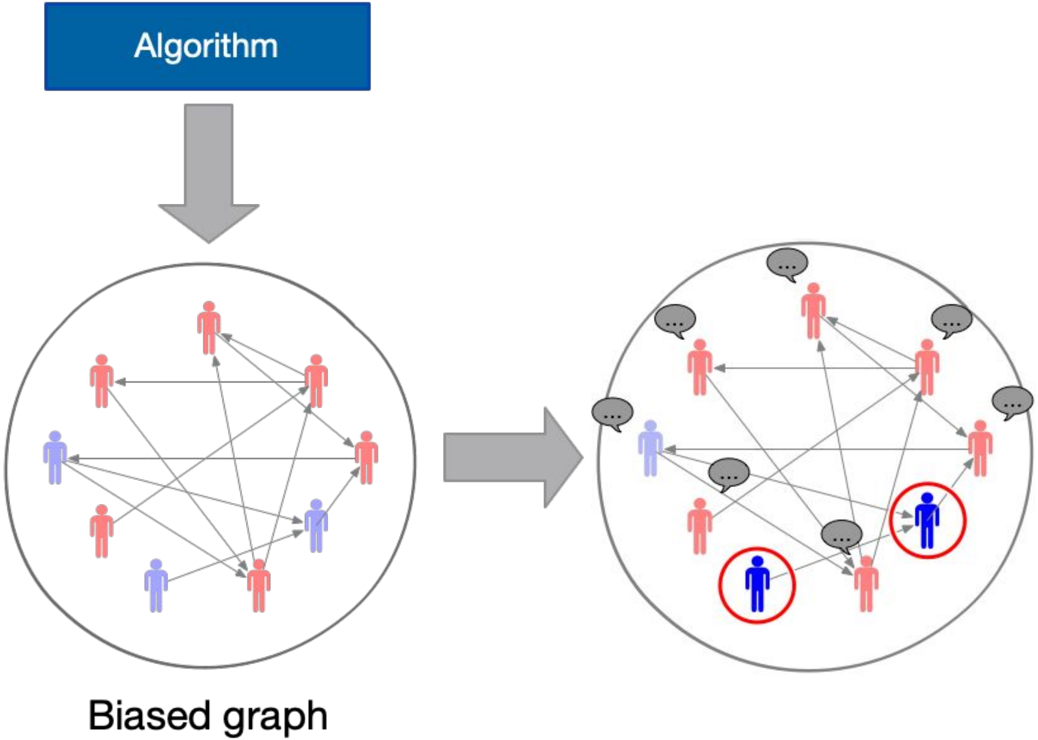
Inequality in graph algorithms: From influence to recommendation

Ana-Andreea Stoica

Columbia University

MGGG Networks & Graphs Workshop
July 2020

Inequality & bias in graphs



Inequality & bias in graphs

- Social influence maximization problem
- Recommendation algorithms

➔ Leverage the network structure in understanding patterns of inequality
➔ Focus on community-based inequality

Social Influence

Information propagated through a social network:

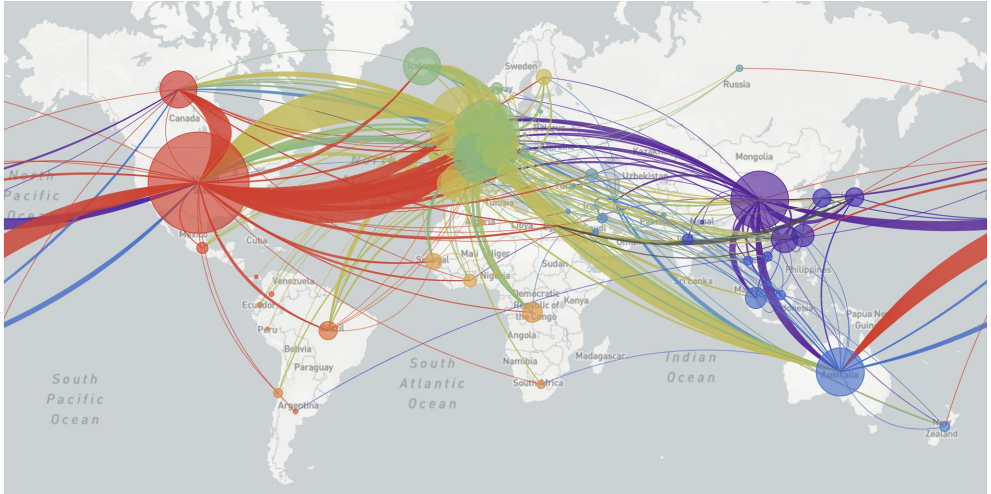
- The news we read
- The technologies we hear about
- Running marketing promotions
- Public health issues

