The Effects of Homophily and Influence On Voting Games

Thesis Defense in Requirements of Thesis Track for Master of Science in Computer Science at Columbia University

Reetahan Mukhopadhyay
Advisor: Augustin Chaintreau

December 2022

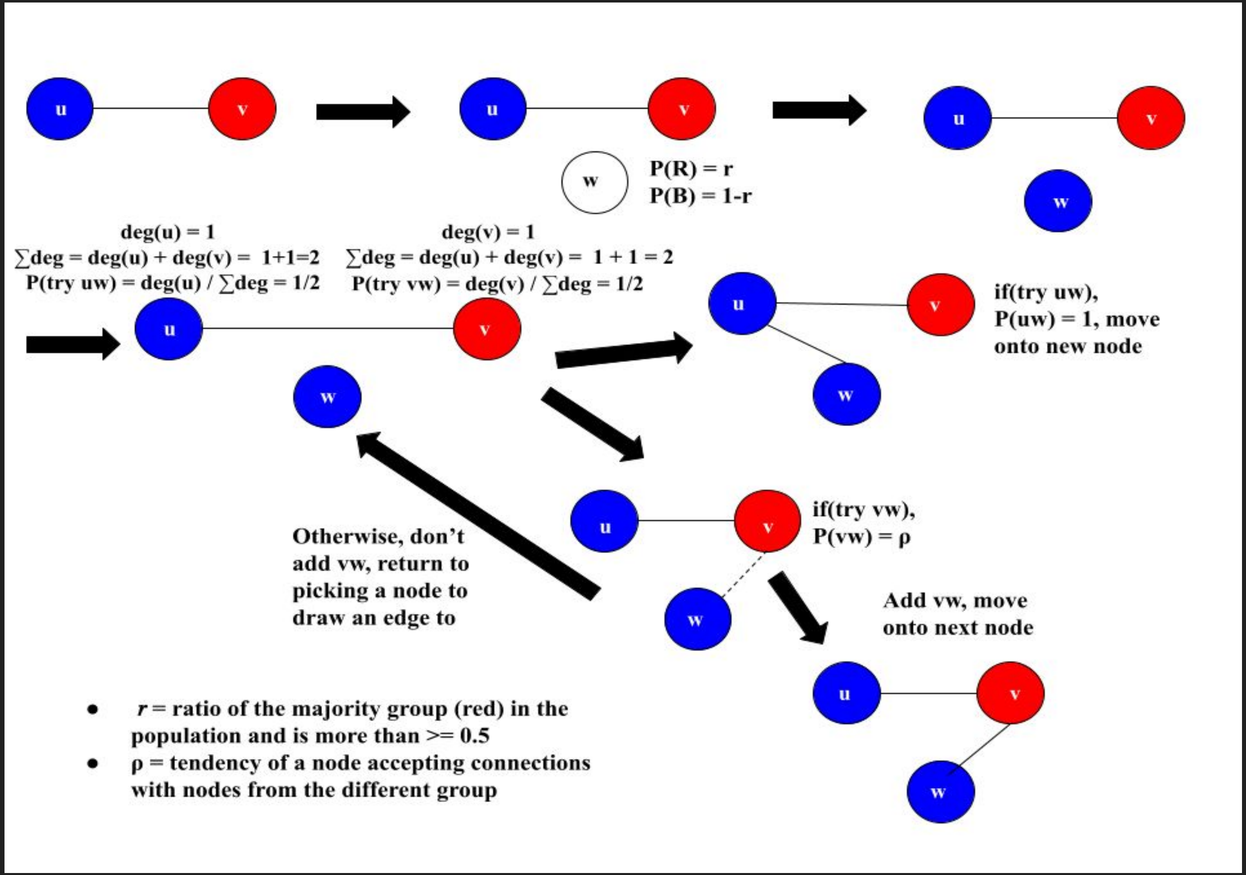
Columbia Engineering
The Fu Foundation School of Engineering and Applied Science

Other Models….

