likes and comments



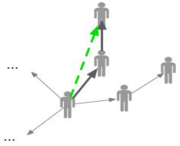
[Stoica et al, 2018]

- Common recommendation algorithms amplify degree inequality between men and women!
- Utility is equivalent to the number of connections after recommendation: $\deg_{RG}(u)$

Adamic Adar index:

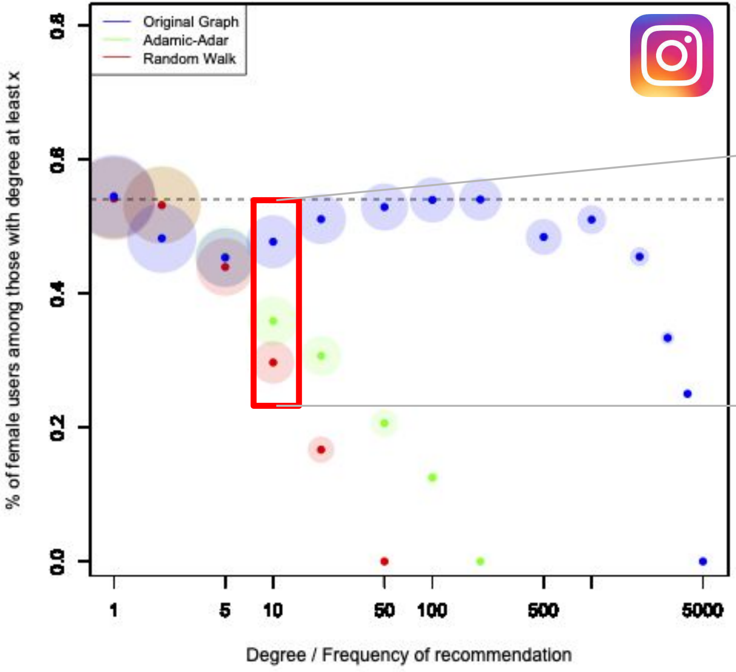
$$A(x,y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$$

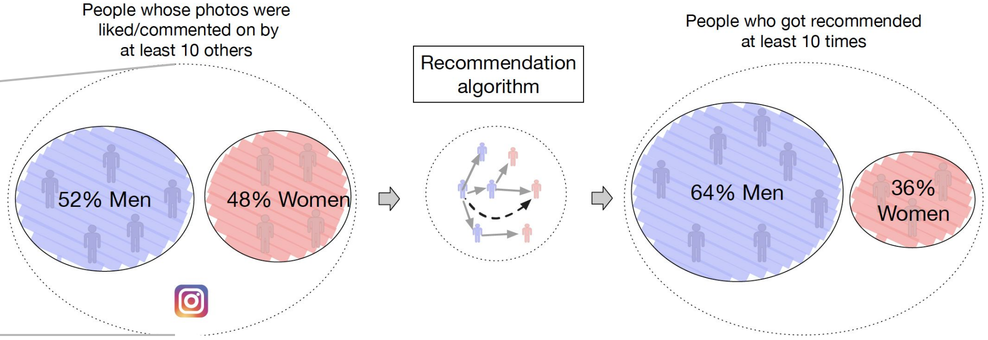
Random walk:



Distributional inequality in social recommendations

Instagram activity graph of likes and comments

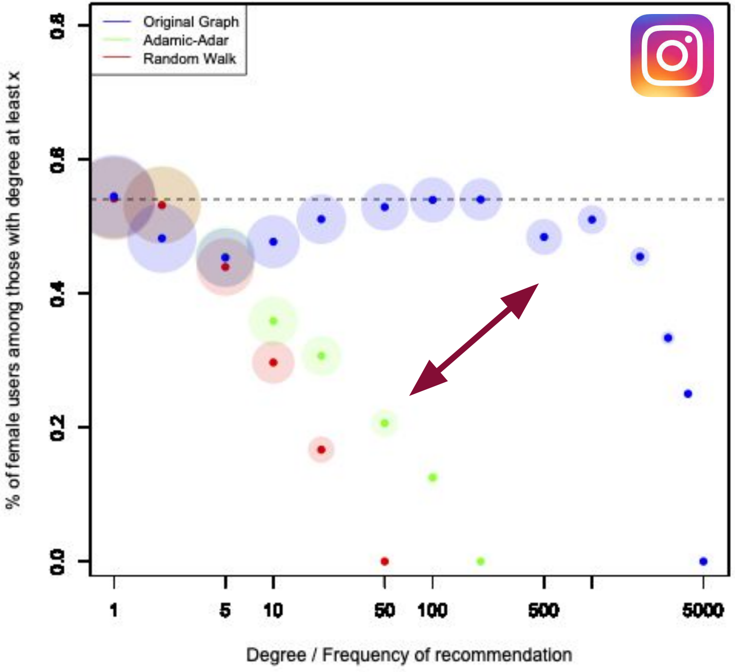




[Stoica et al, 2018]

Distributional inequality in social recommendations

Instagram activity graph of likes and comments



[Stoica et al, 2018]

- Common recommendation algorithms amplify degree inequality between men and women!
- Utility is equivalent to the number of connections after recommendation: $deg_{RG}(u)$

➡ **Algorithmic amplification of bias**

Diagnosing algorithmic bias

Benefit of connections activated by an algorithm:

Recommendation → Receive new connections through recommendations

Information diffusion → Be exposed to an information campaign

Clustering → Be targeted for assistance, help, new products or services, …

Ranking → Receive exposure by showing up in search results

Diagnosing algorithmic bias: is it always a problem?

Benefit of connections activated by an algorithm:

- Recommendation → Receive new connections through recommendations
- Information diffusion → Be exposed to an information campaign
- Clustering → Be targeted for assistance, help, new products or services, …
- Ranking → Receive exposure by showing up in search results

Diagnosing algorithmic bias

Benefit of connections activated by an algorithm:

Recommendation	Receive new connections through recommendations
Information diffusion	Be exposed to an information campaign
Clustering	Be targeted for assistance, help, new products or services, …
Ranking	Receive exposure by showing up in search results

Inequality in information diffusion

Empirical: Internet adoption / job referrals increases among the most prosperous in the presence of homophily [DiMaggio & Garip, 2011][Okafor, 2022]

CS (algorithmic): Defined as the social influence maximization problem

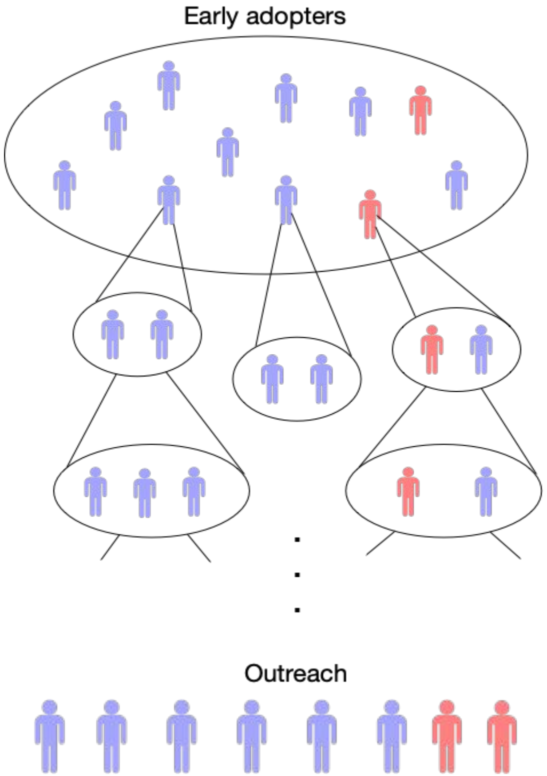
- Algorithms: greedy, centrality based (degree, distance centrality, etc)

Individual:

[Fish et al, 2019]

Group:

[Tsang et al, 2019]
[Ali et al, 2019]
[Stoica et al, 2020]