Social Influence

Information propagated through a social network:

- The news we read
- The technologies we hear about
- Running marketing promotions
- **Public health issues**

Social influence & public health



Spread of COVID-19 by country

Social influence & public health

The Diffusion of Microfinance

Abhijit Banerjee,* Arun G. Chandrasekhar,* Esther Duflo,* Matthew O. Jackson*

Introduction: How do the network positions of the first individuals in a society to receive information about a new product affect its eventual diffusion? To answer this question, we develop a model of information diffusion through a social network that discriminates between information passing (individuals must be aware of the product before they can adopt it, and they can learn from their friends) and endorsement (the decisions of informed individuals to adopt the product might be influenced by their friends' decisions). We apply it to the diffusion of microfinance loans, in a setting where the set of potentially first-informed individuals is known. We then propose two new measures of how "central" individuals are in their social network with regard to spreading information; the centrality of the first-informed individuals in a village helps significantly in predicting eventual adoption.

Social influence & public health

ARTICLE

Bridging the gap between theory and practice in influence maximization: raising awareness about HIV among homeless youth

Authors: Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, Darlene Woo

Authors Info & Affiliations (Less)

Publication: IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence • July 2018 • Pages 5399–5403

Access to information is access to opportunity/healthcare

Social influence & public health

ARTICLE

Bridging the gap between theory and practice in influence maximization: raising awareness about HIV among homeless youth

Authors: Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, Darlene Woo

Authors Info & Affiliations (Less)

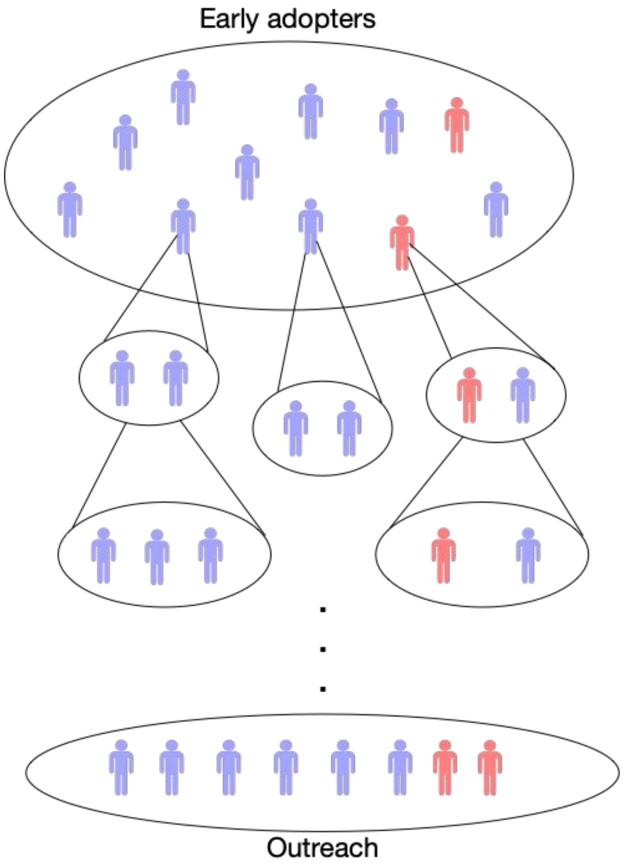
Publication: IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence • July 2018 • Pages 5399–5403

How do we design social influence campaigns that are fair?

Social Influence

- Given a network, pick the best k early-adopters ('seeds') that maximize outreach

- Algorithms that choose in a biased way:
  - Bias in social structure can lead to bias in outcomes[1]
  - Greedy algorithms that ignore labels are prone to reinforcing bias



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

What diversity interventions are helpful?

- Often posed as a parity constraint in an optimization function based on greedy algorithms (assuming full network information)

➡ **Fairness-efficiency trade-off**

Different approach:

- In reality networks are partially known => focus on centrality measures
- Add diversity interventions for early adopters & tap into inactive communities

➡ **More equity increase efficiency (outreach)**

Social Influence [WWW'20]

with Jessy Xinyi Han, Augustin Chaintreau

- **Our vision:** bias as a sign of inefficiency
  - Diversity is beneficial[2]
  - Tapping into inactivated communities

