Interventions for mitigating Algorithmic Inequality in Social Networks

Ana-Andreea Stoica,

Max Planck Institute for Intelligent Systems, Tübingen, Germany

May, 2023

Algorithms may amplify patterns of discrimination

KHARI JOHNSON BUSINESS MAY 16, 2022 10:25 AM

Feds Warn Employers Against Discriminatory Hiring Algorithms

As AI invades the interview process, the DOJ and EECC have provided guidance to protect people with disabilities from bias.

BUSINESS

HUD is reviewing Twitter's and Google's ad practices as part of housing discrimination probe

By Tracy Jan and Elizabeth Dworkin
March 28, 2019 at 6:59 p.m. EDT

TECH POLICY

Facebook's ad algorithms are still excluding women from seeing jobs

Its ad-delivery system is excluding women from opportunities without regard to their qualifications. That would be illegal under US employment law.

By Karen Hao April 9, 2021

The Death and Life of an Admissions Algorithm

U of Texas at Austin has stopped using a machine-learning system to evaluate applicants for its Ph.D. in computer science. Critics say the system exacerbates existing inequality in the field.

By Lilah Burke // December 14, 2020

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
2. Using networks to test algorithms: randomized controlled trials
3. Build interventions to mitigate algorithmic bias
    a. In designing fair information diffusion campaigns
    b. In designing fair committees in opinion aggregation settings
    c. A theoretical framework for navigating trade-offs

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
2. Using networks to test algorithms: randomized controlled trials
3. Build interventions to mitigate algorithmic bias
   a. In designing fair information diffusion campaigns
   b. In designing fair committees in opinion aggregation settings
   c. A theoretical framework for navigating trade-offs

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
2. Using networks to test algorithms: randomized controlled trials
3. Build interventions to mitigate algorithmic bias
    a. In designing fair information diffusion campaigns
    b. In designing fair committees in opinion aggregation settings
    c. A theoretical framework for navigating trade-offs

Overview of published and ongoing projects

1. Diagnosing *when* and *how* an algorithm is amplifying bias

    - A.-A. Stoica, C. Riederer, and A. Chaintreau. *"Algorithmic glass ceiling in social networks: the effects of social recommendations on network diversity"*. The Web Conference. 2018.
    - A.-A. Stoica and A. Chaintreau. *"Bias in spectral embeddings: the case of recommendation algorithms on social networks".* Manuscript in preparation. 2022.

2. Building interventions for mitigating such bias

    - A.-A. Stoica, J.X. Han, and A. Chaintreau. *"Seeding network influence and the benefit of diversity"*. The Web Conference. 2020.
    - A.-A. Stoica, A. Chakraborty, P. Dey, and K.P. Gummadi. *"Minimizing margin of victory for political and educational districting"*. AAMAS. 2020.
    - A.-A. Stoica and C. Papadimitriou. *"Strategic clustering"*. In submission. 2022.

Information diffusion

(Social influence maximization problem)

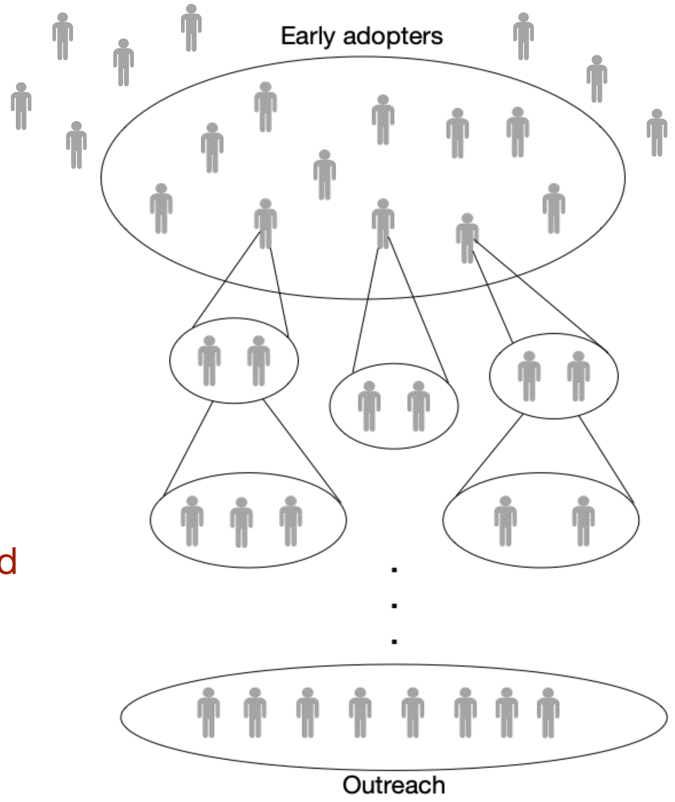
- Given a network G, with diffusion model as independent cascade with probability p, pick the best k early-adopters ('seeds') that maximize outreach:[1]

$$S^* = \text{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|, \quad \text{s.t. } |S| \le k$$

NP-hard

- Algorithms that choose based on:
    - Centrality: degree, distance centrality, …
    - Iteratively: greedy

Agnostic to communities



[1] Kempe, David, Jon Kleinberg, and Éva Tardos. "*Maximizing the spread of influence through a social network.*" In Proceedings of the ninth ACM SIGKDD Conference, pp. 137-146. 2003.

Information diffusion

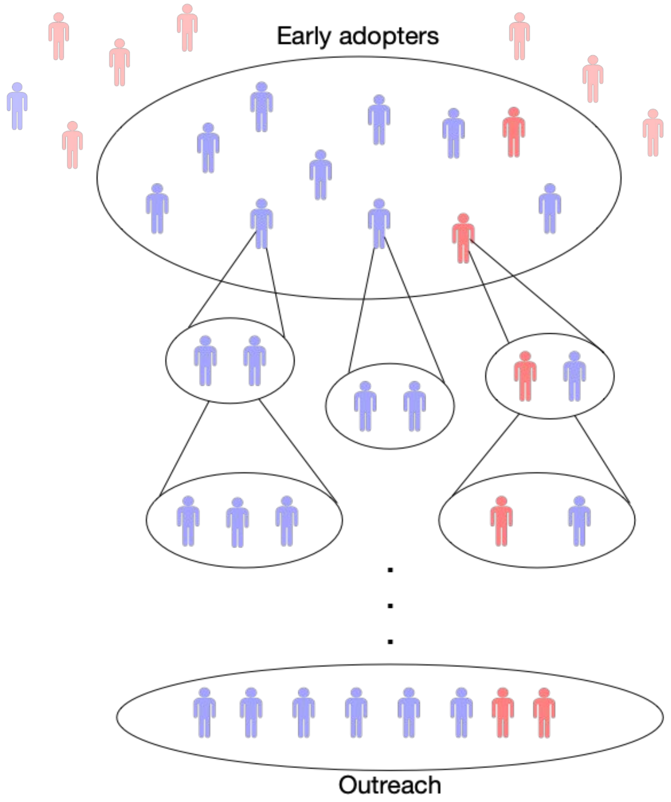
(Social influence maximization problem)

- Given a network G, with diffusion model as independent cascade with probability p, pick the best k early-adopters (‘seeds’) that maximize outreach:

$$S^* = \operatorname{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S, p)|, \\ \text{s.t. } |S| \leq k$$

- Algorithms that choose based on:
  - Centrality: degree, distance centrality, …
  - Iteratively: greedy

⇒ Bias in centrality measures and social structure gets reproduced[2]



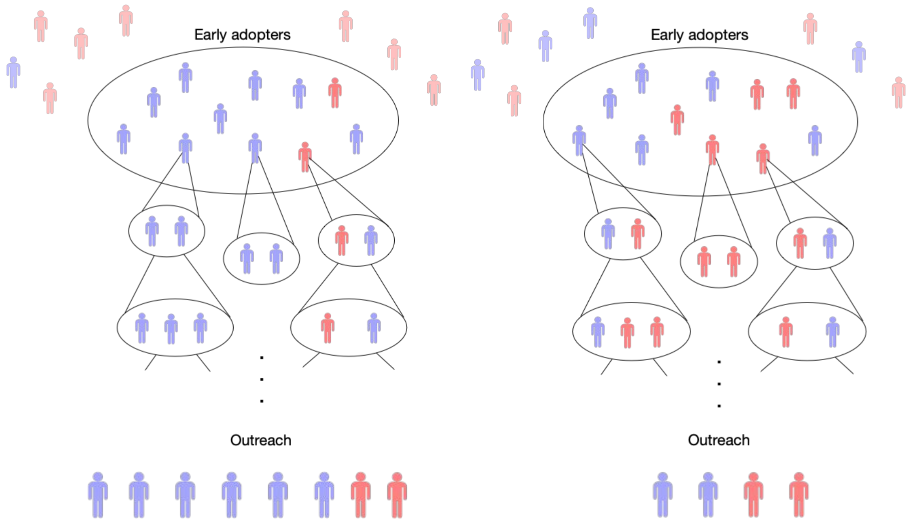
[2] Fish, Benjamin, et al. “Gaps in information access in social networks”. *The World Wide Web Conference.* ACM, 2019.

Information diffusion

- Parity constraint in an optimization function:

$$S^* = \mathrm{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|\phi_G(S,p) \cap R|)}{\mathbb{E}(|\phi_G(S,p) \cap B|)} \simeq \frac{|R|}{|B|}$$

➡ Fairness-efficiency trade-off



no constraint *parity constraint*

Our approach:

- Partially known networks ⇒ centrality measures (# of connections etc)
- Model of network growth & tap into inactive communities
- Theoretical conditions for when **equity increases efficiency (outreach)**

[2] Fish, Benjamin, et al. "Gaps in information access in social networks". *The World Wide Web Conference.* ACM, 2019.

Information diffusion

Just a Few Seeds More:
Value of Network Information for Diffusion*

Mohammad Akbarpour†
Suraj Malladi‡
Amin Saberi§

Random seeding with extra x nodes is comparable to optimal seeding (for small x)

Our approach:

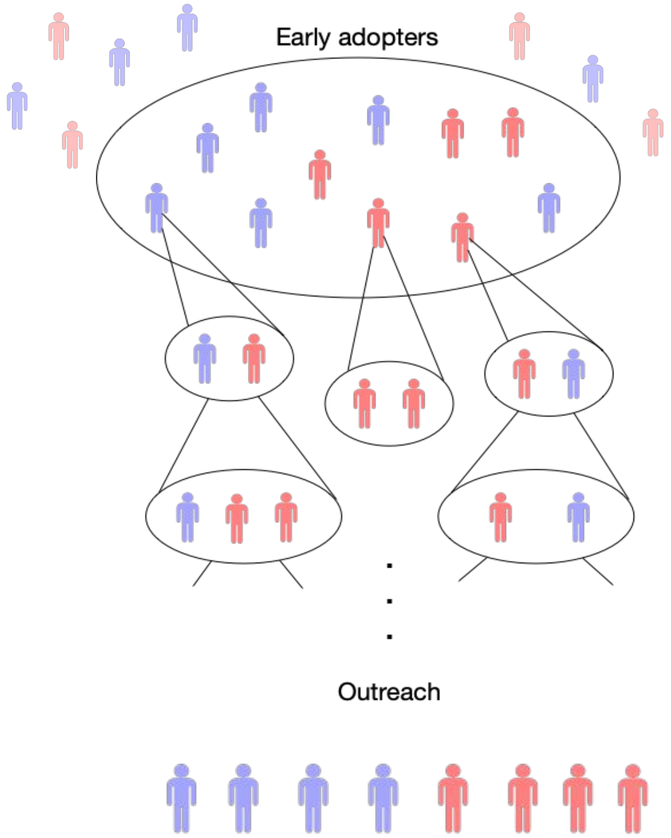
- Partially known networks ⇒ centrality measures (# of connections etc)
- Model of network growth & tap into inactive communities
- Theoretical conditions for when **equity increases efficiency (outreach)**