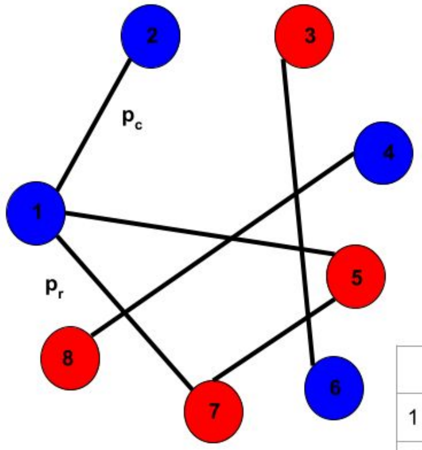
- Voter Model
  - Voters choose opinion proportional to neighbors
  - Must converge at full consensus
- Independent Cascade Model
  - Nodes start as inactive except for starting set, can only be activated
  - Nodes get one shot at activating neighbors
- Linear Threshold Model
  - Nodes start as inactive except for starting set, can only be activated
  - Node is activated when sum of indegree weight from activated set matches threshold

My Model

Biased Preferential Attachment (BPA)

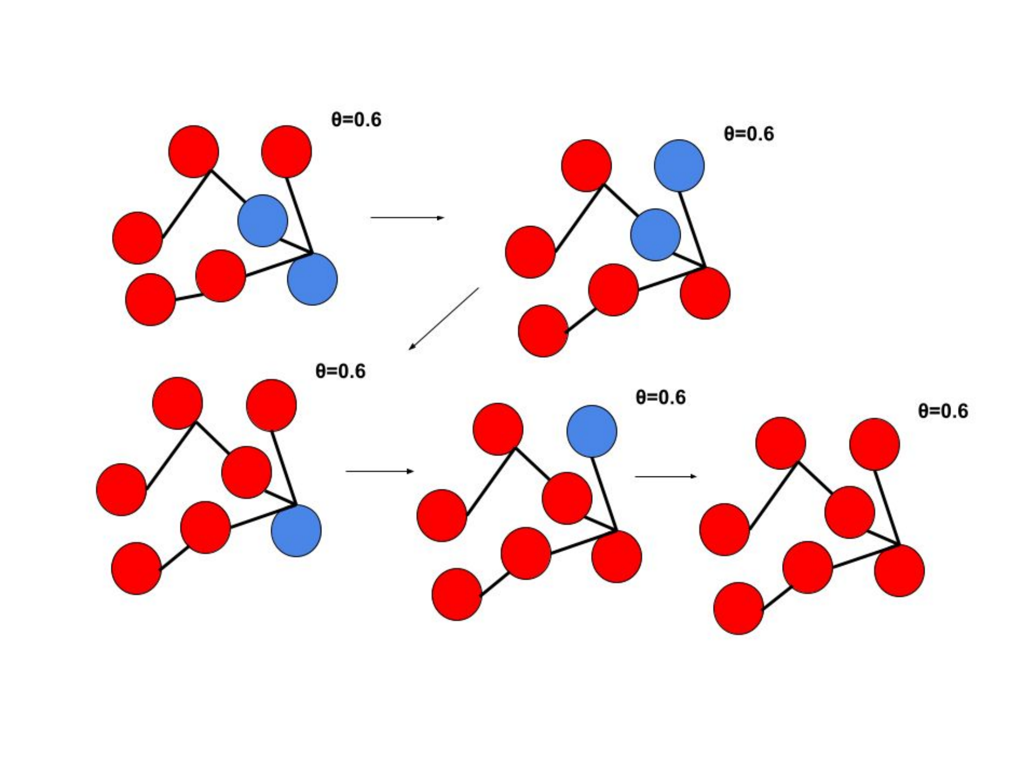


Stochastic Block (SB)



	1	2	3	4	5	6	7	8
1	0	p_c	p_r	p_c	p_r	p_c	p_r	p_r
2	p_c	0	p_r	p_c	p_r	p_c	p_r	p_r
3	p_r	p_r	0	p_r	p_c	p_r	p_c	p_c
4	p_c	p_c	p_r	0	p_r	p_c	p_r	p_r
5	p_r	p_r	p_c	p_r	0	p_r	p_c	p_c
6	p_c	p_c	p_r	p_c	p_r	0	p_r	p_r
7	p_r	p_r	p_c	p_r	p_c	p_r	0	p_c
8	p_r	p_r	p_c	p_r	p_c	p_r	p_c	0