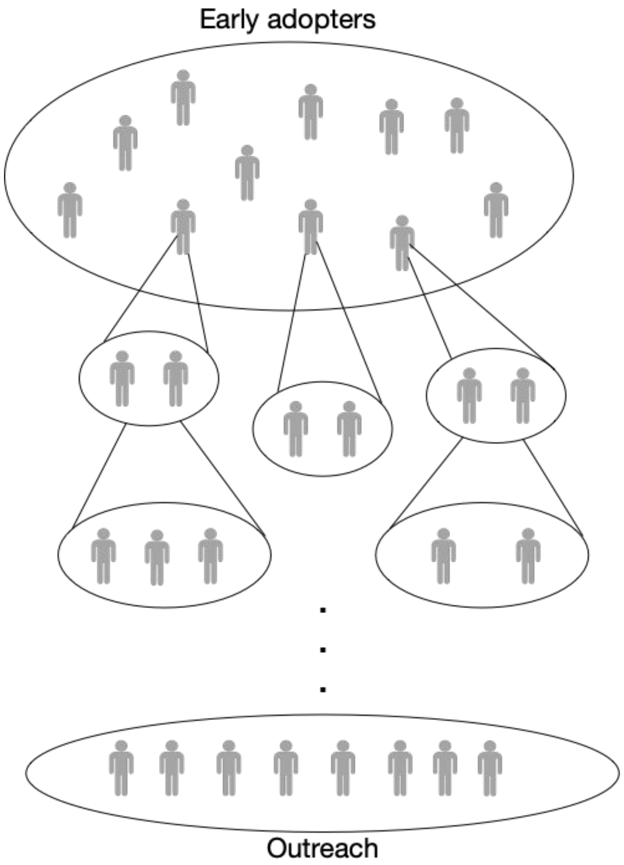
[2] Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press, 2008.

Social Influence [WWW’20]

with Jessy Xinyi Han, Augustin Chaintreau

- **Our vision:** bias as a sign of inefficiency
    - Diversity is beneficial[2]
    - Tapping into inactivated communities

- Seeding can be done **agnostically**: ignore labels, already takes into account network structure

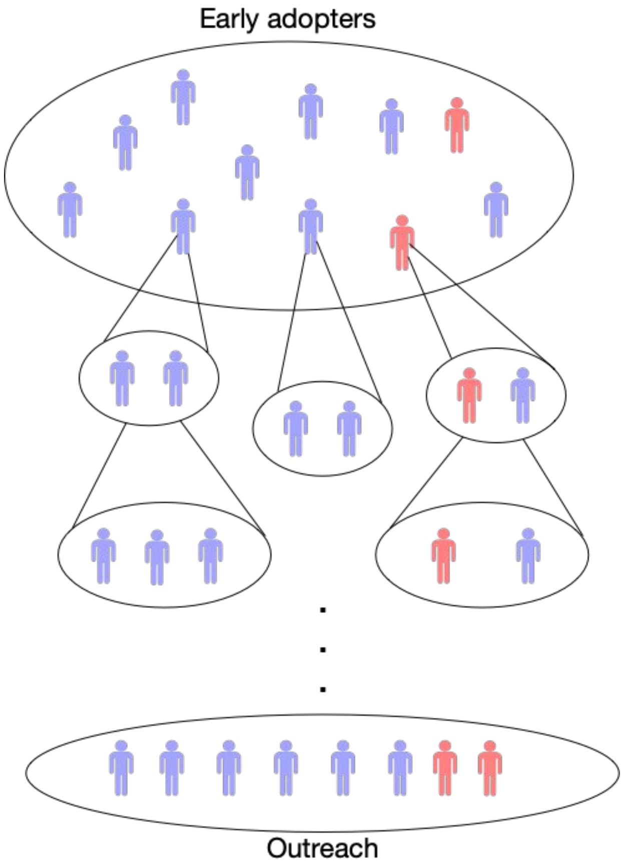


[2] Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press, 2008.

Social Influence [WWW'20]

with Jessy Xinyi Han, Augustin Chaintreau

- **Our vision:** bias as a sign of inefficiency
  - Diversity is beneficial[2]
  - Tapping into inactivated communities
- Seeding can be done **agnostically**: ignore labels, already takes into account network structure



[2] Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press, 2008.

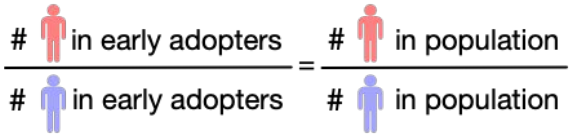
Social Influence [WWW'20]