Information diffusion

- **Our vision:** bias as a sign of inefficiency
  - Diversity: tap into inactivated communities in the *early adopters* set

$$S^* = \mathrm{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \simeq \frac{|R|}{|B|}$$

- Seeding can be done with awareness of labels: **statistical parity** in your campaign (even if choosing less connected people)
  - Parity seeding (strict)
  - Diversity seeding (relaxed)

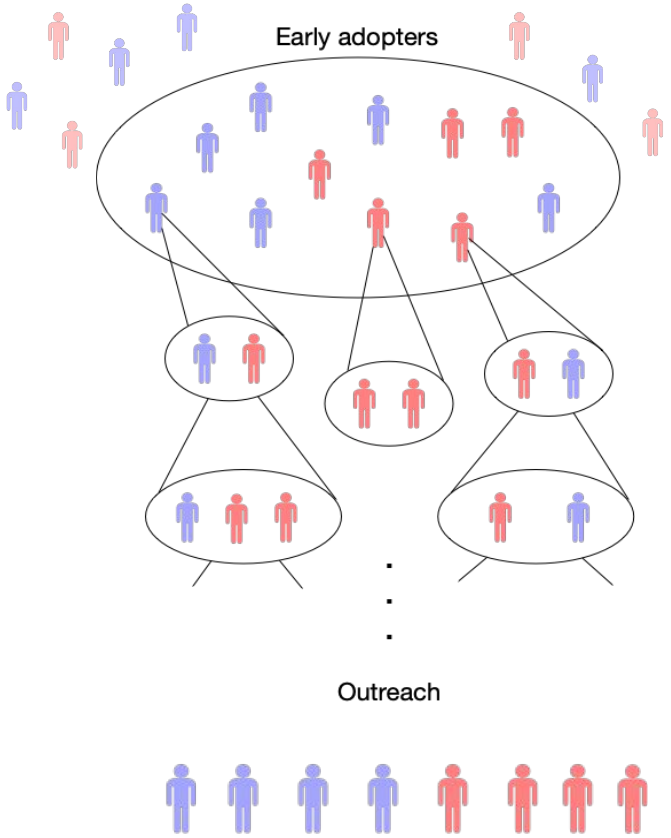


Information diffusion

- **Our vision:** bias as a sign of inefficiency
  - Diversity: tap into inactivated communities in the *early adopters* set

$$S^* = \text{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \simeq \frac{|R|}{|B|}$$

- Seeding can be done with awareness of labels: **statistical parity** in your campaign (even if choosing less connected people)
  - Parity seeding (strict)
  - Diversity seeding (relaxed)

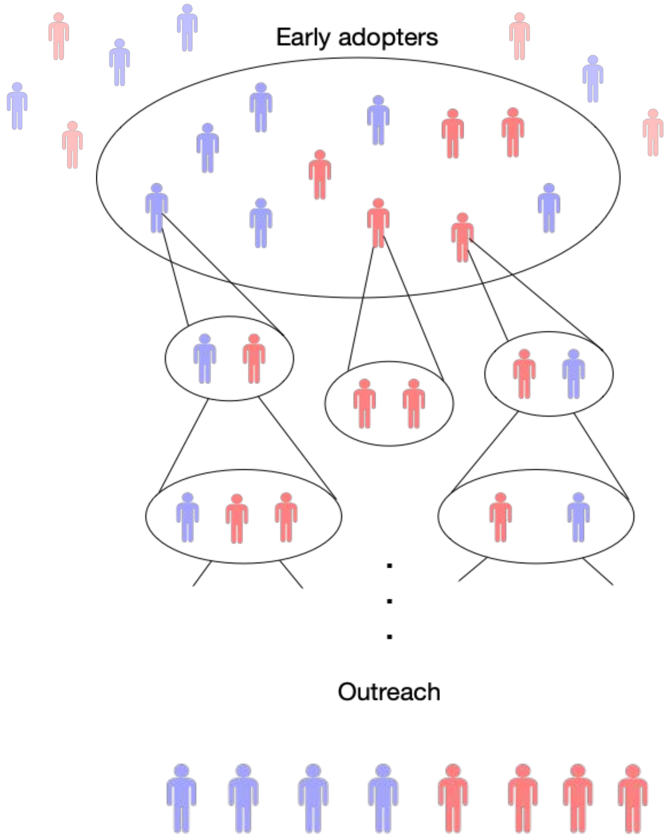


Information diffusion

- **Our vision:** bias as a sign of inefficiency
  - Diversity: tap into inactivated communities in the *early adopters* set

$$S^* = \text{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \simeq \frac{|R|}{|B|}$$

- Seeding can be done with awareness of labels: **statistical parity** in your campaign (even if choosing less connected people)
  - **Parity seeding (strict)** $\frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} = \frac{|R|}{|B|}$
  - Diversity seeding (relaxed)

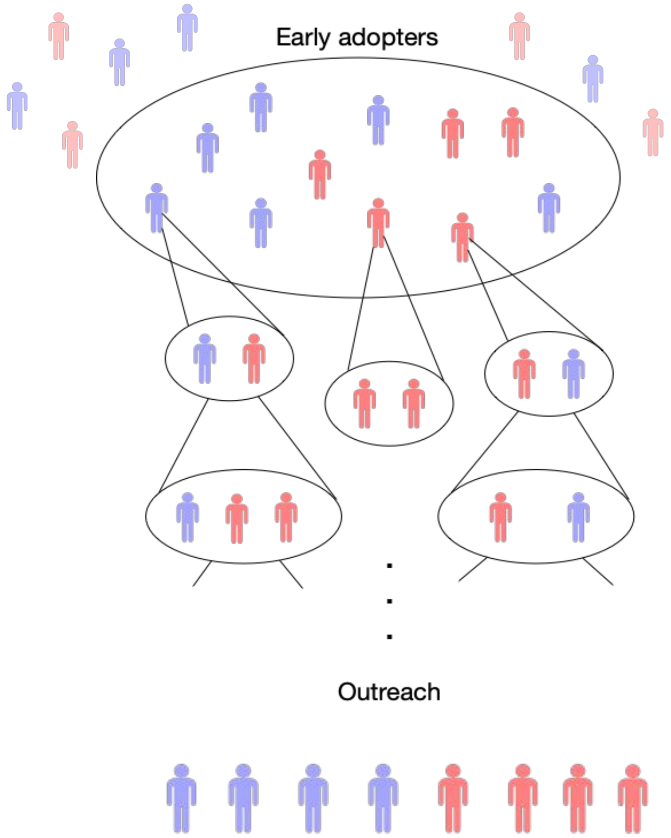


Information diffusion

- **Our vision:** bias as a sign of inefficiency
    - Diversity: tap into inactivated communities in the *early adopters* set

$$S^* = \mathrm{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \simeq \frac{|R|}{|B|}$$

- Seeding can be done with awareness of labels: **statistical parity** in your campaign (even if choosing less connected people)
    - Parity seeding (strict)
    - **Diversity seeding (relaxed)** $\frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \pm \epsilon = \frac{|R|}{|B|}$

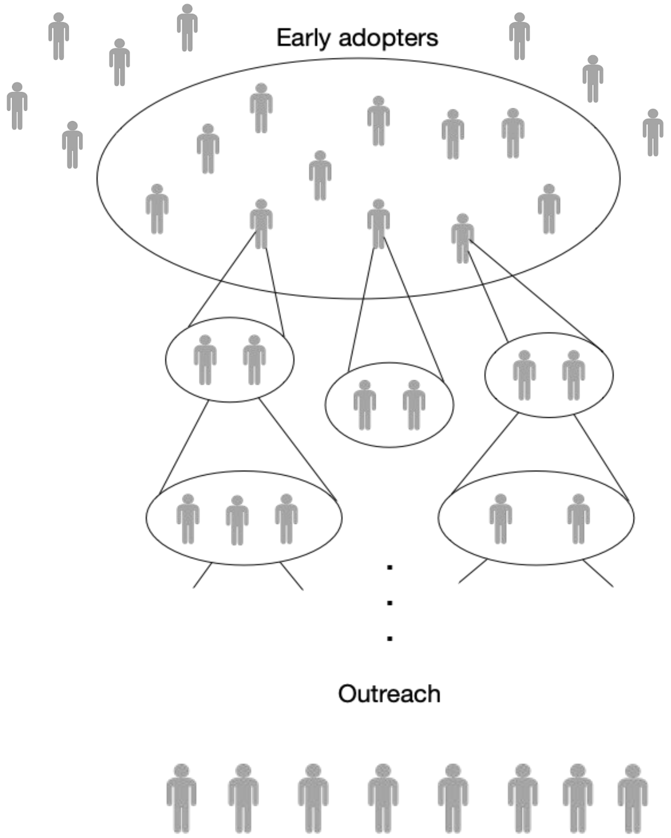


Information diffusion

- **Our vision:** bias as a sign of inefficiency
  - Diversity: tap into inactivated communities in the *early adopters* set

$$S^* = \text{argmax}_{S \subseteq V(G)} \mathbb{E}(|\phi_G(S,p)|,$$
$$\text{s.t. } |S| \leq k \text{ and } \frac{\mathbb{E}(|S \cap R|)}{\mathbb{E}(|S \cap B|)} \simeq \frac{|R|}{|B|}$$

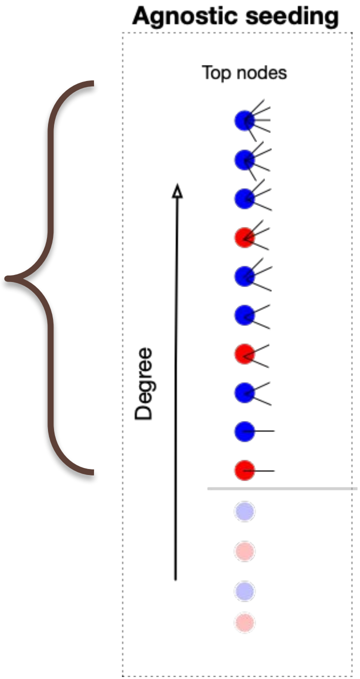
- Seeding can be done with awareness of labels: **statistical parity** in your campaign (even if choosing less connected people)
  - Parity seeding (strict)
  - Diversity seeding (relaxed)
- Baseline: Seeding can be done **agnostically**: ignore labels, already takes into account network structure



[3] Stoica, Ana-Andreea, Jessy Xinyi Han, and Augustin Chaintreau. "Seeding Network Influence in Biased Networks and the Benefits of Diversity." WWW. 2020.