Threshold Influence



Justification and Novelty

- Influence parameter allows for an individual to consider how much they value their own independent thought
- Realistic analogue of convergence process a bit difficult but can think of as agents influencing each other over the course of something such as an election cycle, where discrete time intervals are like polls

- Model does not appear to be discussed or analyzed elsewhere
    - Similar to Linear Threshold but allows nodes to return to “deactivated” state
- Entire Network behavior can be thought of as Markovian process with $O(2^n)$ states

Theoretical Results

Graph Properties

- BPA
  - Node Degree Distribution follows a **Power Law / Scale-Free**
  - **Acyclic**
  - **Sparse**
  - **Rich-get-richer / Preferential Attachment**
  - **Homophilic**
- **Stochastic Block**
  - **Generally dense**
  - **Many cycles**
  - **Homophilic**

We can follow through with modelling each process to understand distribution of red neighbors

$$\mathbb{E}[\alpha_{t+1}|\alpha_t] = \alpha_t + \frac{F(\alpha_t) - \alpha_t}{t+1}, \qquad \textit{where}$$

$$F(x) = \left(1 - (1-r)\frac{(1-x)}{1-x(1-\rho)} + r\frac{x}{1-(1-x)(1-\rho)}\right)/2.$$

Theorem 3.8. *The red fraction distribution $\alpha(u)$ for $u \in V$ of a network of n nodes formed by SB with a red node proportion of r, intra-homophily factor of p_c and inter-homophily factor p_r has a cumulative distribution function as such:*

$$P(\alpha(u) \geq \alpha^*) = \sum_{i=1}^{n-1} \frac{\sum_{k=\lceil \alpha^* i \rceil}^{i} F_{ki}}{\sum_{j=0}^{i} F_{ji}}$$

where

$$F_{xi} = \binom{nr-1}{x} p_c (1-p_c)^{nr-1-x} \binom{n-nr}{i-x} p_r^{i-x} (1-p_r)^{n-nr-i+x}$$

Theorem 3.7. *The red fraction distribution $\alpha(u)$ for a red node $u \in V$ of a network of n nodes formed by BPA with a red node proportion of r, homophily factor of p has a cumulative distribution function as such:*

$$P(\alpha(u) \geq \alpha^*) = \sum_{i=1}^{n} \sum_{j=1}^{|2^M|} \sum_{k=1}^{|2^{Q_j}|} \frac{\prod_{a=1}^{|T_k|} F_{ka} \prod_{b=1}^{|S_k|} 1 - F_{kb}}{\prod_{c=1}^{|Q_j|} W_{jc} \prod_{d=1}^{|R_j|} 1 - W_{jd}}$$

where

$$M = \{i, ..., n\}$$

$$Q = \{q \mid q \in 2^M\} \textit{ for } 2^M = \textit{power set of } M$$

$$T = \{t \mid t \in 2^{Q_j} \wedge |t| \geq \lceil \alpha^* |Q_j| \rceil\} \textit{ for } 2^{Q_j} = \textit{power set of } Q_j$$

$$R = \{\beta \mid \beta \in \mathbf{Z} \wedge i \leq \beta \leq n \wedge \beta \notin Q_j\}$$

$$S = \{\omega \mid \omega \in \mathbf{Z} \wedge i \leq \omega \leq n \wedge \omega \notin T_k\}$$

$$W_{jx} = \frac{r(x+1)}{2(Q_{jx}+1) - (2(Q_{jx}+1) - (x+1))(1-p)} + \frac{(1-r)(x+1)p}{2(Q_{jx}+1) - (x+1)(1-p)}$$

$$F_{kx} = \frac{r(x+1)}{2(T_{kx}+1) - (2(T_{kx}+1) - (x+1))(1-p)}$$

Influence Convergence

Forbidden Subgraphs

Influence Process ***fails*** to converge if we have a forbidden subgraph such that