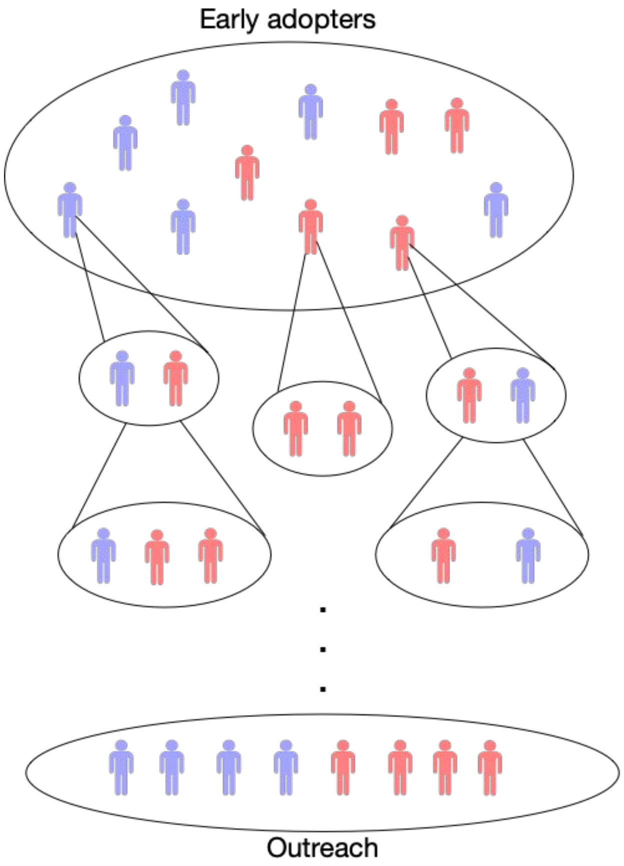
with Jessy Xinyi Han, Augustin Chaintreau

- **Our vision:** bias as a sign of inefficiency
  - Diversity is beneficial[2]
  - Tapping into inactivated communities

- Seeding can be done **agnostically**: ignore labels, already takes into account network structure

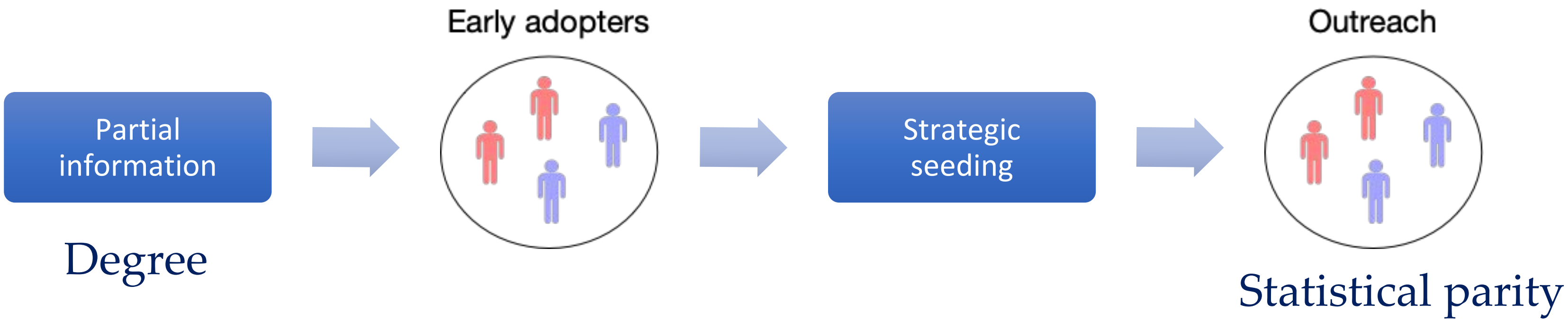
- Seeding can be done with awareness of labels: have **statistical parity** in early adopters





[2] Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press, 2008.

Problem Formulation



- Obtain statistical parity in the seed selection as a **tool** to get parity in outreach
- **Key idea: awareness of communities or sensitive features in a strategic way**

Model for biased networks

- Preferential attachment model with:
    - 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]

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

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

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

Glass ceiling effect

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

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

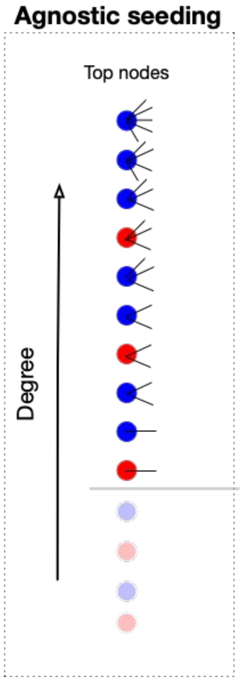
A graph sequence G(n) exhibits a *tail glass ceiling effect* for the R nodes if there exists an increasing function k(n) such that