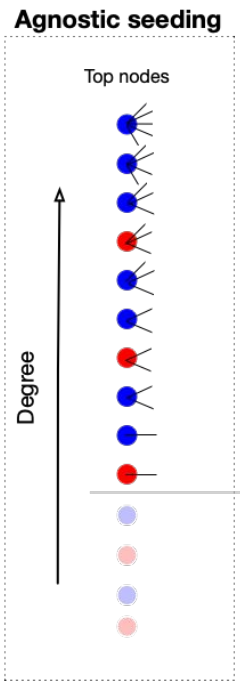
Color-agnostic v. Diversity Seeding

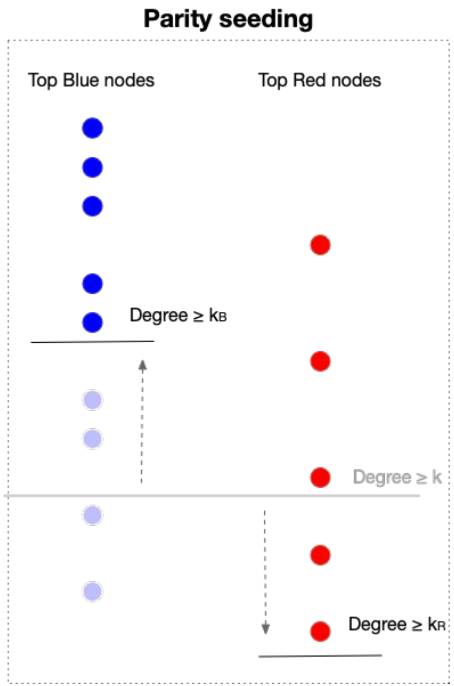
Early
adopters
(seedset)



Color-agnostic v. Diversity Seeding

Early adopters (seedset)

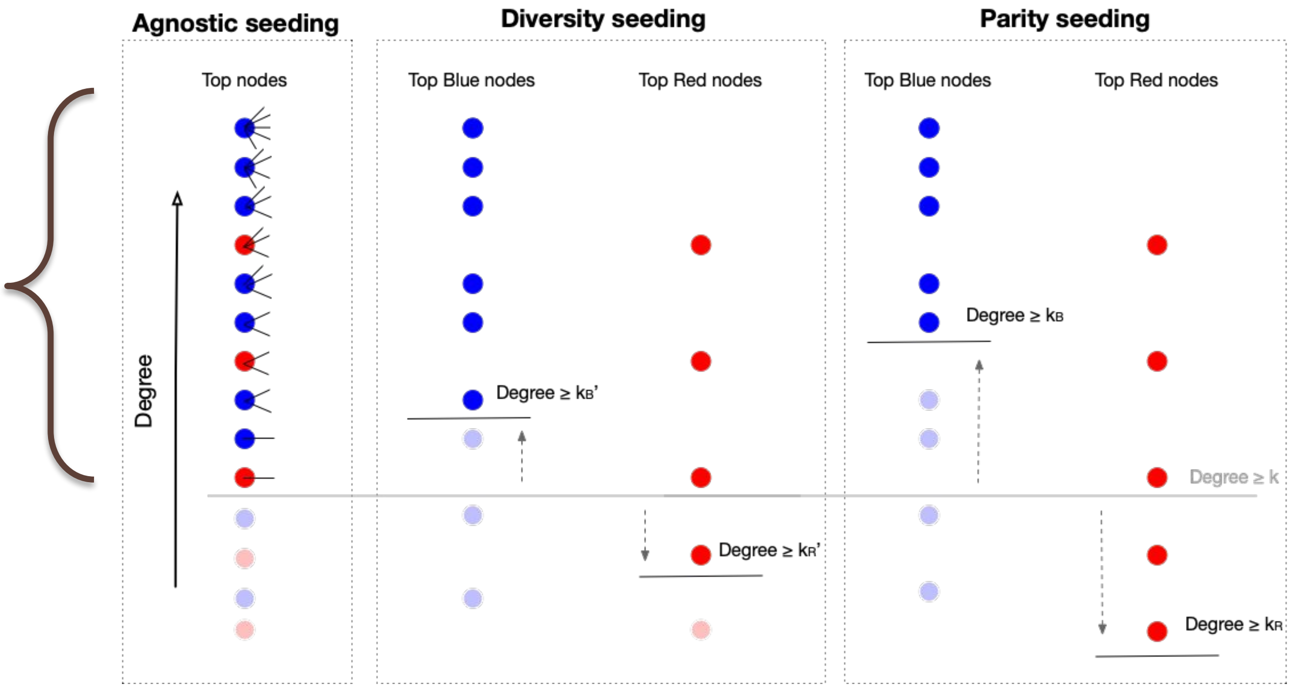




Keeping the same budget!

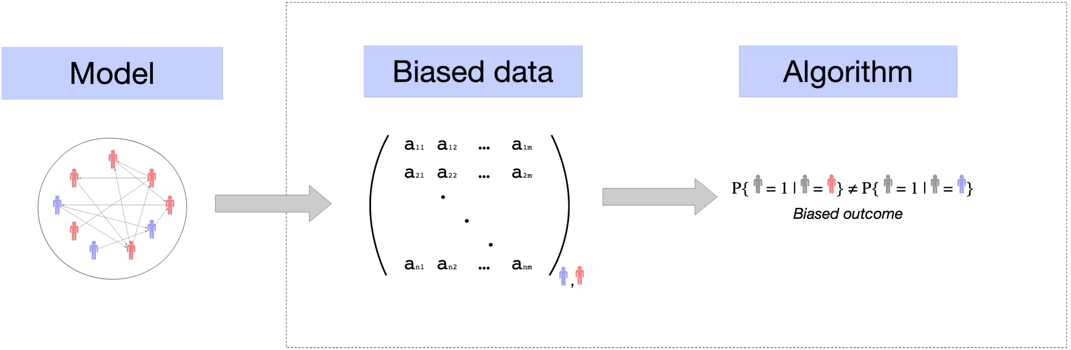
Color-agnostic v. Diversity Seeding

Early adopters (seedset)



Keeping the same budget!

Networks modeling for building more diverse and efficient heuristics



Models of network evolution:

- Explain where inequality or bias originates and how it propagates in an algorithm
- Useful to prove guarantees about interventions to mitigate bias

Biased preferential attachment model (BPAM)

Minority-majority: red label and blue label

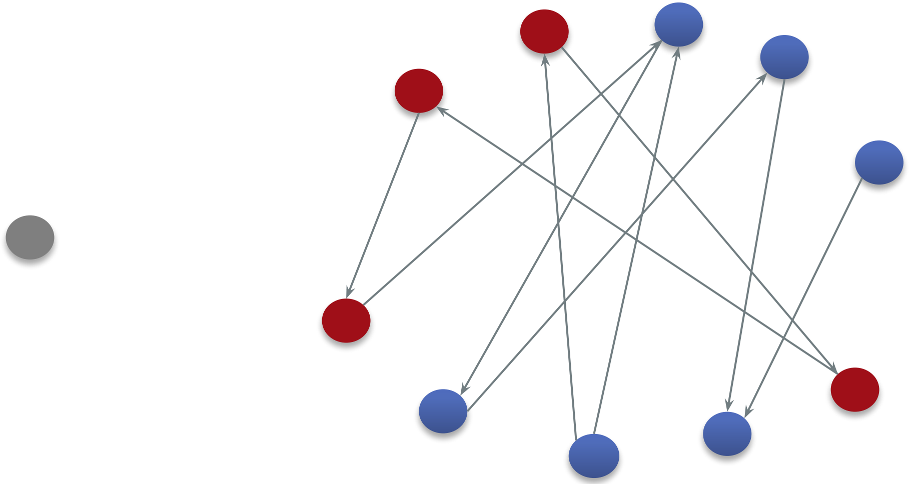
- Fraction of red nodes = $r < ½$

Preferential attachment (rich-get-richer): nodes connect w.p. proportional to degree

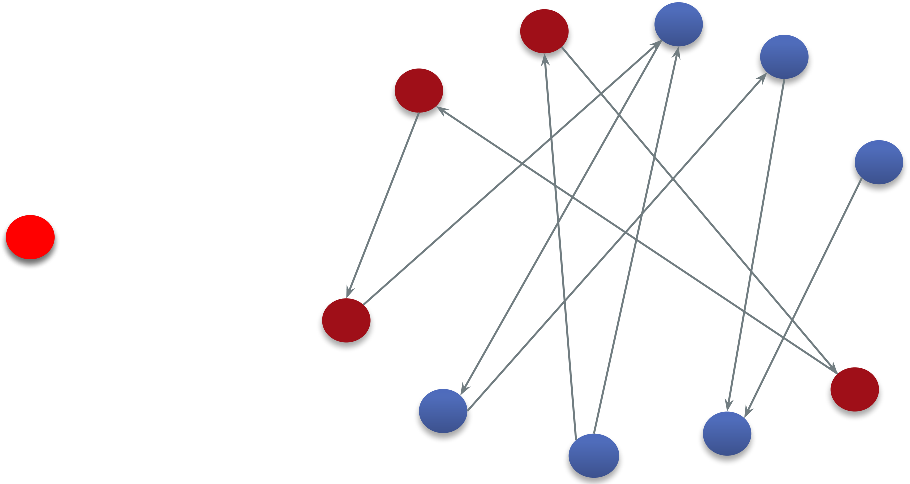
Homophily: if different labels, connection is accepted w.p. ρ

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

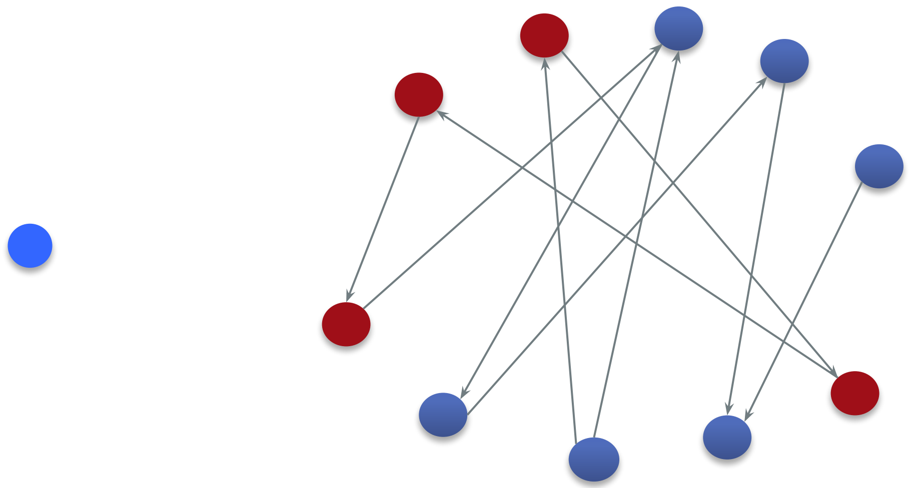
Biased preferential attachment model (BPAM)



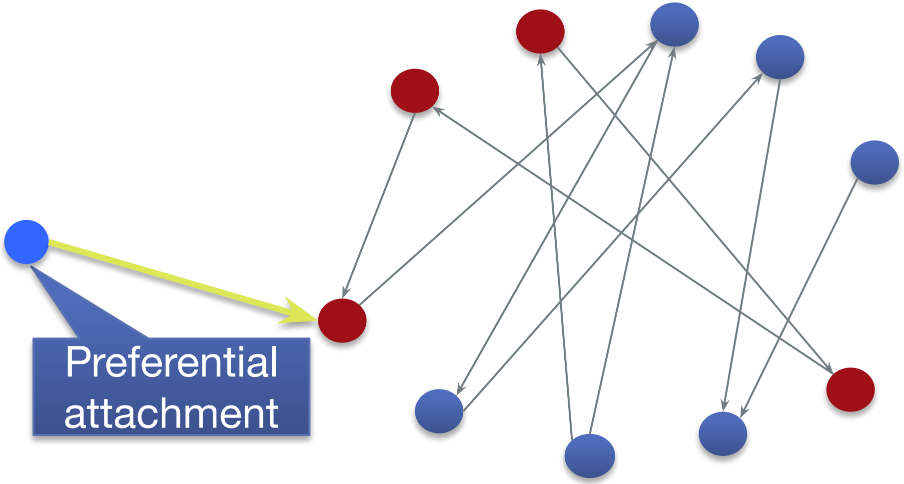
Biased preferential attachment model (BPAM)