- Every node within the subgraph fails to meet the threshold condition for its own current color
- Number of neighbors node has within subgraph (constantly flipping) must exceed number of neighbors outside subgraph (constant)

Lemma 3. *If $G_{\theta,i}$ is such that $\forall u_i \in G_{\theta,i} \mid \frac{|N(u_i)_{C(u_i))}|}{|N(u_i)|} < \theta \wedge \frac{|N(u_i)_{\neg C(u_i))}|}{|N(u_i)|} \geq \theta$ then $\mathbf{I}_{G,\theta}$ will not converge.*

Lemma 4. *If $G_{\theta,i}$ is colored to match a bipartite assignment of its nodes, then $\mathbf{I}_{G,\theta}$ will not converge.*

Influence Maximization

	1.0	r	0
RR	(b,1]	[0,b]	$\emptyset$
RB	(b,1]	[0,r]	(r,b]
BR	(b,r]	[0,b]	(r,1]
BB	$\emptyset$	[0,r]	(r,1]
R$\frac{1}{2}$	(0.5,1]	[0,0.5)	$\emptyset$
B$\frac{1}{2}$	$\emptyset$	[0,0.5)	[0.5,1)

Table 1: Where UV corresponds to node u of color U and neighbor majority of color V, we look at the intervals required to have a weight of 1, the red neighbor fraction r, or 0

Proposed Algorithms

- Optimize for threshold that maximizes sets
- Optimize for degree of the above

Theorem 3.4. *The optimal interval* $\theta_{OPT} \in [\theta_{OPT,min}, \theta_{OPT,max}]$ *that maximizes* $|G_{R,\theta,k}|$ *given that* $\mathbf{I}_{\{G,\theta,k\}}$ *converges at* $i \leq k$ *will be such that* $\theta = argmax_\theta \sum_{u \in BR_\theta} \psi(u) - \sum_{v \in RB_\theta} \psi(v)$ *where* $BR_\theta = \{u_0 \in G_0 \wedge C(u_0) = B \wedge C(u_1) = R\}$ *and* $RB_\theta = \{u_0 \in G_0 \wedge C(u_0) = R \wedge C(u_1) = B\}$

Key claim:

If a node at time i u_i's red fraction rf (u_i) increased from timestep i − 1, the probability of node u_i itself being red at time i + 1 is non-decreasing (increasing for non-R 1/2 - B 1/2 cases).

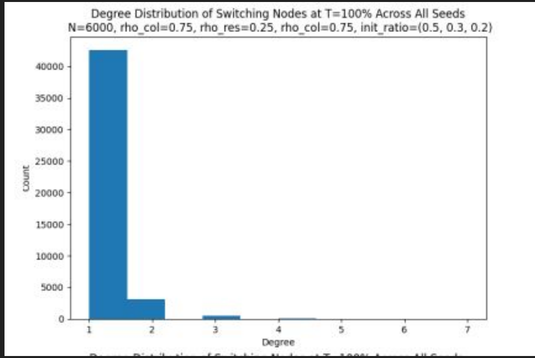
Algorithm to Maximize Red Votes at t=t+1

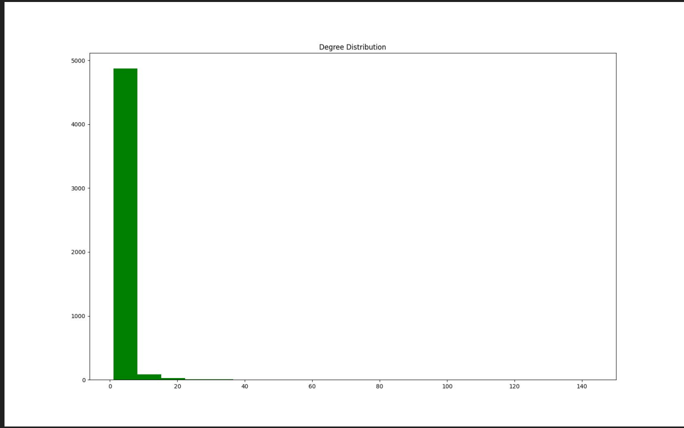
1. For all nodes u, if red fraction *q* is less than 0.5, generate interval [0,q) with weight q. If u itself happens to be red also add interval (1-q,1] with weight 1. If red fraction is exactly 0.5, generate interval [0,0.5] with weight 0.5, if u itself happens to be red, add (0.5,1] with weight 1. If q > 0.5, add interval (q,1] with weight 1 and instead (1-q,1] with weight 1 if u happens to be red, and [0,1-q] with weight q.
2. Mark each interval tuple with (Start,End) tags, then sort all the endpoints of the interval
3. Linearly traverse your sorted list from beginning to end, adding the weight w_i if the current endpoint was the start of interval i, and subtract weight w_i if the current endpoint is the end of interval i. Keep track of your sum.
4. Return the interval in which the sum was maximal as the optimal threshold

