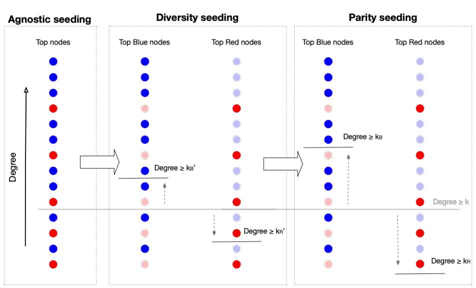
Diversity and Bias in the Influence Maximization Problem

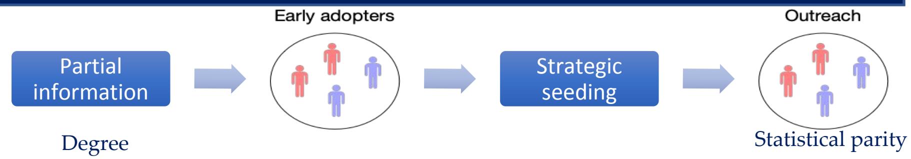
Ana-Andreea Stoica, Jessy Xinyi Han, Augustin Chaintreau

Introduction

We study diversity interventions in the problem of social influence maximization: choosing a set of seeds in a network that maximize the spread of a message, service, or technology. We work with a theoretical model of networks that exhibit community structure and homophily to show that community-aware seeding can lead to a better efficiency in certain conditions, which we analyze analytically in terms of the number of seeds. We propose heuristics to enhance diversity that illustrate the trade-off between diversity and outreach maximization on real networks.



Problem formulation & Results

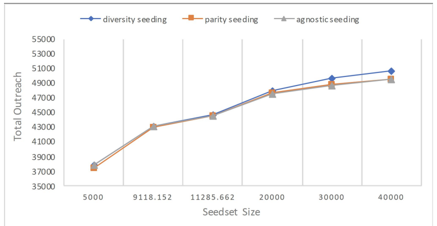


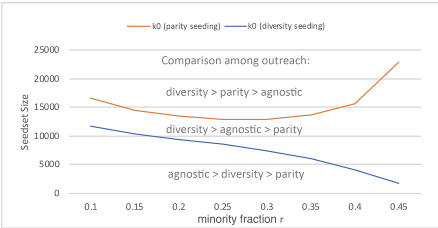
- Obtain statistical parity in the seed selection as a **tool** to get parity in outreach
- Can we be strategic in choosing seeds, while being as efficient?

***Theorem**: For a sufficiently large seedset* we can achieve both parity in outreach and efficiency: strategic seeding can achieve*

- *A better R/B outreach ratio than the color agnostic strategy,*
- *Relaxed-statistical parity in seed selection with the same seed budget,*
- *Better outreach overall.*

***Corollary:** We can analytically compute the seedset size as the root of a function computing marginal influence between **diversity** seeding (k_r) and agnostic, and between **parity** (k_c) seeding k_c and agnostic.*





Theory

Models of network evolution based on a bi-populated network (red and blue nodes) :

- Fraction of red nodes = r < ½

At timestep t, a new edge is formed:

Organic growth model:

A new node connects naturally:

- η: randomly
- 1 – η: preferential attachment

Homophily: if ≠ labels, edge is accepted w.p. ψ

Diffusion model:

- Independent cascade model, each node accepts influence from a neighbor w.p. p

Bias

Key idea: **Homophily** leads to a different degree distribution for the two communities.

Bias in centrality measures affects who receives the message

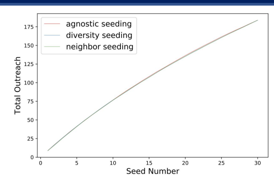
Data

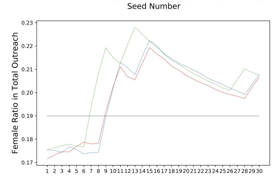
Data from DBLP:

- Co-authorship graph ~50,000 people,
- 81% men, 19% women,
- Homophily: people cluster based on gender.

Results:

- Diversity-enhancing heuristics improve equity in information access and **no cost** to total outreach.
- Can extend notions of equitable access beyond seed selection and outreach, ensuring diversity at different levels of information cascade.





Proof sketch & Conclusion

Power law dynamics of degree distribution:

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

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

The expected size of a seedset with nodes of degree at least k(n) can be computed:

$$\mathbb{E}(|S_{k(n)}|) = n \cdot \alpha \cdot d \cdot \frac{\beta(R)-2}{\beta(R)-1} \cdot k(n)^{1-\beta(R)}$$

$$+n \cdot (1-\alpha) \cdot d \cdot \frac{\beta(B)-2}{\beta(B)-1} \cdot k(n)^{1-\beta(B)},$$

based on the network parameters. The condition from our corollary turns into solving for the roots of a function:

$$F(x) = \mathbb{E}(|\phi_V(S^B_{k^B(x)} \cup S^R_{k^R(x)})|) - \mathbb{E}(|\phi_V(S_{k(n)})|)$$

Key take-aways:

- **Seeding can mirror biases existing in the network:**
  - **In the people chosen as early-adopters,**
  - **In the population that receives information.**
- **We can ensure fairness without losing outreach in certain conditions that are achievable in real datasets.**
- **Our model allows us to compute such conditions analytically and to design better algorithms for information diffusion.**