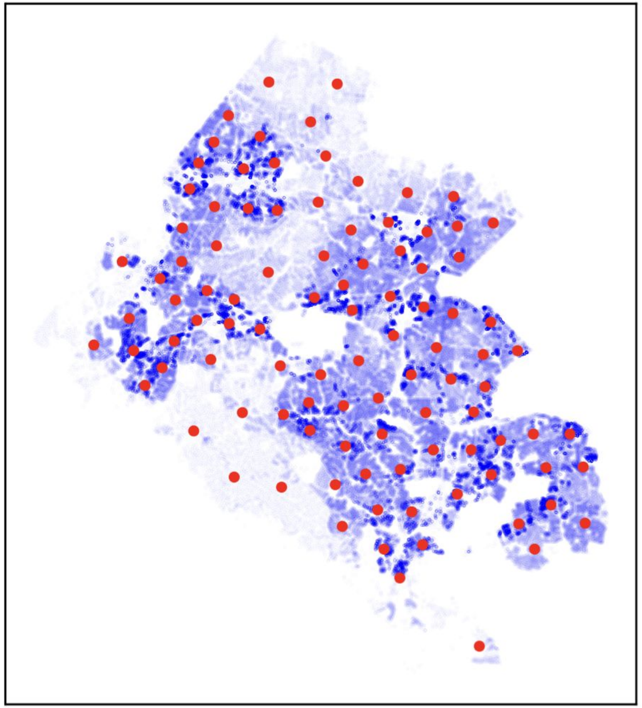
Diagnosing algorithmic bias

Benefit of connections activated by an algorithm:

- Recommendation → Receive new connections through recommendations
- Information diffusion → Be exposed to an information campaign
- Clustering → Be targeted for assistance, help, new products or services, …
- Ranking → Receive exposure by showing up in search results

Inequality in clustering: who benefits from a cluster?

Facility location: [Jung et al, 2019] show that clustering can be beneficial to highly clustered and dense groups, but not so much to others



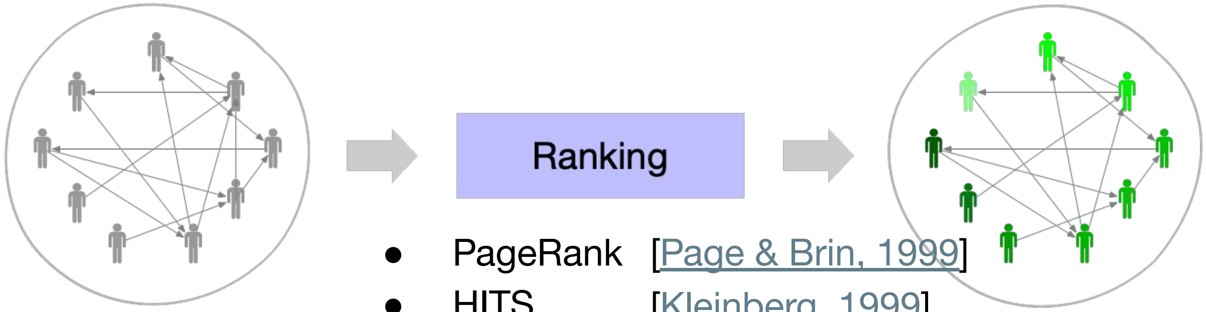
Diagnosing algorithmic bias

Benefit of connections activated by an algorithm:

Recommendation	Receive new connections through recommendations
Information diffusion	Be exposed to an information campaign
Clustering	Be targeted for assistance, help, new products or services, …
Ranking	Receive exposure by showing up in search results

Bias in ranking algorithms

Original graph G = (N,E) Activated graph G' = (N',E')



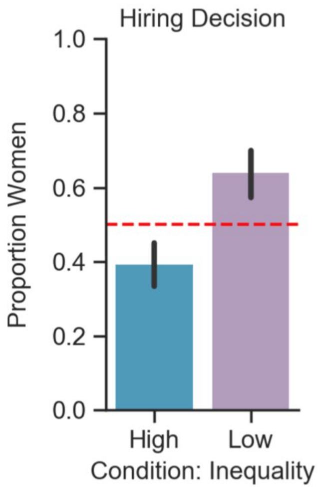
- PageRank [Page & Brin, 1999]
- HITS [Kleinberg, 1999]

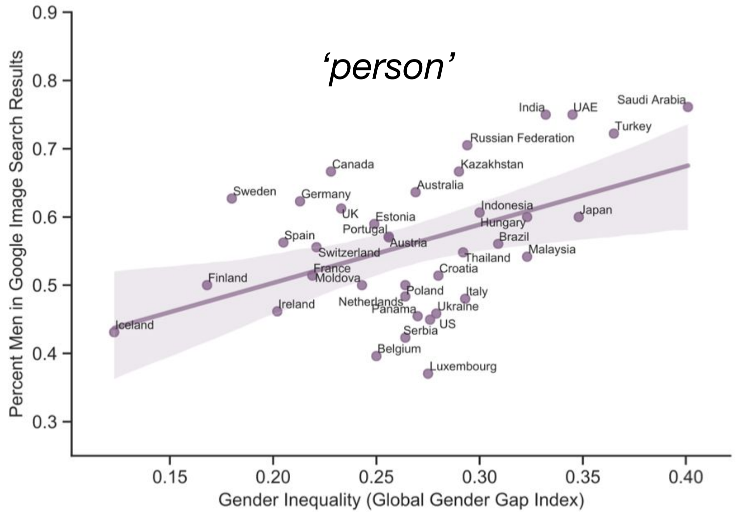
Application to ranking algorithms:

- Content search: Google, Bing, …
- Credibility / popularity metric

Minorities get 'pushed down'

Bias in ranking algorithms



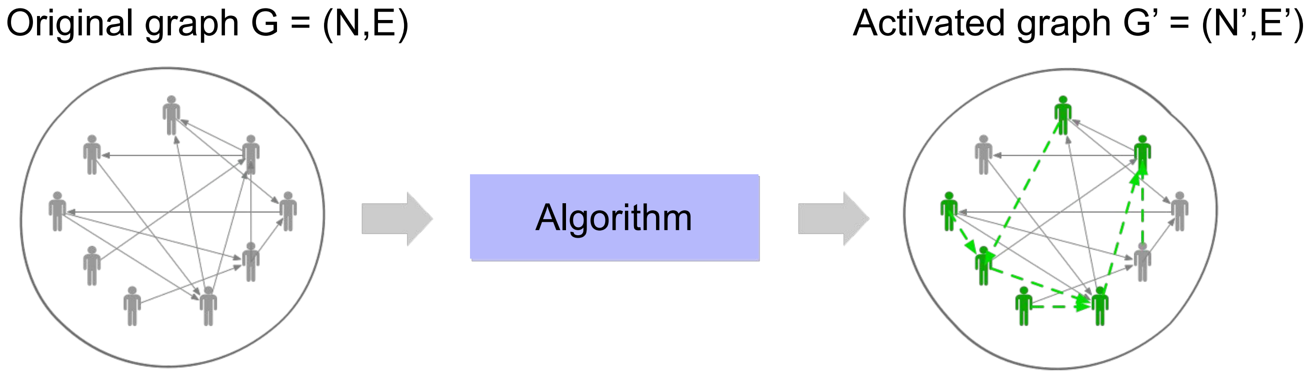


Vlasceanu and Amodio

Minorities get 'pushed down'

[Espin-Noboa et al, 2022]
[Vlasceanu & Amodio, 2022]

Diagnosing algorithmic bias: a unified formulation

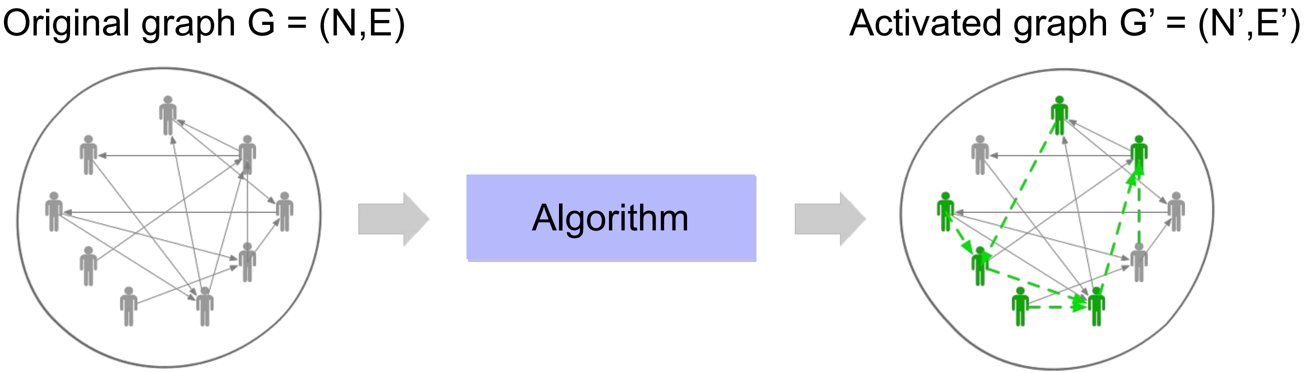


An algorithms outputs a subset of the nodes and a set of edges: $\mathcal{A} : G \rightarrow G', G' = (N', E')$