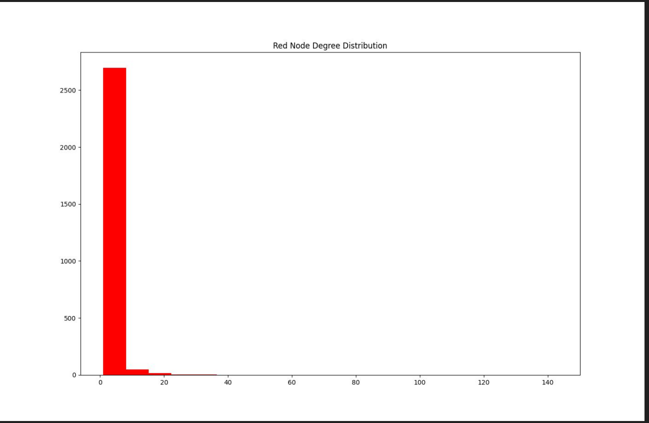
Putting It All Together: Homophily, Population Ratio, Influence Threshold

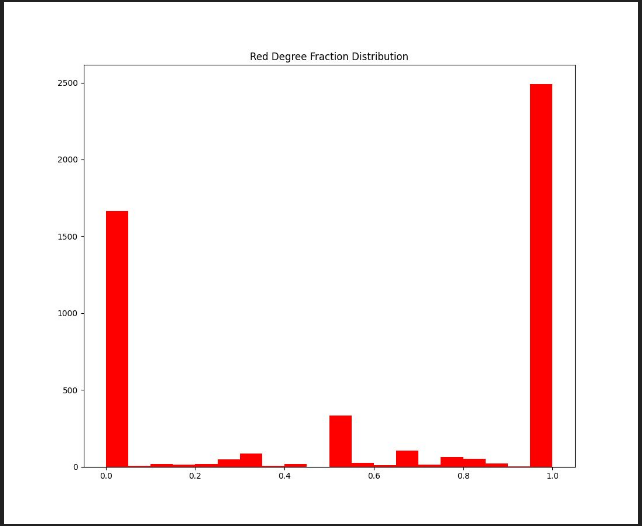
- Homophily + Minority Ratio Affect BPA => BPA Defines $G_{0,\theta}$ => Influence defines $G_{\infty,\theta}$
- Homophily increases interactions between majority and minority group
- Greater impact on increase in majority fraction on minority group than decrease in majority fraction on majority group
- Influence:
    - Threshold very high => Too high for many nodes to change
    - Threshold very low => Approaches voter model, change proportional directly to majority fraction, may encourage perturbations for minority
    - Middle Threshold => Large enough for majority to be valid but minority to miss the cut, has greatest encouragement of minority switches