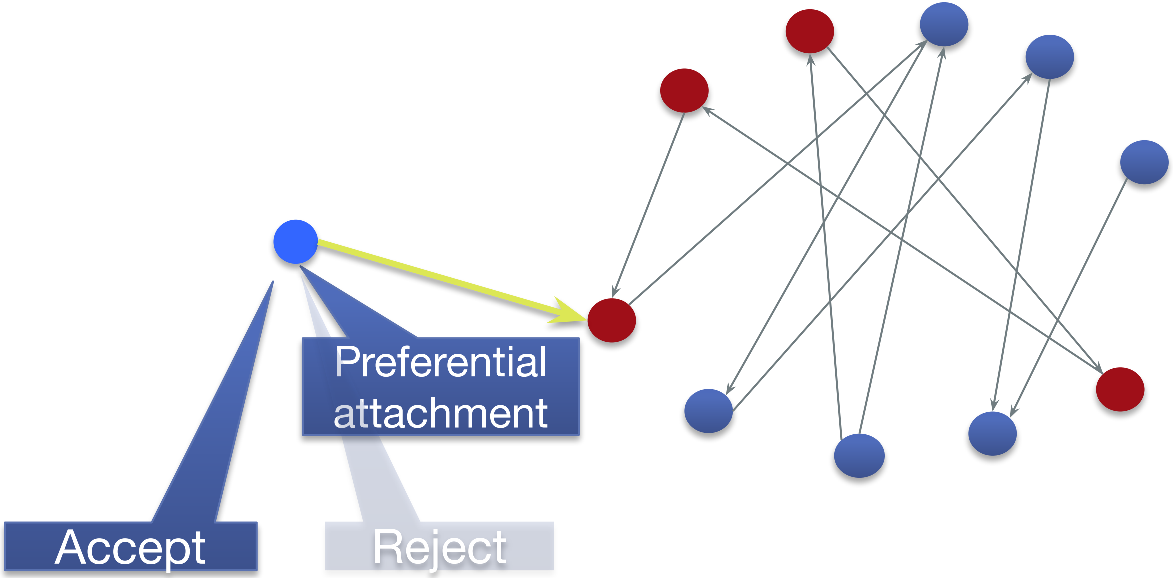
Biased preferential attachment model (BPAM)



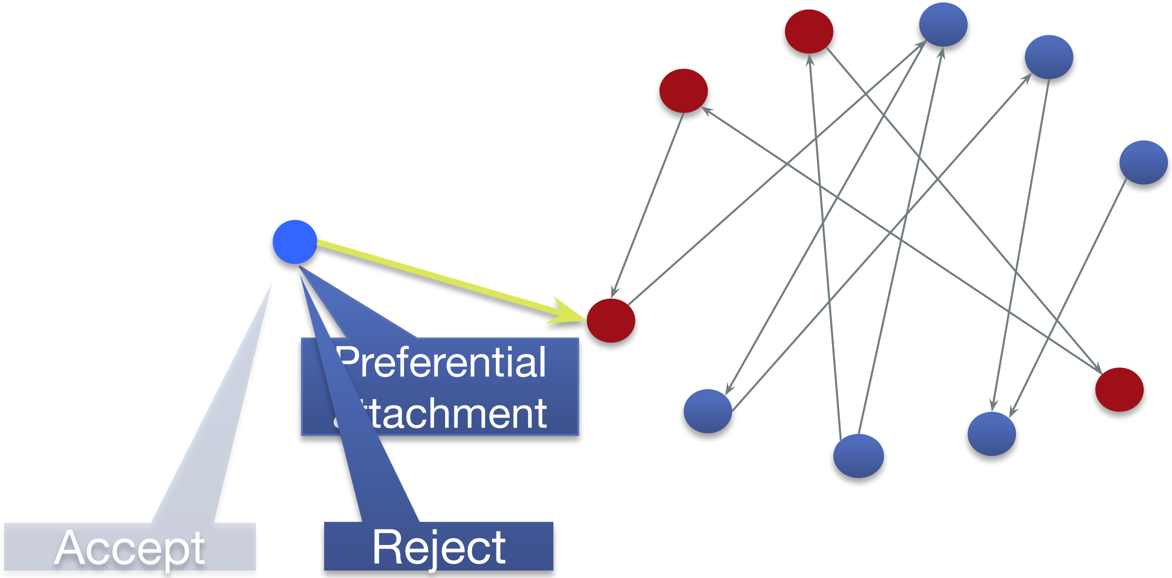
Biased preferential attachment model (BPAM)



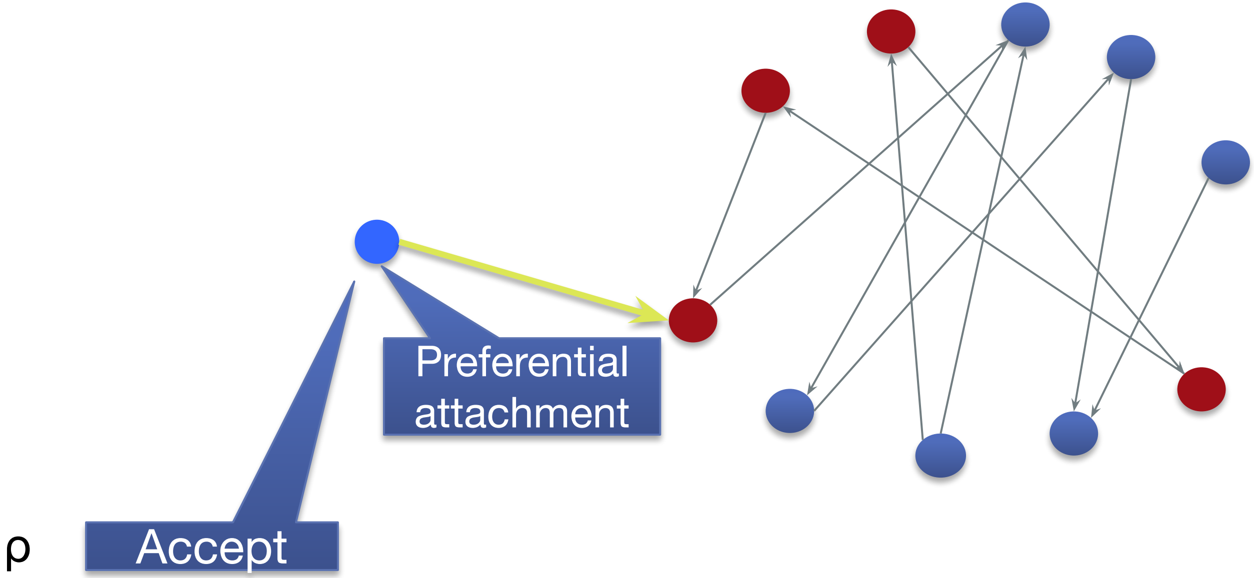
Biased preferential attachment model (BPAM)



Biased preferential attachment model (BPAM)



Biased preferential attachment model (BPAM)



Biased preferential attachment model (BPAM)

Minority-majority: blue label and red label

- Fraction of red nodes = $r < \frac{1}{2}$

Preferential attachment (rich-get-richer): nodes connect w.p. proportional to degree

Homophily: if different labels, connection is accepted w.p. ρ

⇒ known to exhibit inequality in the degree distribution of the two communities[4]

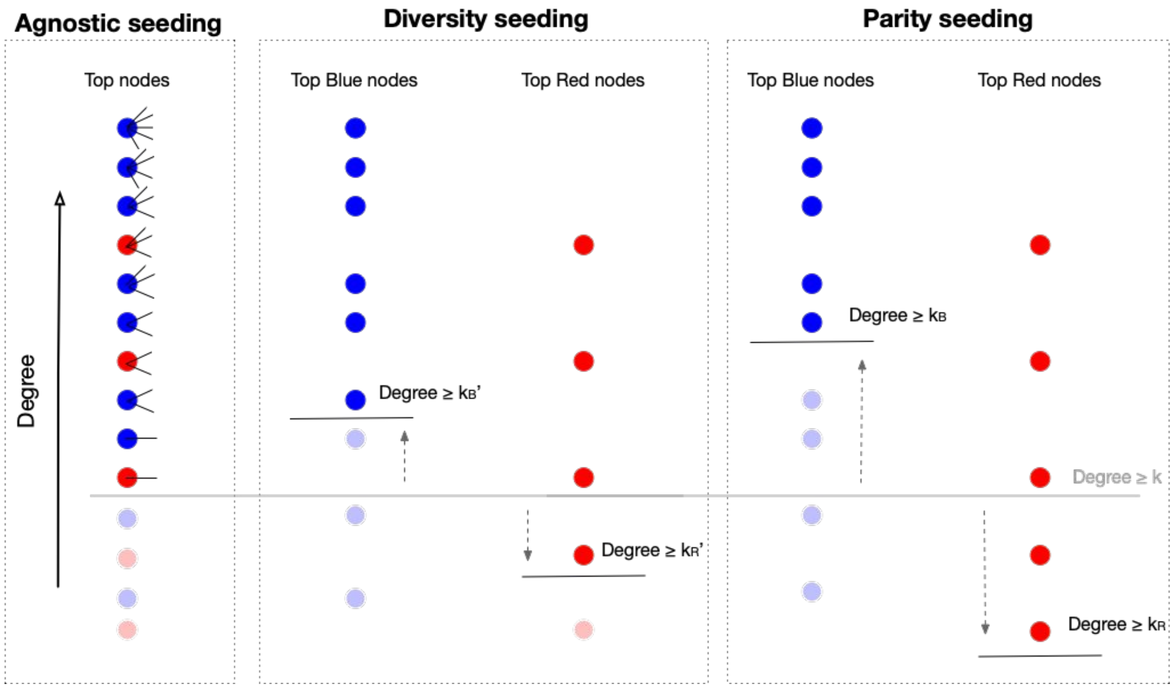
$$top_k(R) \sim k^{-\beta(R)}$$
$$top_k(B) \sim k^{-\beta(B)}$$

Thm [Avin et al]: $\beta(R) > 3 > \beta(B)$

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

Color-agnostic v. Diversity Seeding

Early adopters (seedset)



Keeping the same budget!

Theoretical analysis of diversity interventions

***Theorem**: for the graph sequences G(n) generated from the BPAM:*

1. *Diversity seeding and parity seeding leads to fairer outreach for the same budget*

$$abs\left(\frac{\mathbb{E}(|\phi(S_{\text{diversity,parity}}) \cap R|)}{\mathbb{E}(|\phi(S_{\text{diversity,parity}}) \cap B|)} - \frac{|R|}{|B|}\right) \leq abs\left(\frac{\mathbb{E}(|\phi(S_{\text{agnostic}}) \cap R)|)}{\mathbb{E}(|\phi(S_{\text{agnostic}}) \cap B|)} - \frac{|R|}{|B|}\right)$$

2. *∃ k* (closed form) such that when k > k*, diversity seeding and parity seeding can outperform agnostic seeding in outreach*

$$\mathbb{E}(\phi(S_{\text{diversity}})) > \mathbb{E}(\phi(S_{\text{parity}})) > \mathbb{E}(\phi(S_{\text{agnostic}})),$$
$$\text{given } |S_{\text{diversity}}| = |S_{\text{parity}}| = |S_{\text{agnostic}}| = k$$

Proof sketch

Our goal is to find two thresholds $k^R(n)$and $k^B(n)$that give in expectation the same amount of seeds as a general ("agnostic") threshold k(n) but better influence:

$$\mathbb{E}(\phi(S_{k(n)})) < \mathbb{E}(\phi(S_{k^R(n)} \cup S_{k^B(n)})),$$
$$\text{s.t. } \mathbb{E}(|S_{k(n)}|) = (|S_{k^R(n)} \cup S_{k^B(n)}|)$$

First step: estimate first-step influence size of $S_{k(n)} = \{v \in V | deg(v) \geq k(n)\}$

Second step: extend to an estimation of $\mathbb{E}(\phi(S_{k(n)}))$

Proof sketch

Our goal is to find two thresholds $k^R(n)$and $k^B(n)$that give in expectation the same amount of seeds as a general ("agnostic") threshold k(n) but better influence:

$$\mathbb{E}(\phi(S_{k(n)})) < \mathbb{E}(\phi(S_{k^R(n)} \cup S_{k^B(n)})),$$
$$\text{s.t. } \mathbb{E}(|S_{k(n)}|) = (|S_{k^R(n)} \cup S_{k^B(n)}|)$$