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

$\Rightarrow$ **Degree centrality will be biased!**

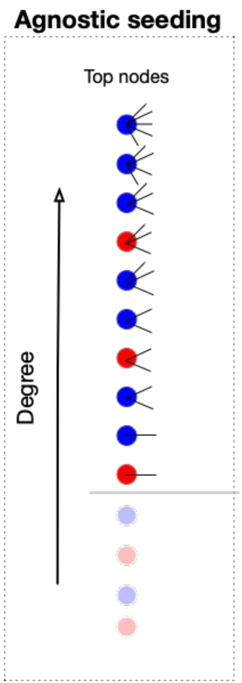
Color-agnostic v. Diversity Seeding

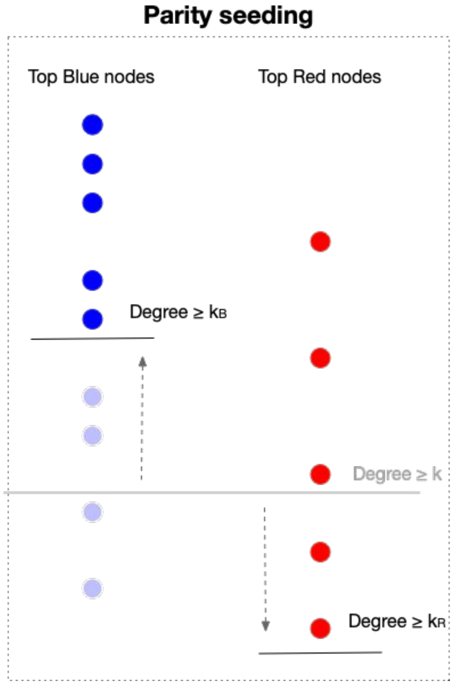
Early adopters (seedset)



Color-agnostic v. Diversity Seeding

Early
adopters
(seedset)

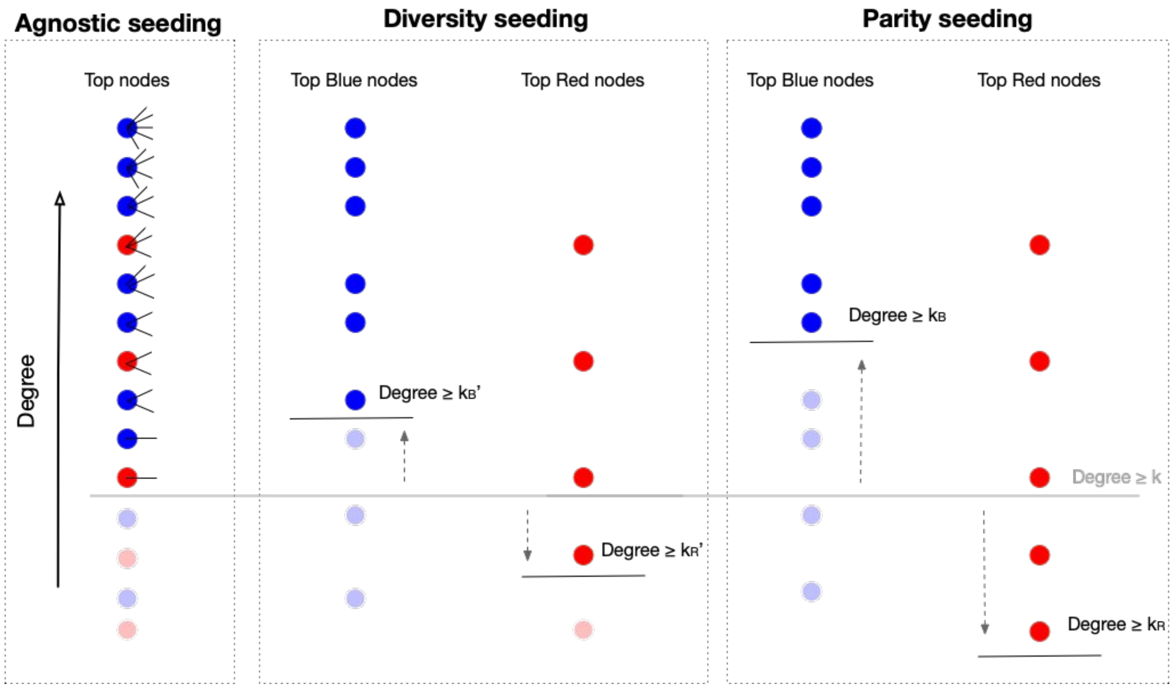




Keeping the same budget!