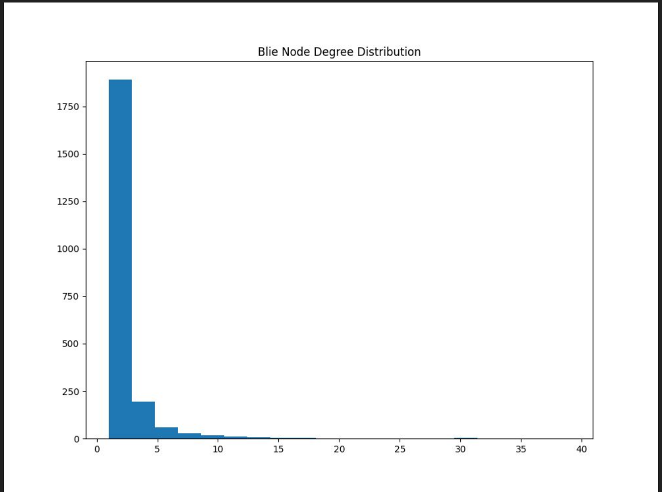
Simulation

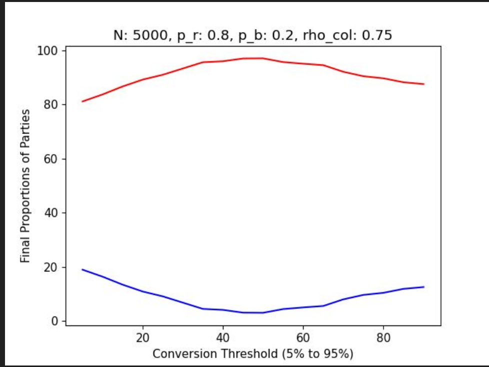

Degree Distribution of Switching Nodes at T=100% Across All Seeds
N=6000, rho_col=0.75, rho_res=0.25, rho_col=0.75, init_ratio=(0.5, 0.3, 0.2)
Count
Degree

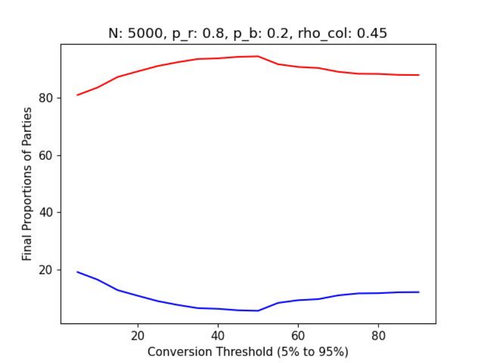

Degree Distribution

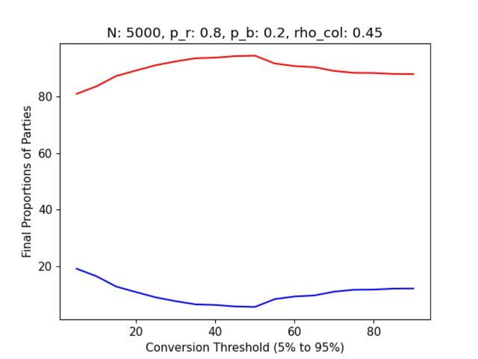

Red Node Degree Distribution

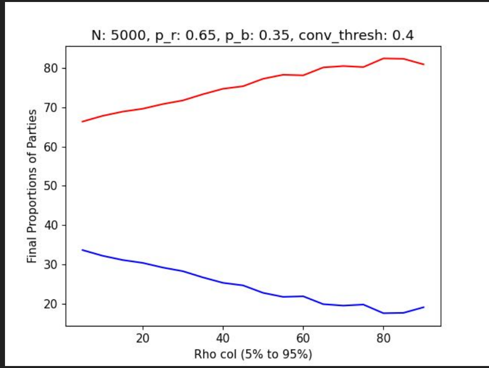

Red Degree Fraction Distribution

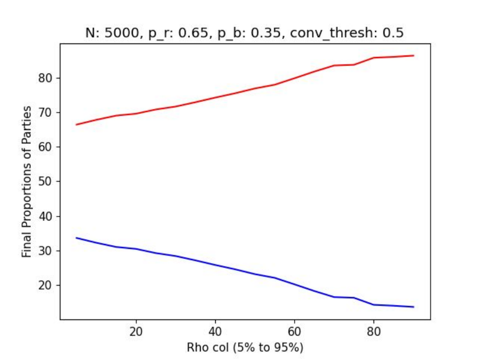

Blie Node Degree Distribution

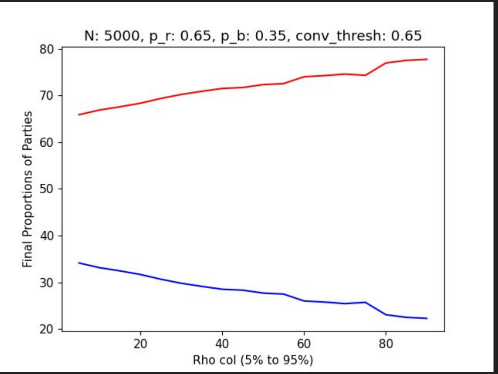

N: 5000, p_r: 0.8, p_b: 0.2, rho_col: 0.75
Final Proportions of Parties
Conversion Threshold (5% to 95%)