First step: estimate first-step influence size of $S_{k(n)} = \{v \in V | deg(v) \geq k(n)\}$

Second step: extend to an estimation of $\mathbb{E}(\phi(S_{k(n)}))$

Proof sketch

Our goal is to find two thresholds $k^R(n)$and $k^B(n)$that give in expectation the same amount of seeds as a general ("agnostic") threshold k(n) but better influence:

$$\mathbb{E}(\phi(S_{k(n)})) < \mathbb{E}(\phi(S_{k^R(n)} \cup S_{k^B(n)})),$$
$$\text{s.t. } \mathbb{E}(|S_{k(n)}|) = (|S_{k^R(n)} \cup S_{k^B(n)}|)$$

First step: estimate first-step influence size of $S_{k(n)} = \{v \in V | deg(v) \geq k(n)\}$

- We know $|S_{k(n)}|$ because the degree distribution follows a power law with coefficients $\beta(R), \beta(B)$
- Can compute first order influence for any threshold by computing $\mathbb{P}(v \text{ influenced by one edge} | v \in B)$ and $\mathbb{P}(v \text{ influenced by one edge} | v \in R)$

Proof sketch

Our goal is to find two thresholds $k^R(n)$ and $k^B(n)$ that give in expectation the same amount of seeds as a general ("agnostic") threshold k(n) but better influence:

$$\mathbb{E}(\phi(S_{k(n)})) < \mathbb{E}(\phi(S_{k^R(n)} \cup S_{k^B(n)})),$$
$$\text{s.t. } \mathbb{E}(|S_{k(n)}|) = (|S_{k^R(n)} \cup S_{k^B(n)}|)$$

Set $k^B(n) = k(n) \cdot x$, compute $k^R(n)$ based on the budget constraint, and solve

$$F(x) = \mathbb{E}(\phi(S_{k^B(n)} \cup S_{k^R(n)})) - \mathbb{E}(\phi(S_{k(n)}))$$

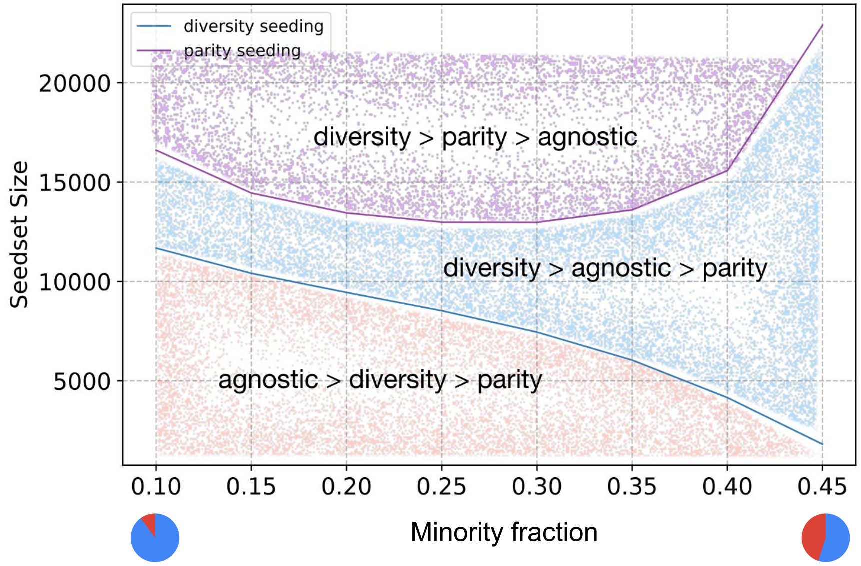
Theoretical analysis of diversity interventions

***Theorem**: for the graph sequences G(n) generated from the BPAM:*

1. *Diversity seeding and parity seeding leads to fairer outreach for the same budget*

2. *∃ k^* (closed form) such that when $k > k^*$, diversity seeding and parity seeding can outperform agnostic seeding in outreach*

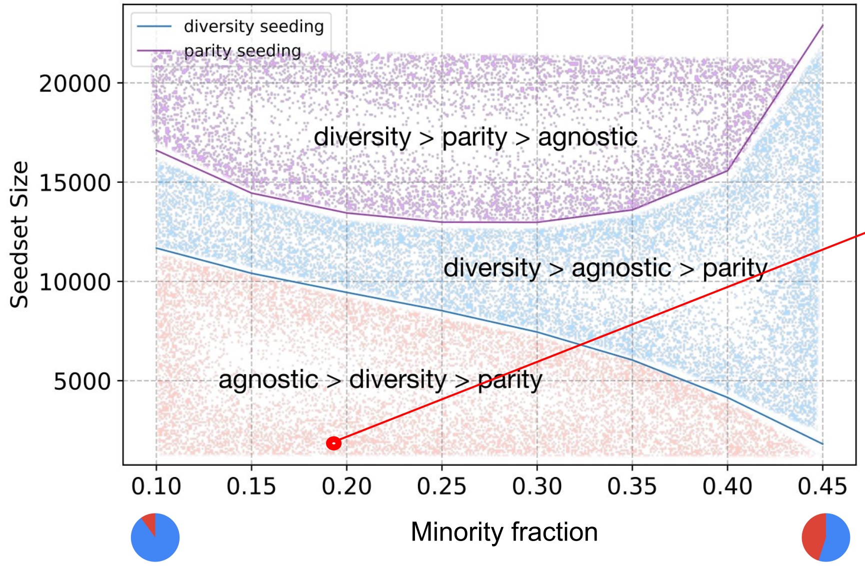
Theoretical analysis of diversity interventions



Network of ~53,000 nodes, 2 communities, homophily $\rho = 0.135$

- Compute regions where each heuristic performs better than the agnostic one
- As communities become more equal, need fewer seeds for diversity heuristic to be more efficient
- Not the same thing happens with the parity heuristic!

Theoretical analysis of diversity interventions

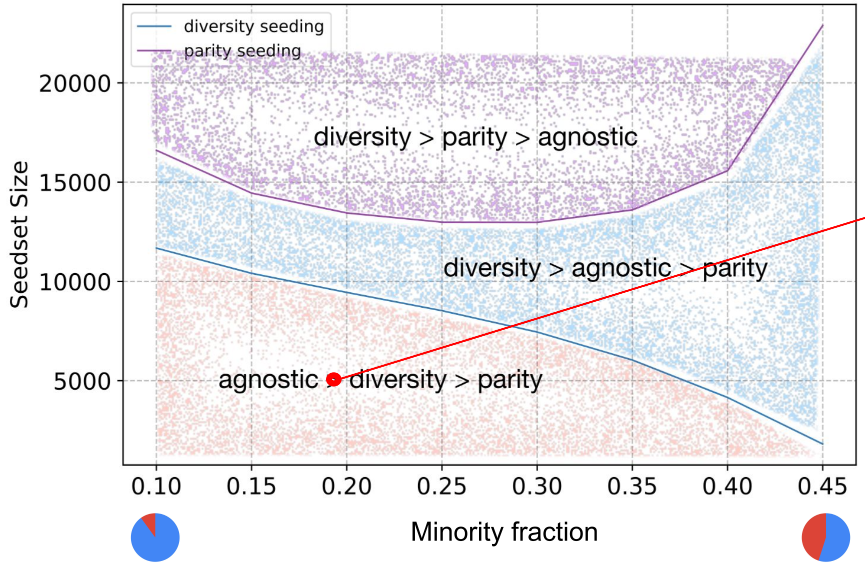


Network of ~53,000 nodes, 2 communities, homophily ρ = 0.135

DBLP citation dataset: men and women

	1,000 seeds		
$p = 0.01$	Agnostic seeding	Parity seeding	Diversity seeding
Total outreach	**1,149.15**	↓1,147.874	↓1,149.1
F outreach	191.95	↑**210.456**	↑196.6
M outreach	**957.2**	↓937.418	↓952.5
F % in outreach	0.167	↑**0.183**	↑0.171

Theoretical analysis of diversity interventions

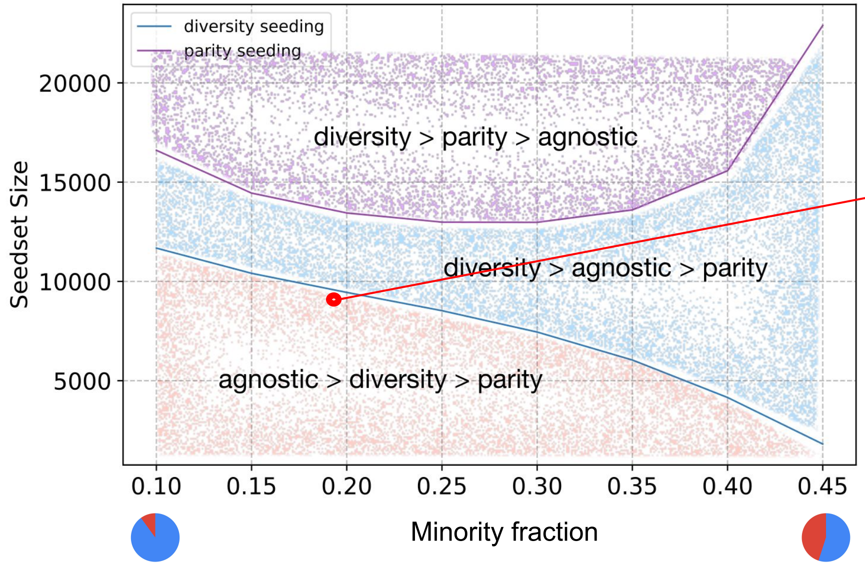


Network of ~53,000 nodes, 2 communities, homophily $\rho = 0.135$

DBLP citation dataset: men and women

	5,000 seeds		
$p = 0.01$	Agnostic seeding	Parity seeding	Diversity seeding
Total outreach	5,410.748	↓5,408.762	↑**5411.191**
F outreach	862.191	↑**1,004.232**	↑892.11
M outreach	**4,548.557**	↓4,404.53	↓4,519.081
F % in outreach	0.15934	↑**0.18567**	↑0.165

Theoretical analysis of diversity interventions



Network of ~53,000 nodes, 2 communities, homophily ρ = 0.135

DBLP citation dataset: men and women

p = 0.01	9,100 seeds		
	Agnostic seeding	Parity seeding	Diversity seeding
Total outreach	9,554.934	↑9,555.559	↑**9,556.349**
F outreach	1,581.842	↑**1,776.037**	↑1,679.423
M outreach	**7,973.092**	↓7,779.522	↓7,876.926
F % in outreach	0.16555	↑**0.186**	↑0.176

Discussion

- Relation to resource-allocation settings:
    - Budgetary constraints ⇔ trade-offs in objectives
- Network formation & causality questions
    - Am I friends with people because we influenced each other or the other way around?[6]
- Is ‘fairness’ transferable to other settings?

[6] Cristali I, Veitch V. Using Embeddings for Causal Estimation of Peer Influence in Social Networks. arXiv preprint arXiv:2205.08033. 2022.

Discussion

- Relation to resource-allocation settings:
  - Budgetary constraints ⇔ trade-offs in objectives
- Network formation & causality questions
  - Am I friends with people because we influenced each other or the other way around?[6]
- **Is ‘fairness’ transferable to other settings?**

[6] Cristali I, Veitch V. Using Embeddings for Causal Estimation of Peer Influence in Social Networks. arXiv preprint arXiv:2205.08033. 2022.