Evaluate the output through a gain function $f : G' \rightarrow \mathbb{R}$ that models one's social capital under $\mathcal{A}$

$$f(u) := \sum_{v \in N} \mathbb{P}((u, v) \in E'), \forall u \in N'$$

Diagnosing algorithmic bias: a unified formulation



Recommendation → E' is the set of newly created edges

$f(\cdot)$ is the number of new connections

$$f(u) := \sum_{v \in N} \mathbb{P}((u,v) \in E'), \forall u \in N'$$

Diagnosing algorithmic bias: a unified formulation

Original graph G = (N,E) Activated graph G' = (N',E')

Recommendation

$$f(u) := \sum_{v \in N} \mathbb{P}((u,v) \in E'), \forall u \in N'$$

Information diffusion

E' is the set of edges that actually transmit information

$f(\cdot)$ is the probability of getting the information

Diagnosing algorithmic bias: a unified formulation

Original graph G = (N,E)

Activated graph G' = (N',E')

Recommendation

Information diffusion

$$f(u) := \sum_{v \in N} \mathbb{P}((u,v) \in E'), \forall u \in N'$$

Clustering

E' is the set of edges within clusters

$f(\cdot)$ is the in-cluster degree

Diagnosing algorithmic bias: a unified formulation

Original graph G = (N,E)

Activated graph G' = (N',E')

Recommendation

Information diffusion

Clustering

Ranking

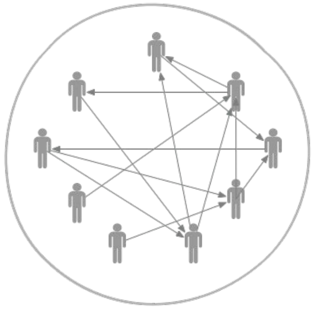
$$f(u) := \sum_{v \in N} \mathbb{P}((u,v) \in E'), \forall u \in N'$$

$E' = \emptyset$

$f(\cdot)$ is the ranking score

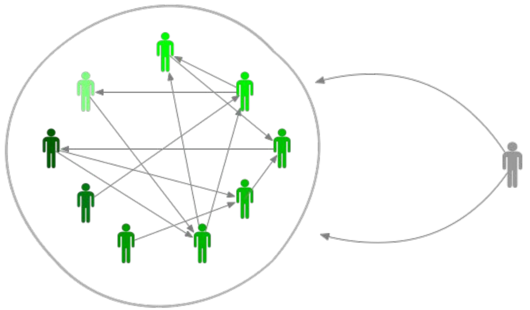
Diagnosing algorithmic bias: a unified formulation

Original graph G = (N,E)



Algorithm

Activated graph G' = (N',E')



Recommendation

Information diffusion

Clustering

Ranking

$$f(u) := \cancel{\sum_{v \in N}} \mathbb{P}((u,v) \in E'), \forall u \in N'$$

E' is the set of edges to be created with the new node

$f(\cdot)$ is the ranking score

How do we use networks to design algorithms?

1. Diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials

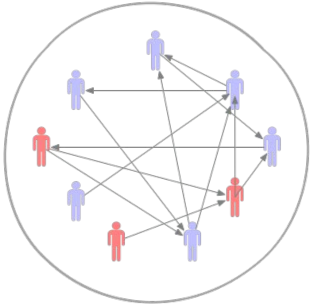
Diagnosing algorithmic bias

Original graph G = (N,E)

Activated graph G’ = (N’,E’)

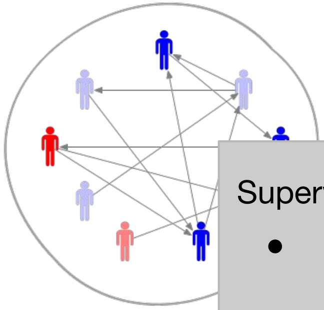
Diagnosing algorithmic bias: impact on different groups

Original graph G = (N,E)



Unsupervised learning

Activated graph G' = (N',E')

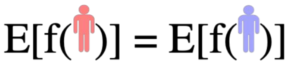


Supervised learning:

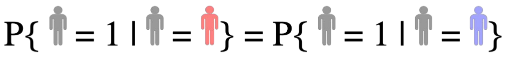
- Decision-making: select people who receive a positive outcome
- Known ground truth

Group fairness:

- Independence (average comparison):

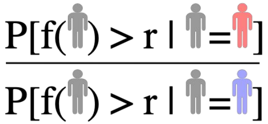


$\Leftrightarrow$



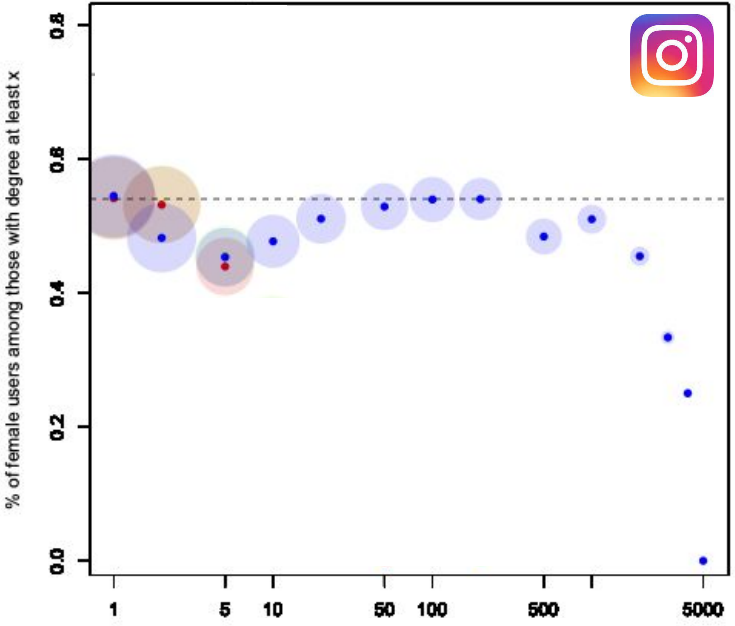
- Analyze distributional inequality in f:

behavior of



Distributional inequality in social capital

Instagram activity graph of likes and comments



Degree

[Stoica et al, 2018]

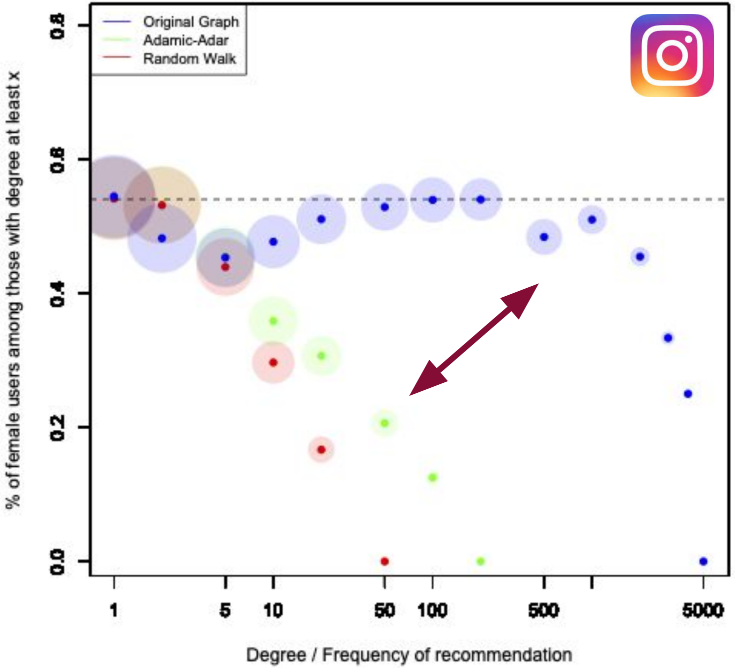
- Groups: men (46%) and women (54%)
- Only **organic** connections
- $f(u) = deg_{OG}(u)$
- Representation of each group on average does not tell the entire story:

$$E[f(\text{women})] = 2.25$$
$$E[f(\text{men})] = 2.52$$

- Representation of women is *increasingly worse* for popular accounts

Distributional inequality in social recommendations

Instagram activity graph of likes and comments



[Stoica et al, 2018]

- Common recommendation algorithms amplify degree inequality between men and women!
- Utility is equivalent to the number of connections after recommendation: $f(u) = deg_{RG}(u)$

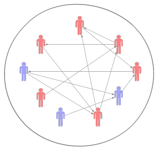
Algorithmic amplification of bias

How do we use networks to design algorithms?