N: 5000, p_r: 0.8, p_b: 0.2, rho_col: 0.45
Final Proportions of Parties
Conversion Threshold (5% to 95%)

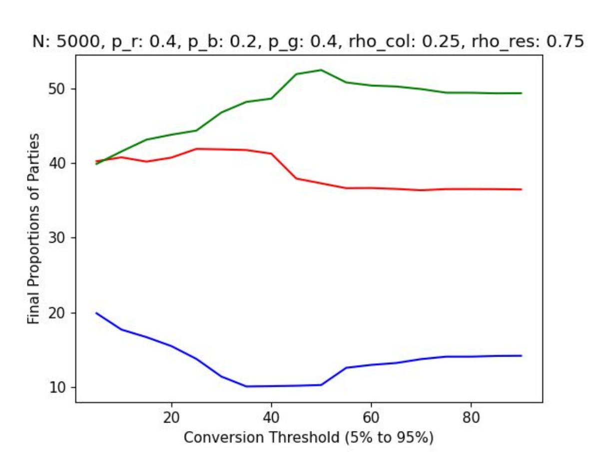

N: 5000, p_r: 0.8, p_b: 0.2, rho_col: 0.45
Final Proportions of Parties
Conversion Threshold (5% to 95%)

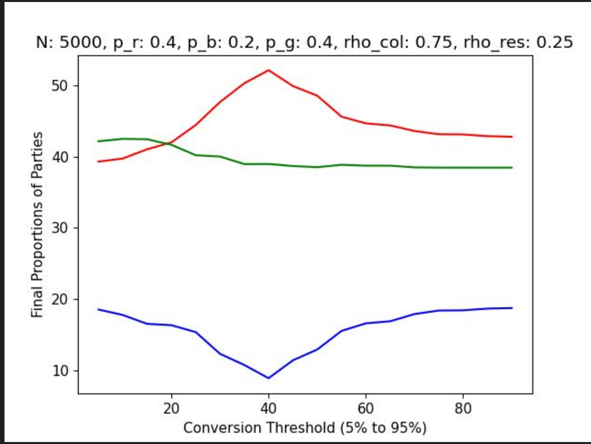

N: 5000, p_r: 0.65, p_b: 0.35, conv_thresh: 0.4
Final Proportions of Parties
Rho col (5% to 95%)

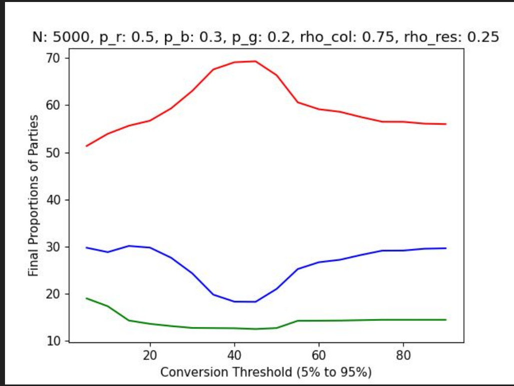

N: 5000, p_r: 0.65, p_b: 0.35, conv_thresh: 0.5
Final Proportions of Parties
Rho col (5% to 95%)