Color-agnostic v. Diversity Seeding

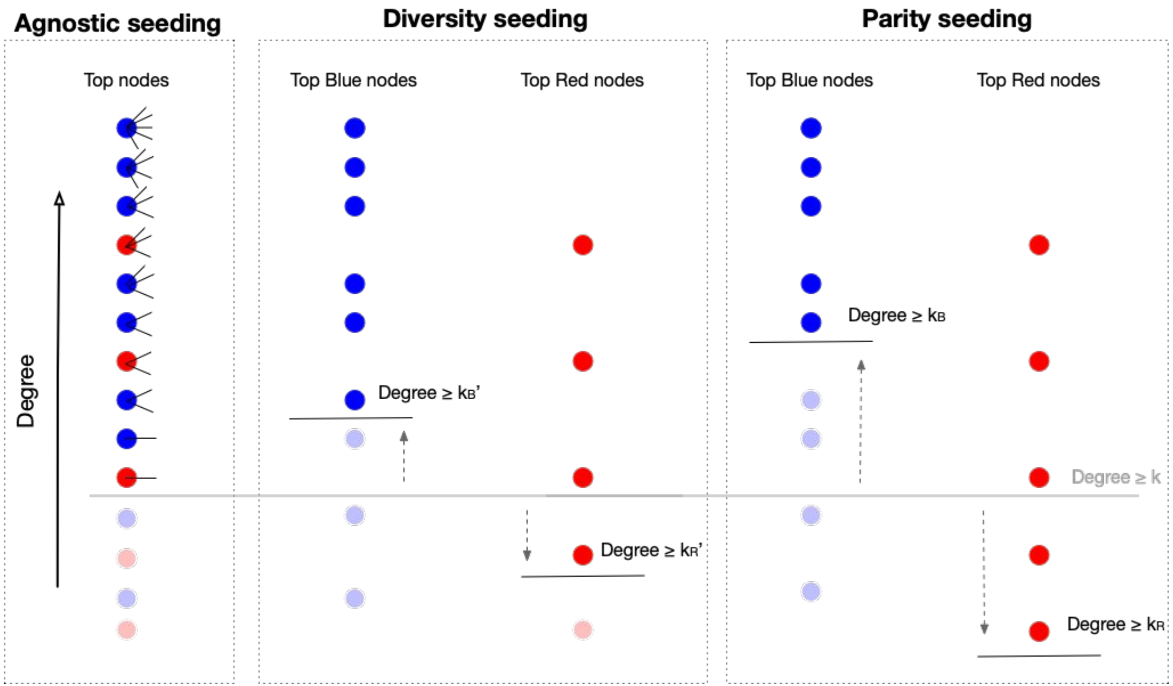
Early adopters (seedset)



Keeping the same budget!

Color-agnostic v. Diversity Seeding

Early adopters (seedset)



How do these three heuristics perform?

Results

Claim #1: *Diversity seeding and parity seeding leads to fairer outreach for the same budget*

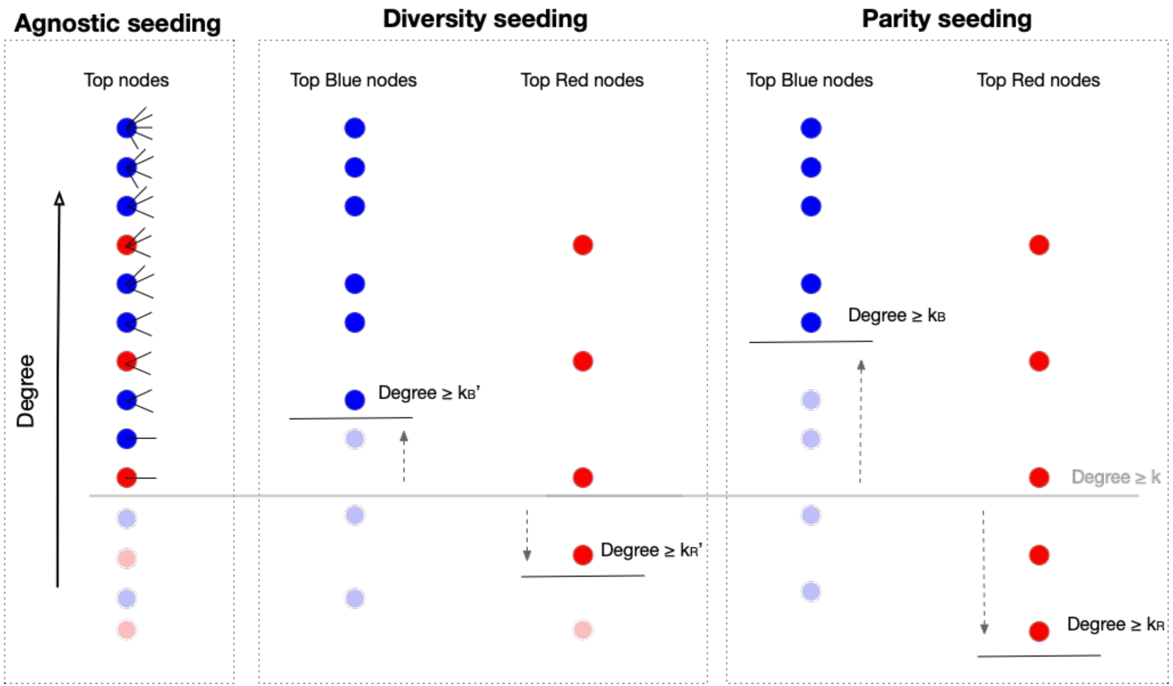
Claim #2: *Diversity seeding and parity seeding can outperform agnostic seeding in outreach*