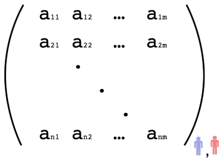
1. Diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials

Networks modeling for finding the root cause of bias

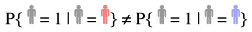
Model



Biased data



Algorithm



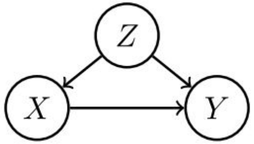
Biased outcome

Bias amplification?

Counterfactual reasoning:

- Infer causal relationships between parameters / variables
- Do not help measure amplification of bias

[Kusner et al, 2017]
[Kilbertus et al, 2017]
[Plecko & Bareinboim, 2022]



Models of network evolution:

- Explain where inequality or bias originates
- Predict a state of the world in the *absence* of a predictive system

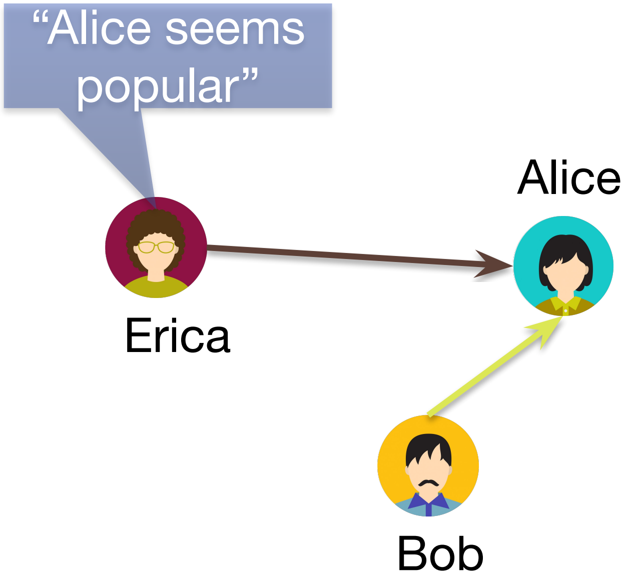
Evaluate the effect of a particular algorithm on the state of the network

Preferential attachment with homophily [Avin et al, 2015]

Model ingredients:

- **Minority-majority**: B label and R label
  - Fraction of R nodes = $r < ½$
- **Preferential attachment** (rich-get-richer): nodes connect w.p. proportional to degree

Organic



Preferential attachment with homophily [Avin et al, 2015]

Model ingredients:

- **Minority-majority**: B label and R label
    - Fraction of R nodes = r < ½
- **Preferential attachment** (rich-get-richer): nodes connect w.p. proportional to degree

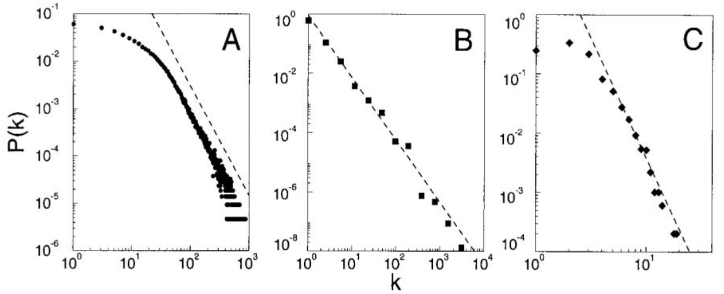


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

[Barabasi-Albert,1999]

Preferential attachment with homophily [Avin et al, 2015]

Model ingredients:

- **Minority-majority**: B label and R label
  - Fraction of R nodes = $r < ½$
- **Preferential attachment** (rich-get-richer): nodes connect w.p. proportional to degree
- **Homophily**: if different labels, connection is accepted w.p. ρ

Preferential attachment with homophily [Avin et al, 2015]

Degree distribution follows a power law at equilibrium:

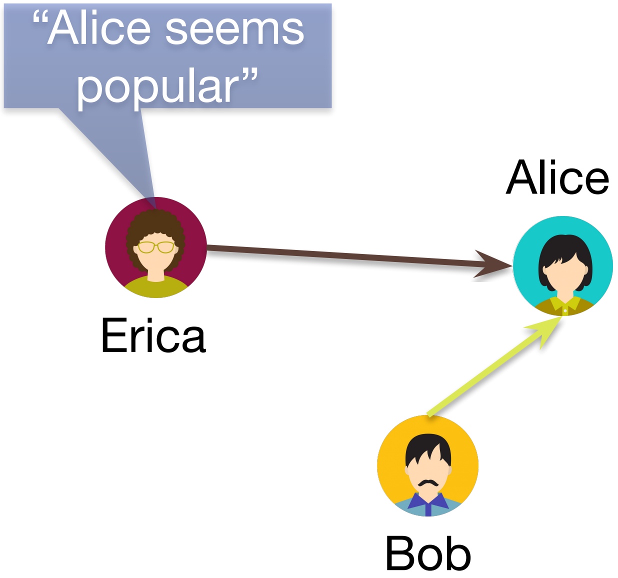
$$top_k(R) \sim k^{-\beta(R)}$$

$$top_k(B) \sim k^{-\beta(B)}$$

Theorem:

$$\underbrace{\beta(R) > 3 > \beta(B)}_{\text{gap}}$$

Organic



Preferential attachment with homophily [Avin et al, 2015]

Data: DBLP dataset of mentors-mentees

- ~400k people, male (79%) and female (21%)
- Female mentors avg. deg: 4.60
- Male mentors avg. deg: 5.25

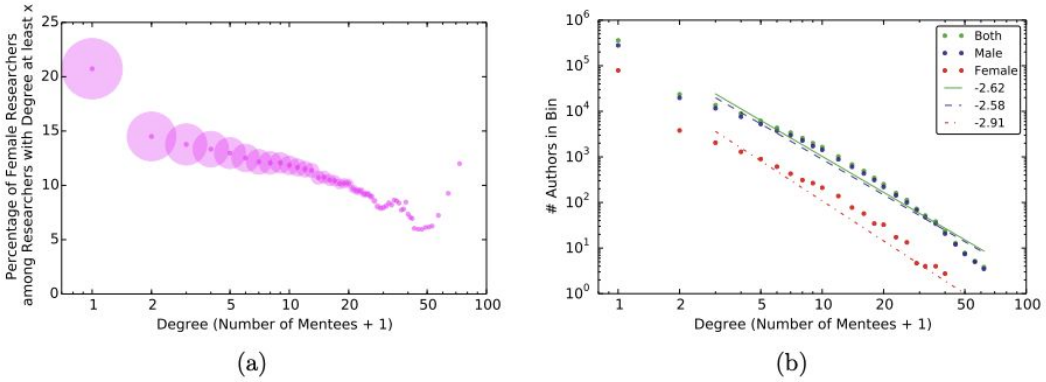


Figure 6: Glass ceiling effect in mentor graph: (a) percentage of females in the mentor population of degree at least k. Female start with 21% in the population and drop to below 15% when considering degree at least 2 (faculty members). It continues to decrease (ignoring small samples at the end, see text). Vertex size and darker color represent larger sample space. (b) The power-law-like degree distribution for both females and males. The exponent β for females is higher than for males, demonstrating the glass ceiling effect.