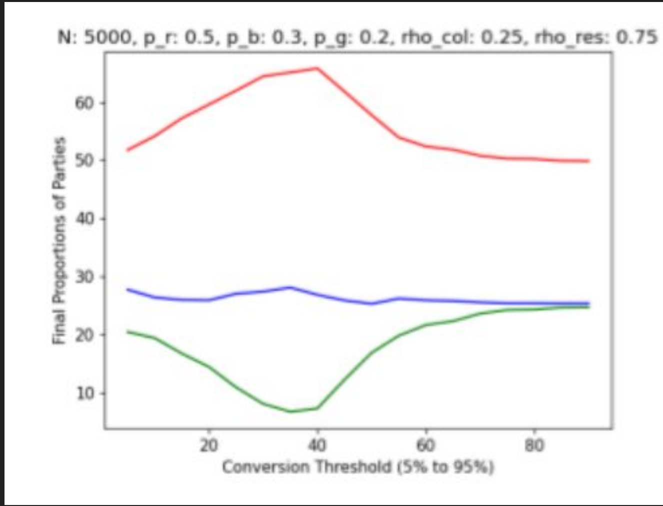

N: 5000, p_r: 0.65, p_b: 0.35, conv_thresh: 0.65
Final Proportions of Parties
Rho col (5% to 95%)

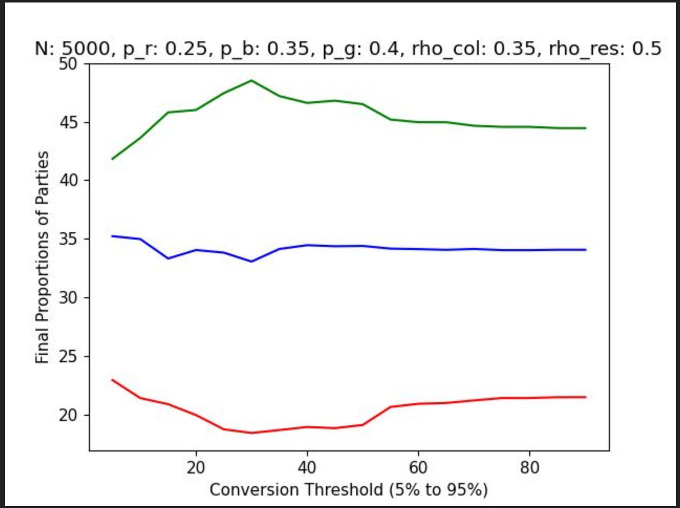
Extension: Three-Party Simulation


N: 5000, p_r: 0.4, p_b: 0.2, p_g: 0.4, rho_col: 0.25, rho_res: 0.75
Final Proportions of Parties
10
20
30
40
50
20
40
60
80
Conversion Threshold (5% to 95%)


N: 5000, p_r: 0.4, p_b: 0.2, p_g: 0.4, rho_col: 0.75, rho_res: 0.25
Final Proportions of Parties
10
20
30
40
50
20
40
60
80
Conversion Threshold (5% to 95%)


N: 5000, p_r: 0.5, p_b: 0.3, p_g: 0.2, rho_col: 0.75, rho_res: 0.25
Final Proportions of Parties
10
20
30
40
50
60
70
20
40
60
80
Conversion Threshold (5% to 95%)


N: 5000, p_r: 0.5, p_b: 0.3, p_g: 0.2, rho_col: 0.25, rho_res: 0.75
Final Proportions of Parties
10
20
30
40
50
60
20
40
60
80
Conversion Threshold (5% to 95%)


N: 5000, p_r: 0.25, p_b: 0.35, p_g: 0.4, rho_col: 0.35, rho_res: 0.5
Final Proportions of Parties
20
25
30
35
40
45
50
20
40
60
80
Conversion Threshold (5% to 95%)

Observations

- **The general 2-party observations seem to generalize to 3 parties**
- Dominant (> 0.5) party still cannot lose
- We can manipulate the parameters to allow the 3rd place party to come in 2nd by the end, or the #2 to overtake 1st place

Further Work and Concluding Remarks

Takeaways

- Developed new model for influence propagation / evolutionary behavior
- Increasing homophily reduces the minority party's power
- Optimal region for influence threshold that minimizes the minority party's power
- Influence method converges under lack of forbidden substructure, several possible ways to optimize it

Further work

- Optimality of Influence Optimization? Average-case rate of convergence?
- **How does the theory generalize to 3+ parties?**
- Can we explore other types of voting systems, ex: Ranked Choice, Condorcet

Questions?