- *When? For a large enough campaign**

Claim #3: **Analytical condition for the size of the seedset needed (size of the campaign)*

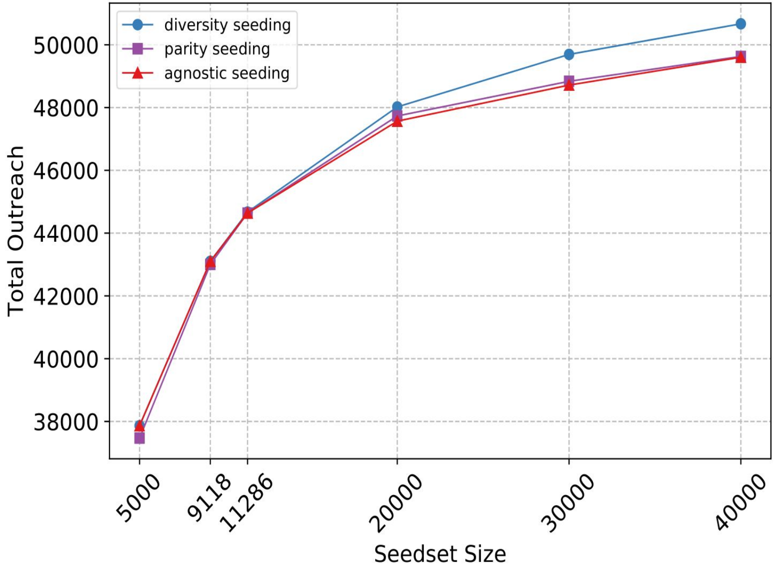
- *Computed from a theoretical model of biased networks*

Color-agnostic v. Diversity Seeding



How do these three heuristics perform?

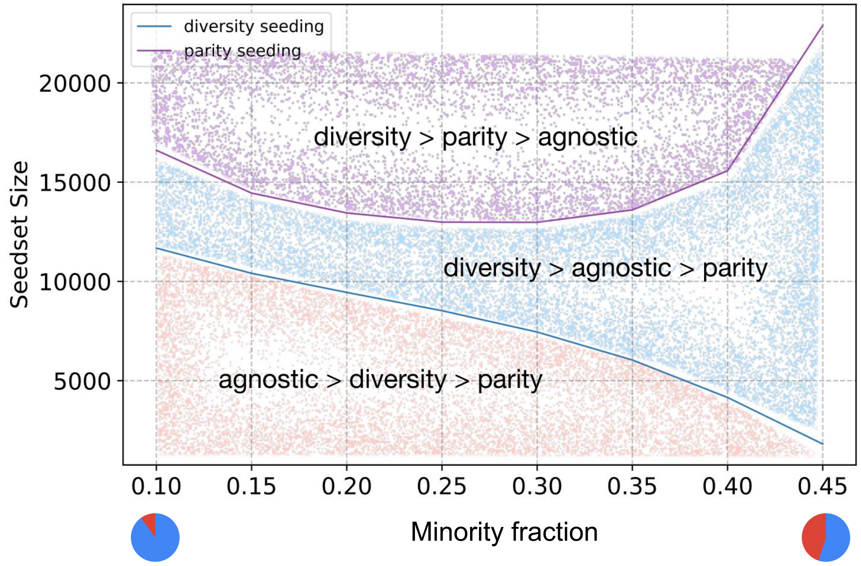
Theoretical analysis of diversity interventions



Network of ~53,000 nodes, two communities (minority 18.6%), homophily $\rho = 0.135$

- Simulate network similar to our dataset
- For more than $k_d = 9{,}118$ seeds, diversity > agnostic > parity
- For more than $k_p = 11{,}286$ seeds, diversity > parity > agnostic

Theoretical analysis of diversity interventions



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

Fairness can reinforce efficiency

- Conditions for which fairness does not come at a cost
  - Test their applicability
- Intuition: tapping into inactive communities
  - Use the biased preferential attachment model degree distribution:

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

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

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

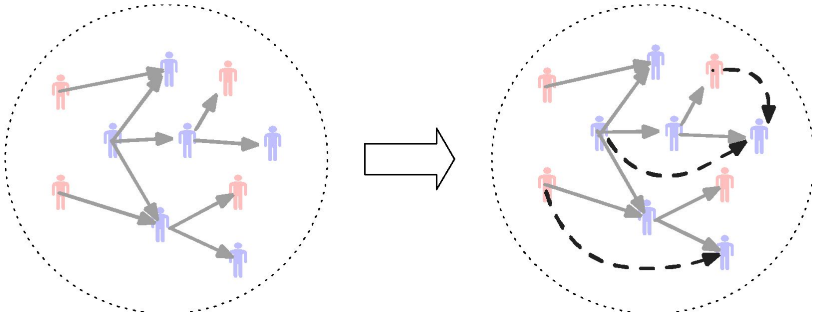
  - Estimate size of cascade for each community

Algorithmic Glass Ceiling in Social Networks [WWW'18]

with Christopher Riederer, Augustin Chaintreau

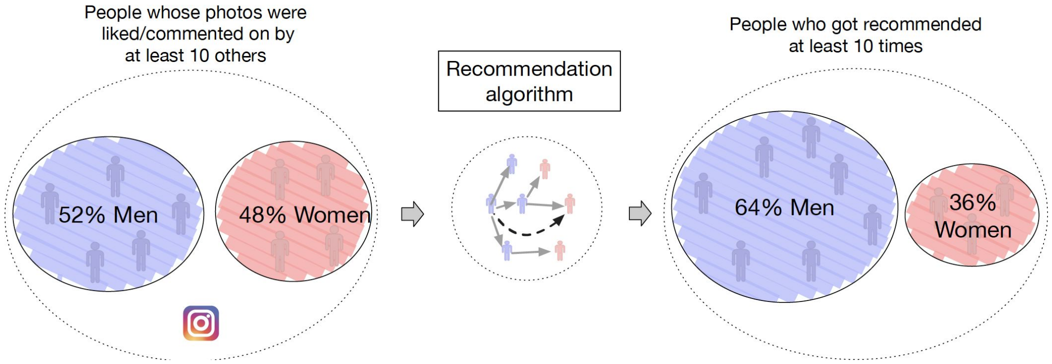
Recommendation algorithms:

- Adamic-Adar (data)
  - Score between each pair of vertices based on # common neighbors
- Random walk of length 2 (data + theory)



Recommendation algorithms on Instagram

- Degree imbalance between men and women with high degrees
  - i.e., although male are a minority, they get most of high degrees
- If you count recommendations instead of degrees, it gets worse



Model evolution

At timestep t, a new edge is formed:

<u>Organic growth:</u>

New node connects:

- randomly
- preferential attachment + homophily

<u>Recommendation model:</u>

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

Degree distribution

Organic growth:

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

Recommendation model:

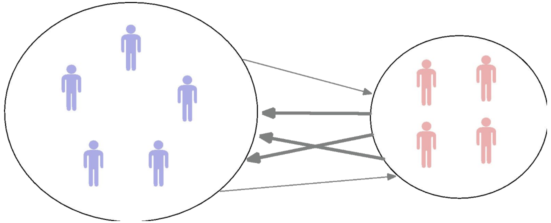
$$top_k'(R) \sim k^{-\beta_{rec}(R)}$$
$$top_k'(B) \sim k^{-\beta_{rec}(B)}$$

***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:*

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

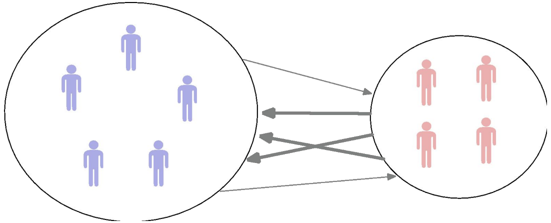
Differential homophily

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

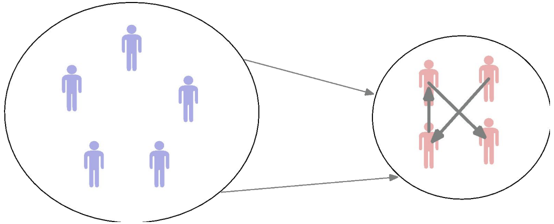


Differential homophily

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



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



Key takeaways

- Seemingly neutral algorithms may reinforce inequality
  - Recommendation & influence algorithms leverage network structure
  - Community structure => homophily
- Design algorithms with *awareness* of network structure

Thank you!

@astoica73
astoica@cs.columbia.edu