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
2. Using networks to test algorithms: randomized controlled trials
3. Build interventions to mitigate algorithmic bias
    a. In designing fair information diffusion campaigns
    b. In designing fair committees in opinion aggregation settings
    c. A theoretical framework for navigating trade-offs

Opinion dynamics models

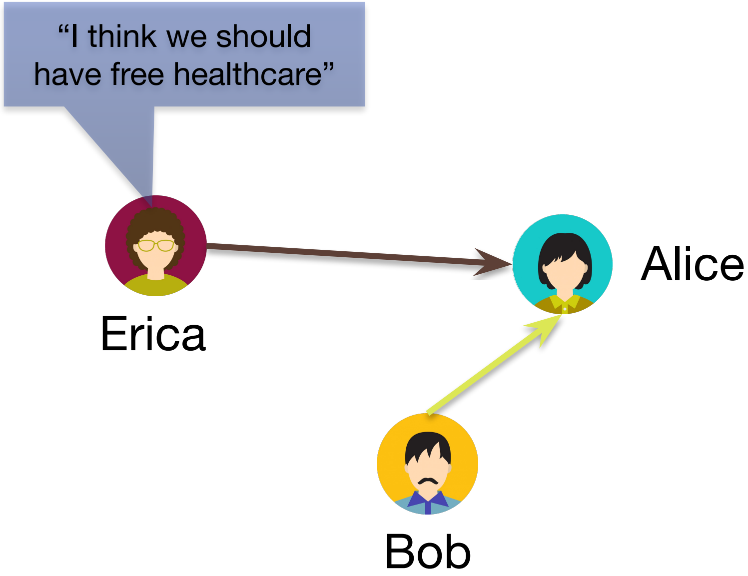
- Political purposes: understanding voting patterns and changes
- Policy purposes:
  - Education policy
  - Healthcare policy
  - Collective action (union formation)
  - Local decisions, e.g. transportation

Opinion **dynamics** models

- Political purposes: understanding voting patterns and changes
- Policy purposes:
  - Education policy
  - Healthcare policy
  - Collective action (union formation)
  - Local decisions, e.g. transportation

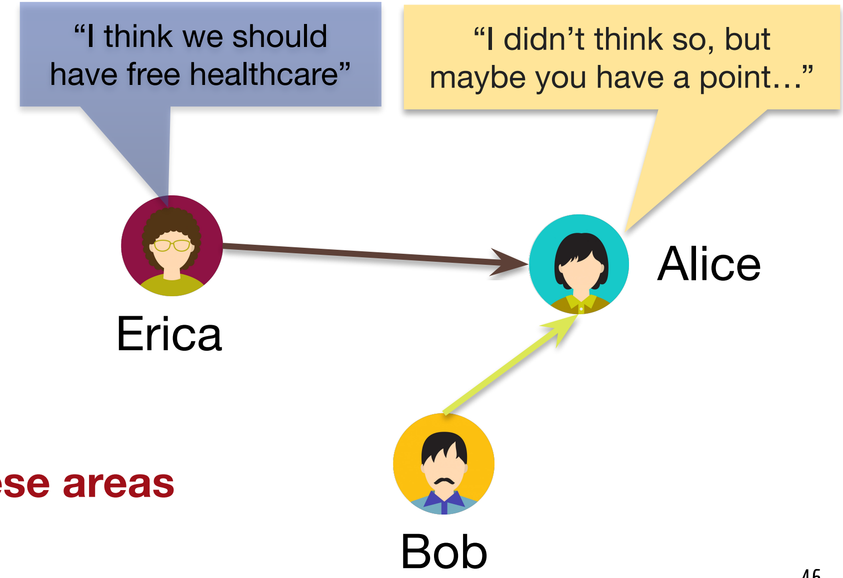
Opinion **dynamics** models

- Political purposes: understanding voting patterns and changes
- Policy purposes:
  - Education policy
  - Healthcare policy
  - Collective action (union formation)
  - Local decisions, e.g. transportation



Opinion **dynamics** models

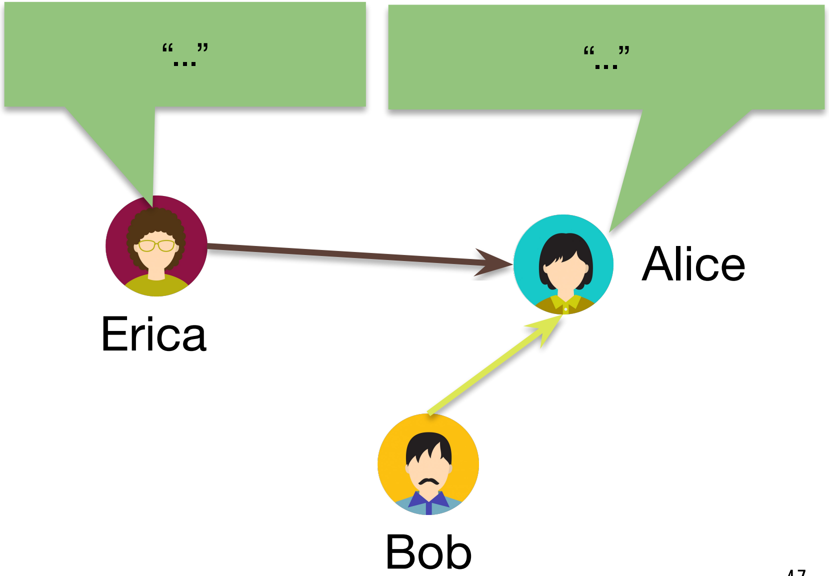
- Political purposes: understanding voting patterns and changes
- Policy purposes:
  - Education policy
  - Healthcare policy
  - Collective action (union formation)
  - Local decisions, e.g. transportation



⇒ final consensus governs decisions in these areas

Opinion **dynamics** models

- [Golub & Jackson][7] describe network conditions to get consensus



[7] Golub, B. and Jackson, M.O., 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), pp.112-149.

Opinion **dynamics** models

- [Golub & Jackson][7] describe network conditions to get consensus
- DeGroot model[8] of opinion aggregation:

For a population of n agents with initial opinions $\{x_1(0), x_2(0), \cdots, x_n(0)\}$

and a network with adjacent matrix A, opinions update at every timestep t:

$$x_i(t+1) = \sum_{j=1}^{n} A_{ij} x_j(t)$$

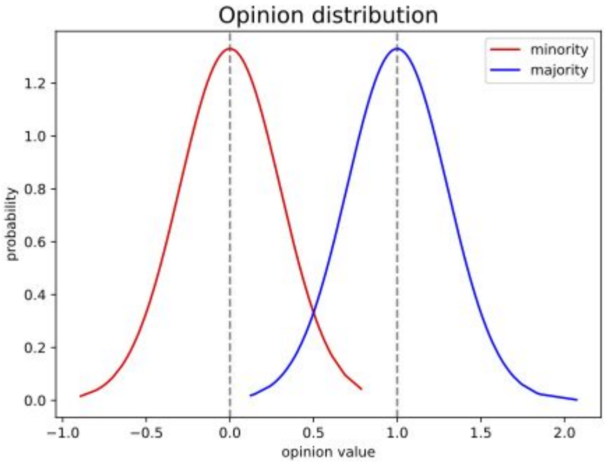
Consensus is reached as $t \to \infty : \mathbf{x}(\infty) = \mathbf{e} \cdot \mathbf{x}(0)$

Eigencentrality matters!

[7] Golub, B. and Jackson, M.O., 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), pp.112-149.
[8] DeGroot, Morris H. (1974). 'Reaching a Consensus', Journal of the American Statistical Association 69(345): 118–121.

Opinion **dynamics** models

- [Golub & Jackson][7] describe network conditions to get consensus
- If different groups have different opinions, how does consensus look like?



[7]Golub, B. and Jackson, M.O., 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), pp.112-149.

How does a committee affect consensus?

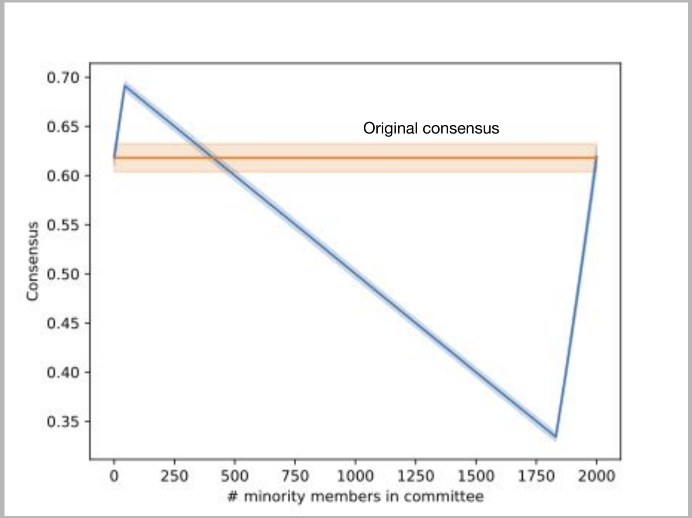
Modeling choices for committees:

- Choose a proportion *p* of the population in the committee
- Assume that consensus **first** occurs in the committee, and then in the general population
  - 2-step process:
    - for a committee $C \subset [n], (x_i(0))_{i \in C} \xrightarrow{t \to \infty} (x_i(\infty))_{i \in C}$ (assume that committee forms a click)
    - initial opinions of the population: $\{(x_i(\infty))_{i \in C}, (x_i(0))_{i \notin C}\}$
- Fairness: how many of each group do we choose?
  - Proportional to their numbers in the population
  - Which individuals do we choose? The most central ones

[9] Ana-Andreea Stoica and Francesca Parise. *Fairness in committee formation: interventions in network opinion aggregation models.* Manuscript in preparation. 2023.

How does a committee affect consensus?

If we choose a committee with proportions equal to the general population (21% minority), we actually skew the consensus more towards the majority!

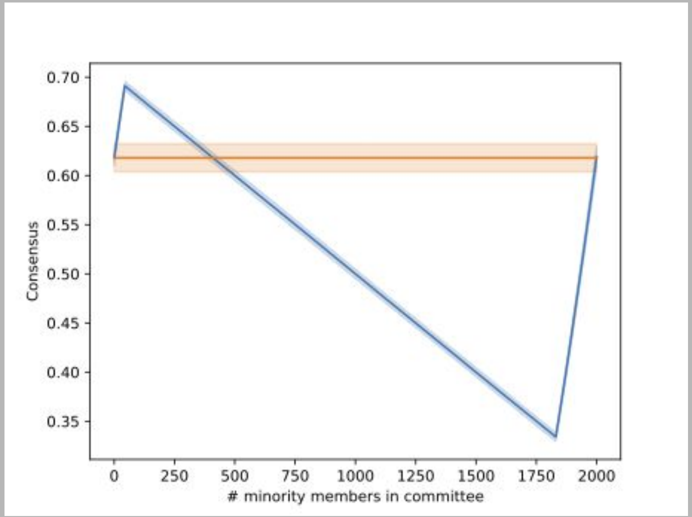


DBLP data, 53,000 nodes, 21% women

What interventions can we enact?

1. Choose more minority members in the committee

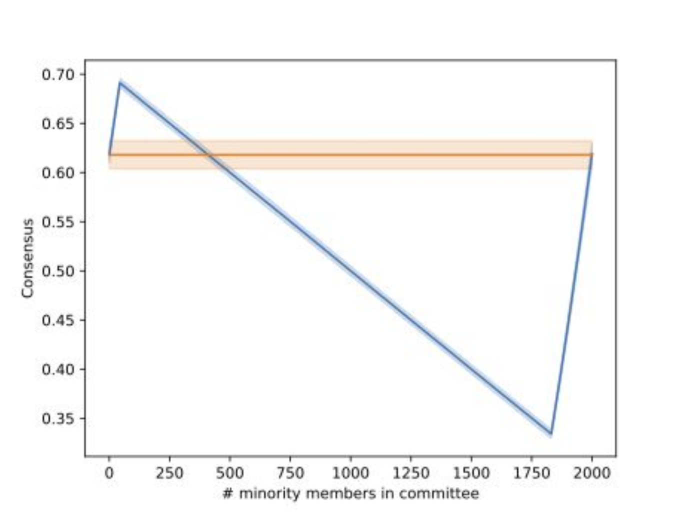
⇒ proportional representation can hurt



DBLP data, 53,000 nodes, 21% women

What interventions can we enact?