Preferential attachment with homophily [Avin et al, 2015]

Measures of inequality between *R and B*:

- Power inequality: $$\lim_{n\to\infty} \frac{\frac{1}{n(\mathrm{R})}\sum_{v\in \mathrm{R}}\delta(v)}{\frac{1}{n(\mathrm{B})}\sum_{v\in \mathrm{B}}\delta(v)} \leq \text{c for some constant c}$$

- Tail glass ceiling effect: there exists an increasing function k(n) such that:

$$\lim_{n\to\infty} top_{k(n)}(B) = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{top_{k(n)}(R)}{top_{k(n)}(B)} = 0$$

- Strong glass ceiling effect: $$\lim_{n\to\infty} \frac{\frac{1}{n(\mathrm{R})}\sum_{v\in \mathrm{R}}\delta(v)^2}{\frac{1}{n(\mathrm{B})}\sum_{v\in \mathrm{B}}\delta(v)^2} = 0$$

Preferential attachment with homophily [Avin et al, 2015]

Main results:

- Minority-majority
- Preferential attachment
- Homophily

- Power inequality
- Tail glass ceiling effect
- Strong glass ceiling effect

Preferential attachment with homophily [Avin et al, 2015]

Main results:

- Minority-majority
- Preferential attachment
- Homophily

→

- Power inequality
- Tail glass ceiling effect
- Strong glass ceiling effect

Preferential attachment with homophily [Avin et al, 2015]

Main results:

- Minority-majority
- Preferential attachment
- Homophily

- Power inequality
- Tail glass ceiling effect
- Strong glass ceiling effect

Preferential attachment with homophily [Avin et al, 2015]

Main results:

- Minority-majority
- Preferential attachment
- Homophily

- Power inequality
- Tail glass ceiling effect
- Strong glass ceiling effect

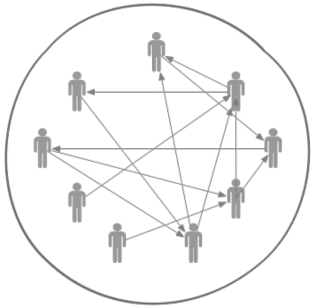
Preferential attachment with homophily [Avin et al, 2015]

Main results:

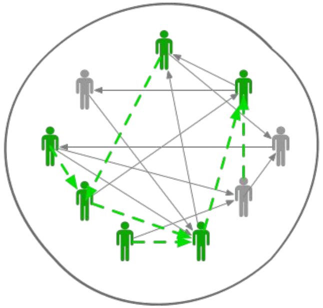
- Minority-majority
- Preferential attachment
- Homophily

- Power inequality?
- Tail glass ceiling effect?
- Strong glass ceiling effect

Diagnosing algorithmic bias



Algorithm



Benefit of connections activated by an algorithm:

Recommendation	Receive new connections through recommendations
Information diffusion	Be exposed to an information campaign
Clustering	Be targeted for assistance, help, new products or services, …
Ranking	Receive exposure by showing up in search results

Bias amplification in recommendation algorithms

Summary of results:

- Experimental results show a **bias amplification**
- Build a **theoretical explanation** for when bias amplifies in recommendation based on an evolving network model
- **Main ingredients** for bias creation and amplification:
  - Disparity in group sizes: minority (**R**), majority (**B**)
  - Preferential attachment (rich-get-richer effect)
  - Homophily (nodes in the same community connect)
  - Recommendations based on random walk of length 2

Model evolution with recommendations

At timestep t, a new edge is formed:

<u>Organic growth:</u>

[Avin et al, 2015]

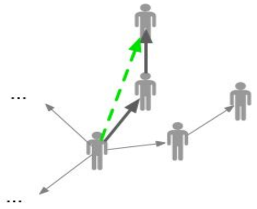
New node connects:

- randomly
- preferential attachment + homophily

Biased Preferential Attachment Model (BPAM)

<u>Recommendation model:</u>

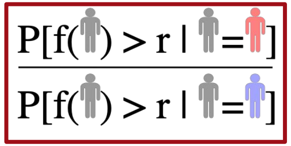
- organic growth
- **existing node connects through a random walk of length 2**



Degree distribution

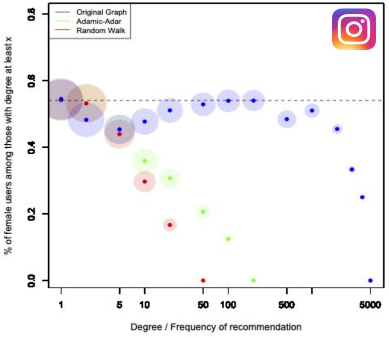
Organic growth:

$$top_k(\color{red}{R}) \sim k^{-\beta(R)}$$
$$top_k(\color{blue}{B}) \sim k^{-\beta(B)}$$



Recommendation model:

$$top_k'(\color{red}{R}) \sim k^{-\beta_{rec}(R)}$$
$$top_k'(\color{blue}{B}) \sim k^{-\beta_{rec}(B)}$$



Theorem*: For 0 < r < ½ and 0 < ρ < 1, for the graph sequences G(n) for the organic model and G'(n) for the recommendation model, the red and blue populations exhibit a power law degree distribution with coefficients:*

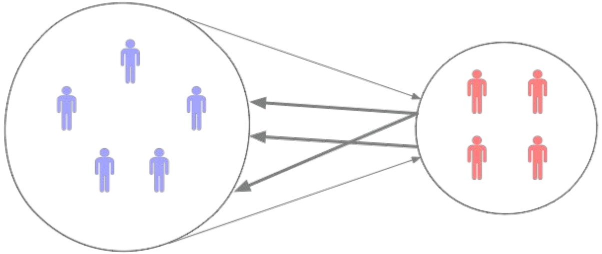
$$\underbrace{\beta_{rec}(\color{red}{R}) > \overbrace{\beta(\color{red}{R}) > 3 > \beta(\color{blue}{B})} > \beta_{rec}(\color{blue}{B})}_{\text{gap}}$$

Bias amplification for whom?

[Okafor, 2022] shows that a more homophilic demographic minority can overcome disadvantage in job referrals

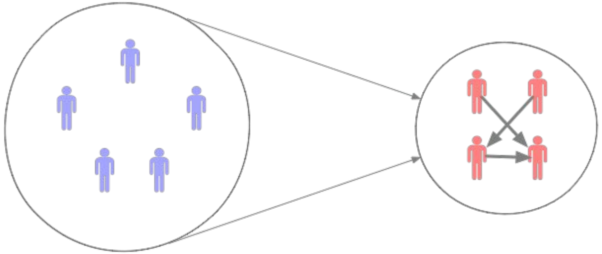
Symmetric homophily predicts majority advantage:

$$\underbrace{\beta_{rec}(R) > \underbrace{\beta(R) > 3 > \beta(B)} > \beta_{rec}(B)}_{\text{gap}}$$



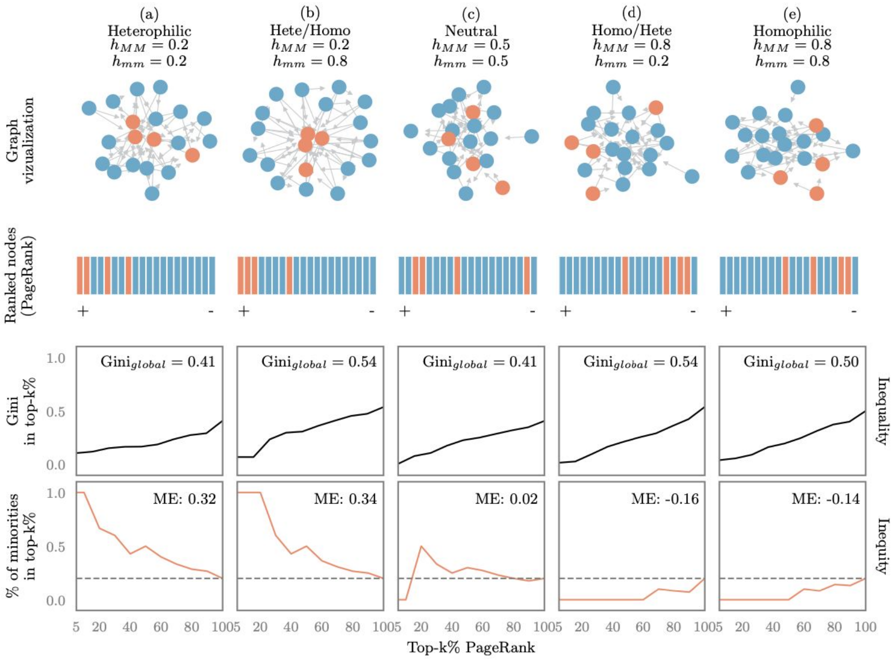
Asymmetric homophily leads to a reversal of bias (amplification):

$$\underbrace{\beta_{rec}(B) > \underbrace{\beta(B) > 3 > \beta(R)} > \beta_{rec}(R)}_{\text{gap}}$$



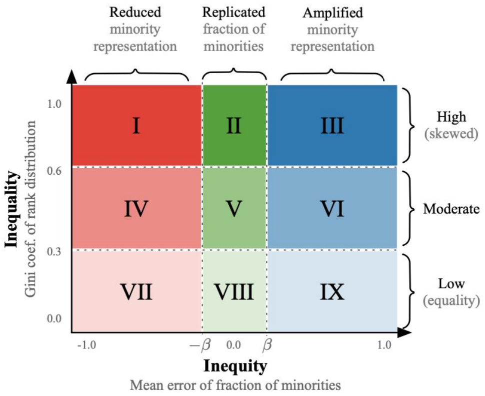
Bias amplification: recommendation and **ranking**

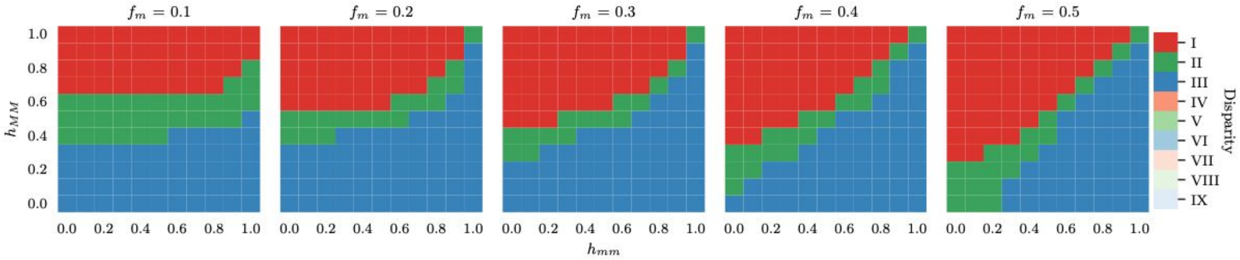
[Espin-Noboa et al, 2022] show the role of homophily/heterophily in the biased preferential attachment model in down-ranking minorities



Bias amplification: recommendation and **ranking**

[Espin-Noboa et al, 2022] show the role of homophily/heterophily in the biased preferential attachment model in down-ranking minorities: differentiated homophily





Knowledge of the network is essential

in diagnosing the impact of an algorithm

on different groups in a population

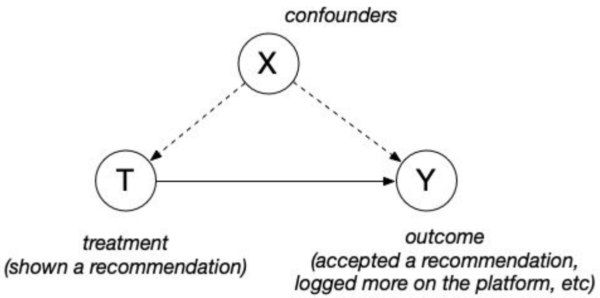
How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
    a. Unify unsupervised graph problems
    b. Define theoretical formulation for capturing distributional inequality
    c. Leverage network models for re-creating the root cause of bias
2. Using networks to test algorithms: randomized controlled trials & interference

Causality inference experiments on networks

Network experiments

- pharmaceutical companies researching the efficacy of a new medication
- policy makers understanding the impact of social welfare programs
- social media companies evaluating the impact of different recommendation algorithms on user engagement across their platforms



Potential outcomes model

Set-up: population of n individuals, a central planner that administers a treatment

- Treatment: binary variable T (let's assume a Bernoulli randomized design, T ~ Bin(n,p))
- Confounders: known attributes (potentially) X
- Outcome: real-valued Y

What are we estimating? $$TTE := \frac{1}{n}\sum_{i=1}^{n}(Y_i(1) - Y_i(0))$$

<u>Classic (non-network) model:</u>

- Stable Unit Treatment Value Assumption (SUTVA)

$$Y_i(T) = c_0 + c_i \cdot T_i \Rightarrow TTE = \frac{1}{n}\sum_{i=1}^{n} c_i$$

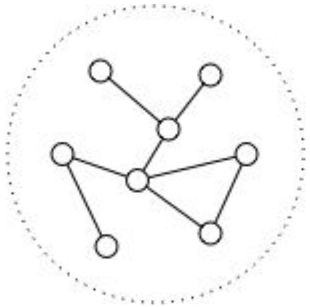
<u>Network interference model:</u>

- No more SUTVA!

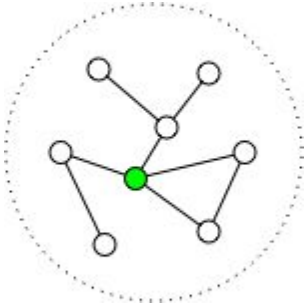
$$Y_i(T) = \sum_{S' \subseteq N_i} c_{i,S'} \prod_{j \in S'} T_j \Rightarrow TTE = \frac{1}{n}\sum_{i=1}^{n}\sum_{S' \subseteq N_i} c_{i,S'}$$

[Cortez-Rodriguez et al, 2022]

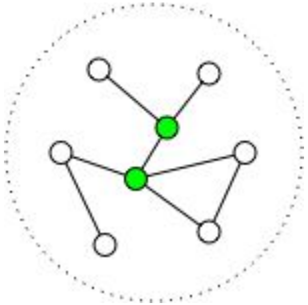
Network interference



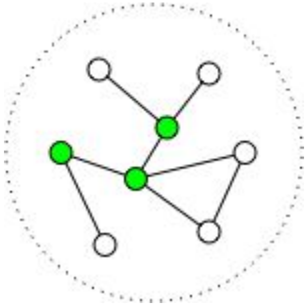
Network interference



Network interference

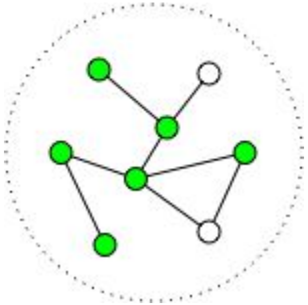


Network interference



Network interference

What is the issue?

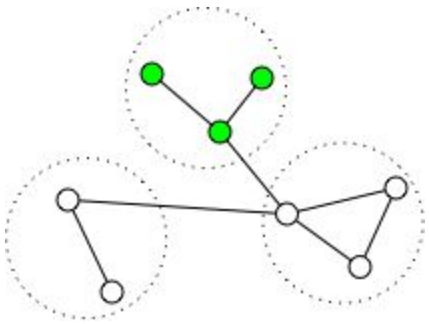


An estimator will have variance as large as the maximal degree: $O\left(\frac{Y_{max}^2 d^2}{np^d}\right)$ [Aronow et al, 2017]

Horvitz-Thompson estimator:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{obs}\left(\frac{\mathbb{I}(T \text{ treats all of } N_i)}{\mathbb{P}(T \text{ treats all of } N_i)} - \frac{\mathbb{I}(T \text{ does not treat all of } N_i)}{\mathbb{P}(T \text{ does not treat all of } N_i)}\right)$$

Network interference: solutions

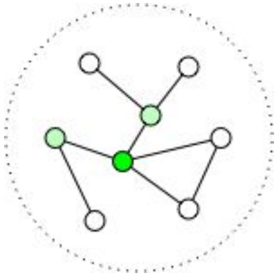


Randomized **clustered** design:

- Cluster the network
- Assume interference only within clusters
- Assign treatment at the level of the cluster

[Ugander et al, 2013]
[Eckles et al, 2016]