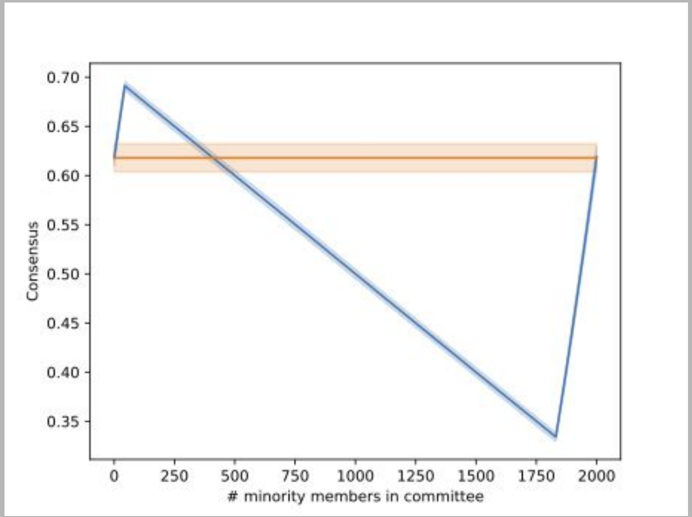
1. Choose more minority members in the committee
2. Choose less central minority members in the committee



DBLP data, 53,000 nodes, 21% women

What interventions can we enact?

1. Choose more minority members in the committee
2. Choose less central minority members in the committee
3. Change the way committee aggregates



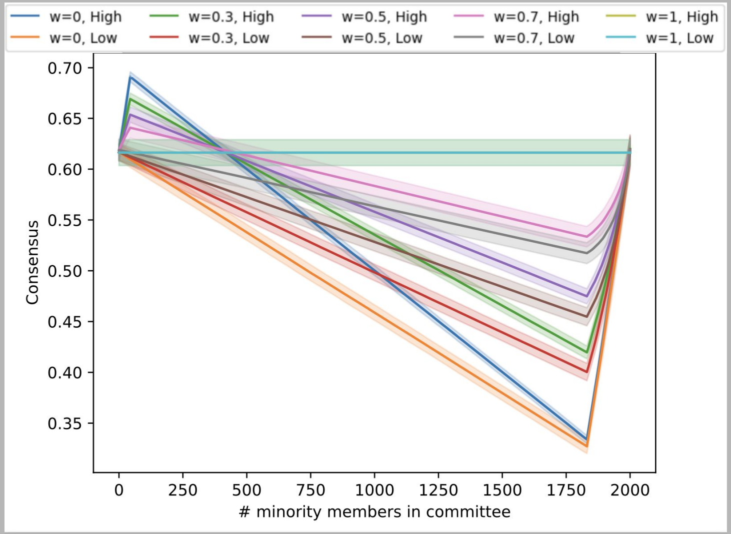
DBLP data, 53,000 nodes, 21% women

What interventions can we enact?

1. Choose more minority members in the committee
2. Choose less central minority members in the committee (Low vs. High)
3. Change the way committee aggregates

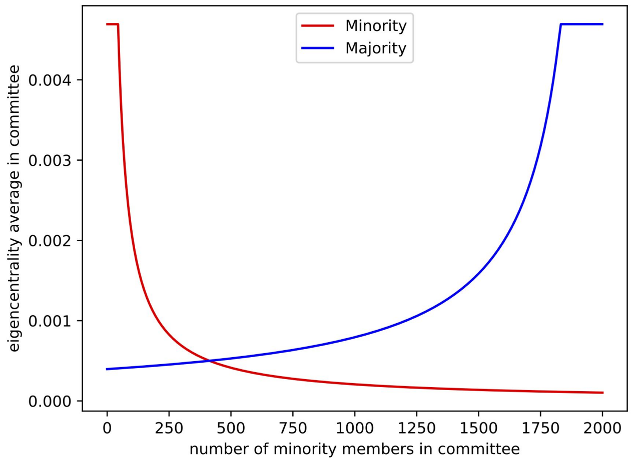
- Committee forms a clique
- Committee aggregates proportional to their network eigencentrality

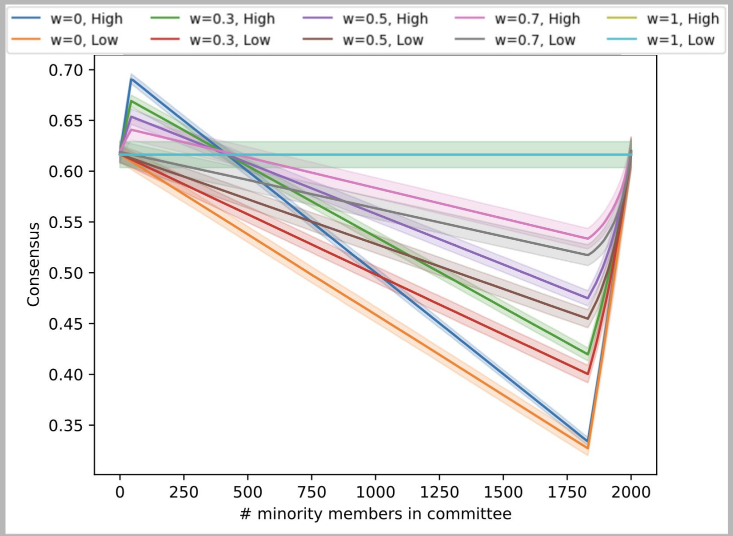
$$\mathbf{e} = w \cdot \mathbf{e}_{\text{original}} + (1 - w) \cdot \mathbf{e}_{\text{clique}}$$



DBLP data, 53,000 nodes, 21% women

What interventions can we enact?

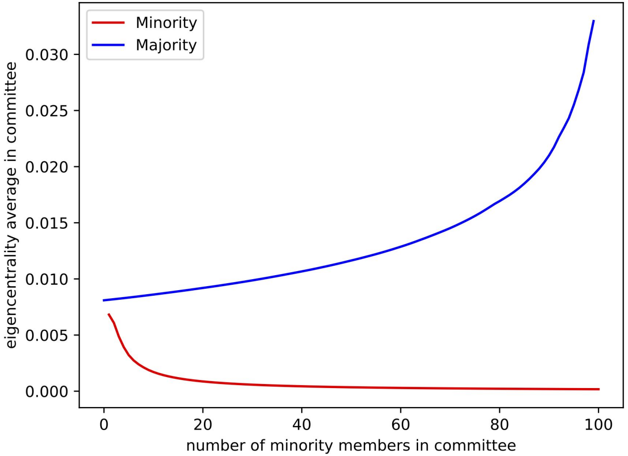


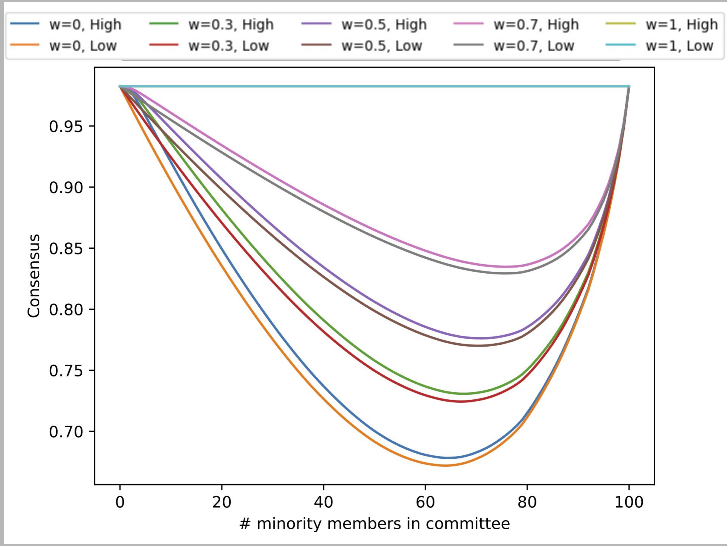


DBLP data, 53,000 nodes, 21% women

⇒ Theoretical explanation for when consensus is skewed towards one of the communities

What interventions can we enact?





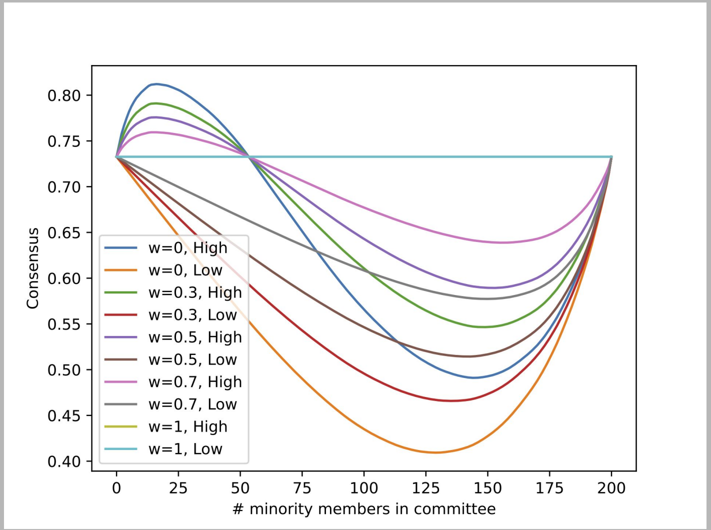
APS citation data, 1,281 nodes, 33% minority (two different fields of physics)

⇒ if the minority eigencentrality is very low, committee impact is the same

[10] E. Lee, F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic. Homophily and minority-group size explain perception biases in social networks. Nature human behaviour, 3(10):1078–1087, 2019.

What interventions can we enact?

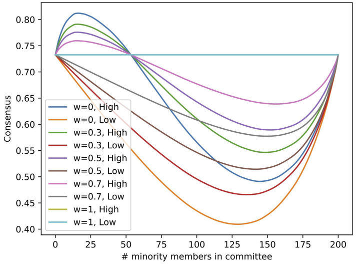
If you want to understand what's wrong with our public schools, you have to look at what is arguably the most powerful force in shaping them: white parents. A five-part series from the makers of Serial and The New York Times. Hosted by Chana Joffe-Walt.



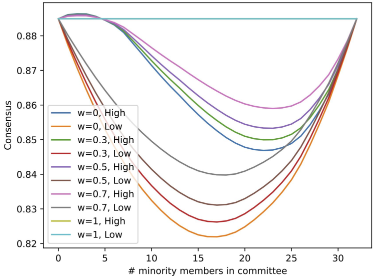
Add Health data: schools with different demographics: hispanic minority of 22% (out of ~1,100 students)

https://addhealth.cpc.unc.edu/

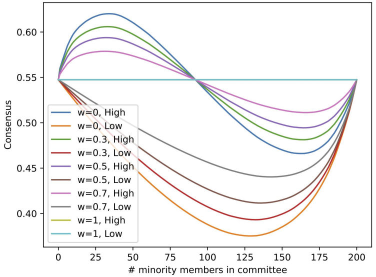
What interventions can we enact?



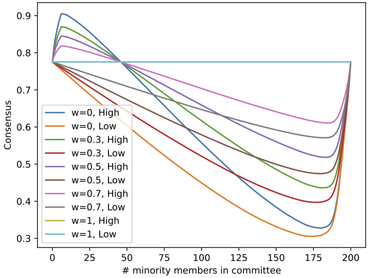
Hispanic minority of 22%



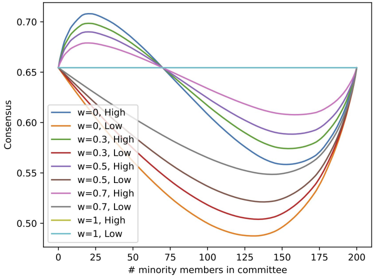
Black minority of 11%



Hispanic minority of 43%



Hispanic minority of 20%



Black minority of 34%

https://addhealth.cpc.unc.edu/

How do we use networks to design algorithms?