Network interference: bounds



	Assumptions on Network Structure		
Assumptions on Model Structure	C Disconnected Subcommunities	κ-restricted Growth	Fully General
Linear			OLS, *Bernoulli RD*; [40, 17, 6, 9, 29, 10]
Generalized Linear	Directions for Future Work		Regression/Machine Learning, *Bernoulli RD*; [10]
β-order Interactions			Pseudoinverse estimator, *Bernoulli RD*; $O\left(\frac{Y_{\max}^2 d^{2\beta+2}}{np^\beta}\right)$
Arbitrary Neighborhood Interference	*Horvitz-Thompson*, Cluster RD; $O\left(\frac{Y_{\max}^2}{Cp}\right)$; [34, 31, 18, 39]	*Horvitz-Thompson*, Randomized Cluster RD; $O\left(\frac{Y_{\max}^2 \kappa^4 d^2}{np}\right)$; [17, 14, 41, 42]	*Horvitz-Thompson, Bernoulli RD*; $O\left(\frac{Y_{\max}^2 d^2}{np^d}\right)$; [1]

[Cortez-Rodriguez et al, 2022]

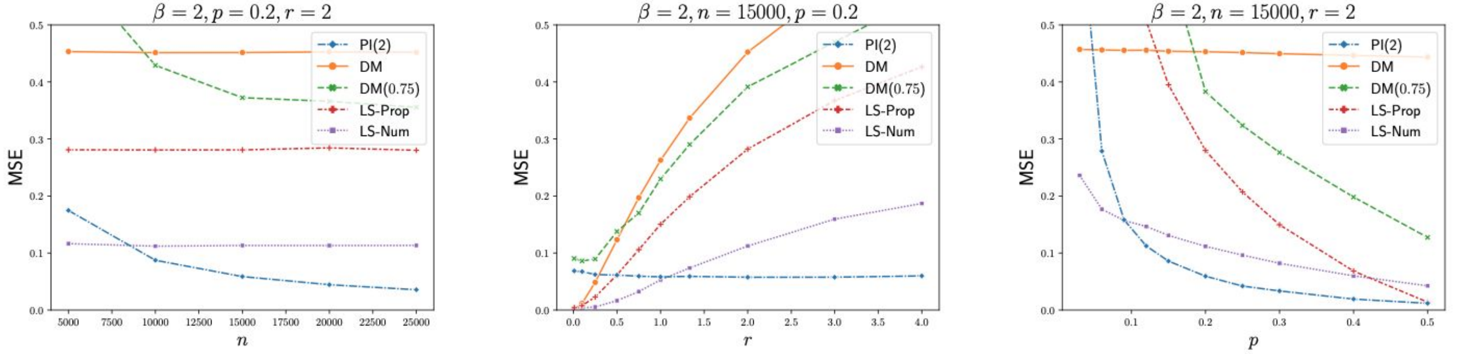
[Ugander et al, 2013]
[Eckles et al, 2016]

Network interference

[Cortez-Rodriguez et al, 2022] proposes a new variant of the Horvitz-Thompson estimator:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{obs} \sum_{S\subseteq N_i, |S|\leq\beta} g(S) \prod_{j\in S}\left(\frac{T_j}{\mathbb{P}(T_j=1)} - \frac{1-T_j}{\mathbb{P}(T_j=0)}\right),$$

$$g(S) = \prod_{s\in S}(1-\mathbb{P}(T_s=1)) - \prod_{s\in S}(-\mathbb{P}(T_s=1)), \forall S\subseteq[n]$$



(a) Varying population size (b) Varying direct:indirect effects (c) Varying treatment budget

DM = difference in means, LS = least squares

Conclusions and open directions

- Generalizing beyond parametric network models
    - What network properties cause bias to be projected onto different embeddings?

Conclusions and open directions

- Generalizing beyond parametric network models
- Bridging causality and fairness
    - How can infer the causal connection between algorithms and bias?

Conclusions and open directions

- Generalizing beyond parametric network models
- Bridging causality and fairness
- Feedback loops and long-term effects
  - Asymptotic analysis? Modeling feedback as strategic behavior?

Conclusions and open directions

- Generalizing beyond parametric network models
- Bridging causality and fairness
- Feedback loops and long-term effects
- Multi-objective optimization
  - How do we balance multiple objectives? How do we incorporate fairness beyond a constraint?

Conclusions and open directions

- Generalizing beyond parametric network models
- Bridging causality and fairness
- Feedback loops and long-term effects
- Multi-objective optimization
- Interdisciplinary studies
  - How can we bridge methods from social sciences, optimization, graph-theoretical modeling to understand patterns of connection / behavior and model the right objectives?

Conclusions and open directions

Thank you!

- Generalizing beyond parametric network models
- Bridging causality and fairness
- Feedback loops and long-term effects
- Multi-objective optimization
- Interdisciplinary studies

MD4SG

Mechanism Design for Social Good

RESEARCH-ARTICLE

Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness

Authors: Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, Stratis Tsirtsis Authors Info & Claims

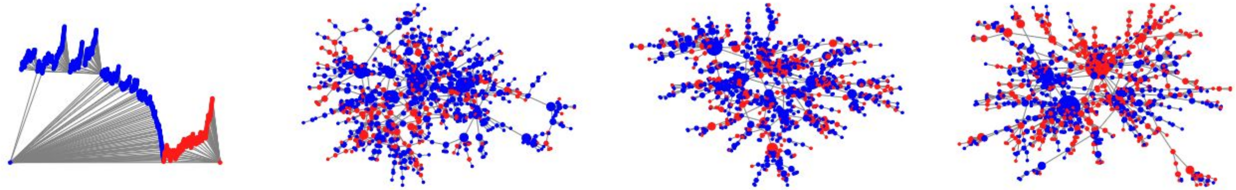
FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency • March 2021 • Pages 489–503 • https://doi.org/10.1145/3442188.3445912

References

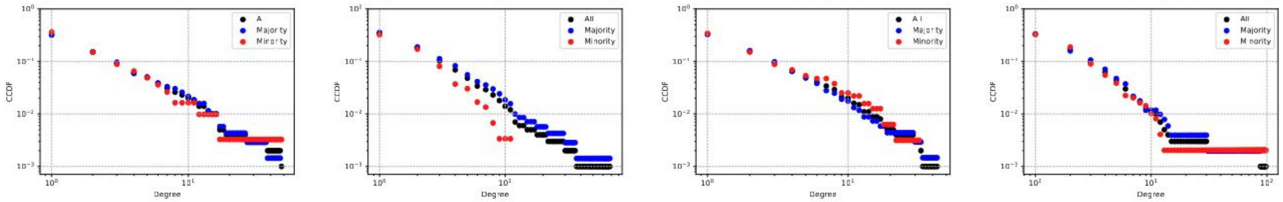
[Schlapher et al, 2014] Schläpfer, Markus, et al. "The scaling of human interactions with city size." *Journal of the Royal Society Interface* 11, no. 98. 2014.
[Smith-Clarke et al, 2014] Smith-Clarke, Christopher, Afra Mashhadi, and Licia Capra. "Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks." *SIGCHI*. 2014.
[Putnam, 2000] Putnam, Robert D. "Bowling alone: America's declining social capital." In *Culture and politics*, pp. 223-234. Palgrave Macmillan, New York, 2000.
[Gündoğdu et al, 2019] Gündoğdu, Didem, et al. "The bridging and bonding structures of place-centric networks: Evidence from a developing country." *PloS one* 14.9 (2019): e0221148.
[Chetty et al, 2022] Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M. and Johnston, D., 2022. Social capital I: measurement and associations with economic mobility. *Nature*, *608*(7921), pp.108-121.
[DiMaggio & Garip, 2011] DiMaggio, Paul, and Filiz Garip. "How network externalities can exacerbate intergroup inequality." *American Journal of Sociology* 116.6 (2011): 1887-1933.
[Barabasi-Albert,1999] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.
[Stoica et al, 2018] Stoica, A.A., Riederer, C. and Chaintreau, A., 2018, April. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In *The Web Conference* (pp. 923-932).
[Plecko & Bareinboim, 2022] Plecko, D. and Bareinboim, E., 2022. Causal fairness analysis. *arXiv preprint arXiv:2207.11385*.
[Kusner et al, 2017] Kusner, M.J., Loftus, J., Russell, C. and Silva, R., 2017. Counterfactual fairness. *Neurips*, *30*.
[Kilbertus et al, 2017] Kilbertus, N., Rojas Carulla, M., Parascandolo, G., Hardt, M., Janzing, D. and Schölkopf, B., 2017. Avoiding discrimination through causal reasoning. *Neurips*, *30*.
[Fish et al, 2019] Fish, B., Bashardoust, A., Boyd, D., Friedler, S., Scheidegger, C. and Venkatasubramanian, S., 2019, May. Gaps in information access in social networks?. In *The Web Conference* (pp. 480-490).
[Tsang et al, 2019] Tsang, A., Wilder, B., Rice, E., Tambe, M. and Zick, Y., 2019. Group-fairness in influence maximization. *arXiv preprint arXiv:1903.00967*.
[Ali et al, 2019] Ali, J., Babaei, M., Chakraborty, A., Mirzasoleiman, B., Gummadi, K. and Singla, A., 2021. On the fairness of time-critical influence maximization in social networks. *IEEE Transactions on Knowledge and Data Engineering*.
[Stoica et al, 2020] Stoica, A.A., Han, J.X. and Chaintreau, A., 2020, April. Seeding network influence in biased networks and the benefits of diversity. In *The Web Conference 2020* (pp. 2089-2098).
[Jung et al, 2019] Jung, C., Kannan, S. and Lutz, N., 2019. A center in your neighborhood: Fairness in facility location. *arXiv preprint arXiv:1908.09041*.
[Page & Brin, 1999] Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. "*The PageRank citation ranking: Bringing order to the web*." Stanford InfoLab.
[Kleinberg, 1999] Kleinberg, Jon M. *"Authoritative sources in a hyperlinked environment." Journal of the ACM (JACM) 46, no. 5*, 604-632. 1999.
[Espin-Noboa et al, 2022] Espín-Noboa, Lisette, Claudia Wagner, Markus Strohmaier, and Fariba Karimi. *"Inequality and inequity in network-based ranking and recommendation algorithms."* Scientific reports 12, no. 1. 2022.
[Vlasceanu & Amodio, 2022] Vlasceanu, Madalina, and David M. Amodio. *"Propagation of Societal Gender Inequality by Internet Search Algorithms."* 2022.
[Avin et al, 2015] Avin, C., Keller, B., Lotker, Z., Mathieu, C., Peleg, D. and Pignolet, Y.A. Homophily and the glass ceiling effect in social networks. In ITCS (pp. 41-50), 2015.
[Okafor, 2022] Okafor, C.O., 2020. Social Networks as a Mechanism for Discrimination. *arXiv e-prints*, pp.arXiv-2006.
[Cortez-Rodriguez et al, 2022] Cortez, M., Eichhorn, M. and Yu, C.L., 2022. Exploiting neighborhood interference with low order interactions under unit randomized design. *arXiv preprint arXiv:2208.05553*.
[Ugander et al, 2013] Ugander, J., Karrer, B., Backstrom, L. and Kleinberg, J., 2013, August. Graph cluster randomization: Network exposure to multiple universes. In *KDD* (pp. 329-337).
[Eckles et al, 2016] Eckles, D., Karrer, B. and Ugander, J., 2016. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, *5*(1).

Additional slides

Biased preferential attachment model illustration



(a) $r = 0.3, \rho = 0$ (b) $r = 0.3, \rho = 0.3$ (c) $r = 0.3, \rho = 1$ (d) $r = 0.5, \rho = 0.3$



(e) $r = 0.3, \rho = 0$ (f) $r = 0.3, \rho = 0.3$ (g) $r = 0.3, \rho = 1$ (h) $r = 0.5, \rho = 0.3$

Figure 3.4: Networks generated from the Biased Preferential Attachment model (top row) and their respective cumulative complementary distribution functions, by community (bottom row), for different parameters.

Model for biased networks

Biased preferential attachment model:

- Minority-majority: blue (B) label and red (R) label (% of red nodes < ½)
- Rich-get-richer: nodes connect w.p. proportional to degree
- Homophily: if different labels, connection is accepted with a certain probability

⇒ known to exhibit inequality in the degree distribution of the two communities[3]

$$top_k(R) \sim k^{-\beta(R)}$$
$$top_k(B) \sim k^{-\beta(B)}$$

$$\beta(\underbrace{R) > 3 > \beta(B}))$$

Necessary and sufficient conditions: **groups, homophily, preferential attachment**

[3]Avin, Chen, et al. "Homophily and the glass ceiling effect in social networks." ITCS. 2015.

Degree distribution

Organic growth:

$$top_k(\mathbf{R}) \sim k^{-\beta(R)}$$

$$top_k(\mathbf{B}) \sim k^{-\beta(B)}$$

Recommendation model:

$$top_k'(\mathbf{R}) \sim k^{-\beta_{rec}(R)}$$

$$top_k'(\mathbf{B}) \sim k^{-\beta_{rec}(B)}$$

***Theorem**: For 0 < r < ½ and 0 < ρ < 1, for the graph sequences G(n) for the organic model and G'(n) for the recommendation model, the red and blue populations exhibit a power law degree distribution with coefficients:*

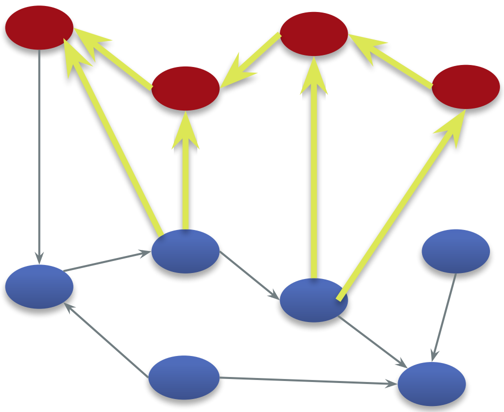
$$\beta_{rec}(\mathbf{R}) > \beta(\mathbf{R}) > 3 > \beta(\mathbf{B}) > \beta_{rec}(\mathbf{B})$$

gap

Proof sketch

r

α



'Wealth' of red nodes:

- Fraction of edges towards R

$$\alpha_t = \sum_{v \in R} in\deg(v)/t$$

Define a function F as the rate of growth of α_t

- F has a fixed point α $\Rightarrow \alpha_t \to \alpha < r$

Organic growth		Recommendation model
α	>	α'

Proof sketch

Evolution equation:

- When does a node of degree k get a new link

Randomly

Preferential attachment

$T_t^R =$ rate at which R nodes receive edges through **randomness**

$k \cdot C_t^R =$ rate at which R nodes receives edges through **preferential attachment**

$$top_k(R) \sim k^{-\beta(R)}$$

$$top_k(B) \sim k^{-\beta(B)}$$

$$\beta(R) = 1 + \frac{1}{C^R}$$

$$\beta(B) = 1 + \frac{1}{C^B}$$

Proof sketch

Goal: compute evolution equation and cl n solutions...

Big mess!

$$\beta_{rec}(\mathbf{R}) > \quad \beta(\mathbf{B}) > \beta_{rec}(\mathbf{B})$$

Key idea: at equilibrium, the rate at which red edges appear must equal the current fraction of red edges, as it does not evolve anymore

Invariant equation modeling asymptotic dynamics of degree distribution

Invariant equation

Organic growth:

$$\alpha \cdot C^R + r \cdot T^R = \alpha$$

Recommendation model:

$$\alpha' \cdot C'^R + r \cdot T'^R = \alpha'$$

$$\alpha > \alpha' \Rightarrow C^R > C'^R \Rightarrow \beta'(R) > \beta(R)$$

$$\beta'(R) > \beta(R) > 3 > \beta(B) > \beta'(B)$$

Degree distribution

Organic growth:

$$top_k(R) \sim k^{-\beta(R)}$$
$$top_k(B) \sim k^{-\beta(B)}$$

Recommendation model:

$$top_k'(R) \sim k^{-\beta_{rec}(R)}$$
$$top_k'(B) \sim k^{-\beta_{rec}(B)}$$

Majority has degree advantage + homophily:

$$\beta_{rec}(R) > \beta(R) > 3 > \beta(B) > \beta_{rec}(B)$$

Minority has degree advantage + homophily:

$$\beta_{rec}(B) > \beta(B) > 3 > \beta(R) > \beta_{rec}(R)$$