1. Using networks to diagnose *when* and *how* an algorithm may amplify bias
2. Using networks to test algorithms: randomized controlled trials
3. Build interventions to mitigate algorithmic bias
    a. In designing fair information diffusion campaigns
    b. In designing fair committees in opinion aggregation settings
    c. A theoretical framework for navigating trade-offs

Navigating trade-offs: clustering problems

Objective(s):

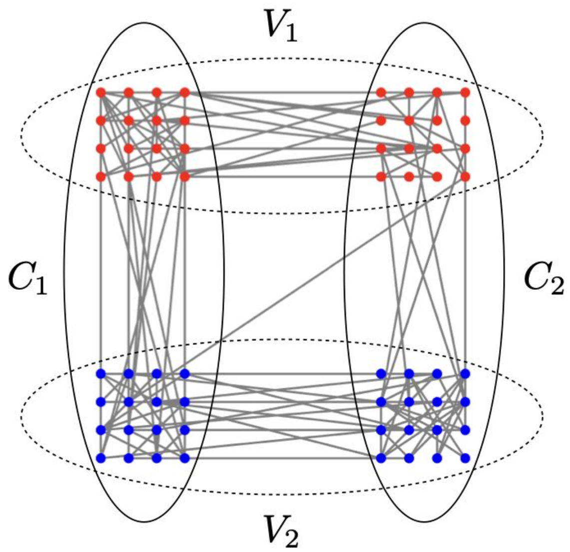
- Quality q(C): minimize the number of edges cross-clusters

$$\mathrm{RatioCut}(S,T) := \frac{cut(S,\overline{S})}{|S|} + \frac{cut(T,\overline{T})}{|T|}$$

$$\mathrm{NCut}(S,T) := \frac{cut(S,\overline{S})}{vol(S)} + \frac{cut(T,\overline{T})}{vol(T)}$$

⇒ Spectral Clustering as an approximation

- Create an embedding of the graph (e.g. the graph Laplacian, L = D - A)
- Take the first 2 dimensions
- Apply k-means on these dimensions



[Kleindessner et al, 2019]

Navigating trade-offs: clustering problems

Objective(s):

- Quality q(C): minimize the number of edges cross-clusters

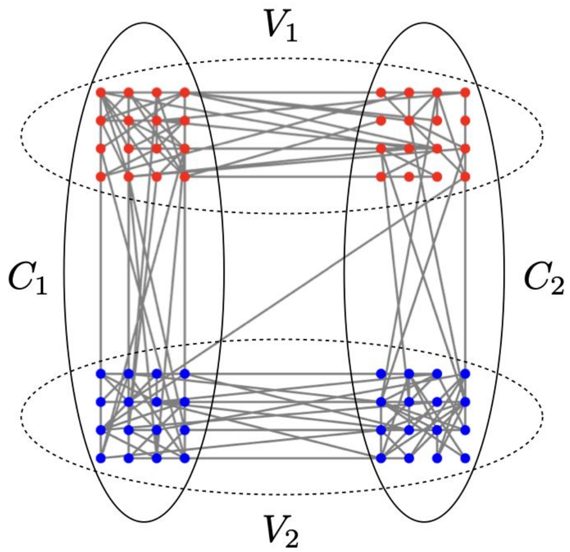
$$\mathrm{RatioCut}(S,T) := \frac{cut(S,\overline{S})}{|S|} + \frac{cut(T,\overline{T})}{|T|}$$

$$\mathrm{NCut}(S,T) := \frac{cut(S,\overline{S})}{vol(S)} + \frac{cut(T,\overline{T})}{vol(T)}$$

- Fairness / utility f(C): some measure of group representation within clusters

$$\mathrm{Balance}(\mathcal{C}) = \min_{C\in\mathcal{C}} \min\left(\frac{\#R(C)}{\#B(C)}, \frac{\#B(C)}{\#R(C)}\right)$$

[Chierichetti et al, 2017]



[Kleindessner et al, 2019]

Navigating trade-offs: clustering problems

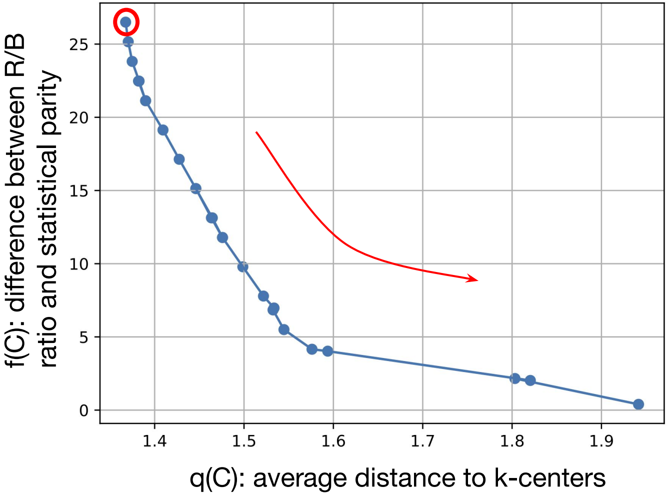
Polynomial time algorithm for tracing the quality-utility trade-off

⇒ find the Pareto frontier between objectives

Next best point on the frontier: starting from a clustering C, find the optimal clustering

$$\mathcal{C}' = \arg \max_{\mathcal{C}' \in \text{changes}(\mathcal{C})} \frac{f(\mathcal{C}') - f(\mathcal{C})}{q(\mathcal{C}) - q(\mathcal{C}')}$$

Key idea: transform this problem into finding an optimal cycle in a doubly-weighted graph



[14] Hakim, Stoica, and Papadimitriou. *"Strategic clustering."* Manuscript in preparation. 2023. Previously at StratML @ Neurips 2021.
[15] Golitschek, M.v. *"Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones".* Numerische Mathematik 39 (1), 65–84. 1982.
[16] Lawler, E.L. *"Optimal cycles in doubly weighted directed linear graphs".* In Proceedings of the International Symposium of Theory of Graphs. 209–232. 1966.

Sketch of algorithm

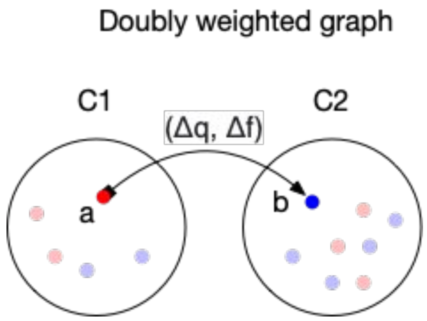
Next best point on the frontier: starting from a clustering C, find the optimal clustering

$$\mathcal{C}' = \arg \max_{\mathcal{C}' \in \text{changes}(\mathcal{C})} \frac{f(\mathcal{C}') - f(\mathcal{C})}{q(\mathcal{C}) - q(\mathcal{C}')}$$

Decompose finding C' into a set of *elementary changes*:

- Nodes u and v switch clusters
- Node u moves to another cluster

Create a doubly-weighted graph where each edge is the quality (fairness) delta from an elementary change



Lawler: we can find the minimum cycle Δf/Δq
⇔ deciding whether there is a neg cycle

[14] Hakim, Stoica, and Papadimitriou. *"Strategic clustering."* Manuscript in preparation. 2023. Previously at StratML @ Neurips 2021.
[15] Golitschek, M.v. *"Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones".* Numerische Mathematik 39 (1), 65–84. 1982.
[16] Lawler, E.L. *"Optimal cycles in doubly weighted directed linear graphs".* In Proceedings of the International Symposium of Theory of Graphs. 209–232. 1966.

Challenges: local optimality

Lawler: we can find the minimum cycle Δf/Δq
⇔ deciding whether there is a neg cycle

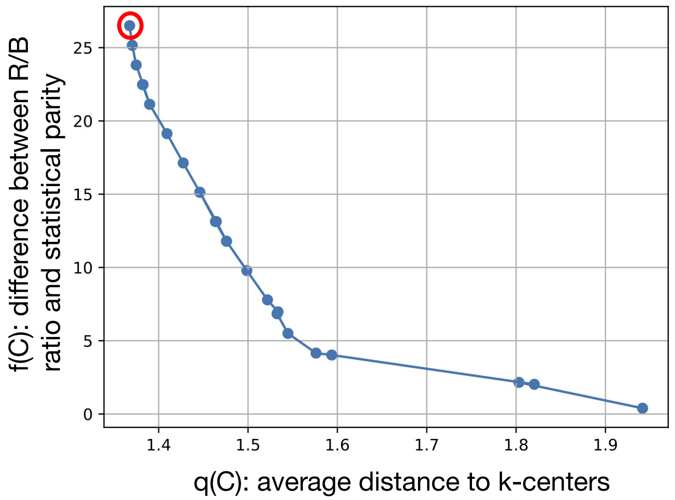
IF there are no negative cycles in the q-weight!

⇔ we are at an optimal clustering in the quality metric

Solution: optimality of $q(C) + \alpha \cdot f(C), \text{ for some } \alpha$

When does this work?

- Linear functions for q and f: great, but we could use a greedy algorithm as well
- Non-linear functions: NP-hard instance for some cases, empirically good results



[14] Hakim, Stoica, and Papadimitriou. *"Strategic clustering."* Manuscript in preparation. 2023. Previously at StratML @ Neurips 2021.
[15] Golitschek, M.v. *"Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones".* Numerische Mathematik 39 (1), 65–84. 1982.
[16] Lawler, E.L. *"Optimal cycles in doubly weighted directed linear graphs".* In Proceedings of the International Symposium of Theory of Graphs. 209–232. 1966.

Challenges: local optimality

Lawler: we can find the minimum cycle Δf/Δq
⇔ deciding whether there is a neg cycle

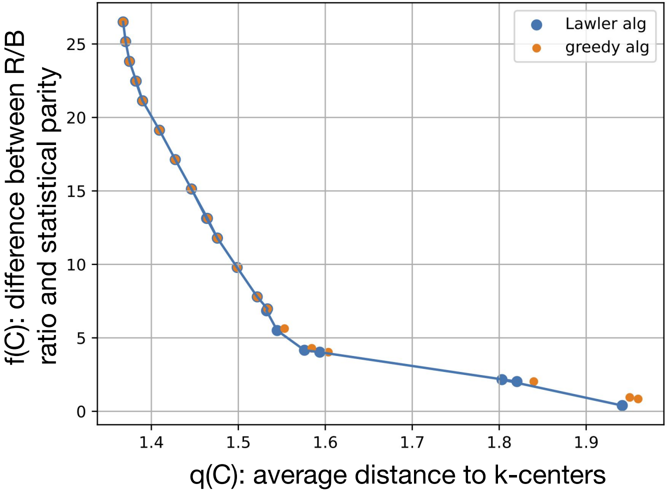
IF there are no negative cycles in the q-weight!

⇔ we are at an optimal clustering in the quality metric

Solution: optimality of $q(C) + \alpha \cdot f(C), \text{ for some } \alpha$

When does this work?

- Linear functions for q and f: great, but we could use a greedy algorithm as well
- Non-linear functions: NP-hard instance for some cases, empirically good results



[14] Hakim, Stoica, and Papadimitriou. *"Strategic clustering."* Manuscript in preparation. 2023. Previously at StratML @ Neurips 2021.
[15] Golitschek, M.v. *"Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones".* Numerische Mathematik 39 (1), 65–84. 1982.
[16] Lawler, E.L. *"Optimal cycles in doubly weighted directed linear graphs".* In Proceedings of the International Symposium of Theory of Graphs. 209–232. 1966.

Future directions

Thank you!

- Normative questions regarding interventions & policy implications
    - Budgetary constraints imply strong trade-offs
    - Constraints vs. multi-objective optimization
- Power & inequality:
    - Bias can be a sign of inefficiency
        - objectives are really hard to achieve and proxies fail
        - long-term dynamics differ from short-term interventions